S$\vec{e}$MA

BOLETÍN NÚMERO 47

Junio 2009

sumario

Editorial 5

Artículos 7

A Realistic Example of Environmental Control: Optimization of River Fishways, por L.J. Alvarez-Vázquez et al. 7

On some difficulties of the numerical approximation of nonconservative hyperbolic systems, por C. Parés y M.L. Muñoz-Ruiz 23

Actas del Workshop Iberoamericano de Matemáticas Aplicadas 53

Mathematical and numerical analysis for reaction-diffusion systems modeling the spread of early tumors, por V. Anaya, M. Bendahmane y M. Sepúlveda 55

Existence of relaxed weak solutions of a generalized Boussinesq system with restriction on the state variables, por J.L. Boldrini, M.A. Rojas-Medar y M. Santos da Rocha 63

Multiresolution simulation of reaction-diffusion systems with strong degeneracy, por R. Bürger y R. Ruiz-Baier 73

Un problema extremal para un conductor de dos fases en una bola, por C. Conca, R. Mahadevan y L. Sanz 81

Método de elementos finitos para la aproximación de un modelo de cristales líquidos nemáticos, por F. Guillén y J.V. Gutiérrez 91

Stationary asymmetric fluids and Hodge operator, por I. Kondrashuk, E.A. Notte-Cuello and M.A. Rojas-Medar 99

Premio S$\vec{e}$MA al Joven Investigador 2008 107

Numerical analysis of some exterior problems, mixed methods and a posteriori error analysis in fluid mechanics and elasticity, por María González Taboada 107

Resúmenes de tesis doctorales 137
Resúmenes de libros 141
Anuncios 143

Boletín de la Sociedad Española de Matemática Aplicada S$\vec{e}$MA

Grupo Editor

Comité Científico

Responsables de secciones

Página web de S$\vec{e}$MA

`http://www.sema.org.es/`

e-mail

`info@sema.org.es`

Dirección Editorial: Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla - La Mancha. Avda. de Camilo José Cela s/n. 13071. Ciudad Real. `boletin.sema@uclm.es`

ISSN 1575-9822.
Depósito Legal: AS-1442-2002.

Imprime: Gráficas Lope. C/ Laguna Grande, parc. 79, Políg. El Montalvo II 37008. Salamanca.

Diseño de portada: Ernesto Aranda
Ilustración de portada: obras de Iñigo Quílez (*Kindercrasher*, *Enmperaltta*, *Baile de Energía* y *Tormenta de Gas*) reproducidas por cortesía del autor.

Estimados socios,

Os hacemos llegar un nuevo número del boletín de nuestra Sociedad en el que en la sección de artículos científicos recogemos dos interesantes trabajos de Lino Álvarez-Vázquez y colaboradores y de Carlos Parés y María Luz Muñoz. También incluímos algunos trabajos presentadas en el segundo "Workshop Iberoamericano de Matemáticas Aplicadas" que tuvo lugar en la ciudad de Chillán (Chile) en Diciembre de 2008.

Por otro lado, publicamos el trabajo de la ganadora del Premio S$\vec{e}$MA al Joven Investigador 2008, nuestra compañera María González Taboada, que recibió éste en el transcurso de la pasada Escuela Hispano-Francesa celebrada en Valladolid.

Os recordamos finalmente que se encuentra próximo a celebrar el XXI CEDYA / XI CMA que tenemos el honor de organizar desde la Universidad de Castilla - La Mancha. Deseando que nos encontremos en dicho evento,

Recibid un cordial saludo,

Grupo Editor
boletin.sema@uclm.es

Bol. Soc. Esp. Mat. Apl.
n^o47(2009), 7–22

A REALISTIC EXAMPLE OF ENVIRONMENTAL CONTROL: OPTIMIZATION OF RIVER FISHWAYS

L.J. ALVAREZ-VÁZQUEZ[1], A. MARTíNEZ[1],
M.E. VÁZQUEZ-MÉNDEZ[2], M.A. VILAR[2]

[1]Departamento de Matemática Aplicada II. ETSI Telecomunicación
Universidad de Vigo. 36310 Vigo. Spain.
[2]Departamento de Matemática Aplicada. EPS
Universidad de Santiago de Compostela. 27002 Lugo. Spain.

{lino,aurea}@dma.uvigo.es {miguelernesto.vazquez,miguel.vilar}@usc.es

Abstract

The main objective of this work is to present an application of mathematical modelling and optimal control theory to an ecological engineering problem related to preserve and enhance natural stocks of salmon and other fish which migrate between saltwater and freshwater. Particularly, we study (a) the design and (b) the management of a hydraulic structure (fishway) that enables fish to overcome stream obstructions as dams or weirs. Both problems are formulated within the framework of the optimal control of partial differential equations. They are approximated by discrete unconstrained optimization problems, and then, solved by using a gradient-free method (the Nelder-Mead algorithm). Finally, numerical results are showed in a standard real-world situation.

1 Introduction

Many types of fish undertake migrations on a regular basis, on time scales ranging from daily to annual, and with distances ranging from a few meters to thousands of kilometers. In this work we take attention on diadromous fish which migrate between salt and fresh water. The best known diadromous fish are salmon (*salmo salar*), trout (*salmo trutta*), eel (*anguilla anguilla*), sturgeon (*acipenser sturio*), lamprey (*lampetra fluviatilis*, *petromyzon marinus*), barbel (*barbus bocagei*), carp (*cyprimus carpio*), perch (*perca fluviatilis*)... There exist three types of diadromous fish, depending on their specific migration patterns: anadromous, catadromous and amphidromous: Anadromous fish spend most of their adult lives in saltwater, and migrate to freshwater rivers and lakes to reproduce. Anadromous fish species include lamprey, sturgeon, salmon, and

Fecha de recepción: 15/04/2009. Aceptado (en forma revisada): 23/04/2009.

trout. More than half of all diadromous fish in the world are anadromous. Catadromous fish spend most of their adult lives in freshwater, and migrate to saltwater to spawn. Juvenile fish migrate back upstream where they stay until maturing into adults, at which time the cycle starts again. One of the main catadromous species is the eel. About one quarter of all diadromous fish are catadromous. Finally, amphidromous species move between estuaries and coastal rivers and streams, usually associated with the search for food or refuge rather than the need to reproduce. Amphidromous fish can spawn in either freshwater or in a marine environment. Less than one fifth of all diadromous fish are amphidromous: An example is the bull shark (*carcharhinus leucas*).

As it is well-known, salmon, for instance, is capable of going hundreds of kilometers upriver. When people construct an artificial barrier in a river (for example, a dam or a weir) European legal regulations force them to install also a fishway in order to allow fish to overcome it.

Fishways are hydraulic structures placed on or around man-made barriers to assist the natural migration of diadromous fish. An exhaustive overview on the design and management of river fishways (also known as fish-ladders or fish-passes) can be found in the interesting book of Clay [10]. In the literature three different types of fishways are studied: the pool and weir type [27], the Denil type [24], and the vertical slot type [26]. In addition to these three types, that enable the fish to swim upstream under their own effort, there is a more recent fourth class: the fish-locks (or fish-elevators), which lift the fish over the obstruction [32].

Pool and weir fishways were the earliest type constructed - the first recorded attempts to construct this type of fishway were made in Europe in the 17th century - and are still built with the addition of orifices in their walls. A pool and weir fishway consists of a number of pools formed by a series of weirs. The fish passes over a weir by swimming at burst speed or by jumping over it. The fish then rests in the pool, then passes over the next weir, and so on, till it completes the ascent. The success of this type of fishway depends on the maintenance of water levels, which can be facilitated by the provision of a set of orifices in the weir walls close to the floor.

The Denil fishway is essentially a straight rectangular flume provided with closely spaced baffles or vanes on the bottom and sides. The first of the classical works of G. Denil on the scientific design of fish-passes was already published in 1909 in *Annales des Travaux Publiques de Belgique*. Of the many types of Denil fishway studied in the scientific literature, the more commonly used are the standard Denil fishway and the more complex "Alaska Steep-pass" [25].

However, we deal here with the third type of fishway, which is the more generally adopted for upstream passage of fish in streams obstructions: the vertical slot fishway. It consists of a rectangular channel with a sloping floor that is divided into a number of pools (see Fig. 1). Water runs downstream in this channel, through a series of vertical slots from one pool to the next one below. The water flow forms a jet at the slot, and the energy is dissipated by mixing in the pool. Fish ascends, using its burst speed, to get past the slot, then it rests in the pool till the next slot is tried [7].

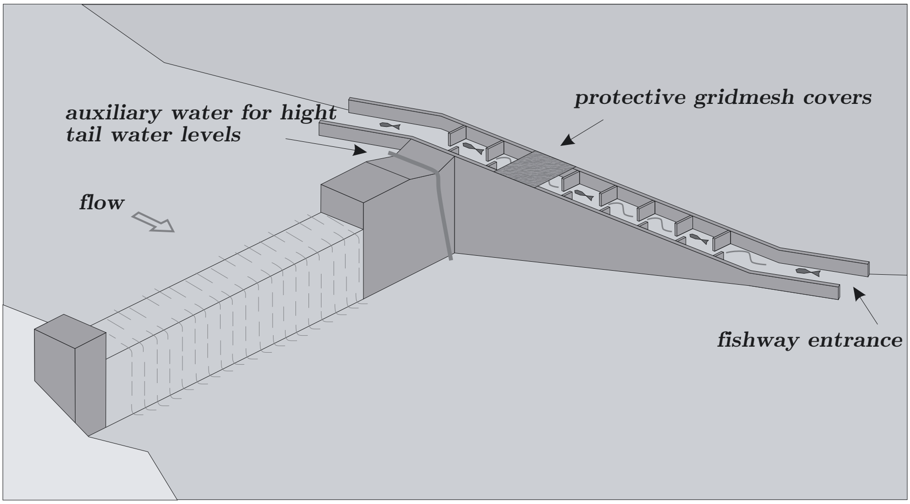


Figure 1: Schematic drawing of a vertical slot fishway

During last decades much attention has been paid, both from the theoretical and the experimental viewpoint, to the hydraulic characteristics, the flow regimes, and the turbulence structures in all types of river fishways. We have to cite here the pioneering works of the team of N. Rajaratnam [28, 29, 34, 11, 12, 18], and, more recently, the interesting works of the team of J. Puertas [23, 31, 9]. Nevertheless, the important role of a correct design in the fishway has been much less studied. We must mention the work of Kim [16] (for the case of pool and weir fishways), the works of Odeh [22] and of Mallen-Cooper and Stuart [19] (for the design of Denil fishways), and - in a more general approach within the field of ecological/environmental engineering - in the papers of Karisch and Power [13], Meselhe et al. [20], Boiten [8], Weber and Joy [33], Yasuda et al. [35] and Richmond et al. [30], among others. However, the optimal design of a vertical slot fishway has not been analyzed, as far as we know, into the scientific literature.

As said above, the objective of a fishway is enabling fish to overcome obstructions. In order to get it, water velocity in the fishway must be controlled. Specifically, that means that in the zone of the channel near the slots, the velocity must be close to a desired velocity suitable for fish leaping and swimming capabilities. In the remains of the fishway, the velocity must be close to zero for making possible the rest of the fish. Moreover, in all the channel, flow turbulence must be minimized.

If a new fishway is going to be built, water velocity can be controlled through the location and length of the baffles separating the pools. On the opposite, if the fishway is already built, it only can be controlled by determining the flux of inflow water. In this work we are going to use mathematical modelling and optimal control theory to study these two situations: first related to the optimal

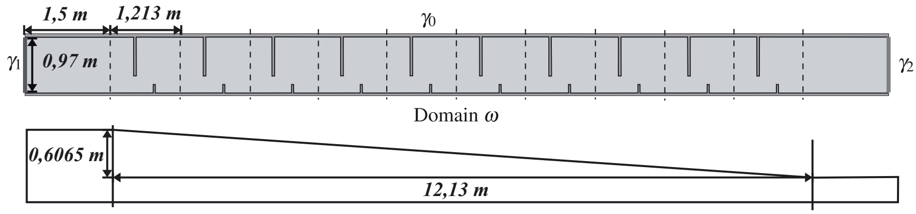


Figure 2: Ground plant (domain ω) and elevation of the fishway under study

design of a new fishway, and second related to the optimal management of a fishway already built.

In order to do it, we begin presenting a mathematical model (shallow water equations) to simulate the water velocity in a fishway and giving a mathematical expression to evaluate the quality of that velocity field in terms of the fish capabilities. Next, we study the first problem related to the optimal design of a new fishway to be built: we describe the problem, formulate it as a shape optimization problem, and show that it can be approximated by a discrete unconstrained optimization problem. In next section we study the second problem, related to the management of a fishway already built: we formulate it as an optimal boundary control problem and, as in the previous case, approximate it by a discretized problem. A gradient-free method (the Nelder-Mead algorithm) to solve this type of optimization problems is proposed in last sections, and numerical results for a standard fishway are presented.

2 Numerical simulation of water velocity in a fishway

Let $\omega \subset \mathbb{R}^2$ be the ground plant of a fishway consisting of a rectangular channel dividing into a reduced number pools with baffles and sloping floor, and two transition pools, one at the beginning and another at the end of the channel, with no baffles and flat floor. A scheme of the fishway used in this paper can be seen in Fig. 2: water enters by the left side and runs downstream to the right side, and fish ascend in the opposite direction [5]. The number of pools (ten) and the dimensions of the full channel correspond to an experimental scale fishway reported by Puertas et al. [23].

Water flow in the domain ω along the time interval $(0, T)$ is governed by the shallow water equations:

$$\left.\begin{array}{ll} \dfrac{\partial H}{\partial t} + \vec{\nabla}.\vec{Q} = 0 & \text{in } \omega \times (0,T) \\ \dfrac{\partial \vec{Q}}{\partial t} + \vec{\nabla}.(\dfrac{\vec{Q}}{H} \otimes \vec{Q}) + gH\vec{\nabla}(H-\eta) = \vec{f} & \text{in } \omega \times (0,T) \end{array}\right\} \quad (1)$$

where

- $H(x, y, t)$ is the height of water at point $(x, y) \in \omega$ at time $t \in (0, T)$,

- $\vec{u}(x,y,t) = (u,v)$ is the averaged horizontal velocity of water,
- $\vec{Q}(x,y,t) = \vec{u}H$ is the areal flow per unit depth,
- g is the gravity acceleration,
- $\eta(x,y)$ represents the bottom geometry of the fishway,
- $\vec{f}$ collects all the effects of bottom friction, atmospheric pressure and so on.

These equations must be completed with a set of initial and boundary conditions. In order to do that, we need to define three different parts in the boundary of ω: the lateral boundary of the channel, denoted by γ_0, the inflow boundary, denoted by γ_1, and the outflow boundary, denoted by γ_2. We also consider $\vec{n}$ the unit outer normal vector to boundary. Thus, we assume the normal flux and the vorticity to be null on the lateral walls of the fishway, we impose an inflow flux in the normal direction, and we fix the height of water on the outflow boundary, that is,

$$\left.\begin{array}{lll} H(0) = H_0, \quad \vec{Q}(0) = \vec{Q}_0 & & \text{in } \omega \\ \vec{Q}.\vec{n} = 0, \quad curl(\dfrac{\vec{Q}}{H}) = 0 & & \text{on } \gamma_0 \times (0,T) \\ \vec{Q} = q\,\vec{n} & & \text{on } \gamma_1 \times (0,T) \\ H = H_2 & & \text{on } \gamma_2 \times (0,T) \end{array}\right\} \tag{2}$$

By using this notation we can give a mathematical expression to evaluate the quality of water velocity in the fishway. We have two objectives:

1. In the zone of the channel near the slots (say the lower third) the velocity must be as close as possible to a typical horizontal velocity c suitable for fish leaping and swimming capabilities, and in the remaining of the fishway, the velocity must be close to zero for making possible the rest of the fish. In short, the velocity must be close to the following target velocity:
$$\vec{v}(x_1,x_2) = \left\{ \begin{array}{ll} (c,0), & \text{if } \ x_2 \leq \frac{1}{3}W \\ (0,0), & \text{otherwise} \end{array} \right. \tag{3}$$
where W is the width of the channel (in our case, as shown in Fig. 2, $W = 0.97\,m$).

2. Flow turbulence must be minimized in all the channel in order to avoid fish disorientation.

According to this, if we fix a weight parameter $\xi \geq 0$ for the role of the vorticity and define the objective function

$$J = \frac{1}{2}\int_0^T \int_\omega \|\frac{\vec{Q}}{H} - \vec{v}\|^2 + \frac{\xi}{2}\int_0^T \int_\omega |curl(\frac{\vec{Q}}{H})|^2, \tag{4}$$

the water velocity $\vec{u} = \vec{Q}/H$ will be better for our purposes as the value of the cost function J becomes smaller.

In order to evaluate J, firstly we have to solve the shallow water equations (1) with initial and boundary conditions (2). In this work we use an implicit discretization in time, upwinding the convective term by the method of characteristics, and Raviart-Thomas finite elements for the space discretization (the whole details of the numerical scheme can be seen in Bermúdez et al. [6]. So, for the time interval $(0, T)$ we choose a natural number N, consider the time step $\Delta t = T/N > 0$ and define the discrete times $t_k = k\Delta t$ for $k = 0, \ldots, N$. We also consider a Lagrange-Galerkin finite element triangulation τ_h of the domain ω. Thus, the numerical scheme provides us, for each discrete time t_k, with an approximated flux $\vec{Q}_h^k$ and an approximated height H_h^k, which are piecewise-linear polynomials and discontinuous piecewise-constant functions, respectively. With these approximated fields we can compute the approximated velocity $\vec{u}_h^k = \vec{Q}_h^k/H_h^k$, and approach the value of J by

$$J_h^{\Delta t} = \frac{\Delta t}{2}\sum_{k=1}^{N}\sum_{E\in\tau_h}\{\int_E \|\vec{u}_h^k - \vec{v}\|^2 + \xi\int_E |curl(\vec{u}_h^k)|^2\} \tag{5}$$

3 Problem 1: Design of a fishway to be built

In this section we are going to study the optimal design of a new fishway. As we have said, if we are going to build a new fishway, we can control the water velocity through the location and length of the baffles in the pools. We take the channel described in previous section, assume that the structure of the ten pools with sloping floor has to be the same (the shape of the complete fishway is given by the shape of the first pool) and then, we take the two midpoints corresponding to the end of the baffles in the first pool (points $a = (s_1, s_2)$ and $b = (s_3, s_4)$ in Fig. 3) as design variables.

We are looking for points a and b which provide the best velocity for fish (i.e. minimizing the function J given by (4)), but, previously, we must impose several design constraints on these points: first, we assume that points a and b are inside the dashed rectangle of Fig. 3, that is, the following eight relations must be satisfied:

$$\left.\begin{array}{rcl} \frac{1}{4}\,1.213 \leq & s_1, s_3 & \leq \frac{3}{4}\,1.213 \\ 0 \leq & s_2, s_4 & \leq \frac{1}{4}\,0.97 \end{array}\right\} \tag{6}$$

The second type of constraints are related to the fact that the vertical slot must be large enough so that fish can pass comfortably through it. This translates into the two additional linear constraints:

$$\left.\begin{array}{l} \Delta_1 = s_3 - s_1 \geq 0.1 \\ \Delta_2 = s_2 - s_4 \geq 0.05 \end{array}\right\} \tag{7}$$

(If the half width of the baffle is $r = 0.0305\,m$, (standard datum as appeared in Puertas et al. [23]) we are actually imposing that the slot width must be, at least, of $\sqrt{(0.1-2r)^2 + 0.05^2} = 0.063\,m$.)

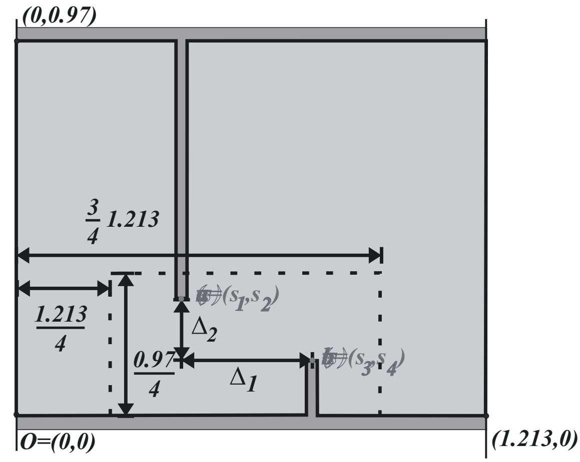


Figure 3: Scheme of the first pool

Then, the first optimization problem can be formulated as follows:

Problem $(\mathcal{P}_1)$: Find the optimal shape of domain ω, that is, find $s = (a, b) = (s_1, s_2, s_3, s_4) \in \mathbb{R}^4$ verifying constraints (6) and (7), in such a way that $\vec{Q}$ and H, given by the solution of the state system (1)-(2) on the fishway $\omega = \omega(s)$, minimize the objective function $J \equiv J(s)$ given by (4).

A mathematical analysis of problem $(\mathcal{P}_1)$ can be found in Alvarez-Vázquez et al. [1, 2]. For its numerical resolution we propose a penalty method. Particularly, for a large enough parameter $\alpha_1 > 0$, we approximate $(\mathcal{P}_1)$ by the discrete unconstrained optimization problem

$$\min \Phi_1(s) \tag{8}$$

where, for $s = (s_1, s_2, s_3, s_4) \in \mathbb{R}^4$, the value of $\Phi_1(s)$ can be computed from the following algorithm:

Step 1. Consider the corresponding domain $\omega(s)$ and its new triangulation $\tau_h(s)$.

Step 2. Solve the state system (1)-(2) on $\omega(s)$ as it has proposed in previous section and compute $J_h^{\Delta t} \equiv J_h^{\Delta t}(s)$ given by (5).

Step 3. Define $\tilde{\Phi}_1(s)$ in such a way that $\tilde{\Phi}_1(s) \leq 0 \Leftrightarrow s$ verifies (6) and (7), that is, compute

$$\begin{aligned}\tilde{\Phi}_1(s) = \max\{&\tfrac{1.213}{4} - s_1, \tfrac{1.213}{4} - s_3, s_1 - \tfrac{3}{4}1.213, s_3 - \tfrac{3}{4}1.213, -s_2,\\ &-s_4, s_2 - \tfrac{0.97}{4}, s_4 - \tfrac{0.97}{4}, 0.1 - s_3 + s_1, 0.05 - s_2 + s_4\}\end{aligned} \tag{9}$$

Step 4. Compute the value of the discrete penalty function

$$\Phi_1(s) = J_h^{\Delta t}(s) + \alpha_1 \max\{\tilde{\Phi}_1(s), 0\} \tag{10}$$

To solve the problem (8) we use a gradient-free method, the Nelder-Mead algorithm, which is summarized in section 5.

4 Problem 2: Management of a fishway already built

The second problem consists of the optimal management of a fishway already built. Now we suppose that the fishway is already built (the domain $\omega \in \mathbb{R}^2$ is known and fixed) and we look for the flux across the inflow boundary providing a suitable water velocity in the fishway (that is, giving a value of the expression (4) as lower as possible).

In this case, in the state system (1)-(2) the domain ω is a datum, and the control variable is the function $q(t)$, the flux across the inflow boundary γ_1 for the time interval $(0, T)$. For this problem, we also have constraints on the control variable: since we need to inject water through the inflow boundary, q must be negative and due to technological reasons q must be bounded by a fixed value. Thus, we are led to consider only the admissible fluxes in the set:

$$U_{ad} \quad = \quad \{l \in L^2(0,T) \ : \ -B \leq l \leq 0\} \tag{11}$$

with $B > 0$ a technological bound for water inflow.

So, the optimal management of the fishway is formulated as:

Problem $(\mathcal{P}_2)$: Find the flux $q \in U_{ad}$ on the fishway inflow in such a way that, verifying the state system $(1)-(2)$, minimizes the cost function $J \equiv J(q)$ given by (4).

From a mathematical point of view, $(\mathcal{P}_2)$ is very different from $(\mathcal{P}_1)$ (a complete mathematical analysis of $(\mathcal{P}_2)$ can be seen in Alvarez-Vázquez et al. [3, 4]), however its numerical resolution can be done in a similar way. In effect, due to technological reasons (flow control mechanisms cannot act upon water flow in a continuous way, but discontinuously at short time periods) we seek the control among the piecewise-constant functions. So, for the time interval $[0, T]$ we choose a number $M \in \mathbb{N}$, we consider the time step $\Delta\tau = T/M > 0$, and we define the discrete times $\tau_m = m\Delta\tau$ for $m = 0, 1, \ldots, M$. Thus, a function $q \in L^2(0,T)$ which is constant at each subinterval determined by the grid $\{\tau_0, \tau_1, \ldots, \tau_M\}$ is completely fixed by the set of values $q^{\Delta\tau} = (q^0, q^1, \ldots, q^{M-1}) \in \mathbb{R}^M$, where $q^m = q(\tau_m),\ m = 0, \ldots, M-1$. For a given $q^{\Delta\tau} \in \mathbb{R}^M$, determining a unique inflow flux q, the shallow water equations can be solved as it was proposed in above sections, and we can compute the value of $J_h^{\Delta t} \equiv J_h^{\Delta t}(q^{\Delta\tau})$ given by (5). Then, for a large enough penalty parameter $\alpha_2 > 0$, the problem $(\mathcal{P}_2)$ can be approximated by

$$\min \Phi_2(q^{\Delta\tau}) \tag{12}$$

where, for $q^{\Delta\tau} = (q^0, q^1, \ldots, q^{M-1}) \in \mathbb{R}^M$, the value of $\Phi_2(q^{\Delta\tau})$ is computed from the two following steps:

Step 1. Solve the state system (1)-(2) with a boundary condition on γ_1 given by $q^{\Delta\tau}$ and compute $J_h^{\Delta t} \equiv J_h^{\Delta t}(q^{\Delta\tau})$ given by (5).

Step 2. Compute the value of the discrete penalty function

$$\Phi_2(q^{\Delta\tau}) = J_h^{\Delta t}(q^{\Delta\tau}) + \alpha_2 \sum_{m=0}^{M-1} \max\{q^m, -B - q^m, 0\} \tag{13}$$

For a small M, we also solve the problem (12) by using the Nelder-Mead algorithm, detailed in next section.

5 Numerical Optimization: The Nelder-Mead Method

The Nelder-Mead simplex method [21] is a direct search method, which merely compares function values; the values of the objective function being taken from a set of sample points (simplex) are used to continue the sampling.

In order to minimize a given function $\Phi : \mathbb{R}^n \to \mathbb{R}$, the algorithm can be easily summarized in the following way: Recall that the convex hull of $n+1$ points in $\mathbb{R}^n$ not contained in the same hyperplane is called a n-simplex. The method constructs a sequence of simplices as approximations to a minimum point. The $n+1$ vertices $y_1, y_2, \dots, y_{n+1}$ of each simplex are sorted according to the objective function values: $\Phi(y_1) \le \Phi(y_2) \le \cdots \le \Phi(y_{n+1})$, and the worst vertex y_{n+1} is replaced with a new point $y(\nu) = (1+\nu)y - \nu\, y_{n+1}$, where y is the centroid of the convex hull of $\{y_1, \dots, y_n\}$, that is, $y = (y_1 + \cdots + y_n)/n$. The value of ν is selected from a sequence $-1 < \nu_\delta < 0 < \nu_\gamma < \nu_\beta < \nu_\alpha$ (typical values are $\nu_\delta = -0.5$, $\nu_\gamma = 0.5$, $\nu_\beta = 1$, $\nu_\alpha = 2$) by rules given in the following algorithm:

While $\Phi(y_{n+1}) - \Phi(y_1)$ is not sufficiently small, compute $y(\nu_\beta)$ and $\Phi_\beta = \Phi(y(\nu_\beta))$. Then:

(a) If $\Phi_\beta < \Phi(y_1)$, compute $\Phi_\alpha = \Phi(y(\nu_\alpha))$. If $\Phi_\alpha < \Phi_\beta$, replace y_{n+1} with $y(\nu_\alpha)$; otherwise replace y_{n+1} with $y(\nu_\beta)$. Go to **(f)**.

(b) If $\Phi(y_1) \le \Phi_\beta < \Phi(y_n)$, replace y_{n+1} with $y(\nu_\beta)$ and go to **(f)**.

(c) If $\Phi(y_n) \le \Phi_\beta < \Phi(y_{n+1})$, compute $\Phi_\gamma = \Phi(y(\nu_\gamma))$. If $\Phi_\gamma \le \Phi_\beta$, replace y_{n+1} with $y(\nu_\gamma)$ and go to **(f)**; otherwise go to **(e)**.

(d) If $\Phi(y_{n+1}) \le \Phi_\beta$, compute $\Phi_\delta = \Phi(y(\nu_\delta))$. If $\Phi_\delta < \Phi(y_{n+1})$, replace y_{n+1} with $y(\nu_\delta)$ and go to **(f)**; otherwise go to **(e)**.

(e) For $k = 2, \dots, n+1$ set $y_k = y_1 + (y_k - y_1)/2$.

(f) Resort the resulting vertices according to Φ values.

Although the Nelder-Mead algorithm is not guaranteed to converge in the general case, it has good convergence properties in low dimensions (see Lagarias et al. [17] for a detailed analysis of the convergence under convexity requirements). Moreover, in order to prevent stagnation at a non-optimal point, we use a modification proposed by Kelley [15]: we define the simplex gradient $D(\Phi) = V^{-T}\, \Delta(\Phi)$, where V and $\Delta(\Phi)$ are the matrices given by:

$$\begin{aligned} &V = (y_2 - y_1, y_3 - y_1, \dots, y_{n+1} - y_1) \\ &\Delta(\Phi) = (\Phi(y_2) - \Phi(y_1), \Phi(y_3) - \Phi(y_1), \dots, \Phi(y_{n+1}) - \Phi(y_1)) \end{aligned} \tag{14}$$

Thus, when stagnation is detected, we modify the simplex by an oriented restart, replacing it by the new smaller simplex $\hat{y}_1 = y_1$, $\hat{y}_j = \hat{y}_1 - \beta_{j-1} e_{j-1}$, $2 \le j \le$

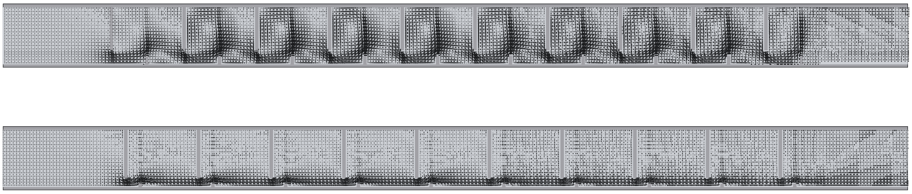


Figure 4: Problem ($\mathcal{P}_1$): Initial (up) and optimal (down) fishways and corresponding horizontal velocity fields at time $t = 300\,s$

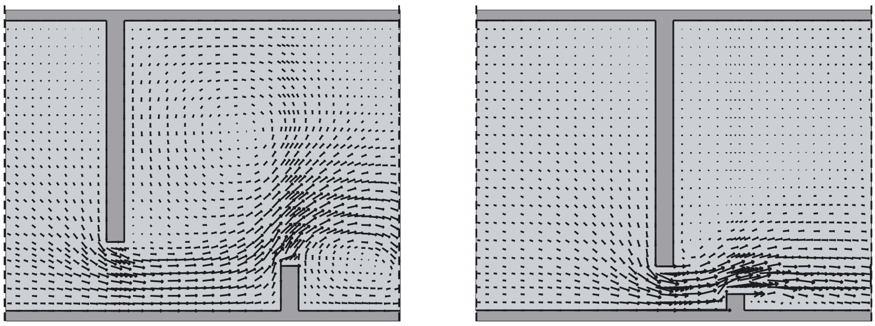


Figure 5: Problem ($\mathcal{P}_1$): Velocity fields in the central pool at time $t = 300\,s$ for the initial shape (left) and the optimal shape (right)

$n+1$, where:

$$\beta_k = \frac{1}{2} \begin{cases} \sigma \ \mathrm{sign}(D(\Phi))_k, & \text{if } (D(\Phi))_k \neq 0 \\ \sigma, & \text{if } (D(\Phi))_k = 0 \end{cases} \tag{15}$$

for the typical length $\sigma = \min_{2 \leq j \leq n+1} \|y_j - y_1\|$.

6 Numerical results

In this section we present the numerical results that we have obtained in a standard situation. We have considered the fishway under study, whose scheme is showed in Fig. 2. Both initial and boundary conditions were taken constant, particularly, $\vec{Q}_0 = (0,0)\,m^2s^{-1}$, $H_0 = 0.5\,m$, $H_2 = 0.5\,m$. The time interval for the simulation was $T = 300\,s$. Moreover, for the sake of simplicity, for the second member $\vec{f}$ we have only considered the bottom friction stress for a Chezy coefficient of $57.36\,m^{0.5}s^{-1}$. For the objective function we have taken a target velocity value $c = 0.8\,ms^{-1}$ and a parameter $\xi = 0$. Finally, for the time discretization we have taken $N = 3000$, that is, a time step of $\Delta t = 0.1\,s$.

Although we have developed many numerical experiences, we present here only one example for each problem.

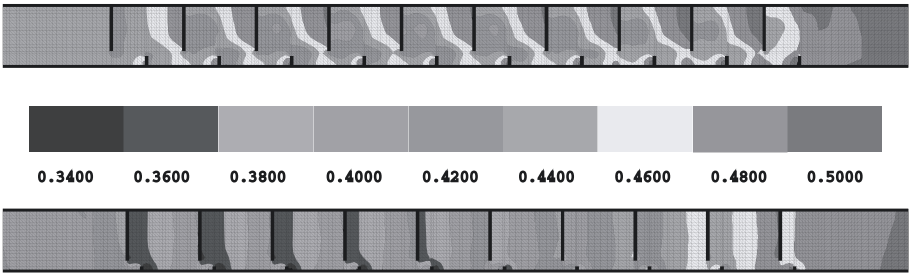


Figure 6: Problem ($\mathcal{P}_1$): Initial (up) and optimal (down) height of water in the whole fishway at time $t = 300\,s$

Problem ($\mathcal{P}_1$): Optimal shape design

We have taken a fixed and constant flux across the inflow boundary $q = -\frac{0.065}{0.97} m^2 s^{-1}$, a penalty parameter $\alpha_1 = 500$ and, for the different space discretizations, we have tried several regular triangulations of about 9500 elements. Thus, applying the Nelder-Mead algorithm, we have passed, after 76 function evaluations, from an initial cost $\Phi_1 = 1046.74$ for a random simplex, to the minimum cost $\Phi_1 = 239.44$, corresponding to the optimal design variables $a = (0.577, 0.147)$, $b = (0.818, 0.054)$. Fig. 4 shows the water velocity at the final time of the simulation (representing the stationary situation) corresponding to the initial random configuration (up) and to the optimal configuration given by a and b (down). We can observe that, in both cases, the flow structure is very similar in all the pools of each fishway, and that, in the optimal case, a clearly defined streamline appears passing through all of the vertical slots. A zoom of the central pools is showed in Fig. 5. In the case of the initial shape (left) we can identify the standard flow patterns presented, for intance, in Rajaratnam et al. [26]: a direct flow region where the flow circulates in a curved trajectory at high velocity from one slot to the next downstream, and two recirculation regions - the larger one located between the long baffles and the smaller one located between the short baffles - flowing in opposite directions. In the case of the optimal shape (right) the horizontal velocity is close to the target velocity $\vec{v}$, and the two large recirculation regions at both sides of the slot are highly reduced. Finally, in Fig. 6 we can also see the height of water at final time $t = 300\,s$. In the initial case (up) the eddy areas create great differences in the height inside each pool, but in the optimal case (down) these variations are much milder. The satisfaction of the boundary condition on γ_2 ($H_2 = 0.5\,m$) can be clearly noticed in both cases.

Problem ($\mathcal{P}_2$): Optimal management

In this case we have taken a fixed fishway given by $a = (0.525, 0.121)$, $b = (0.660, 0610)$, a regular triangulation of 10492 elements, a penalty parameter $\alpha_2 = 10^4$, a technological bound $B = 0.12 m^2 s^{-1}$ and $M = 4$. We have passed, after 66 function evaluations, from an initial cost $\Phi_2 = 612.37$ to the minimum

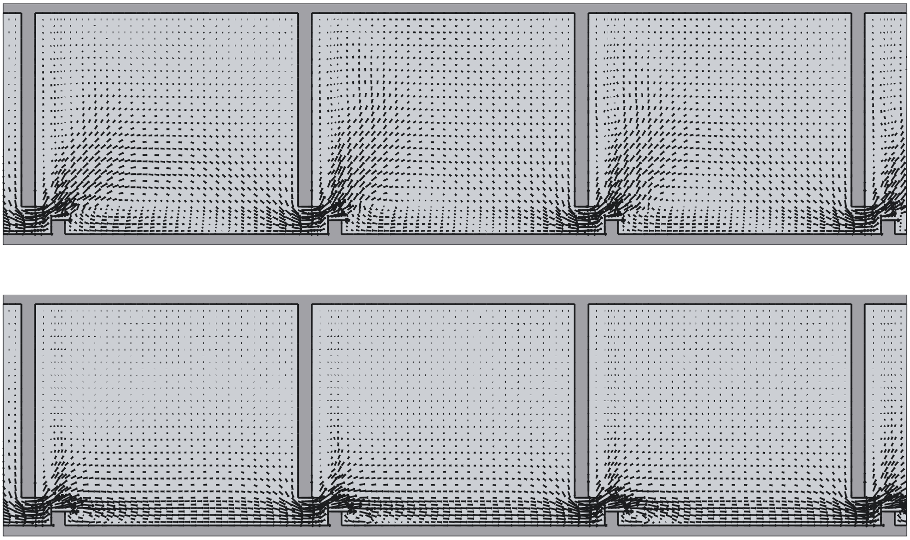


Figure 7: Problem ($\mathcal{P}_2$): Initial (up) and optimal (down) velocity fields in the central pools of the fishway at time $t = 150\,s$

cost $\Phi_2 = 431.27$, corresponding to the optimal flux $q^0 = -0.114$, $q^1 = -0.085$, $q^2 = -0.066$, $q^3 = -0.116$. Fig. 7, 8 and 9 show water horizontal velocities at times $t = 150\,s$, $t = 225\,s$ and $t = 300\,s$, respectively, in the central pools of the fishway, corresponding to the initial random flux (up), and to the optimal flux (down). It can be seen that, for the uncontrolled case (up), the velocity field greatly varies from one time to another, and from one pool to another. However, in the controlled case (down), the velocity field is very similar at all times and in all pools (and also very similar to the desired target velocity $\vec{v}$), although the small eddy region between the short baffles cannot be completely avoided (in contrast to the results obtained for the optimal shape solution in Problem ($\mathcal{P}_1$)).

7 Conclusions

Mathematical modelling has been used to simulate height and velocity of the water in a standard vertical slot fishway. Moreover, optimization and optimal control techniques have been employed to control the water velocity. Particularly, two interesting problems have been formulated and solved: a shape optimization problem which arises in the building of a new fishway and a boundary optimal control problem related to the management of a fishway already built. From the numerical experiences we observe that:

1. Controlling the water velocity in fishways is mandatory, if we want that these hydraulic structures fulfil their task.

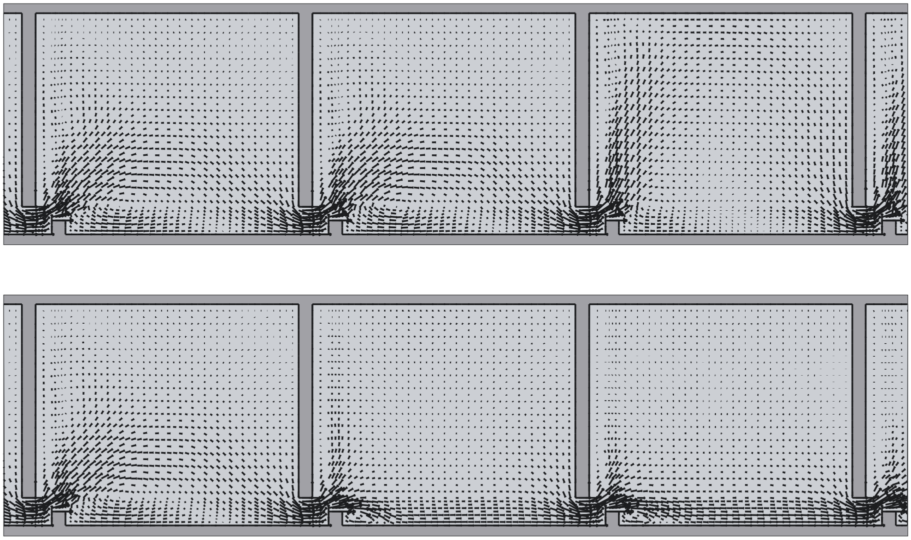


Figure 8: Problem ($\mathcal{P}_2$): Initial (up) and optimal (down) velocity fields in the central pools of the fishway at time $t = 225\,s$

2. Optimal control techniques, joined to numerical simulation, have shown to be very useful tools for this type of hydraulic engineering problems.

3. Controlling the flux of inflow water in a vertical slot fishway can be useful for the management of an already built fishway, but a good shape design is fundamental in order to guarantee a correct hydraulic performance.

Acknowledgments

The research contained in this work was supported by Project MTM2006-01177 of Ministerio de Educación y Ciencia (Spain).

References

[1] Alvarez-Vázquez, L.J., Martínez, A., Rodríguez, C., Vázquez-Méndez, M.E., and Vilar, M.A., 2007. Optimal shape design for fishways in rivers. Math. Comput. Simul. 76, 218-222.

[2] Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., and Vilar, M.A., 2008. An optimal shape problem related to the realistic design of river fishways. Ecological Eng. 32, 293-300.

[3] Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., and Vilar, M.A., 2008. Vertical slot fishways: modeling and optimal management. J. Comput. Appl. Math. 218, 395-403.

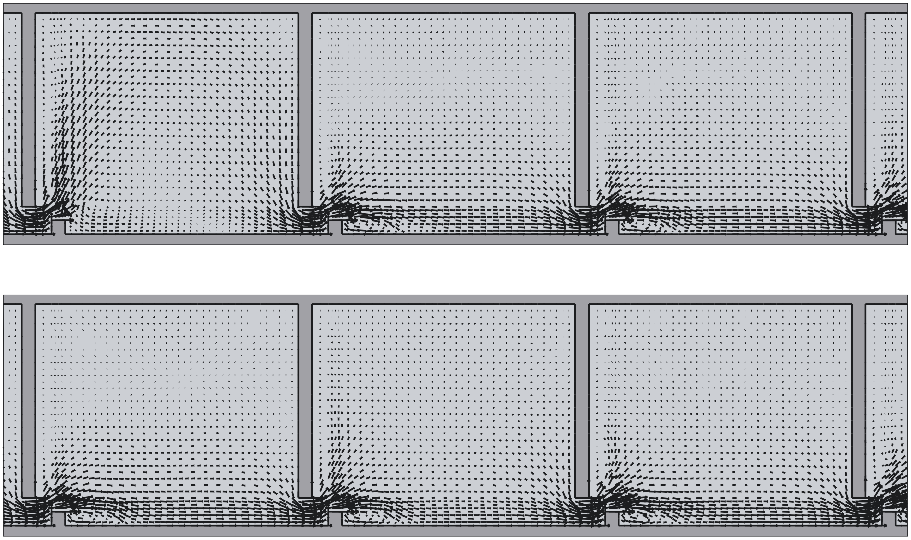


Figure 9: Problem ($\mathcal{P}_2$): Initial (up) and optimal (down) velocity fields in the central pools of the fishway at time $t = 300\,s$

[4] Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., and Vilar, M.A., 2008. Fishway design: An application of the optimal control theory. In: Numerical Mathematics and Advanced Applications (K. Kunish, G. Of, and O. Steinbach, eds.), 581-588, Springer, Berlin.

[5] Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., and Vilar, M.A., 2008. Optimal operation for a river fishway. PAMM: Proc. Appl. Math. Mech., 7, 2060033-2060034.

[6] Bermúdez, A., Rodríguez, C., and Vilar, M.A., 1991. Solving shallow water equations by a mixed implicit finite element method. IMA J. Numer. Anal. 11, 79-97.

[7] Blake, R.W., 1983. Fish locomotion. Cambridge University Press, London.

[8] Boiten, W., 2002. Flow measurement structures. Flow Meas. Instrum. 13, 203-207.

[9] Cea, L., Pena, L., Puertas, J., Vázquez-Cendón, M.E., and Peña, E., 2007. Application of several depth-averaged turbulence models to simulate flow in vertical slot fishways. J. Hydraul. Eng. 133, 160-172.

[10] Clay, C.H., 1995. Design of fishways and other fish facilities. Lewis Publishers, CRC Press, Boca Raton.

[11] Ead, S.A., Rajaratnam, N., and Katopodis, C., 2002. Generalized study of hydraulics of culvert fishways. J. Hydraul. Eng. 128, 1018-1022.

[12] Ead, S.A., Katopodis, C., Sikora, G.J., and Rajaratnam, N., 2004. Flow regimes and structure in pool and weir fishways. J. Environ. Eng. Sci. 3, 379-390.

[13] Karisch, S.E., and Power, M., 1994. A simulation study of fishway design: an example of simulation in environmental problem solving. J. Environ. Management 41, 67-77.

[14] Katopodis, C., Rajaratnam, N., Wu, S., and Towell, D., 1997. Denil fishways of varying geometry. J. Hydraul. Eng. 123, 624-631.

[15] Kelley, C.T., 1999. Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition. SIAM J. Optim. 10, 43-55.

[16] Kim, J.H., 2001. Hydraulic characteristics by weir type in a pool-weir fishway. Ecological Eng. 16, 425-433.

[17] Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex algorithm in low dimensions. SIAM J. Optim. 9, 112-147.

[18] Liu, M., Rajaratnam, N., and Zhu, D.Z., 2006. Mean flow and turbulence structure in vertical slot fishways. J. Hydraul. Eng. 132, 765-777.

[19] Mallen-Cooper, M., and Stuart, I.G., 2007. Optimising Denil fishways for passage of small and large fishes. Fisheries Management Ecol. 14, 61-71.

[20] Meselhe, E.A., Weber, L.J., Odgaard, A.J., and Johnson, T., 2000. Numerical modeling for fish diversion studies. J. Hydraul. Eng. 126, 365-374.

[21] Nelder, J.A., and Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308-313.

[22] Odeh, M., 2003. Discharge rating equation and hydraulic characteristics of standard Denil fishways. J. Hydraul. Eng. 129, 341-348.

[23] Puertas, J., Pena, L., and Teijeiro, T., 2004. Experimental approach to the hydraulics of vertical slot fishways. J. Hydraul. Eng. 130, 10-23.

[24] Rajaratnam, N., and Katopodis, C., 1984. Hydraulics of Denil fishways. J. Hydraul. Eng. 110, 1219-1233.

[25] Rajaratnam, N., and Katopodis, C., 1991. Hydraulics of steeppass fishways. Can. J. Civil Eng. 18, 1024-1032.

[26] Rajaratnam, N., Van de Vinne, G., and Katopodis, C., 1986. Hydraulics of vertical slot fishways. J. Hydraul. Eng. 112, 909-917.

[27] Rajaratnam, N., Katopodis, C., and Mainali, A., 1988. Plunging and streaming flows in pool and weir fishways. J. Hydraul. Eng. 114, 939-944.

[28] Rajaratnam, N., Katopodis, C., and Solanski, S., 1992. New designs for vertical slot fishways. J. Hydraul. Eng. 119, 402-414.

[29] Rajaratnam, N., Katopodis, C., Wu, S., and Sabur, M.A., 1997. Hydraulics of resting pools for Denil fishways. J. Hydraul. Eng. 123, 632-638.

[30] Richmond, M.C., Deng, Z., Guensch, G.R., Tritico, H., and Pearson, W.H., 2007. Mean flow and turbulence characteristics of a full-scale spiral corrugated culvert with implications for fish passage. Ecological Eng. 30, 333-340.

[31] Teijeiro-Rodriguez, T., Puertas, J., Pena, L., and Peña, E., 2006. Evaluating vertical-slot fishway designs in terms of fish swimming capabilities. Ecological Eng. 27, 37-48.

[32] Travade, F., and Larinier, M., 1992. Ecluses et ascenseurs a poissons. Bulletin Français de Peche et Pisciculture 326/327, 95-110.

[33] Weber, N.S., and Joy, D.M., 2002. Use of a scale model to design fishway resting pool improvements. Can. Water Res. J. 27, 401-426.

[34] Wu, S., Rajaratnam, N., and Katopodis, C., 1999. Structure of flow in vertical slot fishway. J. Hydraul. Eng. 125, 351-360.

[35] Yasuda, Y., Ohtsu, I., and Takahashi, M., 2004. New portable fishway design for existing trapezoidal weirs. J. Environ. Eng. Sci. 3, 391-401.

Bol. Soc. Esp. Mat. Apl.
nº47(2009), 23–52

ON SOME DIFFICULTIES OF THE NUMERICAL APPROXIMATION OF NONCONSERVATIVE HYPERBOLIC SYSTEMS

CARLOS PARÉS[1], MARíA LUZ MUÑOZ-RUIZ[2]

[1] Departamento de Análisis Matemático
[2] Departamento de Matemática Aplicada
Universidad de Málaga, 29071 Málaga, Spain

pares@anamat.cie.uma.es, munoz@anamat.cie.uma.es

Abstract

The design of high-order well-balanced shock-capturing numerical methods for nonconservative hyperbolic systems is a very active front of research, as PDE systems of this nature arises in many flow models. The approximated solutions are expected to be consistent with the physics of the real flows to be simulated: in particular (1) they should satisfy the conservation properties prescribed by the physics of the problem and (2) their discontinuities should satisfy some jump conditions consistent with the real phenomena to be simulated. The question addressed here is whether or not it is possible to construct numerical schemes satisfying these two requirements. While for conservative systems the answer to this question is positive, some important difficulties arise in certain nonconservative cases. These difficulties will be discussed and illustrated with some numerical results. Finally, some conclusions will be drawn.

Key words: *Nonconservative hyperbolic system, shock wave, family of paths, equivalent equation, convergence error measure, formally path-consistent scheme*

AMS subject classifications: *65M06 35L65 76L05 76N*

1 Introduction

In this paper we discuss some difficulties related to the numerical discretization of hyperbolic PDE of the form:

$$w_t + f(w)_x + \mathcal{B}(w)w_x = S(w)\sigma_x, \quad x \in \mathbb{R},\ t > 0, \tag{1}$$

where the unknown $w(x,t)$ takes values in an open convex set $\mathcal{O}$ of $\mathbb{R}^N$; f is a regular function from $\mathcal{O}$ to $\mathbb{R}^N$; $\mathcal{B}$, a regular matrix function from $\mathcal{O}$ to

Fecha de recepción: 05/04/2009. Aceptado (en forma revisada): 16/04/2009.

$\mathcal{M}_{N\times N}(\mathbb{R})$; S, a function from $\mathcal{O}$ to $\mathbb{R}^N$; and $\sigma(x)$, a known function from $\mathbb{R}$ to $\mathbb{R}$.

Observe that (1) includes as particular cases systems of conservation laws ($\mathcal{B} \equiv 0$ and $S \equiv 0$) as well as systems of balance laws ($\mathcal{B} \equiv 0$). A number of models with this form have been introduced in fluid dynamics to serve as simplified flow models. As an example, the system of partial differential equations governing the one-dimensional flow of two superposed immiscible layers of shallow water fluids through a straight channel with constant rectangular cross-section (see [13]) is the particular case of (1) corresponding to the choices $N = 4$, $\sigma = H$,

$$w = \begin{bmatrix} h_1 \\ q_1 \\ h_2 \\ q_2 \end{bmatrix}, \quad f(w) = \begin{bmatrix} q_1 \\ \dfrac{q_1^2}{h_1} + \dfrac{g}{2}h_1^2 \\ q_2 \\ \dfrac{q_2^2}{h_2} + \dfrac{g}{2}h_2^2 \end{bmatrix}, \quad S(w) = \begin{bmatrix} 0 \\ gh_1 \\ 0 \\ gh_2 \end{bmatrix}, \tag{2}$$

$$\mathcal{B}(w) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & gh_1 & 0 \\ 0 & 0 & 0 & 0 \\ grh_2 & 0 & 0 & 0 \end{bmatrix}, \tag{3}$$

with

$$r = \frac{\rho_1}{\rho_2}.$$

Index 1 makes reference to the upper layer and index 2 to the lower one. $H(x)$ represents the depth function. Each layer is assumed to have a constant density, ρ_i, $i = 1, 2$ ($\rho_1 < \rho_2$). The unknowns q_i and h_i represent respectively the mass-flow and the thickness of the ith layer. Systems of the form (1) also appear in models of turbulent shallow waters, two-phase flows, sediment transport, turbidity currents, avalanches, submarine avalanches, etc. See for instance [1], [3], [17], [22], [31], [37], [38], [39] for some examples. Some animations corresponding to simulations of geophysical flows obtained by our group of research on the basis of this type of models can be found in the web page `http://anamat.cie.uma.es/animaciones`

In order to avoid the discussion of the choice of boundary conditions, we consider here Cauchy problems consisting of (1) with an initial condition:

$$w(x, 0) = w_0(x), \quad x \in \mathbb{R}. \tag{4}$$

System (1) can be considered in turn as a particular case of the following family of PDE systems:

$$W_t + \mathcal{A}(W)W_x = 0, \quad x \in \mathbb{R},\ t > 0, \tag{5}$$

in which the unknown $W(x, t)$ takes values in an open convex set Ω of $\mathbb{R}^M$, and $W \in \Omega \mapsto \mathcal{A}(W) \in \mathcal{M}_{M\times M}(\mathbb{R})$ is a smooth locally bounded map. In effect, if following [27] the artificial unknown σ and the equation

$$\sigma_t = 0, \tag{6}$$

are added to the system, then (1), (6) can be equivalently written in the form (5) with $M = N + 1$,

$$W = \begin{bmatrix} w \\ \sigma \end{bmatrix} \in \Omega = \mathcal{O} \times \mathbb{R},$$

and $\mathcal{A}(W)$ the matrix-valued function whose block structure is given by

$$\mathcal{A}(W) = \left[\begin{array}{c|c} Df(w) + \mathcal{B}(w) & -S(w) \\ \hline 0 & 0 \end{array}\right], \tag{7}$$

where $Df(w)$ is the Jacobian matrix of f. It can be easily verified that (1), (4) is equivalent to (5) with the initial condition:

$$W(x,0) = W_0(x) = \begin{bmatrix} w_0(x) \\ \sigma(x) \end{bmatrix}, \quad x \in \mathbb{R}. \tag{8}$$

A common feature of systems of the form (5) is the appearance of discontinuities in the solution even for smooth initial conditions. These discontinuities are related to real phenomena in which some of the variables present a sharp variation in a very thin region: this is the case of shocks of transonic airplanes, or hydraulic jumps for free surface water flows.

The design of numerical methods with good properties for problems of the form (1) or the particular cases corresponding to $\mathcal{B} = 0$ or $S = 0$, is a very active front of research. From the point of view of the quality of the numerical solutions, the following properties are usually required to the numerical schemes:

- The discontinuities of the solutions have to be sharply captured and without unphysical oscillations near them.
- The numerical solutions have to provide accurate approximations of the solutions in the smoothness regions.
- The smooth nontrivial stationary solutions (or at least a family of them) have to be exactly or accurately enough approximated in order to avoid unphysical oscillations near equilibria.

Numerical schemes satisfying these properties are called high-order well-balanced shock-capturing methods. A number of methods with good properties for hyperbolic systems with source terms and/or nonconservative products have been recently developed: see for instance [2], [4], [6], [7], [8], [9], [12], [14], [16], [21], [23], [33], [34], [40], [41], [42], [43], [46], [47], ...

Besides these good numerical properties, the approximated solutions must also be consistent with the physics of the real flows to be simulated. In particular:

(P1) The numerical solutions have to satisfy all the conservation properties prescribed by the physics of the problem.

(P2) The appearance and the propagation of discontinuities have to be consistent with the real phenomena that is simulated.

Property (P1) may seem paradoxical for nonconservative systems, but notice that a system of the form (1) may contain a conservative subsystem: this is the case, for instance, for the two mass conservation equations in the two-layer shallow water systems (1st and 3d equations). The numerical solutions are thus expected to satisfy the mass conservation principle.

For conservative systems there is a number of available methods satisfying both properties (P1) and (P2). In this paper, the question addressed is the following: is it possible to design numerical schemes for nonconservative systems (5) having both properties (P1) and (P2)? Unfortunately, this is a very difficult question and we will not be able to give a definitive answer. But we will precisely state the question and discuss which are the main difficulties involved.

The organization is as follows. First, we briefly recall some basic concepts and results concerning pure systems of conservation laws. In Section 3, we discuss the concept of weak solution for nonconservative systems. In Section 4 we introduce a family of numerical schemes satisfying a (P1) property. Next, we discuss to what extent a property (P2) can be proven for this family of schemes. Next, we will show some numerical results that illustrate the theoretical discussion. Finally, some conclusions are drawn.

2 Systems of conservation laws

Many of the difficulties encountered in the study of hyperbolic nonconservative systems are already present in hyperbolic systems of conservation laws. This is the reason why we begin these notes by recalling some concepts and results concerning hyperbolic systems of conservation laws: see [24], [28], [30] for details.

Systems of conservation laws in one space dimension take the form

$$w_t + f(w)_x = 0, \quad x \in \mathbb{R},\ t > 0, \tag{9}$$

where the unknown w is a vector of conserved quantities and the N vector-valued function f is the flux function. System (9) is said to be hyperbolic if the jacobian matrix $Df(w)$ of the flux function is diagonalizable with real eigenvalues $\lambda_1(w), \ldots, \lambda_N(w)$ for each w. In the particular case in which the eigenvalues are all distinct the system is said to be strictly hyperbolic. We are interested in the initial value problem associated with an initial condition (4).

Weak solutions of (9) can be defined by means of a variational formulation: a L^∞_{loc} function w is called a weak solution of the initial value problem (9), (4) if

$$\int_0^\infty \int_{-\infty}^{+\infty} \big(w\,\varphi_t + f(w)\,\varphi_x\big)\,dx\,dt + \int_{-\infty}^{+\infty} w_0(x)\varphi(x,0)\,dx = 0, \tag{10}$$

for any $\mathcal{C}^1$ test function φ with compact support in $\mathbb{R} \times [0, \infty)$.

A weak solution satisfies the system of equations (9) in the sense of distributions. It can be shown that a piecewise $\mathcal{C}^1$ function w is a weak solution if, and only if it is a classical solution where it is smooth and, across a discontinuity, it satisfies the so-called Rankine-Hugoniot condition:

$$\xi[w] = [f(w)], \tag{11}$$

where ξ is the speed of propagation of the discontinuity and square brackets are used to represent the jump of a variable across the discontinuity, i.e.

$$[w] = w^+ - w^-, \quad [f(w)] = f(w^+) - f(w^-),$$

where w^- and w^+ are the left and right limits of the solution at the discontinuity. Essentially, conditions (11) imply that across a discontinuity the conservation laws are also satisfied.

A weak solution for the initial value problem (9), (4) is not necessarily unique. The concept of entropy is used to select the physically relevant solution among all the possible weak solutions. In most cases this concept is given by an entropy pair (η, g), i.e. a pair of regular functions from $\mathcal{O}$ to $\mathbb{R}$, η being convex, such that

$$\nabla g(w) = \nabla \eta(w) \cdot Df(w), \quad \forall\, w \in \mathcal{O}.$$

It can be easily shown that a smooth solution w of (9) also satisfies the conservation law:

$$\eta(w)_t + g(w)_x = 0, \quad x \in \mathbb{R},\ t > 0. \tag{12}$$

A weak solution is said to be an entropy solution if it satisfies the inequality

$$\eta(w)_t + g(w)_x \leq 0, \tag{13}$$

in the distributional sense. For piecewise continuous weak solutions, this entropy criterion is satisfied if, across the discontinuities, the following inequality is satisfied:

$$\xi[\eta(w)] + [g(w)] \leq 0. \tag{14}$$

Functions η and g, which are known as the entropy function and the entropy flux, are usually given by the physics of the system.

By integrating equation (9) in a rectangle $[a, b] \times [t_0, t_1] \subset \mathbb{R} \times [0, \infty)$, it can be easily verified that a smooth solution satisfies the equality:

$$\int_a^b w(x, t_1)\, dx = \int_a^b w(x, t_0)\, dx + \int_{t_0}^{t_1} f(w(a, t))\, dt - \int_{t_0}^{t_1} f(w(b, t))\, dt. \tag{15}$$

In fact, this equality gives an alternative and equivalent way for defining the concept of weak solutions: a L^∞_{loc} function w is called a weak solution of the initial value problem (9), (4) if (15) is satisfied for every rectangle $[a, b] \times [t_0, t_1] \subset \mathbb{R} \times [0, \infty)$. If the flux function is such that $f(0) = 0$ and w is a weak solution of (9), (4) in which w_0 has a compact support, it can be proved that $w(\cdot,t)$ has a compact support for every $t > 0$. Then, by taking $a = -\infty$, $b = \infty$, $t_0 = 0$, and $t_1 = t$ in (15), we obtain:

$$\int_{\mathbb{R}} w(x,t)\,dx = \int_{\mathbb{R}} w_0(x)\,dx, \quad \forall t > 0, \tag{16}$$

which is the global conservation property satisfied by weak solutions of (9), (4).

In the context of computational fluid dynamics, hyperbolic models usually appear as vanishing viscosity limits of parabolic systems. In this case, the discontinuities appearing at the weak solutions of the hyperbolic model are expected to be the limits of the smooth travelling waves appearing in advection dominated problems. More precisely, let us suppose that the conservative system is the limit of the parabolic problems

$$w_t^\epsilon + f(w^\epsilon)_x = \epsilon(\mathcal{R}(w^\epsilon)\,w_x^\epsilon)_x, \tag{17}$$

where the second order term is elliptic, for instance $\mathcal{R}$ may be a constant symmetric positive defined matrix. Let us define a **viscous profile** as a travelling wave

$$w^\epsilon(x,t) = v\left(\frac{x-\xi t}{\epsilon}\right), \tag{18}$$

that is a solution of (17) satisfying

$$\lim_{\chi\to-\infty} v(\chi) = w^-, \quad \lim_{\chi\to+\infty} v(\chi) = w^+, \quad \lim_{\chi\to\pm\infty} v'(\chi) = 0. \tag{19}$$

It can be easily verified that w^ϵ converges a.e. to

$$w(x,t) = \begin{cases} w^- & \text{if } x < \xi t, \\ w^+ & \text{if } x > \xi t, \end{cases} \tag{20}$$

as ϵ tends to 0. Therefore, it is natural to state that a discontinuity linking w^- and w^+ at speed ξ is admissible for the hyperbolic system if there exists a viscous profile satisfying (19). But if a w^ϵ given by (18) solves (17), then v has to be a solution of the ODE:

$$-\xi v' + Df(v)\,v' = (\mathcal{R}(v)\,v')'. \tag{21}$$

By integrating (21) from $-\infty$ to ∞ and taking into account (19), we recover the Rankine-Hugoniot conditions (11). Moreover, it can be shown that every solution w^ϵ of (17) satisfies:

$$\eta(w^\epsilon)_t + g(w^\epsilon)_x \le 0,$$

in the distributional sense, and thus (13) must also be satisfied by (20), as it is the limit of solutions of (17). And this happens if, and only if, (14) is satisfied. Therefore, any admissible discontinuity satisfies both (11) and (14) *independently of the particular form of the viscous term.*

An initial value problem with piecewise constant initial condition

$$w_0(x) = \begin{cases} w_l & \text{if } x < 0, \\ w_r & \text{if } x > 0, \end{cases} \tag{22}$$

is called a Riemann problem. Lax theorem establishes that, for sufficiently close data w_l and w_r, (22) has a unique self-similar weak solution (i.e. $w(x,t) = w(x/t; w_l, w_r)$) composed by at most N simple waves: rarefaction waves, contact discontinuities or shock waves. These problems play a fundamental role in the design of numerical schemes.

In order to discretize the system, computing cells $I_i = [x_{i-1/2}, x_{i+1/2}]$ are considered, whose size Δx is supposed to be constant for simplicity. Let Δt be the constant time step and define $t_n = n\Delta t$. Let us denote by w_i^n the approximation at the cell I_i at time t_n provided by the scheme. The value w_i^n is supposed to be an approximation of the cell averages of the exact solution:

$$w_i^n \cong \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} w(x, t_n)\, dx.$$

The initial cell values are thus given by

$$w_i^0 = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} w_0(x)\, dx. \tag{23}$$

From (15) it can be easily shown that the averages of the exact solution satisfy the equality:

$$\begin{aligned} \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} w(x, t_{n+1})\, dx &= \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} w(x, t_n)\, dx \\ &+ \frac{\Delta t}{\Delta x} \left(\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} w(x_{i-1/2}, t)\, dt - \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} w(x_{i+1/2}, t)\, dt \right). \end{aligned}$$

Therefore it is natural to consider numerical schemes of the form:

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \big(F_{i+1/2}^n - F_{i-1/2}^n \big), \tag{24}$$

where $F_{i+1/2}^n$ is some approximation of the averaged flux through $x_{i+1/2}$:

$$F_{i+1/2}^n \cong \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f(w(x_{i+1/2}, t))\, dt.$$

A numerical scheme is said to be **conservative** if it can be written in the form (24) with

$$F_{i+1/2}^n = F(w_{i-q}^n, \dots, w_{i+p}^n),$$

being F a Lipschitz continuous function, which is called the numerical flux, such that

$$F(w, \dots, w) = f(w), \quad \forall w \in \mathcal{O}.$$

Conservative methods satisfy property (P1): the numerical solutions also satisfy the conservation property (16). In effect, let us denote by w^n the piecewise constant approximation of $w(\cdot, t_n)$ given by

$$w^n(x) = w_i^n, \quad \forall x \in I_i.$$

If the initial condition w_0 has a compact support, the equality

$$\int_{\mathbb{R}} w^n(x)\, dx = \int_{\mathbb{R}} w_0(x)\, dx, \quad \forall n, \tag{25}$$

can be easily obtained by summing up in (24).

Concerning property (P2), the Lax-Wendroff theorem establishes that, when the approximations obtained with a conservative method converge (in a reasonable way), the limit is a weak solution of the initial value problem associated to the system of conservation laws. Therefore, if the numerical solutions provided by a conservative method converge to a discontinuous function, its discontinuities have to satisfy the Rankine-Hugoniot condition (11). Nevertheless, the entropy condition (14) may not be satisfied, i.e. conservative methods may converge to weak solutions which are not entropy solutions. To ensure that, an extra requirement has to be imposed to the numerical scheme. For instance, it is enough to have a *discrete entropy inequality* of the form

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{\Delta x}\big(G_{i+1/2}^n - G_{i-1/2}^n\big), \tag{26}$$

where

$$G_{i+1/2}^n = G\big(w_{i-q}^n, \ldots, w_{i+p}^n\big), \tag{27}$$

G being a Lipschitz continuous function from $\mathcal{O}^{p+q+1}$ to $\mathcal{O}$, consistent with g in the sense that

$$G(w, \ldots, w) = g(w), \quad \forall w \in \mathcal{O}. \tag{28}$$

Summarizing, conservative methods meet the two requirements (P1) and (P2).

Let us finish this section by mentioning four well-known examples of conservative methods.

2.1 Centered scheme

The easiest numerical flux is given by

$$F_C(w_l, w_r) = \frac{f(w_l) + f(w_r)}{2}, \tag{29}$$

but the corresponding conservative method is unconditionally unstable. To obtain stable numerical schemes some information concerning the speed of propagation of small waves (i.e. the eigenvalues of $Df(w)$) has to be added, i.e. some *upwinding* is required.

2.2 Lax-Friedrichs method

The numerical flux is given now by

$$F_{LF}(w_l, w_r) = \frac{f(w_l) + f(w_r)}{2} - \frac{\Delta x}{2\Delta t}(w_r - w_l). \tag{30}$$

This scheme is stable under the CFL condition

$$\Delta t \max_{i,n}\{|\lambda_j(w_i^n)|,\ j = 1, \dots, N\} = cfl \cdot \Delta x, \tag{31}$$

where $cfl \in (0, 1]$. Under the more strict condition $cfl \in (0, 1/2]$, a discrete entropy inequality can also be obtained. Notice that the upwind information is not explicitly used in the expression of the numerical flux but only in the stability condition.

An improvement to the Lax-Friedrichs scheme, called the Rusanov or the local Lax-Friedrichs method can be obtained by replacing the viscosity coefficient $\Delta x/\Delta t$ by a locally chosen value:

$$F_{LLF}(w_l, w_r) = \frac{f(w_l) + f(w_r)}{2} - \frac{\alpha(w_r, w_l)}{2}(w_r - w_l), \tag{32}$$

where

$$\alpha(w_l, w_r) = \max\{|\lambda_j(w_l)|,\ |\lambda_j(w_r)|;\ j = 1, \dots, N\}. \tag{33}$$

2.3 Godunov method

In this method the upwind information is given by the solutions of the Riemann problems: the numerical flux is given by

$$F_G(w_l, w_r) = f(w(0; w_l, w_r)), \tag{34}$$

being $w(x/t; w_l, w_r)$ the self-similar solution of the Riemann problem (9), (22). This method is also stable under the condition (31) with $cfl \in (0, 1]$. If the more strict condition $cfl \in (0, 1/2]$ is imposed, the method can be interpreted as follows: the approximation at the cell I_i at time t_{n+1} is the cell average,

$$u_i^{n+1} = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{u}(x, t_{n+1})\, dx, \tag{35}$$

of the exact entropy solution $\widetilde{u}$ of the Cauchy problem

$$\begin{cases} \widetilde{u}_t + f(\widetilde{u})_x = 0, & x \in \mathbb{R},\ t > t_n, \\ \widetilde{u}(x, t_n) = w^n(x), & x \in \mathbb{R}, \end{cases} \tag{36}$$

where w^n is the piecewise constant given by:

$$w^n(x) = w_i^n \quad \forall x \in I_i, \forall i.$$

A discrete entropy inequality can be also be obtained.

2.4 Roe method

The implementation of Godunov method requires computing the exact solution of Riemann problems (or at least their value at $x = 0$), which may be difficult or very costly for complex systems. In this case, methods based on approximate Riemann solvers are useful. In the case of Roe method, instead of the exact Riemann problems, the following linear problems are considered:

$$\begin{cases} u_t + \mathcal{A}(w_l, w_r)\, u_x = 0, \quad x \in \mathbb{R},\ t > 0, \\ u(x, 0) = \begin{cases} w_l & \text{if } x < 0, \\ w_r & \text{if } x > 0, \end{cases} \end{cases} \tag{37}$$

where $\mathcal{A}(w_l, w_r)$ is a Roe linearization of $Df(w)$, i.e. a function $\mathcal{A}\colon \mathcal{O} \times \mathcal{O} \mapsto \mathcal{M}_{N\times N}(\mathbb{R})$ satisfying the following properties:

- for each $w_l, w_r \in \mathcal{O}$, $\mathcal{A}(w_l, w_r)$ has N distinct real eigenvalues $\lambda_j(w_l, w_r)$, $j = 1, \ldots, N$;
- $\mathcal{A}(w, w) = Df(w)$, for every $w \in \mathcal{O}$;
- for any $w_l, w_r \in \mathcal{O}$,

$$\mathcal{A}(w_l, w_r)\,(w_r - w_l) = f(w_r) - f(w_l). \tag{38}$$

Notice that $\mathcal{A}(w_l, w_r)$ is diagonalizable:

$$\mathcal{A}(w_l, w_r) = \mathcal{K}(w_l, w_r) \cdot \mathcal{L}(w_l, w_r) \cdot \mathcal{K}(w_l, w_r)^{-1}.$$

By proceeding as for Godunov method (solving the linear Riemann problems associated to w_{i-1}^n and w_i^n at every intercell and averaging their solutions at the cells) the numerical scheme can be written as the conservative method whose numerical flux is given by

$$F_R(w_l, w_r) = \frac{f(w_l) + f(w_r)}{2} - \frac{1}{2}|\mathcal{A}(w_l, w_r)|(w_r - w_l), \tag{39}$$

where the following notation has been used:

$$|\mathcal{A}(w_l, w_r)| = \mathcal{K}(w_l, w_r) \cdot |\mathcal{L}(w_l, w_r)| \cdot \mathcal{K}(w_l, w_r)^{-1},$$

with $|\mathcal{L}(w_l, w_r)|$ the diagonal matrix whose coefficients are the absolute value of those of $\mathcal{L}(w_l, w_r)$, i.e. the absolute value of the eigenvalues of $\mathcal{A}(w_l, w_r)$. This method is stable again under the condition (31) with $cfl \in (0, 1]$. Nevertheless, the numerical solutions provided by a Roe method may converge to weak solutions which do not satisfy the entropy criterion. Nevertheless there are some easy entropy-fix techniques available to overcome this drawback (see [25]).

First order conservative methods like Lax-Friedrichs, Roe, or Godunov, may be extended to high-order shock-capturing methods by means of reconstruction operators, ADER techniques or discontinuous Galerkin methods (see [30], [44], [18], [20] and their references).

3 Weak solutions of nonconservative hyperbolic systems

We consider first order quasi-linear PDE systems

$$w_t + \mathcal{A}(w)\, w_x = 0, \quad x \in \mathbb{R},\ t > 0, \tag{40}$$

in which the unknown $w(x,t)$ takes values in an open convex set $\mathcal{O}$ of $\mathbb{R}^N$, and $w \in \mathcal{O} \mapsto \mathcal{A}(w) \in \mathcal{M}_{N\times N}(\mathbb{R})$ is a smooth locally bounded map. The system is supposed to be strictly hyperbolic and the characteristic fields $R_i(w)$, $i = 1,\dots,N$, are supposed to be either genuinely nonlinear:

$$\nabla\lambda_i(w)\cdot R_i(w) \neq 0, \quad \forall\, w \in \mathcal{O},$$

or linearly degenerate:

$$\nabla\lambda_i(w)\cdot R_i(w) = 0, \quad \forall\, w \in \mathcal{O}.$$

Here, $\lambda_1(w),\dots,\lambda_N(w)$ represent the eigenvalues of $\mathcal{A}(w)$ (in increasing order) and $R_1(w),\dots,R_N(w)$ a set of associated eigenvectors.

The formulation (5) of a system of the form (1) is a particular case of (40). Nevertheless in this section we represent the unknown by w instead of W in order to make easier the comparison with the results of the previous section.

A first difficulty to define the weak solutions of system (40) comes from the fact that the usual procedure to obtain a variational formulation (multiply by a regular enough test function and then integrate by parts) does not allow one to pass all the derivatives to the test function.

Let us try the alternative way mentioned in Section 2 for defining weak solutions. First, the integral equation satisfied by a smooth solution in an arbitrary rectangle is obtained: given a rectangle $[a,b]\times[t_0,t_1] \subset \mathbb{R}\times[0,\infty)$, a smooth solution of (40) satisfies the equality

$$\int_a^b w(x,t_1)\,dx = \int_a^b w(x,t_0)\,dx - \int_{t_0}^{t_1}\int_a^b \mathcal{A}(w(x,t))\, w_x(x,t)\,dx\,dt. \tag{41}$$

The idea now is to use this integral equation to define weak solutions. But a new difficulty arises: the integrand of the last term is not defined for discontinuous functions w. At a discontinuity, the product $\mathcal{A}(w)\, w_x$ is expected to produce a Dirac measure whose mass must be related with the jumps of both w and $\mathcal{A}(w)$, but the mathematics of the problem do not determine the expression of such a measure in all the cases. A possible strategy to define weak solutions is based in giving a sense to these integrals for discontinuous functions.

Under some hypotheses of regularity for $\mathcal{A}(w)$, the theory introduced by Dal Maso, LeFloch, and Murat [19] allows one to define the nonconservative product $\mathcal{A}(w)\, w_x$ as a bounded measure for functions w with bounded variation, provided a family of Lipschitz continuous paths $\Phi : [0,1]\times\mathcal{O}\times\mathcal{O} \to \mathcal{O}$ is prescribed, which must satisfy certain regularity and compatibility conditions, in particular

$$\Phi(0; w_l, w_r) = w_l, \qquad \Phi(1; w_l, w_r) = w_r, \tag{42}$$

and

$$\Phi(s; w, w) = w. \tag{43}$$

The interested reader is addressed to [19] for a rigorous and complete presentation of this theory. Here, the family of paths will be just understood as a tool to give a sense to integrals of the form

$$\int_a^b \mathcal{A}(v(x))\, v_x(x)\, dx,$$

for functions v with jump discontinuities. More precisely, given a bounded variation function $v : [a, b] \to \mathbb{R}$, we define:

$$\begin{aligned} \fint_a^b \mathcal{A}(v(x))\, v_x(x)\, dx = \int_a^b &\mathcal{A}(v(x))\, v_x(x)\, dx \\ &+ \sum_m \int_0^1 \mathcal{A}(\Phi(s; v_m^-, v_m^+)) \frac{\partial \Phi}{\partial s}(s; v_m^-, v_m^+)\, ds. \end{aligned} \tag{44}$$

In this definition, v_m^- and v_m^+ represent, respectively, the limits of v to the left and right of its mth discontinuity (remember that the set of discontinuities of a bounded variation function is countable). Observe that, in (44), the family of paths has been used to determine the Dirac measures placed at the discontinuities of v.

According to this definition for the integral, a weak solution of the system can be understood as a bounded variation function satisfying

$$\int_a^b w(x, t_1)\, dx = \int_a^b w(x, t_0)\, dx - \int_{t_0}^{t_1} \fint_a^b \mathcal{A}(w(x, t))\, w_x(x, t)\, dx\, dt, \tag{45}$$

for every $[a, b] \times [t_0, t_1] \subset \mathbb{R} \times [0, \infty)$. It can be shown that, across a discontinuity, weak solutions have to satisfy the generalized Rankine-Hugoniot condition:

$$\xi[w] = \int_0^1 \mathcal{A}(\Phi(s; w^-, w^+)) \frac{\partial \Phi}{\partial s}(s; w^-, w^+)\, ds, \tag{46}$$

where ξ is the speed of propagation of the discontinuity, and w^- and w^+ are the left and right limits of the solution at the discontinuity. If $\mathcal{A}(w)$ is the Jacobian matrix for some function $f(w)$, (46) reduces to the standard Rankine-Hugoniot condition (11).

Again, a weak solution for the initial value problem (40), (4) is not necessarily unique, so a concept of entropy is newly required. It may be given again by an entropy pair (η, g), i.e. a pair of regular functions from $\mathcal{O}$ to $\mathbb{R}$, η being convex, such that

$$\nabla g(w) = \nabla \eta(w) \cdot \mathcal{A}(w), \quad \forall\, w \in \mathcal{O}.$$

A weak solution is said to be an entropy solution if it satisfies the inequality (13). Any smooth solution satisfies the conservation law (12) exactly and it can be verified again that a piecewise continuous solution is an entropy solution if, across a discontinuity, the jump condition (14) is satisfied.

Again, if w_0 has a compact support, by taking $a = -\infty$, $b = \infty$, $t_0 = 0$, and $t_1 = t$ in (45), we obtain:

$$\int_{\mathbb{R}} w(x,t)\,dx = \int_{\mathbb{R}} w_0(x)\,dx - \int_0^\infty \fint_{\mathbb{R}} \mathcal{A}(w(x,t))\,w_x(x,t)\,dx\,dt, \quad \forall t > 0, \quad (47)$$

which is the global conservation property satisfied by weak solutions of (40), (4).

It can be shown again that the Riemann problem (40), (22) has a unique self-similar weak solution (i.e. $w(x,t) = w(x/t; w_l, w_r)$) composed by at most N simple waves (rarefaction waves, contact discontinuities or shock waves) for sufficiently close data w_l and w_r.

Unfortunately, the concept of weak solution depends on the family of paths, which is a priori arbitrary. The crucial question is thus how to choose the 'good' family of paths. In fact, when the hyperbolic system is the vanishing-viscosity limit of the parabolic problems

$$w_t^\epsilon + \mathcal{A}(w^\epsilon)\,w_x^\epsilon = \epsilon(\mathcal{R}(w^\epsilon)w_x^\epsilon)_x, \quad (48)$$

the adequate family of paths should be related to the viscous profiles: let us suppose that the existence of a viscous profile is adopted as a criterion of admissibility for a discontinuity linking the states w^-, w^+ at speed ξ. In this case, a viscous profile is a travelling wave (18) which is a solution of (48) satisfying (19). It can be easily verified that v has to solve now the equation

$$-\xi v' + \mathcal{A}(v)\,v' = (\mathcal{R}(v)\,v')'. \quad (49)$$

By integrating (49) from $-\infty$ to ∞ and taking into account (19), we obtain the jump condition

$$\xi[w] = \int_{-\infty}^{\infty} \mathcal{A}(v(\chi))\,v'(\chi)\,d\chi. \quad (50)$$

Comparing this jump condition with (46), it seems clear that, in this case, the good choice for the path connecting the states w^- and w^+ would be, after a reparameterization, the viscous profile v. For instance, the path

$$\Phi(s; w^-, w^+) = v\left(\tan\big(\pi(s - 1/2)\big)\right), \quad s \in [0,1],$$

would be a natural choice (or any other parameterization: it can be easily verified that the definition of weak solutions is invariant for reparametrizations of the paths). The main difference with the conservative case is that now every choice of viscous term $\mathcal{R}$ leads to different jump conditions, while for conservative systems the Rankine-Hugoniot conditions (11) are always recovered independently of the choice of the viscous term. Unfortunately, the computation of viscous profiles for complex hyperbolic systems is far from being an easy task and thus the computation of a family of paths based on them can be very difficult in practice.

4 Path-conservative methods

Let us denote now by w_i^n the approximation at the cell I_i at time t_n of the cell averages of the exact solution. The initial cell values are given again by (23). From (45), the following equality can be deduced for the cell averages of the exact solution:

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} w(x,t_{n+1})\,dx = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} w(x,t_n)\,dx - \frac{\Delta t}{\Delta x}\frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} \fint_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(w(x,t))\,w_x(x,t)\,dx\,dt.$$

A natural idea would be to design an explicit numerical scheme by a formula such that:

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\fint_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(w^n(x))\,w_x^n(x)\,dx, \tag{51}$$

where w^n represents the piecewise constant function whose value at the cell I_i is the approximation w_i^n. But now the meaning of the weak integral in (51) is ambiguous: as w^n is piecewise constant, its weak integral only consists of the Dirac measures placed at the intercells:

$$\fint_{-\infty}^{\infty} \mathcal{A}(w^n(x))\,w_x^n(x)\,dx = \sum_i \int_0^1 \mathcal{A}(\Phi(s;w_i^n,w_{i+1}^n))\frac{\partial\Phi}{\partial s}(s;w_i^n,w_{i+1}^n)\,ds.$$

What is thus the meaning of the restriction of this integral to the cell I_i? Should the Dirac measure placed in $x_{i+1/2}$ contribute to the weak integral in the cell I_i or to that in its neighbor I_{i+1}? The general idea is to decompose the total mass of this Dirac measure into two summands $D_{i+1/2}^{\pm}$, one contributing to the cell I_i and the other to the cell I_{i+1}. This idea leads to the following definition:

Definition 1 *Given a family of paths Φ, a numerical scheme is said to be Φ-conservative if it can be written in the form:*

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\big(D_{i-1/2}^{n,+} + D_{i+1/2}^{n,-}\big), \tag{52}$$

where

$$D_{i+1/2}^{n,\pm} = D^{\pm}\big(w_{i-q}^n,\ldots,w_{i+p}^n\big),$$

D^- and D^+ being two Lipschitz continuous functions from $\mathcal{O}^{p+q+1}$ to $\mathcal{O}$ satisfying

$$D^{\pm}(w,\ldots,w) = 0, \quad \forall w\in\mathcal{O}, \tag{53}$$

and

$$D^-(w_{-q},\ldots,w_p) + D^+(w_{-q},\ldots,w_p) = \int_0^1 \mathcal{A}\big(\Phi(s;w_0,w_1)\big)\frac{\partial\Phi}{\partial s}(s;w_0,w_1)\,ds, \tag{54}$$

for every set $\{w_{-q},\ldots,w_p\}\subset\mathcal{O}$.

The concept of path-conservative numerical scheme is a generalization of that of conservative scheme: if the system is conservative, that is $\mathcal{A}(w)$ is the Jacobian matrix of $f(w)$, then every path-conservative numerical scheme is equivalent to a conservative scheme with numerical flux

$$\begin{aligned} F(w_{-q},\dots,w_p) &:= D^-(w_{-q},\dots,w_p)+f(w_0) \\ &:= -D^+(w_{-q},\dots,w_p)+f(w_1). \end{aligned}$$

Notice that, due to (54), the two definitions of the numerical flux coincide. Conversely, a conservative numerical scheme is path-conservative for any family of paths: notice that by adding and subtracting $f(w_i)$ in (24), a conservative scheme can be written in the form (52) by definining:

$$\begin{aligned} D^-(w_{-q},\dots,w_p) &:= F(w_{-q},\dots,w_p)-f(w_0), \\ D^+(w_{-q},\dots,w_p) &:= f(w_1)-F(w_{-q},\dots,w_p). \end{aligned}$$

It can be trivially verified using these definitions that (53) and (54) are satisfied for any family of paths.

Path-conservative numerical schemes satisfy a (P1) property: if a Φ-conservative scheme is applied to the initial cell values (23), the equality

$$\int_{\mathbb{R}} w^{n+1}(x)\,dx = \int_{\mathbb{R}} w^n(x)\,dx - \Delta t \fint_{\mathbb{R}} \mathcal{A}(w^n(x))\,w^n_x(x)\,dx, \tag{55}$$

which is the discrete analogous to (47), can be easily obtained by summing up in (52) taking (54) into account. In particular, it can be easily verified that if the nonconservative system (40) has a conservative subsystem, a path-conservative numerical scheme is conservative in the usual sense for that subsystem.

Let us finally show four examples of path-conservative numerical schemes. They extend to nonconservative systems the examples of conservative methods presented in Section 2 in the sense that, if the system is conservative, they are equivalent to their conservative counterpart independently of the choice of the family of paths Φ.

4.1 Centered scheme

The easiest path-conservative method is given by the choice

$$D^{\pm}_C(w_l,w_r)=\frac{1}{2}\int_0^1 \mathcal{A}\big(\Phi(s;w_l,w_r)\big)\frac{\partial\Phi}{\partial s}(s;w_l,w_r)\,ds, \tag{56}$$

but, as expected, the corresponding scheme is unconditionally unstable: some *upwinding* is again required in the definition of $D^{\pm}$.

4.2 Lax-Friedrichs method

We define now

$$D^{\pm}_{LF}(w_l,w_r)=\frac{1}{2}\int_0^1 \mathcal{A}(\Phi(s;w_l,w_r)\frac{\partial\Phi}{\partial s}(s;w_l,w_r)\,ds \pm \frac{\Delta x}{2\Delta t}(w_r-w_l). \tag{57}$$

The corresponding path-conservative scheme is stable under the CFL condition (31) with $cfl \in (0,1]$.

The Rusanov or local Lax-Friedrichs method can also be extended to the nonconservative case:

$$D^{\pm}_{LLF}(w_l,w_r) = \frac{1}{2}\int_0^1 \mathcal{A}(\Phi(s;w_l,w_r)\frac{\partial\Phi}{\partial s}(s;w_l,w_r)\,ds \pm \frac{\alpha(w_l,w_r)}{2}(w_r - w_l), \tag{58}$$

where $\alpha(w_l,w_r)$ is given again by (33).

4.3 Godunov method

We define

$$D^{-}_{G}(w_l,w_r) = \int_0^1 \mathcal{A}(\Phi(s;w_l,w_0))\frac{\partial\Phi}{\partial s}(s;w_l,w_0)\,ds, \tag{59}$$

$$D^{+}_{G}(w_l,w_r) = \int_0^1 \mathcal{A}(\Phi(s;w_0,w_r))\frac{\partial\Phi}{\partial s}(s;w_0,w_r)\,ds, \tag{60}$$

being $w_0 = w(0;w_l,w_r)$ the value at $x = 0$ of the self-similar solution of the Riemann problem (40), (22). It is stable under the condition (31) with $cfl \in (0,1]$. Moreover, if the family of paths satisfies some natural conditions of compatibility with the integral curves of the characteristic fields and the solutions of the Riemann problems (see [32]) and the more strict condition $cfl \in (0,1/2]$ is imposed, the method can be reinterpreted as it was done in the conservative case: w_i^{n+1} is the average (35) of the exact entropy solution of the Cauchy problem

$$\begin{cases} \widetilde{u}_t + \mathcal{A}(\widetilde{u})\,\widetilde{u}_x = 0, & x \in \mathbb{R},\ t > t_n, \\ \widetilde{u}(x,t_n) = w^n(x), & x \in \mathbb{R}, \end{cases} \tag{61}$$

where w^n is the piecewise constant function taking value w_i^n at the cell I_i.

4.4 Roe method

The exact solutions of the Riemann problems are now approximated by those of the linear problems

$$\begin{cases} u_t + \mathcal{A}_\Phi(w_l,w_r)\,u_x = 0, & x \in \mathbb{R},\ t > 0, \\ u(x,0) = \begin{cases} w_l & \text{if } x < 0, \\ w_r & \text{if } x > 0, \end{cases} \end{cases} \tag{62}$$

where now $\mathcal{A}_\Phi(w_l,w_r)$ is a Roe linearization of $\mathcal{A}(w)$ in the sense defined by Toumi in [45], i.e. a function $\mathcal{A}_\Phi\colon \mathcal{O}\times\mathcal{O} \mapsto \mathcal{M}_{N\times N}(\mathbb{R})$ satisfying the following properties:

- for each $w_l, w_r \in \mathcal{O}$, $\mathcal{A}_\Phi(w_l,w_r)$ has N distinct real eigenvalues $\lambda_1(w_l,w_r)$, $\dots$, $\lambda_N(w_l,w_r)$;

- $\mathcal{A}_\Phi(w,w) = \mathcal{A}(w)$, for every $w \in \mathcal{O}$,
- for any $w_L, w_R \in \mathcal{O}$,

$$\mathcal{A}_\Phi(w_l,w_r)\,(w_r - w_l) = \int_0^1 \mathcal{A}(\Phi(s;w_l,w_r))\frac{\partial \Phi}{\partial s}(s;w_l,w_r)\,ds. \tag{63}$$

$\mathcal{A}_\Phi(w_l,w_r)$ is again diagonalizable:

$$\mathcal{A}_\Phi(w_l,w_r) = \mathcal{K}_\Phi(w_l,w_r)\cdot \mathcal{L}_\Phi(w_l,w_r)\cdot \mathcal{K}_\Phi(w_l,w_r)^{-1}.$$

The numerical scheme obtained by solving the linear Riemann problems and averaging their solutions at the cells can be finally written as the path-conservative scheme defined by

$$D_R^{\pm}(w_l,w_r) = \mathcal{A}_\Phi^{\pm}(w_l,w_r)\,(w_r - w_l). \tag{64}$$

Here, the following notation has been used:

$$\mathcal{A}_\Phi^{\pm}(w_l,w_r) = \mathcal{K}_\Phi(w_l,w_r)\cdot \mathcal{L}_\Phi^{\pm}(w_l,w_r)\cdot \mathcal{K}_\Phi(w_l,w_r)^{-1},$$

where $\mathcal{L}_\Phi^{\pm}(w_l,w_r)$ represents the diagonal matrix whose coefficients are the positive or the negative part of those of $\mathcal{L}_\Phi(w_l,w_r)$. Again, this method is stable under the condition (31) with $cfl \in (0,1]$.

If the nonconservative system (40) comes from a PDE system (1), it is always possible to rewrite any of these methods in a closer form to the original formulation of the problem, in which some discretizations of the flux, the source terms, and the nonconservative products appear explicitly. The advantage of this passage by a global formulation is that it ensures the stability under usual CFL conditions and that it makes easier the analysis of the well-balanced properties (see [36], [35]).

Again, first order path-conservative numerical schemes can be extended to high-order either by using reconstruction operators (see [9], [35]), discontinuous Galerkin methods (see [47]) or ADER techniques (see [21]).

5 Convergence Property

Concerning property (P2), the following result can be obtained (see [11]): consider a nonconservative hyperbolic system (40) together with a given family of paths Φ. Consider also a Φ-conservative method. Let $\{w^{\Delta x}\}$ be a sequence of piecewise constant approximate solutions generated by the scheme that satisfies the inequality

$$\sup_{\mathbb{R}} |w^{\Delta x}(\cdot,t)| + TV_{\mathbb{R}}(w^{\Delta x}(\cdot,t)) \le const. \tag{65}$$

uniformly in time. Let us also suppose that $w^{\Delta x}$ converges to a function w *uniformly in the sense of graphs*. Then it can be shown that w is a weak solution of the system according to the family of paths Φ and thus its discontinuities satisfy the jump conditions (46). To ensure that w is also an entropy solution, an extra requirement has to be imposed again to the numerical scheme. The discrete entropy inequality (26) is also enough in this case. In particular, Godunov and Lax-Friedrichs methods satisfy such an inequality under the condition (31) with $cfl \in (0, 1/2]$ (see [32] and [15] respectively) but Roe methods require again the use on an entropy-fix technique.

This theoretical result seems to give a positive answer to the question about whether or not path-conservative schemes satisfy a (P2) property, but in practice this property may fail. To understand the reason, let us briefly recall and illustrate the notion of convergence used to prove the theoretical result: we address the interested readers to [19] for more details.

The family of paths Φ chosen to define the nonconservative products (and the path-conservative method) can also be used to construct what is called the Lipschitz continuous graph completion of a bounded variation function. The idea is as follows: given a function of bounded variation $w : (a, b) \to \mathbb{R}^N$, we define the function $\sigma : (a, b) \to (0, 1)$ by

$$\sigma(x) = \frac{1}{L}\big(x - a + TV_a^x(w)\big),$$

where $TV_a^x(w)$ represents the total variation of w in (a, x) and $L = b - a + TV_a^b(w)$. Let us define $\mathcal{C}(w)$ as the set of points in which w is continuous and $\mathcal{D}(w)$ its complementary set in (a, b). Notice that σ is increasing, left continuous but discontinuous at every point $x \in \mathcal{D}(w)$. $\sigma(x-)$, $\sigma(x+)$, and $[\sigma(x)]$ will denote respectively the limits to the left and to the right and the jump of σ at a point x. We consider next the functions $X : (0, 1) \mapsto (a, b)$ and $W_\Phi : (0, 1) \mapsto \mathbb{R}^N$ defined by the following two conditions:

- for any $x \in \mathcal{C}(w)$,

$$X(\sigma(x)) = x, \quad W_\Phi(\sigma(x)) = w(x);$$

- for any $x \in \mathcal{D}(w)$ and any $s \in [\sigma(x-), \sigma(x+)]$,

$$X(s) = x, \quad W_\Phi(s) = \Phi\left(\frac{s - \sigma(x-)}{[\sigma(x)]}; w(x-), w(x+)\right).$$

The function $V : (0, 1) \mapsto \mathbb{R}^{N+1}$ defined by $V(s) = (X(s), W_\Phi(s))$, which is Lispchitz continuous, is the Φ-graph completion of the function w. It can be understood as a parameterization in $(0, 1)$ of the graph of the function w interpreted as a curve of $\mathbb{R}^{N+1}$. In Figure 2 the graph completion based on the family of straight segments

$$\Phi(s; w_l, w_r) = w_l + s(w_r - w_l)$$

of the discontinuous function w whose graph is depicted in Figure 1 is shown: as it can be seen, in W_Φ the discontinuity is replaced by the path connecting the limits at both sides of the discontinuity.

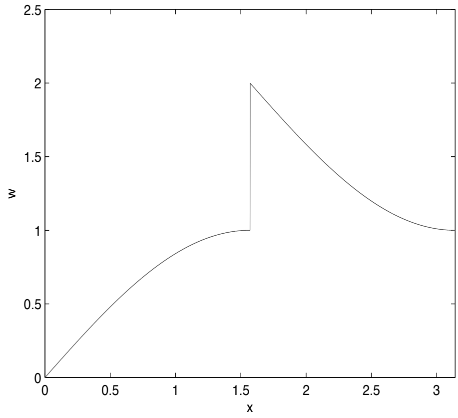


Figure 1: Graph of a bounded variation function w.

According to [5], the distance between two Lipschitz-continuous graphs $V_1, V_2 : (0,1) \mapsto \mathbb{R}^m$ is defined now as follows:

$$\operatorname{dist}(V_1, V_2) = \inf_{\gamma_1, \gamma_2} \left\{ \sup_{s \in (0,1)} |V_1(\gamma_1(s)) - V_2(\gamma_2(s))| \right\},$$

where the infimum is taken over all continuous, nondecreasing and surjective maps $\gamma_1, \gamma_2 : (0,1) \mapsto (0,1)$, i.e. over all the admissible reparameterizations of the graphs.

We can now precisely state the notion of convergence which is required to the numerical solutions to prove the above mentioned theoretical result: it is said that $w^{\Delta x}$ converges to w uniformly in the sense of graphs if

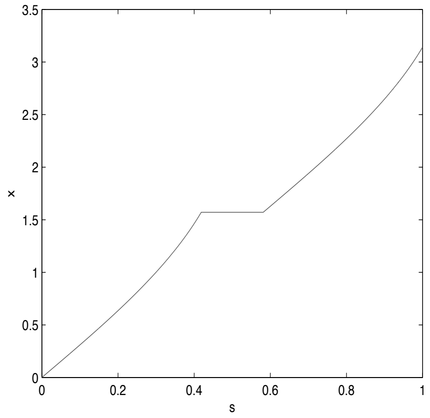
$$\lim_{\Delta x \to 0} \operatorname{dist}\big((X^{\Delta x}, W_\Phi^{\Delta x}), (X, W_\Phi)\big) = 0,$$

where $(X^{\Delta x}, W_\Phi^{\Delta x})$ and (X, W_Φ) are the Φ-completions of $w^{\Delta x}$ and w, respectively.

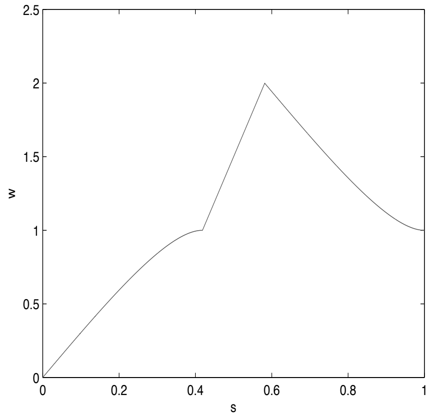
This notion of convergence is weaker than the usual uniform convergence. A simple example of sequence that converges uniformly in the sense of graphs but not in the uniform sense is the sequence of translated step functions

$$w^n(x) = w_l + H(x - y_n)(w_r - w_l),$$

where w_l and w_r are two distinct vectors, $\{y_n\}$ is a sequence that converges to 0 as n tends to ∞, and H is the Heaviside function. The sequence w^n converges



(a) Function X.



(b) Function W_Φ.

Figure 2: Graph completion of w.

uniformly in the sense of graphs to the step function

$$w(x) = w_l + H(x)(w_r - w_l)$$

(see [19] for details).

Nevertheless, even if it is weaker than the uniform convergence, this notion of convergence is still too strong to expect the numerical solutions provided by a finite difference-type scheme to converge in this sense. If the sequence $w^{\Delta x}$ is assumed to converge in a weaker but more realistic sense, the Lax-Wendroff-type result cannot be established for general problems: if, for instance $w^{\Delta x}$ converges almost everywhere to a function w, by extending the arguments in Hou and LeFloch [26] and using the stability results established in [19], it has been proved in [11] that the limit function w solves the following hyperbolic system with source term

$$w_t + [A(w)\, w_x]_\Phi = \nu_w, \tag{66}$$

where ν_w is a bounded measure supported on the discontinuities of w, called the *convergence error measure*. Due to the appearance of this measure, the limit function will be a classical solution of (40) in $\mathcal{C}(w)$ but its discontinuities could not satisfy the generalized Rankine-Hugoniot conditions (46) corresponding to the family of paths Φ.

What are then the Rankine-Hugoniot conditions satisfied at the discontinuities of the limit function? The modified equations are useful to answer this question. For simplicity, let us consider the path-conservative Lax-

Friedrichs scheme defined by (57) which can be also rewritten as follows:

$$\frac{1}{\Delta t}\left(w_i^{n+1} - \frac{1}{2}\left(w_{i-1}^n + w_{i+1}^n\right)\right) + \frac{1}{2\Delta x}\Big(\int_0^1 A\left(\Phi(s; w_{i-1}^n, w_i^n)\right)\frac{\partial \Phi}{\partial s}(s; w_{i-1}^n, w_i^n)\,ds + \int_0^1 A\left(\Phi(s; w_i^n, w_{i+1}^n)\right)\frac{\partial \Phi}{\partial s}(s; w_i^n, w_{i+1}^n)\,ds\Big) = 0.$$

A formal Taylor expansion allows us to show that the numerical solutions solve up to second order the modified equation:

$$w_t + A(w)w_x = \frac{\Delta x^2}{2\Delta t}\Bigg(w_{xx} - \frac{\Delta t^2}{\Delta x^2}\left(A^2(w)w_x\right)_x - \frac{\Delta t^2}{\Delta x^2}\left(DA(w)\left(A(w)w_x, w_x\right) - DA(w)\left(w_x, A(w)w_x\right)\right)\Bigg) - \frac{\Delta x}{2} I_1(w), \quad (67)$$

where

$$I_1(w) = \int_0^1 DA\,(w)\left(D_{w_l}\Phi\cdot w_x, D_{w_l}\Phi_s\cdot w_x\right)\,ds + \int_0^1 DA\,(w)\left(D_{w_r}\Phi\cdot w_x, D_{w_r}\Phi_s\cdot w_x\right)\,ds.$$

The discontinuities of the limit function will then satisfy a generalized Rankine-Hugoniot condition (46) corresponding to a family of paths Ψ related to the viscous profiles of this modified equation. Even if the hyperbolic model is the vanishing viscosity limit of a family of parabolic problems (48) and the family of paths has been constructed on the basis of the viscous profiles of these regularized problems, in general it is not expectable that both the viscous profiles of the modified equation and the regularized problems coincide.

Moreover, as the second order terms of the modified equations depends both on the chosen family of paths Φ (notice that in the case of the Lax-Friedrichs scheme Φ only appears in the term $I_1(w)$) and on the specific form of the viscous terms of the numerical scheme, different numerical schemes based on a same family of paths may produce limit functions satisfying different jump conditions.

In fact, this difficulty potentially affects to any method having some numerical viscosity… which are almost everyone! Random choice methods, as Glimm or the front tracking schemes are viscosity free and, in effect, it can be proved that the numerical solutions provided by these methods converge uniformly in the sense of graphs: see [29]. An example of such a method is the following: once the approximations w_i^n at time t_n have been obtained, the new approximations are given by

$$w_i^{n+1} = \widetilde{u}(x_{i-1/2} + \theta\Delta x, t_{n+1}),$$

where $\widetilde{u}$ is the entropy solution of the problem (61) and θ is a random number in $[0,1]$. If again the CFL condition (31) is imposed with $cfl \in (0,1/2]$, $\widetilde{u}$ can be obtained by solving the Riemann problems at the intercells. Notice that the average stage of Godunov-type methods, which is the source of the numerical viscosity, is avoided. Nevertheless, due to the random stage, this method will not satisfy a global property (P1). Moreover, its implementation may be very difficult or very costly in practice, as the exact solutions of Rieman problems have to be explicitly known.

In certain special situations, the convergence error measure is found to vanish identically. This is the case for systems of balance laws: if the family of paths satisfies some compatibility condition with the integral curves of the linearly degenerate characteristic fields (see [32]), then all of the discontinuities are correctly approximated and the scheme does converge to exact solutions (see [11] for details). To characterize the nonconservative systems for which this convergence error vanishes would be an important issue.

Nevertheless, in general it seems to us a difficult task to construct numerical schemes having both properties (P1) and (P2). (P1), indeed, usually requires an averaging stage. This implies the appearance of certain numerical diffusion, that affects to the viscous profiles captured by the numerical scheme and thus to the jump conditions satisfied by the limits of numerical solutions. In spite of it, as we will see in next section, for some particular problems the numerical results given by a path-conservative scheme may be acceptable even when the convergence error measure is present.

6 Numerical experiments

In this section we summarize the numerical results presented in [11] for the two-layer shallow water system (1), (2), (3) over a flat bottom topography, i.e. H=*const*. This system can be written in the form (5) with:

$$w = \begin{bmatrix} h_1 \\ q_1 \\ h_2 \\ q_2 \end{bmatrix}, \qquad A(W) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -u_1^2 + c_1^2 & 2u_1 & c_1^2 & 0 \\ 0 & 0 & 0 & 1 \\ rc_2^2 & 0 & -u_2^2 + c_2^2 & 2u_2 \end{bmatrix},$$

where $u_i = q_i/h_i$ represents the averaged velocity of the i-th layer and $c_i = \sqrt{gh_i}$, $i = 1, 2$.

The eigenvalues of the system will be denoted by $\lambda_{ext}^- < \lambda_{int}^- < \lambda_{int}^+ < \lambda_{ext}^+$. For density ratios r close to 1 (which is the case in many stratified geophysical flows) one has:

$$|\lambda_{int}^{\pm}| << |\lambda_{ext}^{\pm}|. \tag{68}$$

To give a sense to the nonconservative products, we choose the easy family of straight segments:

$$\Phi(s; w_l, w_r) = w_l + s(w_r - w_l).$$

We have compared the exact and the numerical Hugoniot curves corresponding to one of the internal characteristic fields (i.e. the fields related to the eigenvalues $\lambda_{int}^{\pm}$) using a Roe and a Lax-Friedrichs scheme consistent with this family of paths. We first show a numerical test corresponding to the simulation of an internal dam-break presented in [15]. The axis of the channel is the interval $[0, 10]$. The initial condition is $q_1(x, 0) = q_2(x, 0) = 0$,

$$h_1(x,0) = \begin{cases} 0.6, & \text{if } x < 5, \\ 0.4, & \text{otherwise,} \end{cases}$$

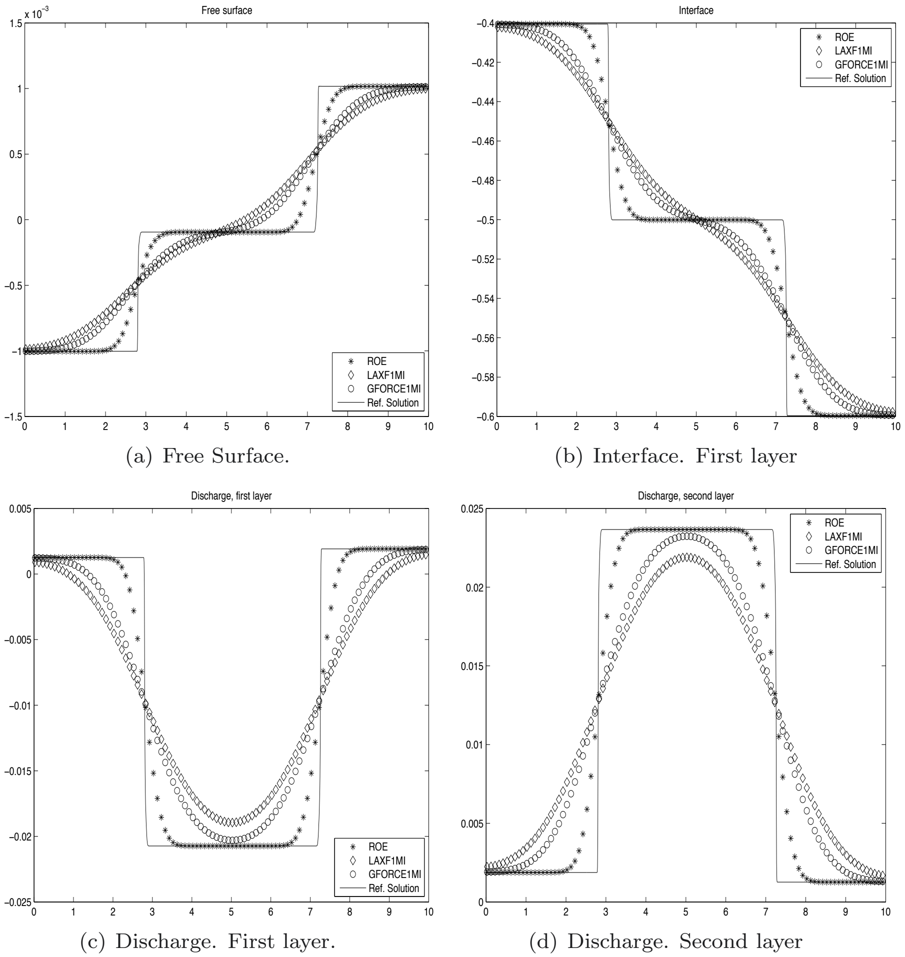
and

$$h_2(x,0) = \begin{cases} 0.4, & \text{if } x < 5, \\ 0.6, & \text{otherwise.} \end{cases}$$

Free boundary conditions are considered. The CFL parameter is set to 0.9. A reference solution is computed with a mesh of 3200 points. Figure 3 shows the comparison of the results obtained with three different first order path-conservative numerical schemes: Roe, Lax-Friedrichs and a nonconservative extension of the GFORCE scheme (see [15] for details). The numerical solutions are compared with the reference solution at time $t = 10$ s: the numerical solutions seem to converge to the same limit function.

Nevertheless, if a more detailed analysis is performed, slight differences between the limits of the different numerical solutions can be found. For instance, in Figure 4 we compare an exact Hugoniot curve with its approximations obtained with Lax-Friedrichs and Roe schemes. To do this, first the Rankine-Hugoniot conditions corresponding to the choice of straight segments have been solved for different values of the speed of the discontinuity ξ for a fixed left state w_l . These computations allow us to draw the Hugoniot curve of states that may be linked to w_l by a 3-entropy shock. Next, the numerical schemes have been applied to solve the Riemann problem corresponding to w_l and any of the right states obtained by solving the jump conditions. In the numerical solutions, the right states of the 3-entropy shock having w_l as left state have been detected. The convergence of these right states as the mesh is refined have been verified. The limit values have been then plotted and compared to the exact Hugoniot curve: see [11] for details.

Due to the inequality (68), the CFL condition adjusts the numerical velocity to the external eigenvalues. As a consequence, the effects of the numerical viscosity are much stronger for internal shocks and thus external shocks are expected to be better captured with Lax-Friedrichs or Roe schemes. In Figure 5 we compare the exact Hugoniot curve with those computed with Lax-Friedrichs and Roe scheme using a mesh with 2000 cells. Note how the curves are now much closer to each other than they were in the previous test case.



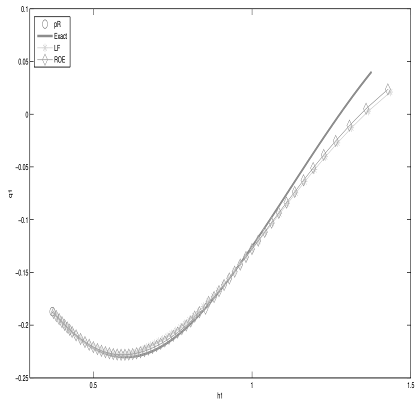
(a) Free Surface.

(b) Interface. First layer

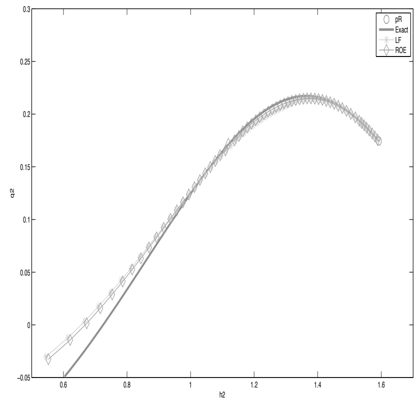
(c) Discharge. First layer.

(d) Discharge. Second layer

Figure 3: Numerical solutions obtained with Roe, Lax-Friedrichs, and GFORCE for an internal dam-brak problem. Comparison with the reference solution at $t = 10$ s.

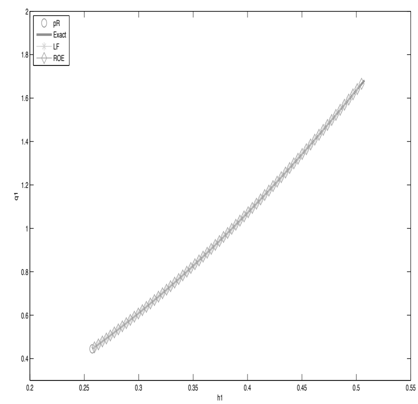


(a) Hugoniot curves (projection onto the plane $h_1 - q_1$): exact (continuous line) and numerical (lines with dots).

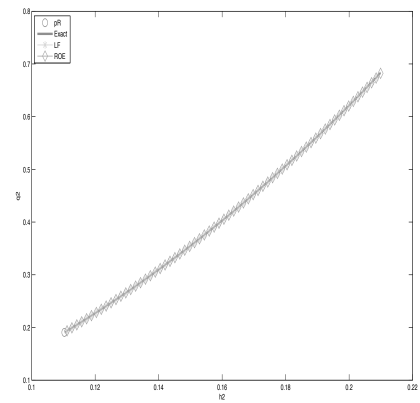


(b) Hugoniot curves (projection onto the plane $h_2 - q_2$): exact (continuous line) and numerical (lines with dots).

Figure 4: Hugoniot curves: exact (continuous line) and numerical (lines with dots).



(a) Hugoniot curves (projection onto the plane $h_1 - q_1$): exact (continuous line) and numerical (lines with dots).



(b) Hugoniot curves (projection onto the plane $h_2 - q_2$): exact (continuous line) and numerical (lines with dots).

Figure 5: Hugoniot curves: exact (continuous line) and numerical (lines with dots).

Some other numerical experiments in which the influence of small changes in the family of paths both in the weak and the numerical solution are studied can be found in [11].

7 Concluding remarks

When a hyperbolic system with nonconservative products and genuinely nonlinear fields is discretized, in order to be sure that the numerical approximations converge to a function which is a classical solution where it is smooth and whose discontinuities are in good agreement with the physics of the problem, the following steps should be taken:

- First, choose a regularization of the system which is consistent with the physics of the problem.
- Next, determine the family of paths consistent with this regularization.
- Finally, design a numerical scheme whose solutions converge to weak solutions associated with this family of paths.

In practice, this strategy may be difficult to follow, since the actual calculation of a family of paths requires calculating viscous profiles. On the other hand, the convergence of the numerical solutions to the correct weak solutions is known for random choices methods only, but these methods do not satisfy good conservation properties and their implementation can be difficult and time consuming since they require the explicit knowledge of the corresponding Riemann solver. In fact, when the nonconservative model under consideration is a simplified version of a more complex (but conservative) model – as it is the case of the two-layer shallow water system, which is obtained by vertically averaging a PDE system composed of two coupled Euler or Navier-Stokes equations – the above strategy may end up being more costly than solving directly the more complex one. In these cases, the use of a numerical strategy based on a direct discretization of the nonconservative system by means of a finite difference scheme which is formally path-consistent is advisable and may have the following advantages:

- The numerical solutions satisfy global properties similar to those satisfied by the exact weak solutions.
- The approximations of the shocks provided by the schemes are consistent with a regularization of the system with higher order terms that vanish as Δx tends to 0.
- This strategy is extendable to high-order methods or to multidimensional problems.

Obviously, the main drawback is that the regularization which is actually used depends both on the chosen family of paths and on the numerical scheme itself. Nevertheless, the following facts should also be considered:

- The convergence error is not present in every nonconservative system.
- Even when it is present, it may only be noticeable for very fine meshes, for discontinuities of great amplitude, and/or for large-time simulations. Otherwise, it may be masked by the discretization errors.
- The convergence error should also be compared with the experimental error: in practice it is very difficult to accurately measure the speed of propagation and the amplitude of shocks in real complex flows.

In the case of the two-layer shallow water system, the shocks captured by Roe scheme and the family of straight segments have been found [10] to be in good agreement with the experimental measurements of internal bores in the Strait of Gibraltar, despite of the simplicity of the family of paths.

Acknowledgments. This research has been partially supported by the Spanish Government Research project MTM2006-08075. The numerical computations have been performed at the Laboratory of Numerical Methods of the University of Málaga.

References

[1] C. Ancey, *Plasticity and geophysical flows: a review*, J. Non-Newt. Fluid Mech. 142 (2007), 4-35.

[2] E. Audusse and M.O. Bristeau, *A well-balanced positivity preserving "second-order" scheme for shallow water flows on unstructured meshes*, J. Comput. Phys. 206 (2005), 311–333.

[3] M.R. Baer and J.W. Nunziato, *A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials*, Int. J. Multiphase Flows 12 (1986), 861–889.

[4] C. Berthon and F. Marche, *A positive preserving high order VFRoe scheme for shallow water equations: a class of relaxation schemes*, SIAM J. Sci. Comput. 30 (2008), 2587–2612.

[5] A. Bressan and F. Rampazzo, *On differential systems with vector-valued impulsive controls*, Boll. Un. Mat. Ital. 7 (1988), 641–656.

[6] S. Bryson and D. Levy, *Balanced central schemes for the shallow water equations on unstructured grids*, SIAM J. Sci. Comput. 27 (2005), 532–552.

[7] M.J. Castro, E. Fernández, A. Ferreiro, J.A. García, and C. Parés, *High order extensions of Roe schemes for two dimensional nonconservative hyperbolic systems*, J. Sci. Comput. 39 (2008), 67–114.

[8] M.J. Castro, J.M. Gallardo, J.A. López, and C. Parés, *Well-balanced high order extensions of Godunov's method for semilinear balance laws*, SIAM J. Num. Anal. 46 (2008), 1012–1039.

[9] M.J. Castro, J.M. Gallardo, and C. Parés, *High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems*, Math. Comput. 75 (2006), 1103–1134.

[10] M.J. Castro, J.A. García, J.M. González, J. Macías, C. Parés, and M.E. Vázquez, *Numerical simulation of two layer shallow water flows through channels with irregular geometry*, J. Comput. Phys. 195 (2004), 202–235.

[11] M.J. Castro, P.G. LeFloch, M.L. Muñoz-Ruiz, and C. Parés, *Why many theories of shock waves are necessary: Convergence error in formally path-consistent schemes*, J. Comput. Phys. 227 (2008), 8107–8129.

[12] M. Castro, J.A. López, and C. Parés. *Finite volume simulation of the geostrophic adjustment in a rotating shallow water system*, SIAM J. Sci. Comput. 31 (2008), 444–477.

[13] M.J. Castro, J. Macías, and C. Parés, *A Q-Scheme for a class of systems of coupled conservation laws with source term. Application to a two-layer 1-D shallow water system*, Math. Mod. Num. Anal. 35 (2001), 107–127.

[14] M.J. Castro, A. Pardo, and C. Parés, *Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique*, Math. Mod. Meth. App. Sci. 17 (2007), 2055–2113.

[15] M.J. Castro, A. Pardo, C. Parés, and E.F. Toro, *On some fast well-balanced first order solvers for nonconservative systems*, submitted to Math. Comp.

[16] C.E. Castro and E.F. Toro, *Solvers for the high-order Riemann problem for hyperbolic balance laws*, J. Comput. Phys. 227 (2008), 2481–2513.

[17] L. Cea, J. Puertas, and M.E. Vázquez-Cendón, *Depth averaged modelling of turbulent shallow water flow with wet-dry fronts*. Arch. Comput. Methods Eng. 14 (2007), 303–341.

[18] B. Cockburn, and C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II: General framework*, Math. Comp. 52 (1989), 411–435.

[19] G. Dal Maso, P.G. LeFloch, and F. Murat, *Definition and weak stability of nonconservative products*, J. Math. Pures Appl. 74 (1995), 483–548.

[20] M. Dumbser, D.S. Balsara, E.F. Toro, and C.D. Munz, *A unified framework for the construction of one-step finite volume and discontinuous Galerkin schemes on unstructured meshes*, J. Comput. Phys. 227 (2008), 8209–8253.

[21] M. Dumbser, C. Enaux, and E.F. Toro, *Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws*, J. Comput. Phys. 227 (2008), 3971–4001.

[22] E. D. Fernández, F. Bouchut, D. Bresch, M. J. Castro, and A. Mangeney, *A new Savage-Hutter type model for submarine avalanches and generated tsunami*, J. Comp. Phys. 227 (2008), 7720-7754.

[23] J.M. Gallardo, C. Parés, and M.J. Castro, *On a well-balanced high-order finite volume scheme for shallow water equations with topography and dry areas*, J. Comput. Phys. (2007), 574–601.

[24] E. Godlewski and P.A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer, 1996.

[25] A. Harten and J.M. Hyman, *Self-adjusting grid methods for one-dimensional hyperbolic conservation laws*, J. Comput. Phys. 50 (1983), 235–269.

[26] T.Y. Hou and P.G. LeFloch, *Why nonconservative schemes converge to wrong solutions: error analysis*, Math. Comp. 62 (1994), 497–530.

[27] P.G. LeFloch, *Shock waves for nonlinear hyperbolic systems in nonconservative form*, Institute for Math. and its Appl., Minneapolis, Preprint # 593, 1989.

[28] P.G. LeFloch, *Hyperbolic systems of conservation laws. The theory of classical and nonclassical shock waves*, Birkhäuser, 2002.

[29] P.G. LeFloch, T.P. Liu, *Existence theory for nonlinear hyperbolic systems in nonconservative form*, Forum Math. 5 (1993), 261–280.

[30] R.J. LeVeque, *Numerical methods for conservation laws*, Birkhäuser, 1992.

[31] D.A. Lyn and M. Altinakar, *St. Venant-Exner equations for near-critical and transcritical flows*, J. Hydraulic Engineering, 128 (2002), 2002.

[32] M.L. Muñoz-Ruiz and C. Parés, *Godunov method for nonconservative hyperbolic systems*, Math. Model. Numer. Anal. 41 (2007), 169–185.

[33] S. Noelle, N. Pankratz, G. Puppo, and J. Natvig, *Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows*, J. Comput. Phys. 213 (2006), 474–499.

[34] S. Noelle, Y. Xing, and C.-W. Shu, *High order well-balanced finite volume WENO schemes for shallow water equation with moving water*, J. Comput. Phys. 226 (2007), 29–59.

[35] C. Parés, *Numerical methods for nonconservative hyperbolic systems: a theoretical framework*, SIAM J. Numer. Anal. 44 (2006), 300–321.

[36] C. Parés and M.J. Castro, *On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to Shallow-Water Systems*, Math. Model. Numer. Anal. 38 (2004), 821– 852.

[37] G.Parker, Y.Fukushima, and H. M. Pantin, *Self-accelerating turbidity currents*, J. Fluid Mech. 171 (1986), 145–181.

[38] M. Pelanti, F. Bouchut, and A. Mangeney, *A Roe-Type Scheme For Two-Phase Shallow Granular Flows Over Variable Topography*, Math. Model. Numer. Anal. 42 (2008) 851– 885.

[39] E.B. Pitman and L. Le, *A two-fluid model for avalanche and debris flows*, Phil. Trans. R. Soc. A 363 (2005), 1573–1601.

[40] G. Puppo and G. Russo, *Central Runge-Kutta schemes for stiff balance laws.* Progress in industrial mathematics at ECMI 2006, 226–230, Math. Ind., 12, Springer, Berlin, 2008.

[41] S. Rhebergen, O. Bokhove, and J.J.W. van der Vegt, *Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations*, J. Comput. Phys. 227 (2008), 1887–1922.

[42] G. Russo, *Central schemes for balance laws*, Int. Ser. Numer. Math. 140, 141, Birkhäuser, Basel, 2001.

[43] P.A. Tassi, S. Rhebergen, C.A. Vionnetb, and O. Bokhove, *A discontinuous Galerkin finite element model for river bed evolution under shallow flows*, Comp. Meth. Appl. Mech. Engrg. 197 (2008), 2930–2947.

[44] V.A. Titarev and E.F. Toro, *ADER: arbitrary high order Godunov approach*, J. Sci. Comput., 17 (2002), 609–618.

[45] I. Toumi, *A weak formulation of Roe's approximate Riemann solver*, J. Comput. Phys. 102 (1992), 360–373.

[46] Y. Xing and C.-W. Shu, *High order finite difference WENO schemes with the exact conservation property for the shallow water equations*, J. Comput. Phys. 208 (2005), 206–227.

[47] Y. Xing and C.-W. Shu, *High order well-balanced finite volume WENO schemes and discontinuous Galerkin methods for a class of hyperbolic systems with source terms*, J. Comput. Phys. 214 (2006), 567–598.

WORKSHOP IBEROAMERICANO DE MATEMÁTICAS APLICADAS CHILLÁN (CHILE), 2–4 DE DICIEMBRE DE 2008

R.C. CABRALES, L. FRIZ, M. ROJAS-MEDAR

Departamento de Ciencias Básicas, Universidad del Bío-Bío, Chile.

rcabrale@ubiobio.cl, lfriz@ubiobio.cl, marko@ueubiobio.cl

Estimado lector:

En este volumen se incluyen algunos de los textos de las conferencias presentadas en la segunda versión del “Workshop Iberoamericano de Matemáticas Aplicadas”, realizado en la ciudad de Chillán (Chile), del 2 al 4 de Diciembre de 2008.

En dicho evento participaron investigadores de varias universidades de Brasil, Chile y España, junto a estudiantes de Pre-Grado interesados por las Matemáticas y sus aplicaciones. En esta segunda ocasión, hemos preservado el espiritu de la primera versión: un ambiente muy distendido y agradable, propicio para el intercambio de información y debatir sobre problemas con origen en teoría del control, ecuaciones en derivadas parciales no lineales, teoría fuzzy, análisis numérico, etc.

Para el comité organizador, la celebración de este evento ha sido una tarea ardua, pero muy gratificante. Muchas son las personas y entidades a las que tenemos que agradecer por su apoyo. Sin ellas no hubiesemos alcanzado el éxito en los objetivos trazados. Nombrarlas a todas es una misión en extremo delicada debido al riesgo de omitir algún nomber. Sin embargo, nos parece justo mencionar de manera especial al profesor Enrique Fernández Cara, quien ha contribuido con recursos y un entusiasmo sin igual. De igual forma, queremos agradecer a los siguientes proyectos y entidades, quienes brindaron su apoyo y financiamiento:

1. Proyectos Fondecyt-Chile números 11060400 y 1080628.
2. Proyecto de cooperación internacional Fondecyt-Chile número 7080074,
3. Proyecto de Investigación número MTM2006-07932 del Ministerio de Educación y Ciencias de España.
4. Departamento de Ciencias Básicas de la Universidad del Bío-Bío,
5. Facultad de Ciencias de la Universidad del Bío-Bío,
6. Vicerrectoría Académica de la Universidad del Bío-Bío,

7. Programa de Magister en Enseñanza de la Ciencia de la Universidad del Bío-Bío.

Los trabajos seleccionados en este volumen son los siguientes:

1. Mathematical and numerical analysis for reaction-diffusion systems modeling the spread of early tumors, V. Anaya, M. Bendahmane, M. Sepúlveda.
2. Existence of relaxed weak solutions of a generalized Boussinesq system with restriction on the state variables, J.L. Boldrini, M. Rojas-Medar, M. Santos da Rocha.
3. Multiresolution simulation of reaction-diffusion systems with strong degeneracy, R. Bürger, R. Ruiz-Baier.
4. Un problema extremal para un conductor de dos fases en una bola, C. Conca, R. Mahadevan, L. Sanz.
5. Método de elementos finitos para la aproximacin de un modelo de cristales líquidos nemáticos, F. Guillén-González, J.V. Gutiérrez-Santacreu.
6. Stationary asymmetric fluids and Hodge operator, I. Kondrashuk, E.Ñotte-Cuello, M. Rojas-Medar.

R.C. Cabrales, L. Friz, M. Rojas-Medar (Comité Organizador)

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 55–62

MATHEMATICAL AND NUMERICAL ANALYSIS FOR REACTION-DIFFUSION SYSTEMS MODELING THE SPREAD OF EARLY TUMORS

V. ANAYA, M. BENDAHMANE, M. SEPÚLVEDA

vanaya@ing-mat.udec.cl,mostafa_bendahmane@yahoo.fr,
mauricio@ing-mat.udec.cl

Abstract

We prove existence results for a reaction-diffusion system modeling the spread of early tumors. The existence result is proved by the Faedo-Galerkin method, a priori estimates and the compactness method. Moreover, we construct a finite volume scheme to our model. Finally, some numerical simulations are reported.

Key words: *Reaction-diffusion system, weak solution, existence, finite volume scheme.*

AMS subject classifications: *35K57 35K55 92B05*

1 Introduction

During the past several years mathematical models have been applied to various aspects of cancer dynamics, in particular avascular and vascular tumor growth, invasion, angiogenesis and metastasis. A better understanding of the dynamics of tumor formation can be expected from a mathematical modelling. Mathematical models could improve the clinical outcome by predicting the results of specific forms of treatment administered at specific time points. During the last decade theoreticians have developed a great variety of tumour models covering various morphological and functional aspects of tumour growth. These advances have been recently reviewed [6, 7, 8] with a focus on the classification of mathematical tools and computational algorithms.
To state our model, we consider a thin sheet of cells in a physical domain $\Omega \subset \mathbb{R}^3$ over a time span $(0,T)$, $T>0$. Let $u=u(x,t)$ and $v=v(x,t)$ represent the densities of cells and growth factor molecules respectively, at time $t \in (0,T)$ and location $x \in \Omega$. Besides, it is assumed that cell proliferation is influenced by the growth factor, which is produced by the medium and bound by cells. A prototype of a nonlinear system that governs the spreading of early tumors in

a spatial domain is the following reaction-diffusion system:

$$\begin{cases} \partial_t u - d_u \Delta u = F(u,v), & \text{in} \quad Q_T, \\ \partial_t v - d_v \Delta v = G(u,v), & \text{in} \quad Q_T, \\ \nabla u \cdot \vec{\eta} = 0, \quad \nabla v \cdot \vec{\eta} = 0, & \text{on} \quad \partial\Omega, \\ u(\cdot,0) = u_0(\cdot) \quad \text{and} \quad v(\cdot,0) = v_0(\cdot), & \text{in} \quad \Omega. \end{cases} \tag{1}$$

where $Q_T := \Omega \times (0,T), T > 0$ is a fixed time, and $\Omega \subset \mathbb{R}^3$ is a bounded domain with smooth boundary $\partial\Omega$ and outer unit normal $\vec{\eta}$. Herein, $d_u > 0$ and $d_v > 0$ are diffusion constants given. The nonlinearities F and G take the form:

$$\begin{aligned} F(u,v) &= \left(\frac{a_1 (uv)^b}{1 + (uv)^b} - \gamma \right) u, \\ G(u,v) &= h_0 \beta_1 \frac{u^{e-1}}{1 + r_1 u^s} - v \left(\alpha + h_0 \alpha \frac{u^e}{1 + r_1 u^s} + \frac{a_1 (uv)^b}{1 + (uv)^b} - \gamma \right), \end{aligned}$$

In this work we assume that the coefficients $a_1, h_0, \beta_1, \alpha, \gamma, s$ and e satisfy

$$a_1, b, \gamma, h_0, \beta_1, r_1, \alpha > 0 \quad \text{and } 1 \le e \le s. \tag{2}$$

In [5], the authors, Marciniak-Czochra and Kimmel, said that the model (the system (1) with $d_u = 0$) can be thought of as representing an early stage of tumor evolution and destabilizacion of the equilibrium in such system represents an initial invasion of cancer. The authors were looking for a transition from a slightly perturbed equilibrium state to uncontrolled and irregular growth. The idea of their paper was to understand how cellular dynamics of tumor cells can generate pattern formation which may be phenomenologically observed at the macroscopic scale. In [2], the authors presented the survey of different models and methods dealing with multiscale modeling of tumor evolution. We mention also that in [4], the authors have shown the derivation of the macroscopic reaction-diffusion models describing the interplay between the cellular dynamics and the signaling molecules diffusing in the intercellular space using homogenization methods of functional analysis.
Before stating our main results, we give the definition of a weak solution.

Definition 1 *A weak solution of (1) is a pair (u,v) of nonnegative functions such that, $u, v \in L^2(0,T; H^1(\Omega)) \cap C([0,T]; L^2(\Omega))$, $\partial_t u, \partial_t v \in L^2(0,T,(H^1(\Omega))')$, $u(0) = u_0$, $v(0) = v_0$ a.e. in Ω, $F(u,v), G(u,v) \in L^2(Q_T)$, satisfying*

$$\begin{aligned} \int_0^T \langle \partial_t u, \varphi_1 \rangle \, dt + d_u \int\int_{Q_T} \nabla u \cdot \nabla \varphi_1 \, dx \, dt &= \int\int_{Q_T} F(u,v) \varphi_1 \, dx \, dt, \\ \int_0^T \langle \partial_t v, \varphi_2 \rangle \, dt + d_v \int\int_{Q_T} \nabla v \cdot \nabla \varphi_2 \, dx \, dt &= \int\int_{Q_T} G(u,v) \varphi_2 \, dx \, dt, \end{aligned}$$

for all $\varphi_1, \varphi_2 \in L^2(0,T; H^1(\Omega))$. Here, $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $H^1(\Omega)$ and $(H^1(\Omega))'$.

Our main result is the following existence theorem for weak solutions.

Theorem 1 *Assume conditions (2) holds. If $u_0, v_0 \in L^2(\Omega)$, then the system (1) possesses at least one weak solution.*

We prove existence of solutions to the system (1) by applying the Faedo-Galerkin method, deriving apriori estimates, and then passing to the limit in the approximate solutions using compactness arguments.

Now we discretize our problem (1). This description follows the framework of [3]. We let Ω be an open bounded polygonal connected subset of $\mathbb{R}^3$ with boundary $\partial\Omega$. Let $\mathcal{T}$ be an admissible mesh of the domain Ω consisting of open and convex polygons called control volumes with diameter h. Let us denote by x_K the center of K, $N(K)$ the set of the neighbors of K. Furthermore, for all $L \in N(K)$ denote by $d(K,L)$ the distance between x_K and x_L, by $\sigma_{K,L}$ the interface between K and L, by $\eta_{K,L}$ the unit normal vector to $\sigma_{K,L}$ outward to K. We denote by $m(K)$ the measure of K. We also need some regularity on the mesh:

$$\min_{K\in\mathcal{T}, L\in N(K)} \frac{d(K,L)}{\text{diam }(K)} \geq \alpha \text{ for some } \alpha \in (0,\infty).$$

We denote by $\mathcal{D}$ an admissible discretization of Q_T, which consists of an admissible mesh of Ω, a time step $\Delta t > 0$, and a positive number N chosen as the smallest integer such that $N\Delta t \geq T$. We set $t^n = n\Delta t$ for $n \in \{0, \dots, N\}$. With the notation introduced above, finite volume discretization of our model takes the following form: Find vectors $(u_K^n)_{K\in\mathcal{T}}$ and $(v_K^n)_{K\in\mathcal{T}}$ for $n \in \{0, \dots, N\}$, such that for all $K \in \mathcal{T}$ and $n \in \{0, \dots, N-1\}$

$$u_K^0 = \frac{1}{m(K)} \int_K u_0(x)\,dx, \quad v_K^0 = \frac{1}{m(K)} \int_K v_0(x)\,dx, \tag{3}$$

$$m(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} - d_u \sum_{L\in N(K)} \frac{m(\sigma_{K,L})}{d(K,L)}(u_L^{n+1} - u_K^{n+1}) - m(K)F(u_K^n, v_K^n) = 0, \tag{4}$$

$$m(K)\frac{v_K^{n+1} - v_K^n}{\Delta t} - d_v \sum_{L\in N(K)} \frac{m(\sigma_{K,L})}{d(K,L)}(v_L^{n+1} - v_K^{n+1}) - m(K)G(u_K^n, v_K^n) = 0. \tag{5}$$

2 Existence of weak solution

Consider the following spectral problem: Find $w \in H^1(\Omega)$ and a number λ such that

$$\begin{cases} (\nabla w, \nabla\varphi)_{L^2(\Omega)} = \lambda(w,\varphi)_{L^2(\Omega)}, & \forall\varphi \in H^1(\Omega), \\ \nabla w \cdot \eta = 0, & \text{on } \partial\Omega. \end{cases} \tag{6}$$

The problem (6) possesses a sequence of eigenvalues $\{\lambda_l\}_{l=1}^\infty$ and the corresponding eigenfunctions form a sequence $\{e_l\}_{l=1}^\infty$ that is orthogonal in $H^1(\Omega)$ and orthonormal in $L^2(\Omega)$. Furthermore, we assume without loss of generality that $\lambda_1 = 0$.

We look for finite dimensional approximate solution to the problem (1) as sequences $(u_n)_{n>1}$, $(v_n)_{n>1}$ defined for $t \geq 0$ and $x \in \overline{\Omega}$ by

$$u_n(t,x) = \sum_{l=1}^{n} b_{n,l}(t)e_l(x), \qquad v_n(t,x) = \sum_{l=1}^{n} c_{n,l}(t)e_l(x). \tag{7}$$

The next step is to determine the coefficients $(b_{n,l}(t))_{l=1}^n$, $(c_{n,l}(t))_{l=1}^n$ such that for $k = 1, \ldots, n$

$$\begin{aligned} (\partial_t u_n, e_k)_{L^2(\Omega)} + d_u \int_\Omega \nabla u_n \cdot \nabla e_k \, dx\, dt &= \int_\Omega F(u_n, v_n) e_k \, dx, \\ (\partial_t v_n, e_k)_{L^2(\Omega)} + d_v \int_\Omega \nabla v_n \cdot \nabla e_k \, dx\, dt &= \int_\Omega G(u_n, v_n) e_k \, dx, \end{aligned} \tag{8}$$

and, with reference to the initial conditions.

By our choice of basis, u_n and v_n satisfy the boundary condition in (1). Observe that since $u_0, v_0 \in L^2(\Omega)$, it is clear that, as $n \to \infty$, $u_{0,n} \to u_0$ and $v_{0,n} \to v_0$ in $L^2(\Omega)$, respectively. Using the normality of the basis, we can write (8) as a system of ordinary differential equations.

Let $\mathcal{F}$ and $\mathcal{G}$ be functions defined as follow:

$$\begin{aligned} \mathcal{F}(t, (b_{n,l}(t))_{l=1}^n, (c_{n,l}(t))_{l=1}^n) &:= \int_\Omega F(u_n, v_n) e_k \, dx - d_u \int_\Omega \nabla u_n \cdot \nabla e_k \, dx, \\ \mathcal{G}(t, (b_{n,l}(t))_{l=1}^n, (c_{n,l}(t))_{l=1}^n) &:= \int_\Omega G(u_n, v_n) e_k \, dx - d_v \int_\Omega \nabla v_n \cdot \nabla e_k \, dx. \end{aligned}$$

Proceeding exactly as in [1], we prove that $\mathcal{F}$ and $\mathcal{G}$ are Caratheodory functions and the existence interval $[0, t')$ for the Faedo-Galerkin solutions u_n and v_n defined by (7).
To prove global existence of the solutions we derive n-independent a priori estimates bounding u_n, v_n in various Banach spaces. Given some continuous coefficientes $d_{1,n,l}(t)$ and $d_{2,n,l}(t)$, we form the functions $\varphi_{1,n}(t,x) := \sum_{l=1}^n d_{1,n,l}(t)e_l(x)$ and $\varphi_{2,n}(t,x) := \sum_{l=1}^n d_{2,n,l}(t)e_l(x)$. Now our Faedo-Galerkin solutions satisfy the following weak formulations:

$$\int_\Omega \partial_s u_n \varphi_{1,n} \, dx + d_u \int_\Omega \nabla u_n \cdot \nabla \varphi_{1,n} \, dx = \int_\Omega F(u_n, v_n) \varphi_{1,n} \, dx, \tag{9}$$

$$\int_\Omega \partial_s v_n \varphi_{2,n} \, dx + d_v \int_\Omega \nabla v_n \cdot \nabla \varphi_{2,n} \, dx = \int_\Omega G(u_n, v_n) \varphi_{2,n} \, dx. \tag{10}$$

Lemma 2 *There exist constants $c_1, c_2 > 0$ not depending on n such that for $t \in [0, t')$*

$$\begin{aligned} \|u_n\|_{L^2(0,t;H^1(\Omega))} + \|v_n\|_{L^2(0,t;H^1(\Omega))} &\leq c_1, \\ \|\partial_s u_n\|_{L^2(0,t;(H^1(\Omega))')} + \|\partial_s v_n\|_{L^2(0,t;(H^1(\Omega))')} &\leq c_2. \end{aligned}$$

Proof. Substituting $\varphi_{1,n} = u_n$ in (9), we get

$$\frac{1}{2}\frac{d}{ds}\int_\Omega |u_n|^2\,dx + d_u\int_\Omega |\nabla u_n|^2\,dx \le (a_1+\gamma)\int_\Omega |u_n|^2\,dx. \tag{11}$$

Using the nonnegativity of the second term of the left-hand side (11) and the Gronwall's inequality, we obtain

$$\int_\Omega |u_n(x,s)|^2\,dx \le c_3 \text{ for all } s\in(0,t], \tag{12}$$

for some constant $c_3 > 0$. Integrating (11) over $(0,t)$ and using (12), we obtain:

$$\int_\Omega |u_n(x,t)|^2\,dx + d_u\int_0^t\int_\Omega |\nabla u_n|^2\,dx \le c_4,$$

for some constant $c_4 > 0$, which proves that $\|u_n\|_{L^2(0,t;H^1(\Omega))} \le c_5$, where $c_5 > 0$ is a constant independent of n.

Reasoning along the same lines as for u_n, substituting $\varphi_{2,n} = v_n$ in (10), we get $\|v_n\|_{L^2(0,t;H^1(\Omega))} \le c_{10}$, for some constant $c_{10} > 0$ independent of n.

Now, we let $\varphi_1 \in L^2(0,t;H^1(\Omega))$. Using the weak formulation (9), we obtain

$$\begin{aligned}
\left|\int_0^t \langle \partial_s u_n, \varphi_1\rangle\,ds\right| &\le \left|\int_0^t\int_\Omega F(u_n,v_n)\varphi_1\,dx\,ds\right| + \left|d_u\int_0^t\int_\Omega \nabla u_n\cdot\nabla\varphi_1\,dx\,ds\right|\\
&\le (a_1+\gamma)\left(\int_0^t\int_\Omega |u_n|^2\,dx\,ds\right)^{1/2}\left(\int_0^t\int_\Omega |\varphi_1|^2\,dx\,ds\right)^{1/2}\\
&\quad + d_u\left(\int_0^t\int_\Omega |\nabla u_n|^2\,dx\,ds\right)^{1/2}\left(\int_0^t\int_\Omega |\nabla\varphi_1|^2\,dx\,ds\right)^{1/2}\\
&\le c_{11}\|\varphi_1\|_{L^2(0,T;H^1(\Omega))},
\end{aligned}$$

where we have used Hölder's inequality. This implies

$$\|\partial_t u_n\|_{L^2(0,t;(H^1(\Omega))')} \le c_{12}. \tag{13}$$

Reasoning along the same lines for u_n, yields (13) for v_n. $\square$

The next is to show that the local solution constructed above can be extended to the whole time interval $[0,T)$ (independent of n) but this can be done as in [1], so we omit the details.

We have the following result:

Lemma 3 *The solution (u_n, v_n) of the system (1) is nonnegative.*

Proof. The proof of Lemma 3 is based on the choice of test functions $\varphi_{1,n} = -u_n^-$, $\varphi_{2,n} = -v_n^-$ in (9) and (10), respectively, where $u_n^- = \max(0,-u_n)$ and $v_n^- = \max(0,-v_n)$. Then integrating over $(0,t)$ with $0 < t \le T$, we obtain

$$\begin{aligned}
\frac{1}{2}\int_\Omega |u_n^-(x,t)|^2\,dx &+ d_u\int_0^t\int_\Omega |\nabla u_n^-|^2\,dx\,dt\\
&= \int_\Omega |u_n^-(x,0)|^2\,dx - \int_0^t\int_\Omega F(u_n,v_n)u_n^-\,dx\,dt \le 0,
\end{aligned}$$

and

$$\frac{1}{2}\int_\Omega \left|v_n^-(x,t)\right|^2 \,dx + d_v \int_0^t \int_\Omega \left|\nabla v_n^-\right|^2 \,dx\,dt$$
$$= \int_\Omega \left|v_n^-(x,0)\right|^2 \,dx - \int_0^t \int_\Omega G(u_n, v_n)u_n^- \,dx\,dt \le 0,$$

where we have used the nonnegativity of the initial condition (u_0, v_0). This implies that $u_n^- = 0$ and $v_n^- = 0$ a.e. in Ω. $\square$

From Lemma 2 we have that $(u_n)_{n>1}$, $(v_n)_{n>1}$, are bounded in $L^2(0,T;H^1(\Omega))$ and $(\partial_t u_n)_{n>1}$, $(\partial_t v_n)_{n>1}$, are bounded in $L^2(0,T;(H^1(\Omega))')$. Therefore, by standard compactness results [9] we can extract subsequences, which we do not relabel and we can assume that there exist limit functions u, v such that as $n \to \infty$

$$\begin{aligned}
&u_n \to u, v_n \to v \text{ strongly in } L^2(Q_T) \text{ and a.e. in } Q_T,\\
&u_n \rightharpoonup u, v_n \rightharpoonup v \text{ weakly in } L^2(0,T;H^1(\Omega)),\\
&\partial_t u_n \rightharpoonup \partial_t u, \partial_t v_n \rightharpoonup \partial_t v \text{ weakly in } L^2(0,T;(H^1(\Omega))')\\
&F(u_n,v_n) \to F(u,v), G(u_n,v_n) \to G(u,v) \text{ a.e. in } Q_T.
\end{aligned} \tag{14}$$

The following lemma is a consequence of (14) and Vitali's theorem.

Lemma 4 *As $n \to \infty$, $F(u_n, v_n)$ and $G(u_n, v_n)$ converge strongly to $F(u, v)$ and $G(u, v)$, respectively, in $L^q(Q_T)$ for all $1 \le q \le 2$.*

Finally, we prove that the limits u and v in (14) obey the initial data in (1) but this can be done as Lemma 5.3 in [1], so we omit the details..

Keeping in mind (14) and Lemma 4, and using the following weak formulation:

$$\int_0^T \langle \partial_t u_n, \varphi_1 \rangle \,dt + d_u \int\int_{Q_T} \nabla u_n \cdot \nabla \varphi_1 \,dx\,dt = \int\int_{Q_T} F(u_n, v_n)\varphi_1 \,dx\,dt,$$

$$\int_0^T \langle \partial_t v_n, \varphi_2 \rangle \,dt + d_v \int\int_{Q_T} \nabla v_n \cdot \nabla \varphi_2 \,dx\,dt = \int\int_{Q_T} G(u_n, v_n)\varphi_2 \,dx\,dt,$$

for all $\varphi_1, \varphi_2 \in L^2(0,T;H^1(\Omega))$, we can let $n \to \infty$ and obtain a weak solution.

3 Numerical Results

We now show some numerical experiments with the scheme proposed. The test problem is the system (1) in the square domain $\Omega =]0,1[\times]0,1[$. We consider a uniform mesh given by a Cartesian grid with $N_x \times N_y$ control volumes and choosing $N_x = N_y = 700$ ($\Delta x = \Delta y = 1/700$). Obviously, it is possible to consider also unstructured meshes, but for simplicity we will use uniform mesh. For our simulations, we use a time step $\Delta t = 0.02$.

Let us precise the initial condition

$$\begin{aligned} u(x,y,0) &= u^*(1+\epsilon_u \cos(n\pi x)\cos(m\pi y)), \quad (x,y)\in\Omega \\ v(x,y,0) &= v^*(1+\epsilon_v \cos(n\pi x)\cos(m\pi y)), \quad (x,y)\in\Omega \end{aligned}$$

with $u^* = 2.711874407, v^* = 0.3542822338, \epsilon_u = 0.1, \epsilon_v = 0.006, n = m = 4.$

Parameters	$a_1 = 1/12, b = 2, \gamma = 0.04, h_0 = 1, \beta_1 = 1, r_1 = 10/3$ $s = 2, \alpha = 0.1, e = 1$

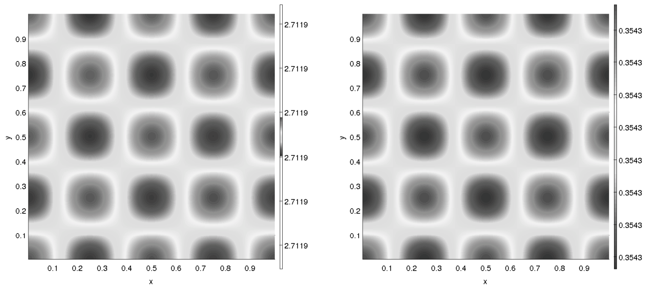


Figure 1: Patterns for cells (left) and growth factor molecules (rigth), with $d_u = 0.005, d_v = 0.0001, T = 200, nT = 10000.$

Now we take the following initial condition and the parameters given before.

$$\begin{aligned} u(x,y,0) &= u^* + \epsilon_u * \omega_u, \quad (x,y)\in\Omega \\ v(x,y,0) &= v^*, \qquad\qquad (x,y)\in\Omega \end{aligned}$$

with $u^* = 2.711874407, v^* = 0.3542822338, \epsilon_u = 0.001, \omega_u \in [0,1]$ is random variable.

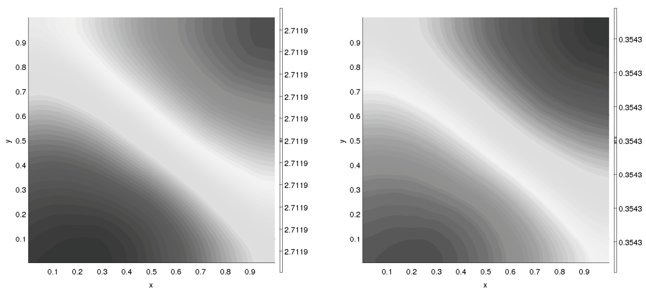


Figure 2: Patterns for cells (left) and growth factor molecules (rigth), with $d_u = 0.005, d_v = 0.0001, T = 200, nT = 10000.$

Acknowledgment

This work has been supported by FONDECYT projects 1070682, 1070694, FONDAP and BASAL projects CMM, Universidad de Chile, and CI^2MA, Universidad de Concepción, and CONICYT fellowship.

References

[1] M. Bendahmane and K. H. Karlsen. Analysis of a class of degenerate reaction-diffusion systems and the bidomain model of cardiac tissue. Networks and Heterogeneous Media, **1**(1):185-218, 2006.

[2] N. Bellomo, E. De Angelis, and L. Preziosi, Multiscale modeling and mathematical problems related to tumor evolution and medical therapy, J. Theor. Med., **5**, (2003), 111–136.

[3] R. Eymard, Th. Gallouët, and R. Herbin. Finite volume methods.In: Handbook of Numerical Analysis, vol. VII, North-Holland, Amsterdam, 2000

[4] A. Marciniak-Czochra, and M. Kimmel *Mathematical model of tumor invasion along linear or tubular structures*, Mathematical and Computer Modelling, **41**, (2005), 1097–1187.

[5] A. Marciniak-Czochra, and M. Kimmel *Reaction-diffusion approach to modelling of the spread of early tumors along linear or tubular structures*, Journal of Theoretical Biology, **244**, (2007), 375–387.

[6] J. Moreira and A. Deutsch *Cellular automaton models of tumor development: A critical review.*, Adv. Compl. Syst., **5**(2-3), (2002), 247–267.

[7] J. D. Murray, Mathematical Biology I: An Introduction, 3rd edition, Springer, 2003.

[8] L. Preziozi, *Cancer modelling and simulation.*, Chapman & Hall/CRC Mathematical Biology & Medicine, 2003

[9] J. Simon, *Compact sets in the space* $L^p(0,T;B)$, Ann. Mat. Pura Appl. (4), **146** (1987), 65–96.

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 63–72

EXISTENCE OF RELAXED WEAK SOLUTIONS OF A GENERALIZED BOUSSINESQ SYSTEM WITH RESTRICTION ON THE STATE VARIABLES

J.L. BOLDRINI[1], M.A. ROJAS-MEDAR[2], M. SANTOS DA ROCHA[3] *

[1]Universidade Estadual de Campinas, Brasil
[2]Universidad del Bío-Bío, Chile
[3]Universidade Estadual de Londrina, Brasil

boldrini@ime.unicamp.br, marko@ueubiobio.cl

Abstract

We analyze the existence of relaxed weak solutions of a generalized Boussinesq model for thermal convection, which allows temperature dependence for both the viscosity and heat diffusion coefficient; the problem is restricted in the sense that the state variables (i.e. velocity and temperature) are required to stay in given closed convex sets of certain functional spaces.

Key words: *Generalized Boussinesq equations, restrictions on the state variables, differential variational inequalities, weak solutions.*
AMS subject classifications: *35K85, 35Q35, 76D03.*

1 Introduction

The classical case for the Boussinesq equations for thermal convection assumes constant viscosity ν and heat diffusion coefficient κ. However, there are fluids for which this is not a good approximation and both the viscosity and heat diffusion coefficient vary with the temperature. and this is very important for the flow behavior.

The generalized Boussinesq equations model such situations and are the following.

$$\mathbf{u}_t - \text{ div } (\nu(\theta)\nabla\mathbf{u}) + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla\pi - \alpha\theta\mathbf{g} = \tag{1}$$

$$\text{div } \mathbf{u} = 0, \tag{2}$$

$$\theta_t - \text{ div } (k(\theta)\nabla\theta) + \mathbf{u}\cdot\nabla\theta = 0, \tag{3}$$

*In memoriam

with initial and boundary conditions:

$$\mathbf{u}(0,\cdot)=\mathbf{u}_0(\cdot),\quad \theta(0,\cdot)=\theta_0(\cdot),$$

$$\mathbf{u}(t,\cdot)=\mathbf{0},\quad \theta(t,\cdot)=T(t,\cdot) \text{ on } (0,T)\times\partial\Omega,$$

where $\Omega\subseteq\mathbb{R}^3$ is a domain bounded in one direction. $\nu(\cdot)$ y $k(\cdot)$ are assumed to be C^1-functions satisfying:

$$\begin{aligned} 0 &< \nu_0\le\nu(\zeta)\le\nu_1<+\infty \\ 0 &< k_0\le k(\zeta)\le k_1<+\infty, \qquad \forall\zeta\in\mathbb{R} \end{aligned}$$

The nonlinearities in those equations are difficult to treat but there are some results on existence and regularity of solutions (see for instance [4, 5].)

There are physical situations, however where the state variables of the flow (velocity and temperature) must be restricted; in particular, in some situations such states must mantained in certain given convex sets in functional spaces.

To explain what we mean with such restrictions, let $\mathcal{V}=\{\mathbf{v}\in(C^\infty(\Omega))^3;\ \operatorname{div}\mathbf{v}=0\}$ and $\mathbf{H}$ and $\mathbf{V}$ respectively the closure of $\mathcal{V}$ with the $L^2(\Omega)^3$=norm and the $(H_0^1(\Omega))^3$-norm. we fix two closed convex sets: Let also $S\in C^2(\Omega)$ whose boundary trace is equal to T and consider two convex closed sets:

$$K_1\subseteq\mathbf{V},\quad K_2\subseteq H_0^1(\Omega).$$

such that $\mathbf{0}\in K_1$, $0\in K_2$, we will look for suitable solutions satisfying $\mathbf{u}(t)\in K_1$ and $\theta(t)\in K_2$ for $t\in(0,T]$.

In this case, in order to comply with such restrictions, in the right-hand sides of (1) and (3) must appear a priori unknown reaction forces. When such restrictions are closed convex sets, as it is the situation considered here, the problem can then be rewritten in terms of variational inequalities.

One possible first mathematical formulation of this problem is the following: by using the change of variable $\tau=\theta-S$, we want solutions $\mathbf{u}\in L^2(0,T;\mathbf{V})\cap L^\infty(0,T;\mathbf{H})$, $\tau\in L^2(0,T;H_0^1)\cap L^\infty(0,T;L^2(\Omega))$, $(\mathbf{u}(t),\tau(t))\in K_1\times K_2$ for $a.e.t\in[0,T]$ satisfying the following differential variational inequalities:

$$\begin{aligned} (\mathbf{u}_t,\mathbf{v}-\mathbf{u}) &+ (\nu(\tau+S)\nabla\mathbf{u},\nabla(\mathbf{v}-\mathbf{u})) \\ &\ge -((\mathbf{u}\cdot\nabla)\mathbf{u},\mathbf{v}-\mathbf{u})+(\alpha(\tau+S)\mathbf{g},\mathbf{v}-\mathbf{u}), \end{aligned} \tag{4}$$

$$\begin{aligned} (\tau_t,\psi-\tau) &+ (k(\tau+S)\nabla\tau,\nabla(\psi-\tau)) \\ &\ge -(\mathbf{u}\cdot\nabla\tau,\psi-\tau)-(S_t,\psi-\tau) \\ &\quad -(k(\tau+S)\nabla S),\nabla(\psi-\tau))-(\mathbf{u}\cdot\nabla S,\psi-\tau) \end{aligned} \tag{5}$$

$\forall(\mathbf{v},\ \psi)\in K_1\times K_2$. Moreover, it is also required that

$$\mathbf{u}(0,\cdot)=\mathbf{u}_0(\cdot)\in K_1,\quad \tau(0,\cdot)=\theta_0(\cdot)-S=\tau_0(\cdot)\in K_2, \tag{6}$$

We remark that under the required regularity of the solutions, it is ease to verify that the initial conditions make sense.

However, due to the difficult interplay among the nonlinearities and the convex restrictions of the state variables, it is difficult to show the existence of solutions of the previous formulations.

Thus, we follow Lions [2] and introduce a weaker relaxed form of the previous variational inequalities for which, under certain conditions, we will be able to show existence of solutions. This relaxed formulation will be described in the next section.

2 Weak relaxed formulation

We start by defining the following:

$$\begin{aligned}\Phi = \{&\mathbf{v} \in L^\infty(0,T;\mathbf{H}) \cap L^2(0,T;\mathbf{V});\ \mathbf{v}_t \in L^2(0,T;\mathbf{V}^*);\\ &\nabla\mathbf{v} \in L^\infty(0,T;L^{3/2}(\Omega)^3),\ v(\cdot,T) = \mathbf{0},\ v(\cdot,t) \in K_1\ a.e.\ t \in [0,T]\}\end{aligned}$$

$$\begin{aligned}\Psi = \{&\psi \in L^\infty(0,T;L^2(\Omega)) \cap L^2(0,T;H_0^1(\Omega)); \psi_t \in L^2(0,T;H^{-1}(\Omega)),\\ &\nabla\psi \in L^\infty(0,T;L^{3/2}(\Omega)),\ \psi(\cdot,T) = 0,\ \psi(\cdot,t) \in K_2\ a.e.\ t \in [0,T]\}\end{aligned}$$

A suitable weak relaxed formulation of the previous problem is obtained, as previously said, by following the ideas similar to the ones in Lions [2]: we formally take $\mathbf{v} \in \Phi$ and put $\mathbf{v}(t)$ in the place of $\mathbf{v}$ in the previous formulation; then, we integrate (4) in time from 0 to T and add and subtract the term $-\int_0^T (\mathbf{v}_t, \mathbf{v}-\mathbf{u})dt$ to the resulting left-hand side; after some integrations by parts in the space variable and some computations, we get:

$$\begin{aligned}&-\int_0^T (\mathbf{v}_t,\mathbf{u})dt + \int_0^T ((\nu(\tau+S)\nabla\mathbf{u}), \nabla(\mathbf{v}-\mathbf{u}))dt + \int_0^T ((\mathbf{u}\cdot\nabla)\mathbf{u},\mathbf{v})dt\\ &\geq \int_0^T (\alpha(\tau+S)\mathbf{g}, \mathbf{v}-\mathbf{u})dt + \frac{1}{2}|\mathbf{v}(0)|^2 - \frac{1}{2}\,|\mathbf{v}(0)-\mathbf{u}_0|_2^2 + \frac{1}{2}\,|\mathbf{u}(T)|_2^2\\ &\geq \int_0^T (\alpha(\tau+S)\mathbf{g}, \mathbf{v}-\mathbf{u})dt + \frac{1}{2}|\mathbf{v}(0)|^2 - \frac{1}{2}\,|\mathbf{v}(0)-\mathbf{u}_0|_2^2\,,\end{aligned}$$

The relaxed formulation is obtained by considering just the last inequality sign. That is, we will look for a function $\mathbf{u}$ satisfying the following differential variational inequality: for any $\mathbf{v} \in \Phi$,

$$\begin{aligned}&-\int_0^T (\mathbf{v}_t,\mathbf{u})dt + \int_0^T ((\nu(\tau+S)\nabla\mathbf{u}), \nabla(\mathbf{v}-\mathbf{u}))dt + \int_0^T ((\mathbf{u}\cdot\nabla)\mathbf{u},\mathbf{v})dt\\ &+\int_0^T (\alpha(\tau+S)\mathbf{g}, \mathbf{v}-\mathbf{u})dt + \frac{1}{2}|\mathbf{v}(0)|^2 - \frac{1}{2}\,|\mathbf{v}(0)-\mathbf{u}_0|_2^2\,.\end{aligned} \tag{7}$$

Similarly, the relaxed formulation for the temperature is the following: for any $\psi \in \Psi$,

$$
\begin{aligned}
&-\int_0^T (\psi_t,\tau)dt - \int_0^T (\text{ div } (k(\tau+S)\nabla\tau),\psi-\tau)dt + \int_0^T (\mathbf{u}\cdot\nabla\tau,\psi)dt \\
&\geq -\int_0^T (S_t,\psi-\tau)dt + \int_0^T (\text{ div } (k(\tau+S)\nabla S),\psi-\tau)dt \qquad (8)\\
&\quad -\int_0^T (\mathbf{u}\cdot\nabla S,\psi-\tau)dt + \frac{1}{2}\,|\tau(0)|_2^2 - \frac{1}{2}\,|\psi(0)-\tau_0|_2^2
\end{aligned}
$$

Now we can define:

Definition: $(\mathbf{u},\tau)$ is a **relaxed weak solution** of the differential variational inequality if $\mathbf{u} \in L^2(0,T;\mathbf{V}) \cap L^\infty(0,T;\mathbf{H})$, $\tau \in L^2(0,T;H_0^1) \cap L^\infty(0,T;L^2(\Omega))$, $(\mathbf{u}(t),\tau(t)) \in K_1 \times K_2$ and verify (7)-(8).

Next, we state the main result of this work:

Theorem *Let $\mathbf{g} \in L^2(0,T;\mathbf{L}^4(\Omega))$. If $||S||_{L^\infty(0,T;L^3(\Omega))}$ and either α or $\mathbf{g}$ are small enough then (4)-(5) admits a relaxed weak solution.*

3 An auxiliary problem

We will use the spectral Faedo-Galerkin method together with a suitable penalization to reduce the differential variational inequality to an equality.

To construct the penalization procedure will consider the usual projection operators on convex $P_{K_1} : \mathbf{L}^2(\Omega) \to K_1$ and $P_{K_2} : L^2(\Omega) \to K_2$ and define

$$Q^1\mathbf{v} = \mathbf{v} - P_{K_1}\mathbf{v}, \qquad Q^2\psi = \psi - P_{K_2}\psi$$

We recall that P_{K_1} has the property that $(\mathbf{v} - P_{K_1}\mathbf{v}, P_{K_1}\mathbf{v} - \mathbf{h}) \geq 0$, $\forall v \in \mathbf{L}^2$; in particular, since $\mathbf{0} \in K_1$, $(\mathbf{v} - P_{K_1}\mathbf{v}, P_{K_1}\mathbf{v}) \geq 0$; similar properties hold for P_{K_2}. For more details on projection operators, see for instance [1].

For the Faedo-Galerkin procedure, let us consider the eigenfunctions $\{\mathbf{v}_1,...,\mathbf{v}_m,...\}$ and $\{\phi_1,...,\phi_m,...\}$ respectively of the Stokes and Laplace operators and define V_m and W_m as the subspaces generated respectively by $\{\mathbf{v}_1,...,\mathbf{v}_m\}$ y $\{\phi_1,...,\phi_m\}$.

Now, we consider the following penalized approximate problem: find $\mathbf{u}_m(t) = \sum_{k=1}^m c_{km}(t)\mathbf{v}_k \in V_m$ and $\tau_m(t) = \sum_{k=1}^m d_{km}(t)\phi_k \in W_m$, satisfying

$$
\begin{aligned}
(\mathbf{u}_{m,t},\mathbf{v}_k) \quad &- \quad (\text{ div } (\nu(\tau_m+S)\nabla\mathbf{u}_m),\mathbf{v}_k) + ((\mathbf{u}_m\cdot\nabla)\mathbf{u}_m,\mathbf{v}_k) \\
&= \quad (\alpha(\tau_m+S)\mathbf{g},\mathbf{v}_k) - m(Q^1\mathbf{u}_m,\mathbf{v}_k), \quad \forall \mathbf{v}_k \in \mathbf{V}_m, \qquad (9)
\end{aligned}
$$

$$
\begin{aligned}
(\tau_{m,t},\phi_k) \quad &- \quad (\text{ div } (k(\tau_m+S)\nabla\tau_m),\phi_k) + (\mathbf{u}_m\cdot\nabla\tau_m,\phi_k) \\
&= \quad -(S_t,\phi_k) + (\text{ div } (k(\tau_m+S)\nabla S),\phi_k) \\
&\quad\;\; -(\mathbf{u}_m\cdot\nabla S,\phi_k) - m(Q^2\tau_m,\phi_k), \quad \forall \phi_k \in W_m, \qquad (10)
\end{aligned}
$$

together with the initial conditions

$$\mathbf{u}_m(0) = \mathbf{u}_{0m}, \quad \tau_m(0) = \tau_{0m},$$

where we can choose $\mathbf{u}_{0m} \in \mathbf{V}_m$ such that $\mathbf{u}_{0m} \to \mathbf{u}_0$ and $|\mathbf{u}_{m0}|_2 \leq |\mathbf{u}_0|_2$; also we can choose $\tau_{0m} \in W_m$ such that $\tau_{0m} \to \tau_0$ in $L^2(\Omega)$ and $|\tau_{m0}|_2 \leq |\tau_0|_2$.

Since the previous expressions for each $m \in \mathbb{N}$ are regular ordinary differential equations, the local existence of solutions are easily guaranteed. The fact that such solutions will be globally defined will be consequence of the estimates to be obtained in the next section.

4 Estimates and convergences

By taking $\mathbf{v}_k = \mathbf{u}_m$ and $\phi_k = \tau_m$ respectively in (9) y (10), we easily obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}|\mathbf{u}_m|_2^2 &+ \nu_0 |\nabla\mathbf{u}_m|_2^2 + m(Q^1\mathbf{u}_m, \mathbf{u}_m)\\ &\leq (\alpha(\tau_m + S)\mathbf{g}, \mathbf{u}_m) \qquad (11)\\ \frac{1}{2}\frac{d}{dt}|\tau_m|_2^2 &+ k_0 |\nabla\tau_m|_2^2 + m(Q^2\tau_m, \tau_m)\\ &\leq -(S_t, \tau_m) + (k(\tau_m + S)\nabla S, \nabla\tau_m) - (\mathbf{u}_m \cdot \nabla S, \tau_m). \qquad (12)\end{aligned}$$

By using Hölder inequality, it is possible to show that

$$\begin{aligned}(\alpha\tau_m\mathbf{g}, \mathbf{u}_m) &\leq 3\alpha |\tau_m|_6 |\mathbf{g}|_{\frac{3}{2}} |\mathbf{u}_m|_6 ,\\ (\alpha S\mathbf{g}, \mathbf{u}_m) &\leq 3\alpha |S|_3 |\mathbf{g}|_2 |\mathbf{u}_m|_6 ,\\ |(\mathbf{u}_m \cdot \nabla S, \tau_m)| &= |(\mathbf{u}_m \cdot \nabla\tau_m, S)| \leq 3 |S|_3 |\nabla\tau_m|_2 |\mathbf{u}_m|_6 ,\\ |(S_t, \tau_m)| &\leq \leq \left|\frac{dS}{dt}\right|_{H^{-1}} |\nabla\tau_m|_2 ,\\ (k(\tau_m + S)\nabla S, \nabla\tau_m) &\leq k_1 |\nabla S|_2 |\nabla\tau_m|_2 .\end{aligned}$$

By using these results in (11) and (12) and adding the results, we get

$$\begin{aligned}&\frac{d}{dt}(|\mathbf{u}_m|_2^2 + |\tau_m|_2^2) + \nu_0(\frac{1}{2} - \frac{3\alpha C_L^2}{2\sqrt{k_0\nu_0}} |\mathbf{g}|_{\frac{3}{2}} - \frac{3C_L}{2\sqrt{k_0\nu_0}} |S|_3) |\nabla\mathbf{u}_m|_2^2\\ &+k_0(\frac{1}{2} - \frac{3\alpha C_L^2}{2\sqrt{k_0\nu_0}} |\mathbf{g}|_{\frac{3}{2}} - \frac{3C_L}{2\sqrt{k_0\nu_0}} |S|_3) |\nabla\tau_m|_2^2) + m((Q^1\mathbf{u}_m, \mathbf{u}_m) + (Q^2\tau_m, \tau_m))\\ &\leq F(t),\end{aligned}$$

where $F(t) = \frac{9\alpha^2}{2\nu_0} |S|_3^2 |\mathbf{g}|_2^2 + \frac{1}{k_0} |S_t|_{H^{-1}}^2 + \frac{k_1^2}{k_0} |\nabla S|_2^2$.

Being

$$\gamma = \inf_t\{\nu_0(\frac{1}{2}-\frac{3\alpha C_L^2}{2\sqrt{k_0\nu_0}} |\mathbf{g}|_{\frac{3}{2}}-\frac{3C_L}{2\sqrt{k_0\nu_0}} |S|_3^2), k_0(\frac{1}{2}-\frac{3\alpha C_L^2}{2\sqrt{k_0\nu_0}} |\mathbf{g}|_{\frac{3}{2}}-\frac{3C_L}{2\sqrt{k_0\nu_0}} |S|_3^2)\},$$

for α or $\mathbf{g}$, and S small enough, we have that $\gamma > 0$.

Thus,

$$\frac{d}{dt}(|\mathbf{u}_m|_2^2+|\tau_m|_2^2)+\gamma(|\nabla\mathbf{u}_m|_2^2+|\nabla\tau_m|_2^2)+m((Q^1\mathbf{u}_m, \mathbf{u}_m)+(Q^2\tau_m, \tau_m)) \leq F(t),$$

and by using the properties of the projection on convex sets and integrating in time this inequality, we get

$$\begin{aligned}
&|\mathbf{u}_m(t)|_2^2 + |\tau_m(t)|_2^2 + \gamma \int_0^T (|\nabla \mathbf{u}_m|_2^2 + |\nabla \tau_m|_2^2) dt \\
&+ m \int_0^T (|\mathbf{u}_m - P_{K_1}\mathbf{u}_m|_2^2 + |\tau_m - P_{K_2}\tau_m)|_2^2) dt \\
&\le |\mathbf{u}_m(0)|_2^2 + |\tau_m(0)|_2^2 + \int_0^T F(t)dt \le N + \int_0^T F(t)dt.
\end{aligned}$$

We thus obtain global existence in time for $(\mathbf{u}_m, \tau_m)$ and the following estimates which are uniform in m:

$$||\mathbf{u}_m||_{L^2(0,T;\mathbf{V})} \le C, \quad ||\tau_m||_{L^2(0,T;H_0^1(\Omega))} \le C,$$
$$m\int_0^T |\mathbf{u}_m - P_{K_1}\mathbf{u}_m|_2^2 \, dt \le C_1, \quad m\int_0^T |\tau_m - P_{K_2}\tau_m|_2^2 \, dt \le C_2.$$

From the last two results, we get in particular that

$$\mathbf{u}_m - P_{K_1}\mathbf{u}_m \to \mathbf{0} \text{ in } L^2(Q_T)^3, \quad \tau_m - P_{K_2}\tau_m \to 0 \text{ in } L^2(Q_T).$$

We can conclude that there are $\mathbf{u} \in L^2(0,T;\mathbf{V})$ and $\theta \in L^2(0,T;H_0^1(\Omega))$ and subsequences $\{\mathbf{u}_m\}$ and $\{\tau_m\}$, for simplicity denoted as before, such that

$$\begin{aligned}
&\mathbf{u}_m \to \mathbf{u} \quad L^2(0,T;\mathbf{V}) - \text{weakly and } L^\infty(0,T;\mathbf{H}) - \text{weakly-*}, \\
&\tau_m \to \tau \quad L^2(0,T;H_0^1) - \text{weakly and } L^\infty(0,T;L^2(\Omega)) - \text{weakly-*}.
\end{aligned}$$

Next, we have to find suitable estimates for $\{\mathbf{u}_{m,t}\}$ and $\{\tau_{m,t}\}$. For this, we consider the following operators:

$$\begin{aligned}
&< A_\nu(\mathbf{u},\tau), \mathbf{v} >= (\nu(\tau+S)\nabla\mathbf{u}, \nabla\mathbf{v}), \quad && < D(\mathbf{u}), \mathbf{v} >= B(\mathbf{u},\mathbf{u},\mathbf{v}), \\
&< E(\tau), \mathbf{v} >= (\tau\mathbf{g}, \mathbf{v}), && < F_m(\mathbf{u}), \mathbf{v} >= m(Q^1(u), \mathbf{v}),
\end{aligned}$$

and follow arguments as in Lions p. 76 [3]. We observe that $\mathbf{L}^2(\Omega)$-the orthogonal projection P_m is over the space generated by eigenfunctions $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$. Due to this choice, we have $P_m \in L(\mathbf{V},\mathbf{V})$ and also $|P_m|_{L(V,V)} \le 1$. By considering P_m^*, we also have $|P_m^*|_{L(V^*,V^*)} \le 1$. From the approximate equations we then get

$$\mathbf{u}_{m,t} = -P_m^*(A_\nu(\mathbf{u}_m,\tau_m) + D(\mathbf{u}_m) + E(\tau_m + S) + F_m(\mathbf{u}_m)) \text{ en } \mathbf{V}^*.$$

The terms in the right-hand side are estimated as follows: $|A_\nu(\mathbf{u}_m,\tau_m)|_{\mathbf{V}^*} \le \nu_1 |\nabla\mathbf{u}_m|_2$, $|D(\mathbf{u}_m)|_{\mathbf{V}^*} \le C|\nabla\mathbf{u}_m|_2$, $|E(\tau_m)|_{\mathbf{V}^*} \le C|\mathbf{g}|_4 |\nabla\tau_m|_2$ and

$$\int_0^T |F_m(\mathbf{u}_m)|_{V^*} \, dt \le \sup_{|\mathbf{v}|_{\mathbf{V}} \le 1} \int_0^T \left|mQ^1(u)\right|_2 dt \, |\nabla\mathbf{v}|_2 \, dt.$$

Since $m\int_0^T |\mathbf{u}_m - P_{K_1}\mathbf{u}_m|_2^2 \, dt \le C_1$, also $m\int_0^T |\mathbf{u}_m - P_{K_1}\mathbf{u}_m|_2 \, dt \le C$.

The previous results imply that $||\mathbf{u}_{m,t}||_{L^1(0,T;V^*)} \leq C$, with a positive constant C independent of m.

We can obtain similar estimates for $||\theta_{m,t}||_{L^1(0,T;H^{-1}(\Omega))}$.

From the previous estimates and a Aubin-Lions type lemma (Simon [6]), we conclude that there are subsequences, again for simplicity denoted as before, and $\mathbf{u}$ and τ such that $\mathbf{u}_m \to \mathbf{u}$ strongly in $L^2(0,T;\mathbf{H})$ and $\tau_m \to \tau$ strongly in $L^2(0,T;L^2(\Omega))$.

We thus conclude that $\mathbf{u}_m - P_{K_1}\mathbf{u}_m \to \mathbf{0}$ strongly in $L^2(0,T;\mathbf{H})$ and $\tau_m - P_{K_2}\tau_m \to 0$ strongly in $L^2(0,T;L^2(\Omega))$. Moreover, since

$$\int_0^T |(\mathbf{u}_m - P_{K_1}\mathbf{u}_m) - (\mathbf{u} - P_{K_1}\mathbf{u})|^2\, dt \leq 4\int_0^T |\mathbf{u}_m - \mathbf{u}|^2\, dt \to 0 \text{ as } m \to \infty,$$

we conclude that $\mathbf{u} - P_{K_1}\mathbf{u} = \mathbf{0}$ and $\tau - P_{K_2}\tau = 0$; thus $(\mathbf{u}(t), \tau(t)) \in K_1 \times K_2$.

Next, fix arbitrary $\mathbf{v} \in \Phi$, $\psi \in \Psi$ and define the sequences $\{\boldsymbol{\gamma}_k(\cdot)\}_{k=1}^{\infty}$ and $\{\sigma_k(\cdot)\}_{k=1}^{\infty}$ by $\boldsymbol{\gamma}_k(t) = \pi_k^{(1)}v(t)$ for $t \in [0,T]$ $\sigma_k(\cdot) = \pi_k^{(2)}\psi$, where $\pi_k^{(1)}$ and $\pi_k^{(2)}$ are respectively the orthogonal projections of $L^2(\Omega)^3$ into the eigen-space $\mathbf{V}_k$ and $L^2(\Omega)$ into the eigen-space W_k. Due to the properties of such operators, $\boldsymbol{\gamma}_k \in V_k$, $\sigma_k \in W_k$, $\boldsymbol{\gamma}_k \to \mathbf{v}$ strongly in the norms stated in Φ, $\sigma_k \to \psi$ strongly in the norms stated in Ψ, $\boldsymbol{\gamma}_{k,t} \to \mathbf{v_t}$ in $L^2(0,T;\mathbf{V}^*)$, $\sigma_{k,t} \to \psi_t$ in $L^2(0,T;H^{-1}(\Omega))$, $\boldsymbol{\gamma}_k(0) \to \mathbf{v}(0)$ in $\mathbf{V}$, $\sigma_k(0) \to \psi(0)$ in $H_0^1(\Omega)$, $\boldsymbol{\gamma}_k(T) = \mathbf{0}$, $\sigma_k(T) = 0$.

Putting $\boldsymbol{\gamma}_k(t) - \mathbf{u}_m(t)$ in the place of $\mathbf{v}_k$ in (9) and $\sigma_k(t) - \tau_m(t)$ in the place of ϕ_k in (10), we get

$$\begin{aligned}(\mathbf{u}_{m,t}, \boldsymbol{\gamma}_k - \mathbf{u}_m) \quad &- \quad (\text{ div } (\nu(\tau_m + S)\nabla\mathbf{u}_m), \boldsymbol{\gamma}_k - \mathbf{u}_m) \\ &= \quad -((\mathbf{u}_m \cdot \nabla)\mathbf{u}_m, \boldsymbol{\gamma}_k - \mathbf{u}_m) + (\alpha(\tau_m + S)\mathbf{g}, \boldsymbol{\gamma}_k - \mathbf{u}_m) \\ &\qquad -m(Q^1\mathbf{u}_m, \boldsymbol{\gamma}_k - \mathbf{u}_m),\end{aligned}$$

$$\begin{aligned}(\tau_{m,t}, \sigma_k - \tau_m) \quad &- \quad (\text{ div } (k(\tau_m + S)\nabla\tau_m), \sigma_k - \tau_m) \\ &\qquad +(\mathbf{u}_m \cdot \nabla\tau_m, \sigma_k - \tau_m) + m(Q^2\tau_m, \sigma_k - \tau_m) \\ &= \quad -(\frac{dS}{dt}, \sigma_k - \tau_m) + (\text{ div } (k(\tau_m + S)\nabla S), \sigma_k - \tau_m) \\ &\qquad -(\mathbf{u}_m \cdot \nabla S, \sigma_k - \tau_m).\end{aligned}$$

Since $m(Q^1\mathbf{u}_m, \boldsymbol{\gamma}_k - \mathbf{u}_m) \leq 0$ y $m(Q^2\tau_m, \sigma_k - \tau_m) \leq 0$, we obtain:

$$\begin{aligned}&(\mathbf{u}_{m,t}, \boldsymbol{\gamma}_k - \mathbf{u}_m) - (\text{ div } (\nu(\tau_m + S)\nabla\mathbf{u}_m), \boldsymbol{\gamma}_k - \mathbf{u}_m) \\ &+((\mathbf{u}_m \cdot \nabla)\mathbf{u}_m, \boldsymbol{\gamma}_k - \mathbf{u}_m) - (\alpha(\tau_m + S)\mathbf{g}, \boldsymbol{\gamma}_k - \mathbf{u}_m) \geq 0,\end{aligned}$$

$$\begin{aligned}&\quad (\tau_{m,t}, \sigma_k - \tau_m) - (\text{ div } (k(\tau_m + S)\nabla\tau_m), \sigma_k - \tau_m) + (\mathbf{u}_m \cdot \nabla\tau_m, \sigma_k - \tau_m) \\ &\geq \quad -(S_t, \sigma_k - \tau_m) + (\text{ div } (k(\tau_m + S)\nabla S), \sigma_k - \tau_m) - (\mathbf{u}_m \cdot \nabla S, \sigma_k - \tau_m).\end{aligned}$$

Next, by working similarly as it was done to obtain the relaxed weak formulation, we get

$$\begin{aligned}&-\int_0^T(\boldsymbol{\gamma}_{k,t},\mathbf{u}_m)dt+\int_0^T((\mathbf{u}_m\cdot\nabla)\mathbf{u}_m,\boldsymbol{\gamma}_k)dt+\int_0^T(\nu(\tau_m+S)\nabla\mathbf{u}_m,\nabla(\boldsymbol{\gamma}_k-\mathbf{u}_m))dt\\&\geq\int_0^T(\alpha(\tau_m+S)\mathbf{g},\boldsymbol{\gamma}_k-\mathbf{u}_m)dt+\frac{1}{2}\left|\boldsymbol{\gamma}_k(0)\right|_2^2-\frac{1}{2}\left|\boldsymbol{\gamma}_k(0)-\mathbf{u}_m(0)\right|_2^2,\end{aligned}$$

and for the temperature we similarly get

$$\begin{aligned}&\int_0^T(\sigma_{k,t},\sigma_k-\tau_m)dt+\int_0^T(\mathbf{u}_m\cdot\nabla\tau_m,\sigma_k-\tau_m)dt\\&\geq-\int_0^T(k(\tau_m+S)\nabla\sigma_k,\nabla(\sigma_k-\tau_m))dt-\int_0^T(S_t,\sigma_k-\tau_m)dt\\&-\int_0^T(k(\tau_m+S)\nabla S,\nabla(\sigma_k-\tau_m))dt-\int_0^T(\mathbf{u}_m\cdot\nabla S,\sigma_k-\tau_m)dt\\&\quad+\frac{1}{2}\left|\tau_m(0)\right|_2^2-\frac{1}{2}\left|\sigma_k(0)-\tau_m(0)\right|_2^2.\end{aligned}$$

To show the existence of weak solutions it is enough to pass to the limit in m. This is done as follows.

Since

$$\left|\int_0^T(\boldsymbol{\gamma}_{k,t},\mathbf{u}_m)dt-\int_0^T(\boldsymbol{\gamma}_{k,t},\mathbf{u})dt\right|\leq\left|\boldsymbol{\gamma}_{k,t}\right|_{L^2(0,T;L^2)}\left|\mathbf{u}_m-\mathbf{u}\right|_{L^2(0,T;L^2)},$$

from the strong convergence of $\mathbf{u}_m$ in $L^2(0,T;\mathbf{L}^2(\Omega))$, we conclude that as $m\to\infty$ then

$$\int_0^T(\boldsymbol{\gamma}_{k,t},\mathbf{u}_m)dt\to\int_0^T(\boldsymbol{\gamma}_{k,t},\mathbf{u})dt.$$

To show that $\int_0^T((\mathbf{u}_m\cdot\nabla)\mathbf{u}_m,\boldsymbol{\gamma}_k)dt\to\int_0^T((\mathbf{u}\cdot\nabla)\mathbf{u},\boldsymbol{\gamma}_k)dt$ as $m\to\infty$, we firstly use the fact that

$$\int_0^T((\mathbf{u}_m\cdot\nabla)\mathbf{u}_m,\boldsymbol{\gamma}_k)dt=-\int_0^T((\mathbf{u}_m\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m)dt.. \tag{13}$$

Next, we observe that

$$((\mathbf{u}_m\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m)-((\mathbf{u}\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u})=(((\mathbf{u}_m-\mathbf{u})\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m)+((\mathbf{u}\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m-\mathbf{u}), \tag{14}$$

which by using Hölder inequality, can be estimated by

$$(((\mathbf{u}_m-\mathbf{u})\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m)\leq\left|\mathbf{u}_m-\mathbf{u}\right|_2\left|\nabla\boldsymbol{\gamma}_k\right|_{\frac{3}{2}}\left|\mathbf{u}_m\right|_6 \tag{15}$$

$$((\mathbf{u}\cdot\nabla)\boldsymbol{\gamma}_k,\mathbf{u}_m-\mathbf{u})\leq\left|\mathbf{u}_m-\mathbf{u}\right|_2\left|\nabla\boldsymbol{\gamma}_k\right|_{\frac{3}{2}}\left|\mathbf{u}_m\right|_6. \tag{16}$$

The convergence then follows from (13)-(16), the fact that $\nabla\boldsymbol{\gamma}_k\in L^\infty(0,T;L^{\frac{3}{2}}(\Omega))$ and the strong convergence of $(\mathbf{u}_m)$ en $L^2(0,T;\mathbf{L}^2(\Omega))$.

To prove that $\int_0^T ((\nu(\tau_m + S)\nabla \mathbf{u}_m, \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m))dt \to \int_0^T ((\nu(\tau + S)\nabla \mathbf{u}, \nabla(\boldsymbol{\gamma}_k - \mathbf{u}))dt$ as $m \to \infty$, we observe that

$$\begin{aligned}
&\int_0^T (((\nu(\tau_m + S)\nabla \mathbf{u}_m, \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m) - ((\nu(\tau + S)\nabla \mathbf{u}, \nabla(\boldsymbol{\gamma}_k - \mathbf{u})))dt \\
&= \int_0^T ((\nu(\tau_m + S) - \nu(\tau + S))\nabla \mathbf{u}_m, \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m))dt \\
&+ \int_0^T (\nu(\tau + S)\nabla(\mathbf{u}_m - \mathbf{u}), \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m))dt \\
&+ \int_0^T (\nu(\tau + S)\nabla \mathbf{u}, \nabla(\mathbf{u} - \mathbf{u}_m))dt = I + II + III
\end{aligned}$$

The term I can be estimated as follows:

$$\begin{aligned}
&\int_0^T ((\nu(\tau_m + S) - \nu(\tau + S))\nabla \mathbf{u}_m, \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m))dt \\
&\le C\,|\tau_m - \tau|_{L^2(0,T;L^2)} \left(|\nabla \mathbf{u}_m|_{L^2(0,T;L^2)}\right) + |\nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m)|_{L^2(0,T;\mathbf{L}^2)}),
\end{aligned}$$

Thus, I goes to zero because $\tau_m - \tau \to 0$ in $L^2(0,T;L^2)$.

The term II goes to zero because $\nu(\tau+S)\nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m)$ is bounded in $H^{-1}(\Omega)$,

$$(\nu(\tau + S)\nabla(\mathbf{u}_m - \mathbf{u}), \nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m)) = (\nabla(\mathbf{u}_m - \mathbf{u}), \nu(\tau + S)\nabla(\boldsymbol{\gamma}_k - \mathbf{u}_m))$$

and also $\nabla(\mathbf{u}_m - \mathbf{u})$ is weakly convergent.

The term III is similarly handled. With the convergences at hand it is easy to pass to limit the other terms and we obtain (7) with $\boldsymbol{\gamma}_k$ in the place of $\mathbf{v}$. By working similarly with the inequality for the approximate temperature, we also obtain (8) with σ_k in the place of ψ.

Finally, by using a standard density argument we can replace $(\boldsymbol{\gamma}_k(t), \sigma_k(t))$ for $(\boldsymbol{\gamma}(t), \sigma(t))$ en $K_1 \times K_2$ in the variational inequality and the theorem is proved.

Final remarks: It is possible to consider different questions concerning the generalized Boussinesq equations with state variable restrictions. For instance, it is possible to prove the existence of very weak solutions (for less regular data) and also the existence of reproductive weak solutions. These results will appear elsewhere.

Acknowledgments

The first and second author were partially supported by DGI-MEC (Spain) Grant MTM2006-07932 and Fondecyt Grant #1080628.

References

[1] Brezis, H., Analyse Fonctionnelle, Masson, Paris, 1987.

[2] Lions, J.L., Partial differential inequalities, Russ. Math. Surv., 27: 91, (1972) 91-159.

[3] Lions, J.L., Quelques Méthodes de Resolution des Problèmes aux Limites Non Linéaires, Dunod, Paris, 1969.

[4] Lorca, S.A., Boldrini, J.L., The initial value problem for a generalized Boussinesq model: regularity and global existence of strong solutions, Mat., Cont., 11 (1996), 71-94.

[5] Lorca, S.A., Boldrini, J.L., The initial value problem for a generalized Boussinesq model, Nonlinear Analysis, 36 (1998),457-480.

[6] Simon, J., Non-homogeneous viscous incompressible fluids: Existence of velocity, density and pressure, Siam J. Math. Anal., 21 (1990), 1093-1117.

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 73–80

MULTIRESOLUTION SIMULATION OF REACTION-DIFFUSION SYSTEMS WITH STRONG DEGENERACY

RAIMUND BÜRGER†, RICARDO RUIZ-BAIER‡

†CI^{2}MA and Departamento de Ingeniería Matemática,
Universidad de Concepción, Casilla 160-C, Concepción, Chile
‡IACS, Chair of Modelling and Scientific Computing, École Polytechnique Fédérale de Lausanne,
Station 8, CH-1015, Lausanne, Switzerland

rburger@ing-mat.udec.cl, ricardo.ruiz@epfl.ch

Abstract

A fully space-adaptive multiresolution method is applied to an explicit finite volume scheme for solving a strongly degenerate reaction-diffusion system. Since a closed mathematical theory is lacking, insight into the behaviour of these systems, in particular into the spatial patterns their solutions may exhibit, can be currently obtained by numerical experimentation only. It is demonstrated that the present space-adaptive scheme is an appropriate tool for this purpose. In particular, the multiresolution method and the classical finite volume scheme are compared and the numerical results show that this strategy provides substantial savings in terms of data storage and computational effort, while giving an accurate approximation to the sought quantities.

Key words: *Adaptive multiresolution schemes, Degenerate diffusion, Pattern-formation, Turing instability.*
AMS subject classifications: *65M06, 35K55.*

1 Introduction

In [1] we present a fully adaptive multiresolution (MR) scheme for spatially 2D, possibly degenerate reaction-diffusion systems, focusing on models of combustion, pattern formation, and chemotaxis. Solutions of these equations in these applications often exhibit steep gradients, and in the degenerate case, sharp fronts and discontinuities. This calls for a concentration of computational effort to zones of strong variation.

In this note we investigate the influence of the form of the diffusion terms, constructed so that the governing equations form a strongly degenerate parabolic system, on the spatial patterns shown by the system. To consider *degenerate* diffusion as a mechanism for the creation of Turing-like instabilities seems to be a novelty in the context of two-dimensional reaction-diffusion systems. The proposed MR scheme is based on finite volume discretizations with explicit time stepping, and the efficiency of the method relies in part on the strategy for storing the solution, namely a dynamic graded tree, whose leaves are the non-uniform finite volumes on the borders of which the numerical divergence is evaluated. By a thresholding procedure, which accounts for the elimination of leaves that are smaller than a threshold value, substantial data compression and CPU time reduction is attained.

We specifically consider the reaction-diffusion system

$$u_t = \gamma f(u,v) + \Delta A(u) \quad \text{on } Q_T := \Omega \times (0,T),\ \Omega := (0,1)^2, \tag{1a}$$
$$v_t = \gamma g(u,v) + d\Delta B(v) \quad \text{on } Q_T, \tag{1b}$$
$$u(x,0) = u_0(x), \quad v(x,0) = v_0(x) \quad \text{for } x \in \Omega, \tag{1c}$$
$$\nabla A(u) \cdot \mathbf{n} = \nabla B(u) \cdot \mathbf{n} = 0 \quad \text{on } \Sigma_T := \partial\Omega \times (0,T). \tag{1d}$$

This system models several phenomena including combustion and chemotaxis [1], but is considered here as the generalization of a well-known model of pattern formation in mathematical biology [6]. Under a number of structural conditions relating the functions $f(u,v)$ and $g(u,v)$ and their derivatives to the parameters γ and d, the system (1) with $A(u) = B(u) = u$ produces stationary solutions with Turing-type spatial patterns [6]. To produce this effect, we could select these diffusion terms along with the kinetics $f(u,v) = a-u+u^2v$ and $g(u,v) = b-u^2v$, with the parameters $a = -0.5$, $b = 1.9$, $d = 4.8$, and $\gamma = 210$ [6]. In this work, however, the diffusion terms are chosen to be strongly degenerate:

$$A(u) = \begin{cases} 0 & \text{for } u \leqslant u_{\mathrm{c}}, \\ u-u_{\mathrm{c}} & \text{otherwise} \end{cases}, \quad B(u) = \begin{cases} 0 & \text{for } u \leqslant v_{\mathrm{c}}, \\ u-v_{\mathrm{c}} & \text{otherwise,} \end{cases} \quad u_{\mathrm{c}}, v_{\mathrm{c}} \geqslant 0. \tag{2}$$

It turns out that even if the stability analysis performed in [6] does *not* apply to the strongly degenerate case, our numerical experiments in Section 4 lead to the formation of spatial patterns. Holden et al. [5] prove existence and uniqueness of entropy solutions of weakly coupled systems of degenerate parabolic equations in an unbounded domain; the well-posedness analysis for (1) is, however, still an open problem due to the boundary condition (1d), which is not covered by the analysis of [5]. Turing instabilities driven by other non-standard diffusion terms, namely by fractional diffusion, have been studied e.g. by Nec and Nepomnyashchy [7].

The remainder of the paper is organized as follows. Section 2 contains a description of the construction of the reference FV formulation used to numerically solve the underlying problem. In Section 3 we detail the main ingredients of the MR framework needed to provide space adaptivity to the overall numerical scheme, and the numerical results provided in Section 4 confirm that the adaptive MR method provides high rates of data compression and CPU time speed-up, while the error remains controlled.

2 Finite volume discretization

An admissible mesh for Ω is formed by a family $\mathcal{T}$ of control volumes (open and convex polygons) of maximum diameter h. For all $K \in \mathcal{T}$, x_K denotes the center of K, $N(K)$ the set of neighbors of K, $\mathcal{E}_{\text{int}}(K)$ is the set of edges of K in the interior of $\mathcal{T}$ and $\mathcal{E}_{\text{ext}}(K)$ the set of edges of K on the boundary $\partial\Omega$. For all $L \in N(K)$ $d(K,L)$ denotes the distance between x_K and x_L, $\sigma_{K,L}$ is the interface between K and L and $\eta_{K,L}$ ($\eta_{K,\sigma}$ respectively) is the unit normal vector to $\sigma_{K,L}$ ($\sigma \in \mathcal{E}_{\text{ext}}(K)$ respectively) oriented from K to L (from K to $\partial\Omega$ respectively). For all $K \in \mathcal{T}$, $|K|$ stands for the measure of the cell K. From the admissibility of $\mathcal{T}$ we have that $\overline{\Omega} = \cup_{K\in\mathcal{T}}\overline{K}$, $K \cap L = \varnothing$ if $K, L \in \mathcal{T}$ and $K \neq L$, and there exists a finite sequence $(x_K)_{K\in\mathcal{T}}$ for which $\overline{x_K x_L}$ is orthogonal to $\sigma_{K,L}$. Now, consider $K \in \mathcal{T}$ and $L \in N(K)$ with common vertices $(a_{\ell,K,L})_{1\leqslant\ell\leqslant I}$ with $I \in \mathbb{N}\backslash\{0\}$ and let $T_{K,L}$ (respectively $T^{\text{ext}}_{K,\sigma}$ for $\sigma \in \mathcal{E}_{\text{ext}}(K)$) be the open and convex polygon with vertices (x_K, x_L) (x_K respectively) and $(a_{\ell,K,L})_{1\leqslant\ell\leqslant I}$. For all $K \in \mathcal{T}$, the approximation $\nabla_h u_h$ of ∇u is defined by

$$\nabla^h_L u_h(x) = \begin{cases} |T_{K,L}|^{-1}|\sigma_{K,L}|(u_L - u_K)\eta_{K,L} & \text{if } x \in T_{K,L}, \\ 0 & \text{if } x \in T^{\text{ext}}_{K,\sigma}. \end{cases}$$

Now we choose an admissible mesh for Ω and a time step size $\Delta t > 0$. We may choose $N > 0$ as the smallest integer such that $N\Delta t \geqslant T$, and set $t^n := n\Delta t$ for $n \in \{0, \dots, N\}$. The discretized reaction terms are defined as $f^{n+1}_K := f(u^{n+1}_K, v^{n+1}_K)$ and $g^{n+1}_K = g(u^{n+1}_K, v^{n+1}_K)$ and the nonlinear diffusions are constructed using the terms $A^{n+1}_K := A(u^{n+1}_K)$ and $B^{n+1}_K := A(v^{n+1}_K)$. Incorporating an explicit first order Euler time integration, the resulting FV scheme reads: Determine $(u^{n+1}_K)_{K\in\mathcal{T}}$, $(v^{n+1}_K)_{K\in\mathcal{T}}$ such that

$$\frac{u^{n+1}_K - u^n_K}{\Delta t} + \sum_{L\in N(K)} \nabla^h_L A^n_K = f^n_K, \quad \frac{v^{n+1}_K - v^n_K}{\Delta t} + \sum_{L\in N(K)} d\nabla^h_L B^n_K = g^n_K, \tag{3}$$

for all $K \in \mathcal{T}$. The boundary condition is taken into account by imposing zero fluxes on external edges. The resulting finite volume scheme has a unique solution that converges to the weak solution of (1) in the non-degenerate case [4]. Moreover, according to [5] this scheme is stable under the CFL condition

$$h^{-1}\Delta t\gamma \max_{K\in\mathcal{T}} \left(|f^u_K| + |f^v_K| + |g^u_K| + |g^v_K|\right) + 4h^{-2}\Delta t \max_{K\in\mathcal{T}}\left(|A'_K| + d|B'_K|\right) \leqslant 1,$$

where $f^u_K := \partial_u f(u_K, v_K)$ and $A'_K := A'(u_K)$, for $K \in \mathcal{T}$.

3 Multiresolution representation

For further details on the one-dimensional theory, we refer to the fairly complete description in [3]. For ease of computations, we only consider rectangular meshes on a rectangular domain, which after a change of variables can be regarded as $\Omega = [0,1]^2$. Nevertheless, the multiresolution analysis could be carried out for non-structured meshes. Firstly, a nested mesh hierarchy $\mathcal{T}_0 \subset \cdots \subset \mathcal{T}_L$ using a partition of Ω is constructed, where each grid $\mathcal{T}_l$ is formed by the control volumes on each level K^l, $l = 0, \dots, L$. Here $l = 0$ corresponds to the coarsest and $l = L$ to the finest resolution level and the so called *refinement sets* are defined by $M_{K,l} = \{L_i^{l+1}\}_i$ and $\overline{K^l} = \cup_{i=1}^{\#M_{K,l}} L_i^{l+1}$. For $x \in K^l$ the *scale box function* is defined as $\tilde{\varphi}_{K,l}(x) := |K^l|^{-1}\chi_{K^l}(x)$ and the average of any function $u(\cdot, t) \in L^1(\Omega)$ in the cell K^l may be written as $u_{K,l} := \langle u, \tilde{\varphi}_{K,l}\rangle_{L^1(\Omega)}$.

It is known that cell averages and box functions satisfy the two-level relation

$$\tilde{\varphi}_{K,l} = \sum_{L_i^{l+1} \in M_{K,l}} \frac{|L_i^{l+1}|}{|K^l|} \tilde{\varphi}_{L_i,l+1}, \quad \bar{u}_{K,l} = \sum_{L_i^{l+1} \in M_{K,l}} \frac{|L_i^{l+1}|}{|K^l|} u_{L_i,l+1}, \tag{4}$$

which defines a *projection* operator needed to move from finer to coarser levels. For $x \in K^{l+1}$ the *wavelet function* is defined by

$$\tilde{\psi}_{K,j,l} = \sum_{L_i^{l+1} \in M_{K,l}} \frac{|L_i^{l+1}|}{|K^l|} (-1)^{ij} \tilde{\varphi}_{L_i,l+1} \quad \text{for } j = 1, \dots, \#M_{K,l},$$

and from (4), a similar inverse two-level relation holds. *Detail coefficients* are defined as $d_{K,j,l} := \langle u, \tilde{\psi}_{K,j,l}\rangle$ for $j = 1, \dots, \#M_{K,l}$. An appealing feature is that a transformation between the cell averages on level L and the cell averages on level zero plus a series of details can be determined and such transformation should be reversible. Therefore

$$\tilde{u}_{K,l+1} = \sum_{T \in \bar{S}_K^l} g_{K,T}^l u_{T,l}, \tag{5}$$

where $\bar{S}_K^l$ is the stencil of interpolation or *coarsening set*, $g_{K,T}^l$ are coefficients, and the tilde over u in the left-hand side of (5) denotes a predicted value. In this way, a *prediction* operator is defined, which is imposed to be local and consistent with the projection and will be necessary to move from coarser to finer resolution levels. For rectangular meshes it corresponds to $\tilde{u}_{L_i,l+1} = u_{L,l} - Q_x - Q_y + Q_{xy}$ for $i = 1, \dots, \#M_{K,l}$, where

$$\begin{aligned} Q_z &:= \sum_{n=1}^{s} \tilde{\gamma}_n \big(u_{S_z,l} - u_{T_z,l}\big), \quad z \in \{x, y\}, \\ Q_{xy} &:= \sum_{n=1}^{s} \tilde{\gamma}_n \sum_{p=1}^{s} \tilde{\gamma}_p \big(u_{S_{x,y},l} - u_{S_{x,-y},l} - u_{S_{-x,y},l} + u_{S_{-x,-y},l}\big). \end{aligned}$$

Here $S_{\pm x,\pm y}$ denote the neighbors of the corner of the control volume S and the corresponding coefficients are $\tilde{\gamma}_1 = -\frac{22}{128}$ and $\tilde{\gamma}_2 = \frac{3}{128}$ (see [8]). Details are related to the regularity of a given function. The more regular u is over K^l, the smaller is the corresponding detail coefficient. Therefore a *thresholding* procedure is also applied, which basically consists in discarding all control volumes corresponding to details that are smaller in absolute value than a level-dependent tolerance. Denoting by α the experimental convergence rate of (3) and by ε_{R} given a reference tolerance determined by (see e.g. [1, 2] for details)

$$\varepsilon_{\mathrm{R}} = \frac{C2^{-(\alpha+2)L}}{|\Omega| \max_{K\in\mathcal{T}} (|f_K^u| + |f_K^v| + |g_K^u| + |g_K^v|) + |\Omega|^{3/2} 2^{L+2} \max_{K\in\mathcal{T}} (|A_K'| + d|B_K'|)},$$

we obtain level-dependent tolerances ε_l defined by $\varepsilon_l = 2^{2(l-L)}\varepsilon_{\mathrm{R}}$, $l = 0,\dots,L$. We organize the cell averages and corresponding details at different levels in a *dynamic graded tree*. The *root* is the basis of the tree, a parent node has four sons, and the sons of the same parent are called *brothers*. A node without sons is a *leaf* and a given node has $s' = 2$ *nearest neighbors* in each spatial direction, needed for the computation of the fluxes of leaves; if these neighbors do not exist, we create them as *virtual leaves*. We denote by Λ the set of all nodes of the tree and by $\mathcal{L}(\Lambda)$ the set of leaves. We apply this MR representation to the spatial part of the pair $\mathbf{u} = (u, v)$, which corresponds to the numerical solution of the underlying problem for each time step, so we need to update the tree structure for the proper representation of the solution during the evolution. To this end, we apply a thresholding strategy, but always keep the graded tree structure of the data. Once the thresholding is performed, we add to the tree a *safety zone*, generated by adding one finer level to the tree in all leaves without violating the graded tree data structure.

The *data compression rate* $\eta := N/(2^{-(2L)}N + \#\mathcal{L}(\Lambda))$ and *speed-up rate* $\mathcal{V} := \mathrm{CPU\,time_{FV}}/\mathrm{CPU\,time_{MR}}$ are used to measure the improvement in data and CPU time compression respectively. Here, N is the number of control volumes in the full finest grid at level L, and $\#\mathcal{L}(\Lambda)$ is the number of leaves.

4 Numerical experiments

As a first numerical result, we consider a computation starting from a random perturbation of the steady state ($u_0 = a + b = 1.4, v_0 = b/(a + b)2 = 0.96939$) and we use $L = 9$ resolution levels and a reference tolerance given by $\varepsilon_{\mathrm{R}} = 7.82 \times 10^{-4}$. The computational domain is the unit square $\Omega = [0, 1]^2$. Figure 1 presents the numerical solution for non-degenerate diffusion, i.e., we choose $A(u) = B(u) = u$. (See [1] for further examples of this case).

Being one of our main interests studying the effect of degenerate diffusion, we present in Example 2 several cases in which all parameters remain the same, except for the test parameters which are the critical concentrations $u_{\mathrm{c}}, v_{\mathrm{c}}$ used in (2). Those cases are $u_{\mathrm{c}} = u_0 + c$, $v_{\mathrm{c}} = v_0 + c$, with $c \in \{0, 0.5, 2.0\}$. In Figure 2 we display the component v of the numerical solution and the leaves of the corresponding tree structure at a transient state at time $t = 1.5$ for all test cases and it is clear that the larger the value of c, the more chaotic the spatial patterns shown by the corresponding system. This behaviour could be explained by the increasing incoherence between solution values at different points.

For Example 3, we select one of the cases from the previous example and perform a study of the error. The effectiveness of the MR method is illustrated

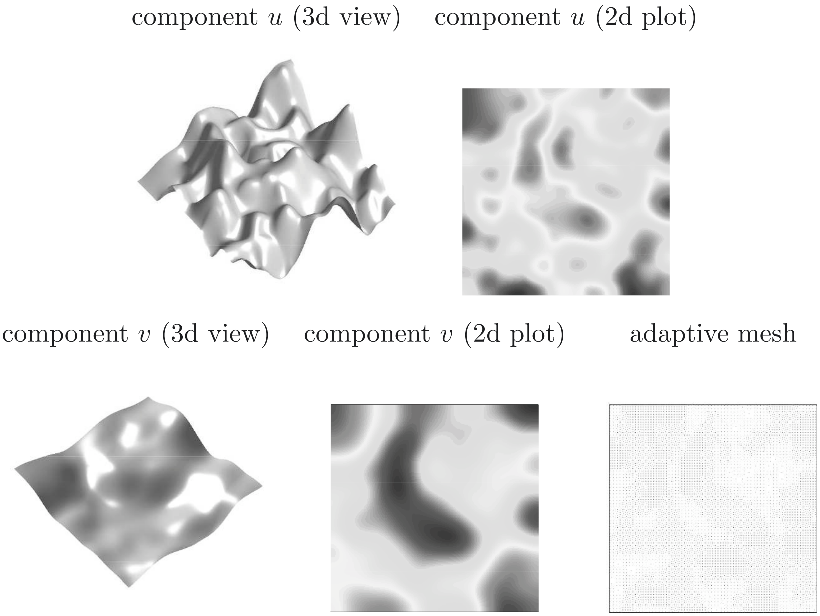


Figure 1: *Example 1: Component* v *and corresponding adaptive mesh at* $t = 1.5$ *for non-degenerate diffusion. The solution assumes values* $1.390 < u < 1.402$ *and* $0.96988 < v < 0.96998$.

in Table 1, specifically displaying the corresponding simulated time, speedup $\mathcal{V}$, data compression rate η, and normalized errors in different norms for both components of the solution. These errors are obtained by comparing with an approximate solution given by a reference FV computation on a fine mesh of 4194304 control volumes. In addition, from Figure 3 a experimental rate of convergence of about 1.9 is noticed for the adaptive MR scheme. As seen in [1], a slightly better rate of convergence may be also obtained by the MR method for the non-degenerate problem.

Acknowledgements

RB acknowledges support by Fondecyt (Chile), project 1090456, Fondap in Applied Mathematics (project 15000001), and BASAL project CMM, Universidad de Chile and Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción.

References

[1] M. Bendahmane, R. Bürger, R. Ruiz-Baier, K. Schneider, *Adaptive multiresolution schemes with local time stepping for two-dimensional degenerate reaction-diffusion systems*, Appl. Numer. Math., to appear.

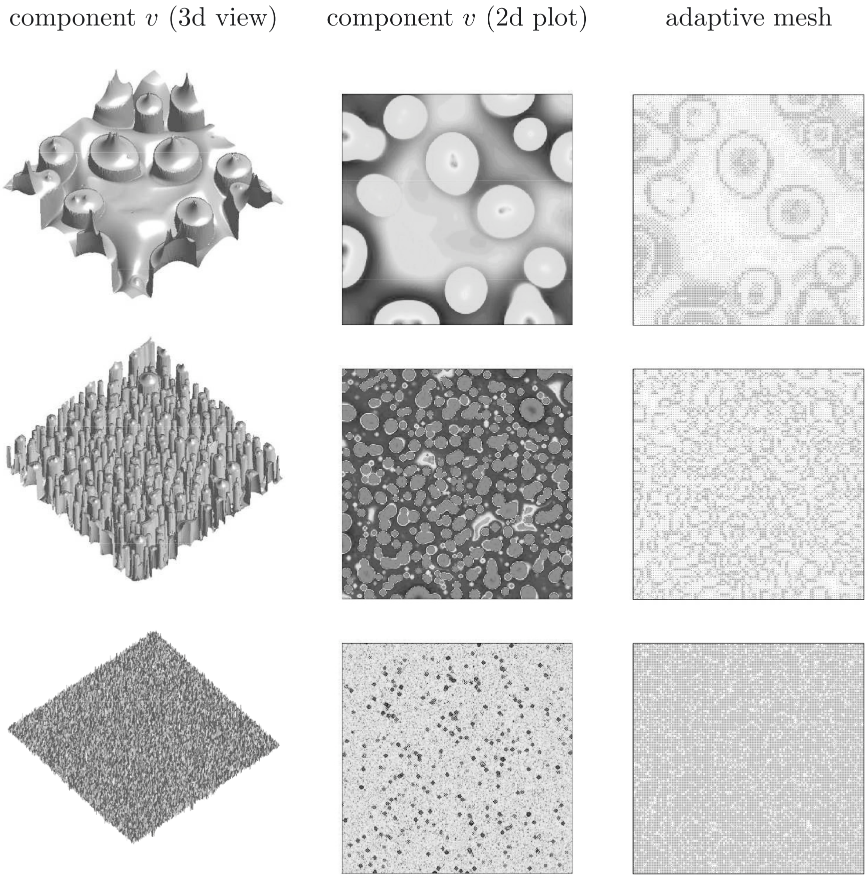


Figure 2: *Example 2: Component v and corresponding adaptive mesh at $t = 1.5$ for $c = 0$ (top), $c = 0.5$ (middle) and $c = 2.0$ (bottom). The solution assumes values $0.96 < v < 0.98$, $0.96 < v < 0.99$ and $0.84 < v < 1.08$ respectively.*

t	$\mathcal{V}$	η	Species	L^1-error	L^2-error	$L^\infty-$error
0.05	9.41	12.7219	u	3.52×10^{-4}	3.65×10^{-4}	6.34×10^{-4}
			v	2.70×10^{-4}	3.08×10^{-4}	5.22×10^{-4}
0.5	12.27	17.9721	u	5.19×10^{-4}	6.13×10^{-4}	6.57×10^{-4}
			v	4.82×10^{-4}	7.06×10^{-4}	1.18×10^{-3}
1.5	12.93	18.7118	u	5.23×10^{-4}	6.65×10^{-4}	9.47×10^{-4}
			v	5.90×10^{-4}	8.31×10^{-4}	2.04×10^{-3}

Table 1: *Example 3: Model with $c = 0$. Corresponding simulated time, speedup $\mathcal{V}$, data compression rate η and errors in different norms.*

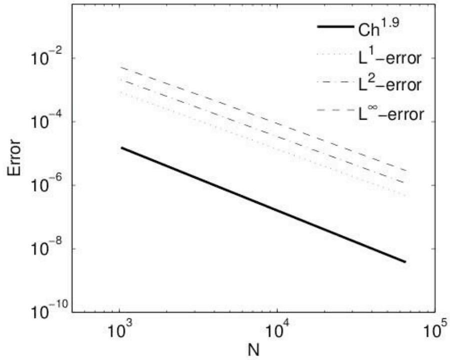


Figure 3: *Example 3: Turing model for* $c = 0$*. Errors in different norms for the MR method at* $t = 1.5$*.*

[2] R. Bürger, R. Ruiz-Baier, K. Schneider, M. Sepúlveda, *Fully adaptive multiresolution schemes for strongly degenerate parabolic equations in one space dimension*, M2AN Math. Model. Numer. Anal. 42 (2008) 535–563.

[3] A. Cohen, S. Kaber, S. Müller, M. Postel, *Fully adaptive multiresolution finite volume schemes for conservation laws*, Math. Comp. 72 (2003) 183–225.

[4] R. Eymard, Th. Gallouët, R. Herbin, *Finite Volume Methods*. In: P.G. Ciarlet, J.L. Lions (eds.), Handbook of Numerical Analysis, vol. VII, North-Holland, Amsterdam, (2000) 713–1020.

[5] H. Holden, K.H. Karlsen, N.H. Risebro, *On uniqueness and existence of entropy solutions of weakly coupled systems of nonlinear degenerate parabolic equations*, Electron. J. Differential Equations 46 (2003), pp. 1–31.

[6] J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, Third Edition, Springer-Verlag, New York (2003).

[7] Y. Nec, A.A. Nepomnyashchy, *Turing instability of anomalous reaction-anomalous diffusion systems*, Eur. J. Appl. Math. 19 (2008), pp. 329–349.

[8] O. Roussel, K. Schneider, A. Tsigulin, H. Bockhorn, *A conservative fully adaptive multiresolution algorithm for parabolic PDEs*, J. Comput. Phys. 188 (2003) 493–523.

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 81–90

UN PROBLEMA EXTREMAL PARA UN CONDUCTOR DE DOS FASES EN UNA BOLA

CARLOS CONCA, RAJESH MAHADEVAN, LEÓN SANZ

Departamento de Matemáticas Universidad de Chile.
Departamento de Matemáticas Universidad de Concepción.

cconca@dim.uchile.cl
rmahadevan@udec.cl
lsanz@dim.uchile.cl

Resumen

En este artículo científico estudiamos el problema de minimizar el primer valor propio de un conductor compuesto por dos materiales homogéneos, que son distribuidos en proporciones fijas dentro de un dominio.

Los trabajos pioneros de F. Murat y L. Tartar (1970-1980) muestran que esta clase de problemas del cálculo de variaciones podrían tener existencia de minimizadores sólo en una clase más grande, llamada clase de materiales homogeneizados o con micro-estructura, excluyendo a priori distribuciones clásicas de material como soluciones optimales.

Para dominios en una dimensión, M. G. Kreĭn (1955) probó la existencia de una solución clásica. En dimensiones más altas, cuando el problema se restringe a una bola, A. Alvino, P. L. Lions y P. L. Trombetti (1989) probaron que se pueden obtener soluciones clásicas radialmente simétricas. Sin embargo, estos resultados han sido vistos como excepcionales, atribuidos a la completa simetría del dominio.

S. Cox y R. Lipton (1996), sólo estudiaron condiciones para un diseño óptimo del problema, asumiendo soluciones homogeneizadas. Aún es desconocido si en dominios con simetría parcial es posible o no obtener una solución clásica que respete la simetría del dominio.

Esperamos revivir el interés a esta pregunta dando una nueva prueba del resultado en una bola. Creemos además que, para este caso en particular, distribuir el material de mayor conductividad en el centro es una solución óptima. Apoyamos esta conjetura con un resultado parcial demostrado utilizando las técnicas de derivación con respecto al dominio.

1 Introducción

Sea Ω un dominio acotado de $\mathbb{R}^N$ y m un número positivo, $0 < m < |\Omega|$, donde Ω es el volumen total (medida de Lebesgue) de la región Ω. Dos materiales de conductividad α y β $(0 < \alpha < \beta)$ son distribuidos en subconjuntos arbitrarios de Ω disjuntos ω y $\Omega\backslash\omega$ respectivamente, de tal manera que $|B| = m$. Para cada una de estas formas de distribuir los materiales α y β, analicemos el **primer valor propio** del problema espectral

$$\left\{ \begin{array}{rcll} -div((\alpha\chi_{\Omega\backslash\omega} + \beta\chi_\omega)\nabla u) & = & \lambda u & \text{en} \quad \Omega \\ u & = & 0 & \text{sobre} \quad \partial\Omega \end{array} \right. , \tag{1}$$

que se expresa mediante el cuociente de Rayleigh, por

$$\lambda_1(\omega) := \min_{u \in H_0^1(\Omega)} \frac{\int_\Omega (\alpha\chi_{\Omega\backslash\omega} + \beta\chi_\omega) |\nabla u|^2 \, dx}{\int_\Omega |u|^2 \, dx} \tag{2}$$

Sea $\mathcal{A} := \{\omega : \omega \subset \Omega, B \text{ medible}, |B| = m\}$ la clase de dominios admisibles asociado a esta configuración. Nos interesa el problema de minimizar el primer valor propio de (1), esto es,

$$\inf \{\lambda_1(\omega) : \omega \in \mathcal{A}\} . \tag{\textbf{PG}}$$

En este trabajo, estudiamos el problema (**PG**) cuando el dominio es una bola en $\mathbb{R}^N$. En virtud de la geometría radial del problema, utilizaremos las técnicas de rearreglos como estrategia para resolverlo. En efecto, en este trabajo se desarrolla una demostración más simple de un resultado de existencia debido a Alvino, Trombetti y Lions [1], el cual dice que, cuando el dominio es una bola, existe una solución radial para (**PG**) en el conjunto de dominios admisibles $\mathcal{A}$.

Uno de los problemas que nos interesa es el de caracterizar la solución. Se conjetura que la solución óptima consiste en distribuir todo el material de conductividad β en el centro de la bola. Se espera poder potenciar esta conjetura en base a un análisis del signo de esta derivada, utilizando las herramientas de derivación con respecto al dominio.

2 Notación y Preliminares

En lo que sigue Ω será la bola unitaria en $\mathbb{R}^N$ centrada en el origen. Cuando f sea una función real evaluada que no se anule, su recíproco lo denotaremos f^{-1}. Dada una función medible $f : \Omega \to \mathbb{R}$ y un número real c, $\Omega_{f,c}$ denotará el conjunto de nivel

$$\Omega_{f,c} := \{x \in \Omega : f(x) \geq c\}$$

el cual es un subconjunto medible de Ω. Denotamos por $\Omega^*_{f,c}$ la bola concéntrica a Ω que tiene la misma medida de Lebesgue que $\Omega_{f,c}$.

La **simetrización de Schwarz** de la función f es la función radialmente simétrica f^* definida en Ω, dada por

$$f^*(z) := \sup\{c \in \mathbb{R} : z \in \Omega^*_{f,c}\} .$$

Se sigue de la definición que la función f y su simetrización de Schwarz f^* que estas son *equimedibles*, en el sentido que

$$|x \in \Omega : f(x) \geq c| = |z \in \Omega : f^*(z) \geq c|,$$

donde $|\cdot|$ corresponde a la medida de Lebesgue. Esta propiedad tiene varias consecuencias importantes. Para una función medible $h : \mathbb{R} \to \mathbb{R}$, se satisface:

$$\int_\Omega h(f(x))\, dx = \int_\Omega h(f^*(z))\, dz\,. \tag{3}$$

En particular, se tiene

$$\int_\Omega |f(x)|^2\, dx = \int_\Omega |f^*(z)|^2\, dz\,. \tag{4}$$

Otra propiedad fundamental es la desigualdad isoperimétrica

$$P\{f \geq c\} \geq P\{f^* \geq c\} \tag{5}$$

donde P, cuando está definido, denota el perímetro de un subconjunto en Ω.

Nota 2.1 *Otras versiones de las propiedades dadas anteriormente son tambien válidas en otros tipos de simetrización. Un tratamiento extensivo de simetrización y sus aplicaciones pueden ser encontradas en las monografías [8, 9].*

3 Existencia

Re-escribamos el funcional $\lambda^1(\omega)$ definido en (2), que depende del lugar que ocupa el material β, definiendo $\nu := \alpha\chi_{\Omega\setminus B} + \beta\chi_B$ y escribiendo ahora $\lambda^1(\nu)$. Sea $\theta := \alpha\chi_{\Omega\setminus B_0} + \beta\chi_{B_0}$ una distribución fija de materiales, donde B_0 es la bola centrada en el origen centrada en 0 que tiene medida de Lebesgue $|\omega|$. En lo que sigue mostraremos un resumen de los resultados de existencia descritos en el artículo [12].

Proposición 3.1 *El problema de minimización (**PG**) puede ser re-escrito como*

$$\operatorname{ínf}\left\{\lambda^1(\nu) : \nu^* = \theta\right\}\,. \tag{6}$$

De la misma manera, si definimos $\eta(\xi) = \lambda^1\left(\xi^{-1}\right)$, el problema de minimización se puede re-escribir nuevamente como

$$\operatorname{ínf}\left\{\eta(\xi) : \xi^* = (\theta)^{-1}\right\}^*\,. \tag{7}$$

Nota 3.1 *Para calcular el ínfimo en las formulaciones anteriores, en general se necesita calcular la clausura del conjunto factible con respecto a una topología adecuada y luego, la envoltura semi-continua inferior del funcional objetivo con respecto a esta topología.*

Notemos que el conjunto de restricciones ya sea en la formulación (6) o (7) es de la forma

$$C(\varphi) = \{f : f^* = \varphi\} \tag{8}$$

donde φ es una función no negativa, acotada, medible y radialmente simétrica, definida en la bola Ω . Conjuntos de este tipo son relativamente compactos para la topología débil-$*$, mas no son cerrrados. Su clausura $K(\varphi)$ tiene la siguiente caracterización:

Proposición 3.2 *[1]. La clausura débil-$*$ de $C(\varphi)$, denotada $K(\varphi)$, es un conjunto compacto y convexo de funciones $f \in L^\infty(\Omega)$ caracterizadas por la propiedad:*

$$\int_{B(0,r)} f(x)\,dx \leq \int_{B(0,r)} \varphi(z)\,dz \;\forall r \qquad y \int_\Omega f(x)\,dx = \int_\Omega \varphi(z)\,dz$$

y lo anterior se mantiene válido si consideramos $C^s(\varphi)$ y $K^s(\varphi)$ que consiste en las funciones radialmente simétricas en $C(\varphi)$ y $K(\varphi)$, respectívamente.

Además, se tiene que:

- *El conjunto $C(\varphi)$ es el conjunto de puntos extremos de $K(\varphi)$.*
- *El conjunto $C^s(\varphi)$ es el conjunto de puntos extremos de $K^s(\varphi)$.*

Nota 3.2 *El funcional λ^1 de la formulación (6) no es semi-continuo para la topología débil-$*$. La teoría de la Homogeneización de Murat-Tartar nos permite obtener su envoltura semi-continua inferior. Sin embargo, a partir de ésta, no es fácil obtener la existencia de una solución clásica. (Ver Cox y Lipton [5]).*

Por otra parte, debido al siguiente resultado de simetrización:

Proposición 3.3 *[1, 12] Dado cualquier $\nu \in C(\theta)$ y cualquier $u \in H_0^1(\Omega)$, existe $\tilde{\nu} \in K\left((\theta^{-1})^*\right)$ radialmente simétrico tal que*

$$\int_\Omega \nu|\nabla u|^2\,dx \geq \int_\Omega \tilde{\nu}|\nabla u^*|^2\,dx$$

y la débil-$*$ continuidad del funcional η cuando es restringido a $K^s(\varphi)$ [12], podemos obtener minimizadores radialmente simétricos en la formulación (7). Como el recíproco del funcional η es convexo, y la formulación (7) es equivalente a maximizar η^{-1} sobre un conjunto de restricciones convexo (y compacto), se tiene una solución, la cual es un punto extremo. De esta manera, tenemos el siguiente teorema.

Teorema 3.1 *[1, 12] Sea Ω una bola en $\mathbb{R}^n$. Dados dos materiales conductores con coeficientes de conductividad $\alpha < \beta$, el problema (**PG**) de minimizar el primer valor propio, admite una solución radialmente simétrica.*

4 Caracterización

Finalmente queremos saber el lugar exacto en el cual distribuir los dos materiales para obtener el valor más bajo. Tenemos la siguiente conjetura.

Conjetura 4.1 *Cuando el dominio es una bola, entre todas las distribuciones radialmente simétricas con volumen fijo, la configuración donde todo el material β está distribuido en el medio de manera radialmente simétrica entrega el primer valor propio más bajo.*

Para buscar esta distribución, procedemos con la teoría de optimización de forma, estudiando el problema mediante las herramientas de derivación con respecto al dominio. En los siguientes párrafos se muestra el desarrollo que nos permite obtener una medida de la variación que sufre el primer valor propio cuando se perturba levemente la distribución de los materiales. Esta cantidad es la derivada con respecto al dominio del primer valor propio asociado al problema de conductividad. Más detalles de la teoría se pueden encontrar en [15, 14].

Recordemos el problema espectral asociado al primer valor propio. Sea ω la configuración de referencia con frontera suave donde el material β es distribuido. Tomamos ω un abierto cuya clausura está contenida en Ω. Dada la distribución de los conductores $\sigma(\omega) = \alpha\chi_{\Omega\setminus\omega} + \beta\chi_\omega$, consideramos el problema de valores propios

$$\left\{ \begin{array}{rcl} -\mathrm{div}\,(\sigma(\omega)\nabla u) & = & \lambda(\omega)\; u \text{ en } \Omega \\ u & = & 0 \text{ sobre } \partial\Omega \end{array} \right. \tag{9}$$

Sea $\lambda_1(\omega)$ el primer valor propio. Se sabe que es simple y la primera función propia asociada está caracterizada por su signo constante [7]. Normalizamos la primera función propia asumiendo que está es no negativa y tomándola que satisfaga

$$\int_\Omega |u|^2\, dx = 1\,. \tag{10}$$

Nos interesa perturbar el problema espectral. Sea θ un campo vectorial suave con su soporte dentro de Ω. Las perturbaciones admisibles de la región a ocupar por el material β son las regiones $\omega_t := \Phi_t\,(\omega)$ que corresponden a las imágenes de ω bajo las transformaciones $\Phi_t := \mathrm{id} + t\theta$ para $t > 0$ suficientemente pequeño. Para tales t, Φ_t es un difeomorfismo en Ω y se tiene $\omega_t \subset\subset \Omega$. Sea $(\lambda_1(\omega_t), u_t)$ el par propio normalizado correspondiente al primer valor propio del problema (9) asociado a la configuración $\sigma(\omega_t)$.

Definición 4.1 *La derivada con respecto al dominio de λ, la derivada local (también llamada derivada de forma) y la derivada total (también llamada derivada de material) de u en la dirección θ en el la configuración ω son, respectívamente, cuando están definidas,*

$$\begin{aligned} \lambda' &= \lim_{t\to 0} \frac{\lambda_t - \lambda}{t} \\ u'(\cdot) &= \lim_{t\to 0} \frac{u_t(\cdot) - u(\cdot)}{t} \\ \dot{u}(\cdot) &= \lim_{t\to 0} \frac{u_t(\cdot + t\theta) - u(\cdot)}{t}, \end{aligned}$$

donde los últimos dos límites se entienden en un espacio adecuado al que pertenezcan las funciones.

Teorema 4.1 *La derivada con respecto al dominio del primer valor propio λ_1 existe. La derivada de material $\dot{u}$ de la primera función propia normalizada u existe y $\dot{u} \in H_0^1(\Omega)$. Su derivada u' también existe y es tal que su restricción ω y $\Omega \setminus \overline{\omega}$ pertenecen a $H^1(\omega)$ y $H^1(\Omega \setminus \overline{\omega})$ respectivamente.*

El argumento para probar la existencia de estas cantidades utiliza el Teorema de la Función Implícita siguiendo un procedemiento establecido el cual es bien explicado en en el documento [14]. La existencia de la derivada con respecto al dominio del primer valor propio y la existencia de la derivada de material pueden ser vistas como un resultado de existencia de una familia suficientemente suave de soluciones para una reformulación del problema perturbado de valores propios, el cual se escribe como

$$\left\{ \begin{array}{rcl} -\mathrm{div}\,(\sigma(\omega_t)\nabla u_t) & = & \lambda(\omega_t)\; u_t \text{ en } \Omega \\ u_t & = & 0 \text{ sobre } \partial\Omega\,. \end{array} \right. \tag{11}$$

el cual después de un cambio de coordenadas Φ_t^{-1} se traduce a

$$\left\{ \begin{array}{rcl} -\mathrm{div}\,((\sigma(\omega_t)\circ\Phi_t)\, A_t\nabla\,(u_t\circ\Phi_t)) & = & \lambda(\omega_t)\,(u_t\circ\Phi_t)\, J(\Phi_t) \text{ en } \Omega \\ u_t\circ\Phi_t & = & 0 \text{ sobre } \partial\Omega \end{array} \right. \tag{12}$$

donde $A_t := D\Phi_t^{-1}\left(D\Phi_t^{-1}\right)^T J(\Phi_t)$ y $J(\Phi_t)$ es el Jacobiano de la transformación Φ_t. Hacemos referencia a [14, 13] para ver más detalles. Con esta transformación, $(\lambda(\omega_t)\,,\, u_t)$ es un par propio de (11) si y sólo si $(\lambda(\omega_t)\,,\, u_t\circ\Phi_t)$ satisface la ecuación (12). Además, u_t satisface la condición $\displaystyle\int_\Omega u_t^2\; dx = 1$ si y solo si se tiene

$$\int_\Omega |u_t\circ\Phi_t|^2\, J(\Phi_t)dx = 1\,. \tag{13}$$

Definimos $F : \mathbb{R}\times\mathbb{R}\times H_0^1(\Omega) \to H^{-1}(\Omega)\times\mathbb{R}$

$$\begin{aligned} F(t,\lambda,v) \quad &:= \left(-\mathrm{div}\,((\sigma(\omega_t)\circ\Phi_t)\, A_t\nabla v) - \lambda v\,,\; \int_\Omega |v|^2\, J(\Phi_t)dx - 1\right) \\ &= \left(-\mathrm{div}\,(\sigma(\omega)A_t\nabla v) - \lambda v\,,\; \int_\Omega |v|^2\, J(\Phi_t)dx - 1\right) \end{aligned} \tag{14}$$

en una vecindad de $(0,\lambda_1(\omega),u_0)$. Notar que la ultima igualdad en (14) es debida al hecho que $\sigma(\omega_t)\circ\Phi_t \equiv \sigma(\omega)$.

Así, verificando las hipótesis del teorema de la función implícita se puede demostrar la existencia de una curva suave de ceros $t \mapsto (t,\lambda_t,v_t)$ para F en torno a $(0,\lambda_1,u_0)$ [13].

Para concluir la existencia de las derivadas se prosigue de la siguiente manera. Notando que el primer valor propio es simple, se deduce que λ_t corresponde al primer valor propio del problema perturbado $\lambda^1(\omega_t)$. Así, $u_t = v_t\circ\Phi_t^{-1}$ es la primera función propia de (11) la cual también satisface la condición de normalización. La diferenciabilidad de la trayectoria de soluciones (λ_t, v_t) con respecto a t nos permite concluir inmediatamente, que las derivada de $\lambda_1(\omega_t)$ y $u_t \circ \Phi_t$ existen, esto es, la derivada con respecto al dominio λ_1 y la derivada de material de u existen.

Finalmente, la existencia de la derivada con respecto al dominio de u existe debido a la relación entre la derivada local y total (ver Simon [15])

$$u'(x) = \dot{u}(x) - \theta \cdot \nabla u(x) \tag{15}$$

donde u es la función en el dominio sin perturbar. Por otro lado hemos visto que $\dot{u} \in H_0^1(\Omega)$. Además, cuando ω es un dominio suave, u satisface un problema de valores propios a coeficientes suaves en ω y $\Omega \setminus \overline{\omega}$, y por la teoría estándar de regularidad, u y ∇u es suave en cada uno de estos dominios. De la relación (15), podemos concluir que $u' \llcorner_{\omega} \in H^1(\omega)$ y $u' \llcorner_{\Omega \setminus \overline{\omega}} \in H^1(\Omega \setminus \omega)$. □

Nota 4.1 *El teorema anterior muestra la Gateaux diferenciabilidad de la primera función propia u en la dirección del campo vectorial θ. La misma prueba modificada, considerando ahora las deformaciones* $\mathrm{id} + \theta$, *con θ suficientemente pequeño, mostrarían la Fréchet diferenciabilidad con respecto a θ.* □

Además, tenemos la siguiente caracterización de la derivada con respecto al dominio del primer valor propio. Para los detalles, ver [13].

Teorema 1 *La derivada con respecto al dominio λ_1, dada una perturbación admisible θ, está dada por la siguiente fórmula*

$$\lambda_1'(\omega;\theta) = \int_{\partial\omega} \left[\sigma \left|\nabla u\right|^2\right] \theta \cdot n dS - \lambda^1(\omega) \int_{\partial\omega} \left[u^2\right] \theta \cdot n dS + \int_{\partial\omega} \left[\theta \cdot \nabla u\right] (\sigma \nabla u) \cdot n dS$$

donde $[\varphi]$ es el salto de φ sobre $\partial\omega$, esto es, $[\varphi](x) = \left(\varphi\llcorner_{\partial\omega^-} - \varphi\llcorner_{\partial\omega^+}\right)(x)$ con $\varphi\llcorner_{\partial\omega^-}$ y $\varphi\llcorner_{\partial\omega^+}$ denotando, respectivamente la interior y exterior traza de φ en $\partial\omega$.

La siguiente proposición, la cual parcialmente justifica la conjetura (4.1) es demostrada en [13] utilizando la formula de la derivada con respecto al dominio dada arriba.

Proposición 4.1 *Cuando el dominio es una bola y ubicamos el material β en un anillo concéntrico a la bola, el primer valor propio se puede disminuir, eligiendo una perturbación θ que mueva el anillo hacia adentro o hacia afuera, haciendo que la cantidad*

$$\lambda_1'(\omega_0;\theta) = 3\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)\left\{\left((\sigma\left|\nabla u\right|)_{S_1}\right)^2 - \left((\sigma\left|\nabla u\right|)_{S_2}\right)^2\right\} \theta \cdot n_{S_1} \mathrm{per}(S_1) \tag{16}$$

sea positiva, donde S_1 y S_2 corresponden a la frontera interna y externa del anillo donde se ubica el material β.

En los gráficos 1 y 2 mostramos evidencia numérica de la conjetura. Para el experimento, consideramos el dominio Ω el disco unitario en $\mathbb{R}^2$. El material de conductividad β se distribuye en un anillo concéntrico, el cual es parametrizado por el radio r_1 de la frontera interna, manteniendo constante el área de distribución de materiales. Para valores de $\beta = 2, 200$ y proporciones de material 0,1, 0,5 y 0,9, se muestra el comportamiento del primer valor propio y de su derivada con respecto al dominio.

Observamos de los gráficos que el valor propio más pequeño se ubica distribuyendo el material de conductividad β en el centro. Además, el comportamiento del signo de la derivada con respecto al dominio del primer valor propio, se condice con el comportamiento de crecimiento del primer valor propio.

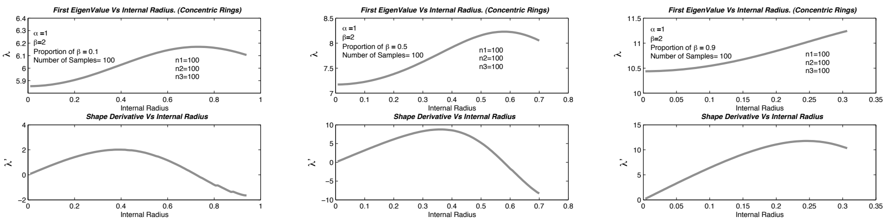


Figura 1: Discos concéntricos para $\beta = 2$

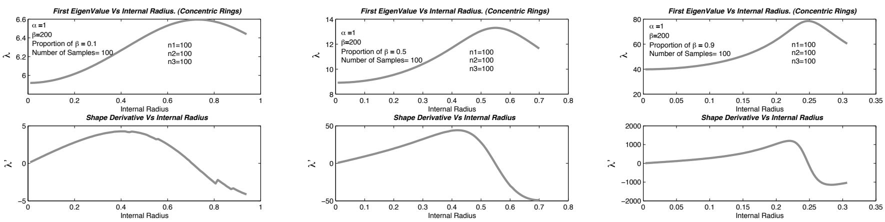


Figura 2: Discos concéntricos para $\beta = 200$

5 Preguntas Abiertas

Si tenemos un configuración finita de anillos de interior no vacío en los cuales distribuimos el material de conductividad β, en caso de tener el valor de flujo $\sigma\,|\nabla u|$ en cada frontera anular, la fórmula (16) nos entrega una dirección de decrecimiento para el primer valor propio. Dicho de otra manera, la proposición (4.1) nos permite tener un algoritmo de búsqueda de configuraciones minimizantes para nuestro problema.

Las experiencias numéricas confirman la conjetura, sin embargo, de los resultado obtenidos no es posible obtener una demostración analítica de este comportamiento. Algún estudio de mínimos locales para el funcional podría fortalecer la conjetura. Por otro lado, todavía no es claro si se pueden excluir soluciones minimizantes que no permitan ser estudiadas con el cálculo de la derivada con respecto al dominio. Este el caso de configuraciones que poseen una cantidad infinita no-numerable de anillos de interior vacío, tales que su unión es un conjunto de medida positiva. Configuraciones con conjuntos de nivel radialmente simétricos, con cortes radiales tipo Cantor, pertenecen a esta configuración, patológica. Finalmente, invitamos a estudiar la existencia y caracterización de soluciones optimales para otro tipo de simetrías, como por ejemplo, dominios cuadrados o estrellados.

Referencias

[1] Alvino A, Lions PL, Trombetti G (1989) Optimization problems with prescribed rearrangements. Nonlinear Analysis TMA 13(2): 185–220.

[2] Alvino A, Trombetti G (1983) A lower bound for the first eigenvalue of an elliptic operator. Jl. of Math. Anal. Appl. 94: 328–337.

[3] Cox S, McLaughlin JR(1990) Extremal eigenvalue problems for composite membranes I. Appl. Math. Optim. 22(2): 153–167.

[4] Cox S, McLaughlin JR(1990) Extremal eigenvalue problems for composite membranes II. Appl. Math. Optim. 22(2): 169–187.

[5] Cox S, Lipton R(1996) Extremal eigenvalue problems for two-phase conductors. Arch. Rational Mech. Anal. 136: 101–117.

[6] Henrot H(2006) Extremum Problems for Eigenvalues of Elliptic Operators. Birkhaüser.

[7] Kreĭn MG(1955) On certain problems on the maximum and minimum of characteristic values and on the Lyapunov zones of stability. AMS Translations Ser. 2(1): 163–187.

[8] Kawohl B Rearrangement and Convexity of Level Sets in PDE, LNM 1150. Springer-Verlag.

[9] Kesavan S(2006) Symmetrization and Applications. World Scientific.

[10] Murat F(1978) H-convergence. Séminaire d'analyse fonctionelle et numérique, Univ. d'Alger., mimeographed notes.

[11] Murat F, Tartar L(1997) Calculus of Variations and Homogenization (engl. transl. of original french article) in "Topics in the Mathematical Modelling of Composite materials" Eds. A. Cherkaev and R.V. Kohn, PNLDE 31. Birkhaäuser.

[12] Conca C, Mahadevan R, Sanz L(2008) An extremal eigenvalue problem for a two-phase conductor in a ball, published online in *Appl. Math. Optim.*

[13] Conca C, Mahadevan R, Sanz L(2008) Shape derivative for a two-phase eigenvalue problem and optimal configurations in a ball. To Appear in *ESAIM Proceedings* .

[14] Henrot A, Pierre M (2005) Variation et Optimisation de Formes. Mathématiques et Applications 48. Springer.

[15] Simon J (1980) Differentiation with respect to the domain in boundary value problems. Numer. Funct. and Optimiz. 2(7-8), 649–687.

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 91–98

MÉTODO DE ELEMENTOS FINITOS PARA LA APROXIMACIÓN DE UN MODELO DE CRISTALES LíQUIDOS NEMÁTICOS

F. GUILLÉN-GONZÁLEZ, J.V. GUTIÉRREZ-SANTACREU

Dpto. Ecuaciones Diferenciales y Análisis Numérico
Dpto. Matemática Aplicada I
Universidad de Sevilla

guillen@us.es, juanvi@us.es

Resumen

En esta charla analizamos la aproximación numérica con elementos finitos en espacio y diferencias finitas en tiempo de un modelo de cristales líquidos nemáticos (de tipo Eriksen-Leslie) y de un modelo penalizado de tipo Ginzburg-Landau. Después de describir los principales antecedentes del tema, se propone un esquema lineal totalmente acoplado y condicionalmente estable. La convergente (respecto de los parámetros de discretización y del parámetro de penalización) hacia una solución débil del problema de Eriksen-Leslie queda como problema abierto.

Palabras clave: *Cristales líquidos, Navier-Stokes, estabilidad, convergencia, elementos finitos*
Clasificación por materias AMS: *35Q35 65M12 65M60*

1 Introducción

Supongamos $\Omega \subset \mathbb{R}^3$ un dominio acotado y de frontera Γ poliédrica tal que el problema de Stokes tenga regularidad $\boldsymbol{H}^2(\Omega) \times L^2(\Omega)$ en velocidad y presión. Denotamos $Q = \Omega \times (0,T)$ y $\Sigma = \Gamma \times (0,T)$, donde $[0,T]$ es el intervalo temporal de observación, para $T > 0$. Las incógnitas son: $\boldsymbol{u} : Q \to \mathbb{R}^3$, el campo de velocidades, $p : Q \to \mathbb{R}$, la presión del fluido, y $\boldsymbol{d} : Q \to \mathbb{R}^3$, el vector orientación de las macromoléculas de cristales líquidos, verificando el problema en derivadas parciales (de tipo Eriksen-Leslie):

$$\left\{\begin{array}{rcll} |\boldsymbol{d}| = 1, \quad \partial_t \boldsymbol{d} + \boldsymbol{u}\cdot\nabla\boldsymbol{d} - \gamma\Delta\boldsymbol{d} - \gamma|\nabla\boldsymbol{d}|^2\boldsymbol{d} & = & \boldsymbol{0} & \text{en } Q, \\ \partial_t\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} - \nu\Delta\boldsymbol{u} + \nabla p + \lambda\nabla\cdot((\nabla\boldsymbol{d})^t\nabla\boldsymbol{d}) & = & \boldsymbol{0} & \text{en } Q, \\ \nabla\cdot\boldsymbol{u} & = & 0 & \text{en } Q, \\ \boldsymbol{u} = 0, \quad \boldsymbol{d} & = & \boldsymbol{l} & \text{en } \Sigma, \\ \boldsymbol{u}_{|t=0} = \boldsymbol{u}_0, \quad \boldsymbol{d}_{|t=0} & = & \boldsymbol{d}_0 & \text{en } \Omega, \end{array}\right. \qquad (1)$$

donde $\boldsymbol{u}_0 : \Omega \to \mathbb{R}^3$ y $\boldsymbol{d}_0 : \Omega \to \mathbb{R}^3$ son los datos iniciales, $\boldsymbol{l} : \partial\Omega \to \mathbb{R}^3$ es el dato de Dirichlet para $\boldsymbol{d}$, $\nu > 0$ es una constante dependiente de la viscosidad del fluido, $\lambda > 0$ es una constante de elasticidad y $\gamma > 0$ es una constante del tiempo de relajación. $(\nabla\boldsymbol{d})^t$ denota la matriz traspuesta de $\nabla\boldsymbol{d} = (\partial_j d_i)_{i,j}$.

Para este trabajo se supone el dato de Dirichlet para $\boldsymbol{d}$ independiente del tiempo y la condición de compatibilidad $\boldsymbol{d}_0\big|_{\partial\Omega} = \boldsymbol{l}$. En principio los argumentos que desarrollaremos no son válidos para el caso dependiente del tiempo.

Este modelo se estudia a través del modelo penalizado, de tipo Ginzburg-Landau, que se obtiene de (1) relajando la restricción $|\boldsymbol{d}| = 1$ por $|\boldsymbol{d}| \le 1$, y en el sistema para $\boldsymbol{d}$ cambiando el término más no lineal $|\nabla\boldsymbol{d}|^2\boldsymbol{d}$ (que es el multiplicador de Lagrange asociado a la restricción $|\boldsymbol{d}| = 1$) por el término de penalización $\boldsymbol{f}_\varepsilon(\boldsymbol{d}) = \varepsilon^{-2}(|\boldsymbol{d}|^2 - 1)\boldsymbol{d}$, asociado al parámetro $\varepsilon > 0$, llegando al sistema

$$\left\{\begin{array}{rcll} |\boldsymbol{d}| \le 1, \quad \partial_t\boldsymbol{d} + \boldsymbol{u}\cdot\nabla\boldsymbol{d} + \gamma(\boldsymbol{f}_\varepsilon(\boldsymbol{d}) - \Delta\boldsymbol{d}) & = & \boldsymbol{0} & \text{en } Q, \\ \partial_t\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} - \nu\Delta\boldsymbol{u} + \nabla p + \lambda\nabla\cdot((\nabla\boldsymbol{d})^t\nabla\boldsymbol{d}) & = & \boldsymbol{0} & \text{en } Q, \\ \nabla\cdot\boldsymbol{u} & = & 0 & \text{en } Q, \end{array}\right. \tag{2}$$

junto con las condiciones iniciales y de contorno de (1). Obsérvese que $\boldsymbol{f}_\varepsilon(\boldsymbol{d}) = \nabla_{\boldsymbol{d}}(F_\varepsilon(\boldsymbol{d}))$ para cada $\boldsymbol{d} \in \mathbb{R}^3$ con F_ε la función potencial:

$$F_\varepsilon(\boldsymbol{d}) = \frac{1}{4\varepsilon^2}(|\boldsymbol{d}|^2 - 1)^2.$$

El problema (2) tiene la siguiente igualdad de energía (usada por *F. H. Lin* y *C. Liu* en [7] para obtener solución débil de (2)), imponiendo que $|\boldsymbol{d}_0| = 1$ en Q y $|\boldsymbol{l}| = 1$ sobre Σ:

$$\left\{\begin{array}{l} \dfrac{d}{dt}\left(\dfrac{1}{2}\|\boldsymbol{u}\|^2_{L^2(\Omega)} + \dfrac{\lambda}{2}\|\nabla\boldsymbol{d}\|^2_{L^2(\Omega)} + \lambda\displaystyle\int_\Omega F_\varepsilon(\boldsymbol{d})d\boldsymbol{x}\right) \\ +\nu\|\nabla\boldsymbol{u}\|^2_{L^2(\Omega)} + \lambda\gamma\|\boldsymbol{f}_\varepsilon(\boldsymbol{d}) - \Delta\boldsymbol{d}\|^2_{L^2(\Omega)} = 0, \end{array}\right. \tag{3}$$

que se obtiene eligiendo $\lambda(\boldsymbol{f}_\varepsilon(\boldsymbol{d}) - \Delta\boldsymbol{d})$ y $\boldsymbol{u}$ como funciones test en $(2)_a$ y $(2)_b$, respectivamente. Esta expresión refleja la relación entre la energía cinética $\frac{1}{2}\|\boldsymbol{u}\|^2_{L^2(\Omega)}$ y la elástica $\frac{\lambda}{2}\|\nabla\boldsymbol{d}\|^2_{L^2(\Omega)} + \lambda\int_\Omega F_\varepsilon(\boldsymbol{d})$ e indica que la energía total del sistema decrece con el tiempo. Luego, una solución débil para (1) tendrá la regularidad

$$\boldsymbol{d} \in L^\infty(0,T;\boldsymbol{H}^1(\Omega)), \quad \mathbf{u} \in L^\infty(0,T;\boldsymbol{L}^2(\Omega)) \cap L^2(0,T;\boldsymbol{H}^1(\Omega)). \tag{4}$$

2 Antecedentes

Los primeros autores que tratan desde el punto de vista numérico el modelo (2) fueron C. Liu y N. J. Walkington en [6]. Proponen un método mixto para aproximar el sistema de momentos usando un par de elementos finitos globalmente C^0 que verifican la condición de estabilidad de tipo Babuska-Brezzi, y una aproximación de elementos finitos globalmente C^1 para el vector de orientación. Pero resulta que la implementación de los elementos finitos globalmente C^1 es muy complicada.

En un trabajo posterior [9], estos mismos autores construyen un esquema mixto para el vector de orientación para evitar los elementos finitos de tipo C^1. Ello lo logran introduciendo una nueva variable $\boldsymbol{w} = \nabla \boldsymbol{d}$, la cual incrementa el número de grados de libertad a calcular en cada etapa de tiempo.

Además, hay que decir que ambos esquemas son no lineales lo cual hacen aún más costoso el cálculo de la solución aproximada. Estos esquemas son analizados usando técnicas de estimaciones de error; suponiendo una solución suficientemente regular del problema continuo (2).

En [4], V. Girault y F. Guillén-González introducen un esquema (lineal) con una variable auxiliar $\boldsymbol{w} = -\Delta \boldsymbol{d}$ para el problema penalizado (2). Dicho esquema resulta ser incondicionalmente estable, respecto de los parámetros de discretización, y convergente hacia (2); obteniéndose además estimaciones de error óptimales y convergencia de métodos iterativos para desacoplar $(\boldsymbol{u}, p)$ de $(\boldsymbol{w}, \boldsymbol{d})$ en cada etapa de tiempo ya que éste resulta totalmente acoplado. Este esquema reduce los grados de libertad a calcular respecto a los esquemas anteriores de [6] y [9] para (2) lo que implica una coste computacional menor, además de ser un esquema lineal.

En [8], P. Lin y C. Liu presentan dos esquemas numéricos lineales que sólo usan elementos finitos globalmente C^0 y localmente P_2 para la velocidad y el vector de orientación, y elementos finitos P_1 para la presión. El primero de ellos discretiza la derivada temporal usando diferencias finitas y el segundo con el método de las características. Estos esquemas son justificados observando el decaimiento de la energía en las experiencias numéricas.

En [2], se proponen un esquema numérico para (2) y otro para (1), considerando la condición frontera para el vector de orientación de tipo Neumann. Para construir el esquema de (2) definen la variable auxiliar $\boldsymbol{w} = -\Delta \boldsymbol{d} + \boldsymbol{f}_\varepsilon(\boldsymbol{d})$ reescribiendo el tensor

$$\lambda \nabla \cdot ((\nabla \boldsymbol{d})^t \nabla \boldsymbol{d}) = \lambda \nabla (\frac{1}{2} |\nabla \boldsymbol{d}|^2 + F_\varepsilon(\boldsymbol{d})) - \lambda (\nabla \boldsymbol{d})^t (-\Delta \boldsymbol{d} + \boldsymbol{f}_\varepsilon(\boldsymbol{d})) = \nabla q - \lambda (\nabla \boldsymbol{d})^t \boldsymbol{w}, \tag{5}$$

lo que introduce una presión modificada. La discretización de la parte no lineal de $\boldsymbol{f}_\varepsilon$ se realiza de forma implícita, resultando un esquema no lineal. Se prueba la estabilidad incondicional, respecto de los parámetros de dicretización y penalización, y convergencia hacia una solución débil de (2). A posteriori, hacen tender a cero el parámetro de penalización obteniendo una solución débil en el sentido con valores medidas de (1), debido a que no consiguen compacidad para $\nabla \boldsymbol{d}$. Para (1) desarrollan un esquema no lineal cuya idea principal es la reescritura del término no lineal

$$-|\nabla \boldsymbol{d}|^2 \boldsymbol{d} - \Delta \boldsymbol{d} = \boldsymbol{d} \times (\boldsymbol{d} \times \Delta \boldsymbol{d}) \quad (\text{usando } |\boldsymbol{d}| = 1)$$

y una discretización temporal de punto medio para este término no lineal. Este esquema es condicionalmente estable, pero la convergencia hacia una solución débil de (1) queda como problema abierto.

Destacamos que ninguno de los esquemas anteriormente descritos para el problema (2) convergen hacia una solución débil del problema (1), si $(h,k,\varepsilon) \to 0$.

3 Esquema numérico

En este trabajo aportamos la construcción un esquema lineal para (2) que sea estable, en el sentido que se tenga una versión discreta de la igualdad de energía (3). Hasta el momento no se sabe pasar al límite en el sistema de momentos hacia una solución débil de (1) con la regularidad (4), haciendo tender ε a cero junto con los parámetros discretos en espacio y tiempo (h,k). La principal dificultad es la compacidad de $\nabla \boldsymbol{d}$ en $\boldsymbol{L}^2(Q)$.

Denotaremos por $\Big(\cdot,\cdot\Big)$ al producto escalar en $L^2(\Omega)$.

El esquema que proponemos para aproximar las incógnitas (velocidad, presión y vector de orientación) está basado en la siguiente formulación variacional mixta del problema (2): $\forall\, \bar{\boldsymbol{u}} \in \boldsymbol{H}_0^1, \bar{\boldsymbol{w}} \in \boldsymbol{L}^2, \bar{q} \in L_0^2, \bar{\boldsymbol{d}} \in \boldsymbol{H}_0^1$,

$$
\begin{array}{rcl}
\Big(\partial_t \boldsymbol{u}, \bar{\boldsymbol{u}}\Big) + \nu\Big(\nabla \boldsymbol{u}, \nabla \bar{\boldsymbol{u}}\Big) + \Big((\boldsymbol{u}\cdot\nabla)\boldsymbol{u}, \bar{\boldsymbol{u}}\Big) - \Big((\nabla \boldsymbol{d})^t \boldsymbol{w}, \bar{\boldsymbol{u}}\Big) - \Big(p, \nabla\cdot\bar{\boldsymbol{u}}\Big) & = & 0,\\
\Big(\partial_t \boldsymbol{d}, \bar{\boldsymbol{w}}\Big) + \Big((\boldsymbol{u}\cdot\nabla)\boldsymbol{d}, \bar{\boldsymbol{w}}\Big) + \gamma\Big(\boldsymbol{f}_\varepsilon(\boldsymbol{d}), \bar{\boldsymbol{w}}\Big) + \gamma\Big(\boldsymbol{w}, \bar{\boldsymbol{w}}\Big) & = & 0,\\
\Big(\nabla\cdot\boldsymbol{u}, \bar{q}\Big) & = & 0,\\
\Big(\nabla \boldsymbol{d}, \nabla \bar{\boldsymbol{d}}\Big) - \Big(\boldsymbol{w}, \bar{\boldsymbol{d}}\Big) & = & 0,
\end{array}
$$

donde hemos usado previamente la igualdad (5) e introducido la variable auxiliar $\boldsymbol{w} = -\Delta \boldsymbol{d} + \boldsymbol{f}_\varepsilon(\boldsymbol{d})$.

Suponemos por simplicidad, una partición uniforme de $[0,T]$ siendo $t_n = nk$, donde $k = T/N$ es el paso de tiempo con $N \in \mathbb{N}$. En cada etapa de tiempo, la velocidad, la presión y el vector de orientación $(\boldsymbol{u}, p, \boldsymbol{d})$ son aproximados en espacios de elementos finitos de funciones globalmente continuas $(\boldsymbol{X}_h, Q_h, \boldsymbol{D}_h) \subset (\boldsymbol{H}_0^1(\Omega), L_0^2(\Omega), \boldsymbol{H}^1(\Omega))$ y el laplaciano del vector de orientación $\boldsymbol{w}$ es aproximado en un espacio de elementos finitos $\boldsymbol{W}_h \subset \boldsymbol{L}^2(\Omega)$. Los elementos finitos están asociados a una familia regular y quasi-uniforme de triangulaciones $\{\mathcal{T}_h\}_{h>0}$ de Ω tales que $(\boldsymbol{X}_h, Q_h)$ verifican la condición *inf-sup* discreta y $(\boldsymbol{W}_h, \boldsymbol{D}_h)$ satisfacen $\boldsymbol{D}_h \subset \boldsymbol{W}_h$. Por ejemplo, la siguiente aproximación verifica las hipótesis anteriores:

$$P_1 + bubble/P_1 \quad \text{para } (\boldsymbol{X}_h, Q_h), \qquad \text{y} \qquad P_1/P_1 \quad \text{para } (\boldsymbol{W}_h, \boldsymbol{D}_h).$$

Además, suponemos las restricciones (de estabilidad) sobre los parámetros (k,h,ε):

$$\text{(S)} \qquad \lim_{(h,k,\varepsilon)\to 0} \frac{k}{h^2\varepsilon^6} = 0 \quad \text{y} \quad \lim_{(h,k,\varepsilon)\to 0} \frac{h}{\varepsilon^2} = 0.$$

Entonces, el algoritmo numérico que presentamos consiste en:

Inicialización: Sea $(\boldsymbol{u}_h^0, \boldsymbol{d}_h^0) \in (\boldsymbol{X}_h, \boldsymbol{D}_h)$ determinadas aproximaciones de $(\boldsymbol{u}_0, \boldsymbol{d}_0)$, con $\boldsymbol{d}_h^0|_{\partial\Omega} = \boldsymbol{l}_h$ para $\boldsymbol{l}_h$ una aproximación frontera de $\boldsymbol{l}$.

Etapa $n+1$**:** Dado $(\boldsymbol{u}_h^n, \boldsymbol{d}_h^n) \in (\boldsymbol{X}_h, \boldsymbol{D}_h)$ con $\boldsymbol{d}_h^n|_{\partial\Omega} = \boldsymbol{l}_h$, encontrar $(\boldsymbol{u}_h^{n+1}, \boldsymbol{w}_h^{n+1}) \in \boldsymbol{X}_h \times \boldsymbol{W}_h$ y $(p_h^{n+1}, \boldsymbol{d}_h^{n+1}) \in Q_h \times \boldsymbol{D}_h$ (con $\boldsymbol{d}_h^{n+1}|_{\partial\Omega} = \boldsymbol{l}_h$) tal que:

$$\begin{aligned}&\Big(\frac{\boldsymbol{u}_h^{n+1} - \boldsymbol{u}_h^n}{k}, \bar{\boldsymbol{u}}_h\Big) + c\Big(\boldsymbol{u}_h^n, \boldsymbol{u}_h^{n+1}, \bar{\boldsymbol{u}}_h\Big) + \nu\Big(\nabla \boldsymbol{u}_h^{n+1}, \nabla \bar{\boldsymbol{u}}_h\Big)\\ &-\lambda\Big((\nabla \boldsymbol{d}_h^n)^t \boldsymbol{w}_h^{n+1}, \bar{\boldsymbol{u}}_h\Big) - \Big(p_h^{n+1}, \nabla\cdot\bar{\boldsymbol{u}}_h\Big) = 0 \quad \forall\, \bar{\boldsymbol{u}}_h \in \boldsymbol{X}_h,\end{aligned} \tag{6}$$

$$\Big(\bar{p}_h, \nabla\cdot \boldsymbol{u}_h^{n+1}\Big) = 0 \quad \forall\, \bar{p}_h \in Q_h, \tag{7}$$

$$\Big(\frac{\boldsymbol{d}_h^{n+1} - \boldsymbol{d}_h^n}{k}, \bar{\boldsymbol{w}}_h\Big) + \Big((\boldsymbol{u}_h^{n+1}\cdot\nabla)\boldsymbol{d}_h^n, \bar{\boldsymbol{w}}_h\Big) + \gamma\Big(\boldsymbol{w}_h^{n+1}, \bar{\boldsymbol{w}}_h\Big) = 0 \quad \forall\, \bar{\boldsymbol{w}}_h \in \boldsymbol{W}_h, \tag{8}$$

$$\Big(\nabla \boldsymbol{d}_h^{n+1}, \nabla \bar{\boldsymbol{d}}_h\Big) + \Big(\boldsymbol{f}_\varepsilon(\boldsymbol{d}_h^n), \bar{\boldsymbol{d}}_h) = \Big(\boldsymbol{w}_h^{n+1}, \bar{\boldsymbol{d}}_h\Big) = 0 \quad \forall\, \bar{\boldsymbol{d}}_h \in \boldsymbol{D}_{0h} := \boldsymbol{D}_h \cap \boldsymbol{H}_0^1(\Omega), \tag{9}$$

donde hemos introducido la forma trilineal $c\Big(\cdot,\cdot,\cdot\Big)$ definida por

$$c\Big(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\Big) = \Big((\boldsymbol{u}\cdot\nabla)\boldsymbol{v}, \boldsymbol{w}\Big) + \frac{1}{2}\Big(\nabla\cdot\boldsymbol{u}\, \boldsymbol{v}, \boldsymbol{w}\Big) \quad \forall\, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}_0^1(\Omega),$$

la cual muestra la propiedad de antisimetría $c\Big(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}\Big) = 0$ aunque $\boldsymbol{u}$ no verifique la condición de incompresibilidad puntualmente. Notar que $\boldsymbol{d}_h^{n+1} - \boldsymbol{d}_h^n \in \boldsymbol{D}_{0h}$ gracias a que el dato de contorno $\boldsymbol{l}$ para el vector de orientación no depende del tiempo.

4 Estabilidad condicional

En líneas generales, tomando primero $\bar{\boldsymbol{u}}_h = 2\,k\,\boldsymbol{u}_h^{n+1}$ en (6), $\bar{p}_h = p_h^{n+1}$ en (7), y luego $\bar{\boldsymbol{w}}_h = 2\,\lambda\, k \boldsymbol{w}_h^{n+1}$ en (8) conjuntamente con $\bar{\boldsymbol{d}}_h = \boldsymbol{d}_h^{n+1} - \boldsymbol{d}_h^n \in \boldsymbol{D}_{0h}$ en (9) y usando las igualdades

$$-\Big((\nabla \boldsymbol{d}_h^n)^t \boldsymbol{w}_h^{n+1}, \boldsymbol{u}_h^{n+1}\Big) + \Big(\boldsymbol{u}_h^{n+1}\cdot\nabla \boldsymbol{d}_h^n, \boldsymbol{w}_h^{n+1}\Big) = 0,$$

$$\Big(\boldsymbol{u}_h^{n+1}\cdot\nabla \boldsymbol{d}_h^n, \boldsymbol{f}_\varepsilon(\boldsymbol{d}_h^n)\Big) + \Big(\nabla\cdot \boldsymbol{u}_h^{n+1}, F_\varepsilon(\boldsymbol{d}_h^n)\Big) = 0,$$

llegamos al siguiente resultado (ver [5] para los detalles). Denotamos por $|\cdot|$ a la norma en $L^2(\Omega)$.

Lema 1 (Desigualdad de energía discreta) *Supongamos que existe una constante $C_d > 0$ independiente de (h, k, ε) tal que $|\boldsymbol{u}_h^n|^2 + \lambda|\nabla \boldsymbol{d}_h^n|^2 \le C_d$. Entonces, existen $h_0 > 0$, $k_0 > 0$ y $\varepsilon_0 > 0$ tales que para todo $h \le h_0$, $k \le k_0$ y $\varepsilon \le \varepsilon_0$ satisfaciendo la hipótesis (S), la solución correspondiente $(\boldsymbol{u}_h^{n+1}, \boldsymbol{d}_h^{n+1}, \boldsymbol{w}_h^{n+1})$ del problema discreto (6)-(9) verifica la siguiente desigualdad:*

$$\begin{aligned}&\Big(|\boldsymbol{u}_h^{n+1}|^2 - |\boldsymbol{u}_h^n|^2\Big) + \nu k |\nabla \boldsymbol{u}_h^{n+1}|^2 + \lambda\Big(|\nabla \boldsymbol{d}_h^{n+1}|^2 - |\nabla \boldsymbol{d}_h^n|^2\Big)\\ &+2\lambda \int_\Omega (F_\varepsilon(\boldsymbol{d}_h^{n+1}) - F_\varepsilon(\boldsymbol{d}_h^n)) + \lambda\gamma k |\boldsymbol{w}_h^{n+1}|^2 \le 0.\end{aligned} \tag{10}$$

Por un proceso de inducción sobre la etapa de tiempo [5] se puede acotar las aproximaciones $(\boldsymbol{u}_h^{n+1}, \boldsymbol{d}_h^{n+1})$ en los espacios discretos de (4) y $\boldsymbol{w}_h^{n+1}$ en $L^2(0,T;\boldsymbol{L}^2(\Omega))$.

5 Compacidad para $\boldsymbol{u}$ y $\boldsymbol{d}$. Convergencia del sistema en $\boldsymbol{d}$

Eligiendo como función test $\bar{\boldsymbol{w}}_h = P_h \bar{\boldsymbol{w}}$ en (8), con $\boldsymbol{w} \in \boldsymbol{L}^3(\Omega)$ y P_h el proyector ortogonal en L^2 sobre $\boldsymbol{W}_h$, y usando las estimaciones del Lema 1 y la estabilidad en $L^3(\Omega)$ de P_h ([3]), obtenemos

$$k \sum_{n=0}^{N-1} \left\| \frac{\boldsymbol{d}_h^{n+1} - \boldsymbol{d}_h^n}{k} \right\|_{L^{3/2}(\Omega)}^2 \leq C.$$

Como consecuencia de esta estimación y de las estimaciones de estabilidad del Lema 1, conseguimos la compacidad de la sucesión $(\boldsymbol{d}_{h,k,\varepsilon})$ en $L^q(0,T;L^r(\Omega))$ con $1 \leq r < 6$ y $1 \leq q < \infty$, donde $\boldsymbol{d}_{h,k,\varepsilon}$ es la función continua y lineal a trozos tal que $\boldsymbol{d}_{h,k,\varepsilon}(t_n) = \boldsymbol{d}_h^n$.

Para la compacidad de la velocidad discreta $\{\boldsymbol{u}_h^{n+1}\}$ consideramos el espacio de velocidades de divergencia discreta cero

$$\boldsymbol{V}_h = \{\boldsymbol{v}_h \in \boldsymbol{X}_h : \Big(\nabla \cdot \boldsymbol{v}_h, q_h\Big) = 0 \;\forall\, q_h \in Q_h\}.$$

Primero, se llega a las siguientes estimaciones de una derivada fraccionaria para $\boldsymbol{u}_{h,k,\varepsilon}$ (la función constante a trozos que vale $\boldsymbol{u}_{h,k,\varepsilon}(t) = \boldsymbol{u}_h^{n+1}$ en $(t_n, t_{n+1}]$):

$$\int_0^{T-\delta} \|\boldsymbol{u}_{h,k,\varepsilon}(t+\delta) - \boldsymbol{u}_{h,k,\varepsilon}(t)\|_{\boldsymbol{V}_h'}^2 \, dt \leq C\,\delta^{1/2} \quad \forall\, \delta : 0 < \delta < T.$$

Nótese que la derivada fraccionaria en tiempo para la velocidad discreta ha sido acotada en la norma $\boldsymbol{V}_h'$ la cual "se mueve" con respecto al parámetro de espacio h, por lo que no podemos aplicar los resultados de compacidad dados por *J. Simon* en [10]. La siguiente idea es encontrar una norma, que no dependa de los parámetros de discretización, donde la derivada fraccionaria en tiempo pueda ser acotada. Para ello, consideramos el espacio $\boldsymbol{V} = \{\boldsymbol{u} \in \boldsymbol{H}_0^1(\Omega) \;:\; \nabla \cdot \boldsymbol{u} = 0\}$ y la proyección ortogonal $R_h : \boldsymbol{V}_h \to \boldsymbol{V}$ tal que $\Big(\nabla(R_h \boldsymbol{u}_h - \boldsymbol{u}_h), \nabla \boldsymbol{v}\Big) = 0, \;\; \forall\, \boldsymbol{v} \in \boldsymbol{V}$. Usando este operador R_h se prueba [5] que $\|R_h \boldsymbol{u}_h\|_{\boldsymbol{V}'} \leq C\Big(h|\nabla \cdot \boldsymbol{u}_h| + \|\boldsymbol{u}_h\|_{\boldsymbol{V}_h'}\Big)$ y la estimación

$$\int_0^{T-\delta} \|R_h \boldsymbol{u}_{h,k,\varepsilon}(t+\delta) - R_h \boldsymbol{u}_{h,k,\varepsilon}(t)\|_{\boldsymbol{V}'}^2 \, dt \leq C\,\delta^{1/2} + C\,h.$$

Entonces, la convergencia fuerte de una subsucesión $(R_h \boldsymbol{u}_{h,k,\varepsilon})$ en $L^2(\boldsymbol{L}^2)$ se deduce a partir de un resultado obtenido (de compacidad por perturbación) por P. Azérad y F. Guillén-González [1].

Finalmente, gracias a la aproximación (externa) de $\boldsymbol{V}_h$ a $\boldsymbol{V}$, se tiene [5] la convergencia fuerte de $(\boldsymbol{u}_{h,k,\varepsilon})$ en $L^2(\boldsymbol{L}^2(\Omega))$. Con las convergencias obtenidas hasta el momento es posible pasar al límite en el sistema (8)-(9) y obtener $(\boldsymbol{u}, \boldsymbol{d})$ verificando $(1)_1$.

La convergencia para el sistema de momentos discreto queda condicionada a la compacidad del $\nabla \boldsymbol{d}$. Sin embargo, es posible obtener una convergencia más débil tal y como se hace en [2].

6 Experiencias numéricas

La siguiente experiencia numérica muestra el comportamiento dinámico de cristales líquidos en presencia de singularidades (o puntos de defectos). Consideremos $\Omega = (-1, 1) \times (-1, 1)$, una triangulación uniforme de Ω de talla $h = \frac{1}{16}$ y un paso de tiempo $k = \frac{1}{400}$. Los parámetros son $\lambda = \nu = \gamma = 1$ y $\varepsilon = 0{,}05$.

Tomamos las condiciones iniciales $\boldsymbol{u}_0 = 0$ y $\boldsymbol{d}_0 = \widehat{\boldsymbol{d}}/\sqrt{|\widehat{\boldsymbol{d}}|^2 + 0{,}05^2}$, donde $\widehat{\boldsymbol{d}} = (x^2 + y^2 - 0{,}25, y)$. Obsérvese que el vector orientación tiene dos singularidades en el tiempo inicial en $(\pm 1/2, 0)$ (puntos en los cuales $\widehat{\boldsymbol{d}} = (0, 0)$) como se muestra en la Figura 1. A continuación, mostramos la evolución de las dos singularidades en $t = 0$, 0.2, 0.3 y 0.4, respectivamente. Su aniquilación se produce en torno a $t = 0{,}33$.

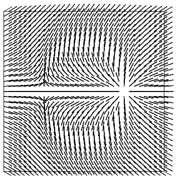


Figura 1: Vector orientación t=0

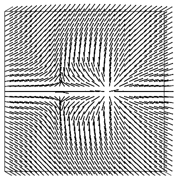


Figura 2: Vector orientación t=0.2

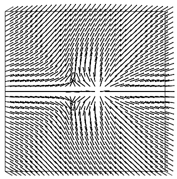


Figura 3: Vector orientación t=0.3

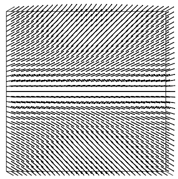


Figura 4: Vector orientación t=0.4

En las próximas figuras mostramos el comportamiento de la energía cinética $E_{cin} = \frac{1}{2}\|\boldsymbol{u}_h^{n+1}\|^2_{L^2(\Omega)}$, la energía elástica $E_{elas} = \frac{\lambda}{2}\|\nabla \boldsymbol{d}_h^{n+1}\|^2_{L^2(\Omega)}$, la energía de penalización $E_{pen} = \frac{\lambda}{2}\int_\Omega F_\varepsilon(\boldsymbol{d}_h^{n+1})$ y la energía total $E_{tot} = E_{kin} + E_{elas} + E_{pen}$ que decae con el tiempo tal y como ocurre para la energía continua (3). En particular, se puede observar el cambio significativo de las energías que se produce en torno al tiempo de aniquilación.

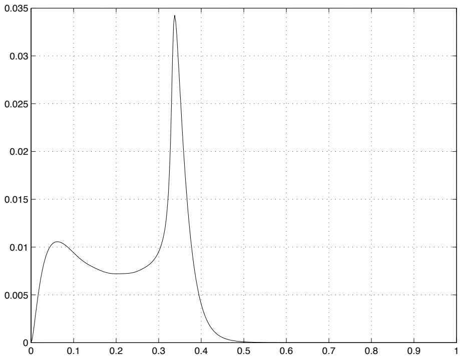


Figura 5: Energía cinética

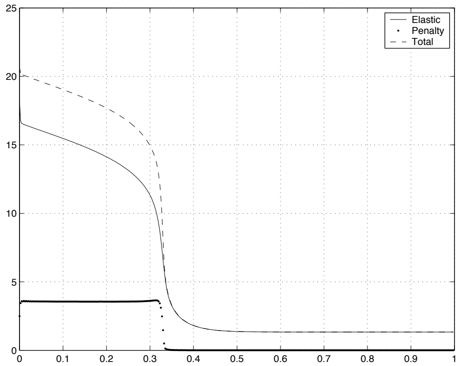


Figura 6: Energía cinética, de penalización y total

Referencias

[1] P. Azérad, F. Guillen-González. *Mathematical justification of the hidrostatic approximation in the primitive equations of geophysical fluid dynamics.* SIAM J. Math. Anal. 33 (2001), no. 4, 847–859.

[2] R. Becker, X. Feng, A. Prohl. *Finite element approximations of the Ericksen-Leslie model for nematic liquid crystal flow.* SIAM J. Numer. Anal. 46 (2008), no. 4, 1704–1731.

[3] J. J. Douglas, T. Dupont, L. Wahlbin. *The stability in L^q of the L^2-projection into finite element function spaces.* Numer. Math. 23 (1974/75), 193–197.

[4] V. Girault, F. Guillén-González. *Mixed formulation, approximation and decoupling algorithm for a nematic liquid crystals model.* In preparation.

[5] F. Guillén-González and J.V. Gutiérrez-Santacreu. In preparation.

[6] C. Liu, N.J. Walkington. *Mixed methods for the approximation of liquid crystal flows.* M2AN Math. Model. Numer. Anal. 36 (2002), no. 2, 205–222.

[7] F.H. Lin, C. Liu. Non-parabolic dissipative systems modelling the flow of liquid crystals. Comm. Pure Appl. Math. 48, (1995), 501-537.

[8] P. Lin, C. Liu. *Simulations of singularity dynamics in liquid crystal flows: A C^0 finite element approach.* Journal of Computational Physics 215 (2006) 348-362.

[9] C. Liu, N.J. Walkington. *Approximation of liquid crystal flows.* SIAM J. Numer. Anal. 37 (2000), no. 3, 725–741.

[10] J. Simon. *Compact sets in $L^p(0,T;B)$.* Ann. Mat. Pura Appl., sér. IV, CXLVI (1987), 65–96.

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 99–106

STATIONARY ASYMMETRIC FLUIDS AND HODGE OPERATOR

IGOR KONDRASHUK[1], EDUARDO A. NOTTE-CUELLO[2],
MARKO A. ROJAS-MEDAR[1]

[1]Departamento de Ciencias Básicas, Universidad del Bío-Bío, Chile.
[2]Departamento de Matemáticas, Universidad de La Serena, Chile.

igor.kondrashuk@ubiobio.cl, enotte@userena.cl, marko@ueubiobio.cl

Abstract

In this work we study a boundary value problem for a system of equations modeling the stationary flow of a incompressible asymmetric fluid. Based on methods of Clifford analysis, we write the system of asymmetric fluid in the hypercomplex formulation and represent its solution in Clifford operator terms.

Key words: *Asymmetric fluids, Clifford algebra*
AMS subject classifications: *15A66 35Q30 76D05*

1 Introduction

In this work we consider a boundary value problem for a system of equations modeling the stationary flow of a incompressible asymmetric fluid. Based on methods of Clifford analysis and following the work of P. Cereijeiras and U. Kähler [2], where they develop a Clifford operator calculus over unbounded domains, we write the system of asymmetric fluid in the hypercomplex formulation and we represent the solutions in term of these Clifford operators. The main difference between the Navier-stokes equations studied in [2] and the system of the asymmetric fluid studied in this work is the term curl $\boldsymbol{w}^*$, where $\boldsymbol{w}^*$ is the angular velocity of rotation of the fluid particles respectively. To write this term in the hypergeometric formulation we use the Hodge star operator.

2 Hodge operator in the Clifford Algebra approach

Let V be a vector space over the real field $\mathbb{R}$ of finite dimension, i.e., $\dim V = n, n \in \mathbb{N}$. By V^* we denote the dual space of V.

We recall that the space of k-tensors (denoted $T_k(V^*)$) are the set of all k-linear mappings τ_k such that

$$\tau_k : V^* \times \cdots \times V^* \to \mathbb{R}$$

and a multitensor τ of order $m \in \mathbb{N}$ is an element of $T(V)$ where

$$T(V) \equiv \sum_{k=0}^{\infty} \oplus T_k(V^*)$$

of the form $\tau = \sum_{k=0}^{m} \oplus \tau_k$, with $\tau_k \in T_k(V^*)$, such that all the components $\tau_k \in T_k(V^*)$ of τ are null for $k > m$. $T(V)$ is called the space of multitensors.

The Clifford algebra $\mathcal{C}l(V,g)$ of a metric vector space (V,g) is defined as the quotient algebra

$$\mathcal{C}l(V,g) = \frac{T(V)}{J_g},$$

where $J_g \subset T(V)$ is the bilateral ideal of $T(V)$ generated by the elements of the form $u \otimes v + v \otimes u - g(u,v)$, with $u, v \in V \subset T(V)$. The elements of $\mathcal{C}l(V,g)$ are sometime called *Clifford numbers.*

Let $\rho_g : T(V) \to \mathcal{C}l(V,g)$ be the natural projection of $T(V)$ onto the quotient algebra $\mathcal{C}l(V,g)$. Multiplication in $\mathcal{C}l(V,g)$ is called Clifford product and defined as

$$AB = \rho_g(A \otimes B),$$

for all $A, B \in \mathcal{C}l(V,g)$. In particular, for $u, v \in V \subset \mathcal{C}l(V,g)$, we have

$$u \otimes v = \frac{1}{2}(u \otimes v - v \otimes u) + g(u,v) + \frac{1}{2}(u \otimes v + v \otimes u) - g(u,v)$$

and then

$$\rho_g(u \otimes v) \equiv uv = \frac{1}{2}(u \otimes v - v \otimes u) + g(u,v) = u \wedge v + g(u,v).$$

From here we get the standard relation characterizing the Clifford algebra $\mathcal{C}l(V,g)$,

$$uv + vu = 2g(u,v).$$

In that follows we take $V = \mathbb{R}^n$, and we denote by $\mathbb{R}^{p,q}$ $(n = p + q)$ the real vector space $\mathbb{R}^n$ endowed with a non-degenerated metric $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, such that, if $\{e_i\}$, $(i = 1, 2, ..., n)$ is an orthonormal basis of $\mathbb{R}^{p,q}$, we have

$$g(e_i, e_j) = g_{ij} = g_{ji} = \begin{cases} +1, & i = j = 1, ..., p \\ -1, & i = j = p+1, ..., p+q = n \\ 0, & i \neq j. \end{cases}$$

The Clifford algebra $\mathcal{C}l(\mathbb{R}^{p,q}, g) = \mathbb{R}_{p,q} = Cl_{p,q}$, is the Clifford algebra over $\mathbb{R}$, generated by 1 and the $\{e_i\}$, $(i = 1, 2, ..., n)$ such that $e_i^2 = g(e_i, e_i)$, $e_i e_j = -e_j e_i$ $(i \neq j)$, and $e_A = e_1 e_2 \cdots e_n \neq \pm 1$.

Therefore the universal Clifford algebra $Cl_{p,q}$ has the dimension 2^n. Henceforth, each element $a \in Cl_{p,q}$ shall be written in the form

$$a = \sum_A a_A e_A$$

where the coefficients a_A are real numbers.

Now, we briefly describe Hodge operator, which will be used throughout these article. The Hodge star operator (or Hodge dual) is the linear mapping $\star : \bigwedge^r V \to \bigwedge^{n-r} V$ such that

$$A \wedge \star B = (A \cdot B)\tau_g,$$

for every $A, B \in \bigwedge^r V^*$ and where τ_g is the volume element in $\bigwedge^n V^*$. The inverse $\star^{-1} : \bigwedge^{n-r} V^* \to \bigwedge^r V^*$ of the Hodge star operator is given by:

$$\star^{-1} = (-1)^{r(n-r)} \operatorname{sgn}(g)\star,$$

where $\operatorname{sgn} g = \det g / |\det g|$ denotes the sign of the determinant of the matrix $(g_{ij} = g(e_i, e_j))$. A property of the Hodge star operator is

$$\star A_r = \widetilde{A}_r \lrcorner \tau_g = \widetilde{A}_r \tau_g \tag{1}$$

for any $A_r \in \bigwedge^r V^*$. Here $\bigwedge^r V^*$ denotes the space of k-forms, but the same results are obtained for k-vectors, for details see [5].

Let $\Omega \subset \mathbb{R}^n$ and $\Gamma = \partial\Omega$. Then functions u defined in Ω with values in $Cl_{0,n}$ ($p = 0$ and $q = n$) are considered. These functions may be written as

$$u(x) = \sum_A e_A u_A(x), \qquad x \in \Omega.$$

Properties such as continuity, differentiability, integrability, and so on, which are ascribed to u have to be possessed by all components $u_A(x)$. In this way, the usual Banach space of these functions are denoted by $\mathcal{C}^\alpha(\Omega, Cl_{0,n}), \mathcal{L}_q(\Omega, Cl_{0,n})$ and $\mathcal{W}_q^k(\Omega, Cl_{0,n})$ or in abbreviated form $\mathcal{C}^\alpha(\Omega), \mathcal{L}_q(\Omega)$ and $\mathcal{W}_q^k(\Omega)$.

Let us now introduce the Dirac operator by

$$D = \sum_{K=1}^{n} e_k \frac{\partial}{\partial x_k}.$$

is easy prove that $D^2 = -\Delta$, where Δ is the Laplacian.

We remember that the subspace of $Cl_{0,n}$ generated by the basic element e_A with equal length k is denoted by $Cl_{0,n}^k$ its elements being called k-vectors. It follows that $Cl_{0,n}^1$ is isomorphic to $\mathbb{R}^n$ ($Cl_{0,n}^1 \approx \mathbb{R}^n$). In this sense, we can identify each vector $u(x) \in \mathbb{R}^n$ with

$$u(x) = u_1(x)e_1 + \cdots + u_n(x)e_n \in Cl_{0,n}^1 \hookrightarrow Cl_{0,n}.$$

Then we can calculated $Du(x)$ when $u(x) \in Cl_{0,3}^1 \hookrightarrow Cl_{0,3}$, i.e., if $u(x) = u_1(x)e_1 + u_2(x)e_2 + u_3(x)e_3$

$$\begin{aligned}
Du(x) &= \sum_{k=1}^{3} e_k \tfrac{\partial u(x)}{\partial x_k} \\
&= e_1 \tfrac{\partial}{\partial x_1}(u_1(x)e_1 + u_2(x)e_2 + u_3(x)e_3) \\
&+ e_2 \tfrac{\partial}{\partial x_2}(u_1(x)e_1 + u_2(x)e_2 + u_3(x)e_3) \\
&+ e_3 \tfrac{\partial}{\partial x_3}(u_1(x)e_1 + u_2(x)e_2 + u_3(x)e_3),
\end{aligned}$$

now, we can computing term to term the above equation

$$\begin{aligned} Du(x) &= -\frac{\partial u_1}{\partial x_1} - \frac{\partial u_2}{\partial x_2} - \frac{\partial u_3}{\partial x_3} + e_1 \wedge e_2 (\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}) \\ &+ e_1 \wedge e_3 (\frac{\partial u_3}{\partial x_1} - \frac{\partial u_1}{\partial x_3}) + e_2 \wedge e_3 (\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3}) \\ &= -\operatorname{div}(u^*(x)) + \star \operatorname{curl}(u^*(x)) \end{aligned} \tag{2}$$

where $u^*(x) \in \mathbb{R}^3$. Thus, from equation above we have

$$\operatorname{curl}(u^*(x)) = \star^{-1} \operatorname{div}(u^*(x)) + \star^{-1} Du(x). \tag{3}$$

Now, we will show that $\star^{-1} Du(x) = D \star^{-1} u(x)$. From eq. (3) we get (we remember that $\star^{-1} = (-1)^{r(n-r)} \operatorname{sgn}(g)\star$),

$$\star^{-1} Du(x) = (-1)^{r(n-r)} \operatorname{sgn}(g)(-\operatorname{div}(u^*(x))) e_1 \wedge e_2 \wedge e_3 + \operatorname{curl}(u^*(x)). \tag{4}$$

On the other hand,

$$D \star^{-1} u(x) = (-1)^{r(n-r)} \operatorname{sgn}(g) D \star u(x) \tag{5}$$

and

$$\star u(x) = (e_3 \wedge e_2) u_1(x) + (e_1 \wedge e_3) u_2(x) + (e_2 \wedge e_1) u_1(x) \tag{6}$$

then

$$\begin{aligned} D \star u(x) &= \sum_{k=1}^{3} e_k \frac{\partial}{\partial x_k} ((e_3 \wedge e_2) u_1(x) + (e_1 \wedge e_3) u_2(x) + (e_2 \wedge e_1) u_{3\,\mathrm{sgn}}(x)) \\ &= (e_1 \wedge e_3 \wedge e_2) \frac{\partial u_1}{\partial x_1} - (e_2 \wedge e_3 \wedge e_1) \frac{\partial u_2}{\partial x_2} + (e_3 \wedge e_2 \wedge e_1) \frac{\partial u_3}{\partial x_3} \\ &+ e_3 (\frac{\partial u_1}{\partial x_2} - \frac{\partial u_2}{\partial x_1}) + e_2 (\frac{\partial u_3}{\partial x_1} - \frac{\partial u_1}{\partial x_3}) + e_1 (\frac{\partial u_2}{\partial x_3} - \frac{\partial u_3}{\partial x_2}) \end{aligned}$$

or

$$D \star u(x) = -\operatorname{div}(u^*(x)) \tau_g - \operatorname{curl}(u^*(x)) \tag{7}$$

where $\tau_g = e_1 \wedge e_2 \wedge e_3$. Thus from equation above, we have

$$\begin{aligned} D \star^{-1} u(x) &= (-1)^{r(n-r)} \operatorname{sgn}(g)(-\operatorname{div}(u^*(x))) \tau_g \\ &- ((-1)^{r(n-r)} \operatorname{sgn}(g)) \operatorname{curl}(u^*(x)) \\ &= (\operatorname{div}(u^*(x))) \tau_g + \operatorname{curl}(u^*(x)) \end{aligned} \tag{8}$$

where

$$\begin{aligned} ((-1)^{r(n-r)} \operatorname{sgn}(g)) \operatorname{curl}(u^*(x)) &= -\operatorname{curl}(u^*(x)) \\ ((-1)^{r(n-r)} \operatorname{sgn}(g)) \operatorname{div}(u^*(x)) &= -\operatorname{div}(u^*(x)) \end{aligned}$$

Finally, from eqs. (4) and (8) we get that

$$\star^{-1} Du(x) = D \star^{-1} u(x) \tag{9}$$

then, the eq. (3) can be write as

$$\begin{aligned} \operatorname{curl}(u^*(x)) &= \star^{-1} \operatorname{div}(u^*(x)) + D \star^{-1} u(x) \\ &= -Sc(Du(x))\tau_g + D \star^{-1} u(x) \end{aligned} \tag{10}$$

Note that the eq.(2) have scalar and bivector part, e.g., the scalar part of Du, denoted by $Sc(Du)$, is

$$Sc(Du) = -\frac{\partial u_1(x)}{\partial x_1} - \frac{\partial u_2(x)}{\partial x_2} - \frac{\partial u_3(x)}{\partial x_3}. \tag{11}$$

In the same way we obtain $Sc(uD) = -\frac{\partial u_1}{\partial x_1} - \frac{\partial u_2}{\partial x_2} - \frac{\partial u_3}{\partial x_3}$.

3 Asymmetric Fluid and Hodge Operator

Now, we consider the stationary incompressible asymmetric fluid with density constant, a detailed study on this system in bounded domain can be viewed in [3] (see also [4]) and exterior domains in [1]. Thus, let us denote by $\boldsymbol{u}^*, \boldsymbol{w}^*$ and p the velocity field, the angular velocity of rotation of the fluid particles and the pressure distribution, respectively. The governing equations are the following:

$$\begin{aligned} &-\Delta \boldsymbol{u}^* + \frac{\rho}{\eta l_1}(\boldsymbol{u}^* \cdot \nabla)\boldsymbol{u}^* + \frac{1}{\rho l_1}\nabla p = \frac{2\mu_r}{l_1} \operatorname{curl} \boldsymbol{w}^* + \boldsymbol{f}_1^* \\ &\operatorname{div} \boldsymbol{u}^* = 0 \\ &-\Delta \boldsymbol{w}^* + \frac{1}{l_2}(\boldsymbol{u}^* \cdot \nabla)\boldsymbol{w}^* - \frac{l_3}{l_2}\nabla \operatorname{div} \boldsymbol{w}^* + \frac{4\mu_r}{l_2}\boldsymbol{w}^* = \frac{2\mu_r}{l_2} \operatorname{curl} \boldsymbol{u}^* + \boldsymbol{g}_1^*. \end{aligned} \tag{12}$$

For simplicity, they will be completed with the following boundary conditions

$$\boldsymbol{u}^*(x) = 0, \quad \boldsymbol{w}^*(x) = 0 \quad \text{on } \partial\Omega = \Gamma. \tag{13}$$

In (12) $\boldsymbol{f}^*$ and $\boldsymbol{g}^*$ are known density functions of external sources for the linear and the angular momentum of particles, respectively. The positive constants l_1, l_2 and l_3 are given by

$$l_1 = \frac{\mu + \mu_r}{\rho}; l_2 = \frac{c_a + c_d}{\rho}; l_3 = \frac{c_0 + c_d - c_a}{\rho}$$

where μ, μ_r, c_0, c_a and c_d characterize the physical properties of the fluid. Thus, μ is the usual Newtonian viscosity; μ_r, c_0, c_a and c_d are additional viscosities related to the lack of symmetry of the stress tensor and, consequently, to the fact that the field of internal rotation w does not vanish. These constants must satisfy the inequality

$$c_0 + c_d > c_a.$$

Now, we can write the system (12,13) in the Clifford formalism with

$$\boldsymbol{u}(x), \boldsymbol{w}(x) \in Cl^1_{0,3} \hookrightarrow Cl_{0,3}$$

as

$$\begin{aligned}
&-\Delta\boldsymbol{u} + \frac{\rho}{\eta l_1} M(\boldsymbol{u}) + \tfrac{1}{\rho l_1} Dp = \frac{2\mu_r}{l_1}(D \star^{-1} \boldsymbol{w} - Sc(D\boldsymbol{u})\tau_g) \\
&ScD\boldsymbol{u} = 0 \\
&-\Delta\boldsymbol{w} + \frac{1}{l_2} N(\boldsymbol{u}, \boldsymbol{w}) - \frac{l_3}{l_2} D(Sc(D\boldsymbol{w})) + \frac{4\mu_r}{l_2}\boldsymbol{w} = \frac{2\mu_r}{l_2} D \star^{-1} \boldsymbol{u} \\
&\boldsymbol{u} = 0, \quad \boldsymbol{u} = 0 \quad \text{on } \partial\Omega = \Gamma.
\end{aligned} \tag{14}$$

where we using the eq.(10) and

$$M(\boldsymbol{u}) = [Sc(\boldsymbol{u}D)]\boldsymbol{u} - \boldsymbol{f}_1; \qquad N(\boldsymbol{u}, \boldsymbol{u}) = [Sc(\boldsymbol{u}D)]\boldsymbol{u} - \boldsymbol{g}_1.$$

4 Hodge operator and Projections

Now, we recall without proof the theorems and operators introduced in [2]. Let a fixed point z lying in the complement of the closure of Ω, which contains a non-empty open set. Then we can consider the operator

$$\widetilde{T}f(y) = \textstyle\int_\Omega K_z(x,y) f(x) d\Omega_x, \tag{15}$$

with $K_z(x,y) = G(x-y) - G(x-z)$, and where $G(x)$ is the so-called generalized Cauchy kernel, which is left-and right-monogenic function, i.e., $(DG)(x) = (GD)(x) = 0$. This operator is a continuous mapping of $\mathcal{W}^k_q(\Omega)$ in $\mathcal{W}^{k+1}_q(\Omega)$, $1 < q < \infty$, $k = 0, 1, ...$ and is bounded operator of $\mathcal{W}^{-1}_q(\Omega)$ in $\mathcal{L}_q(\Omega,\ 1 < q < \infty$.

Theorem 1 (Borel-Pompeiu's formula) *If $f \in \mathcal{W}^1_q(\Omega)$, $1 < q < \infty$ then we have*

$$\widetilde{F}_\Gamma f = f - \widetilde{T}Df,$$

with

$$\widetilde{F}_\Gamma f = \textstyle\int_\Gamma K_z(x,y)\alpha(x) f(x) d\Gamma_x$$

where $\alpha(x)$ is the outward pointing normal unit vector to Γ at the point x.

Proposition 2 *If $k \in \mathbb{N}$ then the operator*

$$\widetilde{F}_\Gamma : \mathcal{W}^{k-1/q}_q(\Gamma) \to \mathcal{W}^k_q(\Omega) \cap \ker D$$

is a continuous operator.

Theorem 3 (Plemelj-Sokhotzki's formula) *If $f \in \mathcal{W}_q^1(\Gamma)$, $1 < q < \infty, l > 0$, then we have*

$$tr\widetilde{F}_\Gamma f = \frac{1}{2}f + \frac{1}{2}\widetilde{S}_\Gamma f,$$

whereby

$$\widetilde{S}_\Gamma f = 2\int_\Gamma K_z(x,y)\alpha(x)f(x)d\Gamma_x$$

is the singular integral operator of Cauchy type over the boundary.

Theorem 4 *The space $\mathcal{L}_q(\Omega)$, $1 < q < \infty$, allows the direct decomposition*

$$\mathcal{L}_q(\Omega) = \ker D(\Omega) \cap \mathcal{L}_q(\Omega) \oplus D(\overset{0}{\mathcal{W}_q^1}(\Omega)).$$

The above theorem allows get the projections

$$\mathbf{P} : \mathcal{L}_q(\Omega) \to \ker D(\Omega) \cap \mathcal{L}_q(\Omega)$$

and

$$\mathbf{Q} : \mathcal{L}_q(\Omega) \to D(\overset{0}{\mathcal{W}_q^1}(\Omega)),$$

for $q = 2$ these projections are orthoprojections. Moreover, in [2] show that

$$\mathbf{Q}f = D\Delta_0^{-1}Df$$

where Δ_0^{-1}, is the solution operator of the Dirichlet problem of the Poisson equation with homogeneous boundary data

$$\begin{array}{lll} -\Delta u & = f & \text{in } \Omega, \\ u & = 0 & \text{on } \Gamma \end{array}$$

for $f \in \mathcal{W}_q^{-1}(\Omega)$, $1 < q < \infty$.

Theorem 5 *Suppose $\boldsymbol{f}_1, \boldsymbol{g}_1 \in \mathcal{W}_q^{-1}(\Omega)$, $p \in \mathcal{L}_q(\Omega, \mathbb{R})$, $1 < q < \infty$; then any solution of the system (14) has the representation*

$$\boldsymbol{u} + \frac{\rho}{\eta l_1}\widetilde{T}Q\widetilde{T}M(\boldsymbol{u}) + \tfrac{1}{\rho l_1}\widetilde{T}Qp = \frac{2\mu_r}{l_1}(\widetilde{T}Q \star^{-1} \boldsymbol{w} - \widetilde{T}Q\widetilde{T}(Sc(D\boldsymbol{w}))\tau_{\boldsymbol{g}})$$

$$\boldsymbol{w} + \frac{1}{l_2}\widetilde{T}Q\widetilde{T}N(\boldsymbol{u},\boldsymbol{w}) - \frac{l_3}{l_2}\widetilde{T}Q(Sc(D\boldsymbol{w})) + \frac{4\mu_r}{l_2}\widetilde{T}Q\widetilde{T}\boldsymbol{w} = \frac{2\mu_r}{l_2}\widetilde{T}Q \star^{-1} \boldsymbol{u}$$

$$Sc\frac{\rho}{\eta l_1}Q\widetilde{T}M(\boldsymbol{u}) + Sc\tfrac{1}{\rho l_1}Qp = Sc\frac{2\mu_r}{l_1}Q\widetilde{T}(\star^{-1}\boldsymbol{w}) - Sc\frac{2\mu_r}{l_1}Q\widetilde{T}(ScD\boldsymbol{w})\tau_{\boldsymbol{g}}$$

Proof. Recall that $Qf = D\Delta_0^{-1}Df$ and $D\widetilde{T}f(y) = f(y)$, and the Borel-Pompeiu's formula

$$\widetilde{T}D\boldsymbol{u} = \boldsymbol{u} - \widetilde{F}_\Gamma\boldsymbol{u} = \boldsymbol{u}, \qquad \boldsymbol{u} \in \overset{0}{\mathcal{W}}{}_q^1(\Omega)$$

we can write

$$\widetilde{T}Q\widetilde{T}DD\boldsymbol{u} = \widetilde{T}(D\Delta_0^{-1}D)\widetilde{T}DD\boldsymbol{u} = \widetilde{T}D\Delta_0^{-1}DD\boldsymbol{u} = \widetilde{T}D\boldsymbol{u} = \boldsymbol{u}. \tag{16}$$

On the other hand,

$$\begin{aligned}\widetilde{T}Q\widetilde{T}(D\star^{-1}\boldsymbol{w} - Sc(D\boldsymbol{w})\tau_{\boldsymbol{g}}) &= \widetilde{T}Q\widetilde{T}D\star^{-1}\boldsymbol{w} - \widetilde{T}Q\widetilde{T}(Sc(D\boldsymbol{w}))\tau_{\boldsymbol{g}} \\ &= \widetilde{T}Q\star^{-1}\boldsymbol{w} - \widetilde{T}Q\widetilde{T}(Sc(D\boldsymbol{w}))\tau_{\boldsymbol{g}},\end{aligned} \tag{17}$$

then by applying the $\widetilde{T}Q\widetilde{T}$ operator to system (14) and using the formulas (16) and (17) we obtain the expected result. □

Acknowledgments

I. Kondrashuk was supported by Fondecyt (Chile) grants #1040368, #1050512 and by DIUBB grant (UBB, Chile) #082609. E.A. Notte-Cuello was supported by Dirección de Investigación de la Universidad de La Serena, DIULS. M.A. Rojas-Medar was partially supported by DGI-MEC (Spain) Grant MTM2006-07932 and Fondecyt Grant #1080628.

References

[1] Durán, M.; Ortega-Torres, E.; Rojas-Medar, M. *Stationary solutions of magneto-micropolar fluid equations in exterior domains.* Proyecciones 22 (2003), no. 1, 63–79.

[2] P. Cerejeiras and U. Kähler: Math. Meth. Appl.Sci., **23**, 81-101, (2000).

[3] G. Lukaszewicz, *On stationary flows of asymmetric fluids,* Volume XII, Rend. Accad. Naz. Sci. detta dei XL, **106**, 1988, 35-44.

[4] G. Lukaszewicz *Micropolar Fluids. Theory and Applications,* Modeling and Simulation in Science, Engineering and Thecnology, Birkhauser, Boston (1999).

[5] W. A. Rodrigues, Jr. and E. Capelas Oliveira: *The Many Faces of Maxwell, Dirac and Einstein Equations. A Clifford Bundle Approach*, Lecture Notes in Physics **722**, Springer, New York, (2007).

Bol. Soc. Esp. Mat. Apl.
n^{o}47(2009), 107–135

PREMIO S$\vec{e}$MA JOVEN INVESTIGADOR

NUMERICAL ANALYSIS OF SOME EXTERIOR PROBLEMS, MIXED METHODS AND A POSTERIORI ERROR ANALYSIS IN FLUID MECHANICS AND ELASTICITY

MARÍA GONZÁLEZ TABOADA

Departamento de Matemáticas
Universidad de A Coruña

mgtaboad@udc.es

Abstract

This paper contains a brief description of various problems in the field of numerical analysis of partial differential equations. First, some results concerning the numerical analysis of boundary value problems in exterior domains of the plane are reviewed. Then, the derivation of dual-dual mixed variational formulations in fluid mechanics is explored. Finally, some results related to a posteriori error analysis in linear elasticity and a new augmented formulation in elasticity are discussed.

Key words: *Finite element method, boundary element method, symmetric coupling, mixed finite elements, twofold saddle point formulation, augmented formulation, a posteriori error analysis, parabolic-elliptic problem, quasi-linear problem, quasi-Newtonian flow, Stokes equation, linear elasticity, hyperelasticity.*

AMS subject classifications: *65N30 65N22 65N15 65N50 76D07 76M10 74B05*

1 Introduction

The aim of our research is to design and analyze efficient numerical methods that could be used to solve boundary value problems for partial differential equations in practice. We have dealt with problems in exterior domains, including linear and nonlinear elliptic and parabolic-elliptic equations and some problems in elasticity. We also coped with some models from fluid mechanics and elasticity in bounded domains.

The numerical solution of boundary value problems in exterior domains combining boundary elements and finite elements present several difficulties

The research of this author is supported by projects MTM2007-67596-C02-01 and PGIDIT06PXIB105230PR.

Fecha de recepción: 11/05/2009. Aceptado: 12/05/2009.

in practice. Indeed, these methods lead to ill-conditioned and badly structured systems of equations, and the computation of the boundary terms is cumbersome when the coupling boundary is a polygonal curve. Moreover, in this case we do not know how to control the effect of numerical quadratures on convergence. These facts motivated the work described in section 2. We present there a parametrized version of the standard symmetric method of coupling boundary elements and finite elements. This technique offers some advantages to solve a class of problems posed in exterior domains of the plane and has been applied to several nonlinear problems and to the elasticity system (cf. [43, 58, 57, 44]).

Later, we became interested in the derivation of mixed methods in fluids mechanics when the constitutive law cannot be inverted explicitly. Dual-dual methods are of special interest in this situation since no inversion process is required in their derivation. They had been applied to some linear and nonlinear problems in potential theory and elasticity (cf. [32, 35, 3, 40]). In [29] we derived a low-order mixed finite element method based on a dual-dual formulation for a class of quasi-Newtonian Stokes flows. In particular, we obtained, as a by-product, a new mixed finite element method for the usual Stokes problem. We carried out an a posteriori error analysis, based on local problems, and obtained fully explicit and reliable a posteriori error estimates of the accuracy of the computed numerical solution (see [30]). Then, we applied the approach from [29] to derive a low-order mixed finite element method for the generalized Stokes problem and developed the corresponding a posteriori error analysis (see [11]). In section 3, we review the derivation of dual-dual mixed formulations for quasi-Newtonian Stokes flows and for the generalized Stokes problem.

More recently, we turned our attention to an augmented mixed finite element method proposed in [27] for the linear elasticity system in the plane. We developed a residual-based a posteriori error analysis (see [4, 5]), combining a technique used in mixed finite element schemes with the usual procedure applied to primal finite element methods, and obtained a posteriori error estimators of residual type that are both efficient and reliable. In the last section, we outline the proof in the case of pure homogeneous Dirichlet boundary conditions.

2 Symmetric coupling of boundary elements and finite elements for solving exterior problems in 2D

Many physical and engineering problems are naturally posed in the exterior of a bounded domain; typical applications arise in electromagnetism and acoustics. The finite element method (FEM) can be used to solve nonlinear and nonhomogeneous boundary value problems; however, it can only be applied in bounded regions. On the other hand, the boundary element method (BEM) is well suited to solve problems in unbounded domains since it is based on the idea of reducing the partial differential equation to an integral equation on the boundary, yet it has the drawback that the equation must be linear, homogeneous and with constant coefficients. BEM-FEM methods were conceived with the aim of making the most of both techniques to solve boundary value problems in exterior domains (cf. for instance [74, 75]).

When a boundary value problem in an exterior domain is solved using a BEM-FEM method, an artificial boundary –called *the coupling boundary*– is introduced in order to divide the domain of the original problem in two regions: a bounded interior region and an unbounded exterior one, so that the equation is linear, homogeneous and with constant coefficients in the latter. Then, the problem can be written equivalently as a *transmission problem*, demanding that the solution satisfy appropriate conditions on the coupling boundary. Next, the BEM is applied in the unbounded region and the problem there is reduced to an integral equation on the coupling boundary. Then, the original problem reduces to a problem in the bounded region with *non-local* boundary conditions on the coupling boundary, and can be solved using the FEM. Finally, the solution in the exterior region is recovered through the integral representation formula.

To fix ideas, let us consider the Poisson problem in an exterior domain of $\mathbb{R}^2$. Let $\Omega_0 \subset \mathbb{R}^2$ be a bounded domain. We assume, for simplicity, that $\Gamma_0 := \partial\Omega_0$ is a polygonal curve, and denote by $\Omega_0' := \mathbb{R}^2 \setminus \overline{\Omega}_0$. Given $f \in L^2(\Omega_0')$, of compact support, we look for a function $u\colon \Omega_0' \to \mathbb{R}$ such that

$$\left\{\begin{array}{rcll} -\Delta u & = & f & \text{in } \ \Omega_0', \\ u & = & 0 & \text{on } \ \Gamma_0, \\ u & = & \mathcal{O}(1) & \text{as } \ |\mathbf{x}| \to +\infty. \end{array}\right. \tag{1}$$

Let Γ be a simple closed curve such that the support of f and the domain $\overline{\Omega}_0$ are contained in the region bounded by Γ. The coupling boundary Γ divides the domain of the original problem, Ω_0', in two regions: a bounded interior region, that we denote Ω^-, and the unbounded region exterior to Γ, that we denote Ω^+. The limit or trace over Γ of a function v defined in Ω^+ (resp., Ω^-) is denoted

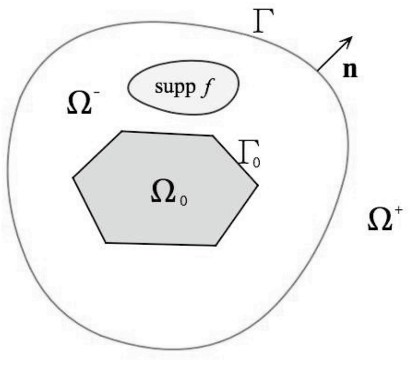


Figure 1: Domain of the transmission problem

by v^+ (resp., v^-). Finally, $\mathbf{n}$ denotes the unit normal vector to Γ, pointing from

Ω^- to Ω^+, and $\mathbf{t}$ denotes the tangent vector.

Then, problem (1) is equivalent to a transmission problem, which consists of a problem posed in Ω^-:

$$\left\{\begin{array}{rcll} -\Delta u &=& f & \text{in} \quad \Omega^-, \\ u &=& 0 & \text{on} \quad \Gamma_0, \end{array}\right. \tag{2}$$

and an *homogeneous* problem in Ω^+:

$$\left\{\begin{array}{rcll} -\Delta u &=& 0 & \text{in} \quad \Omega^+, \\ u &=& \mathcal{O}(1) & \text{as} \quad |\mathbf{x}| \to +\infty, \end{array}\right. \tag{3}$$

coupled by means of *transmission conditions* on the coupling boundary Γ:

$$u^- = u^+, \qquad\qquad \frac{\partial u}{\partial \mathbf{n}}^- = \frac{\partial u}{\partial \mathbf{n}}^+. \tag{4}$$

The variational formulation of problem (2) reads: find $u \in V$ such that

$$a(u,v) - \int_\Gamma \frac{\partial u}{\partial \mathbf{n}}^- v^- = \int_{\Omega^-} f v \qquad \forall\, v \in V, \tag{5}$$

where $V := \{v \in H^1(\Omega^-) : v|_{\Gamma_0} = 0\}$ and $a(u,v) := \displaystyle\int_{\Omega^-} \nabla u \cdot \nabla v\,.$

On the other hand, Green's formula applied in Ω^+ to the solution to problem (3) and the fundamental solution of the bidimensional Laplacian, $G(\mathbf{x},\mathbf{y}) := -\frac{1}{2\pi}\log|\mathbf{x}-\mathbf{y}|$, yields the following integral representation formula for u in Ω^+ (cf. for instance [19]):

$$u(\mathbf{x}) = \int_\Gamma u^+(\mathbf{y}) \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial \mathbf{n}(\mathbf{y})}\, d\sigma_{\mathbf{y}} - \int_\Gamma G(\mathbf{x},\mathbf{y})\, \frac{\partial u}{\partial \mathbf{n}}^+ (\mathbf{y})\, d\sigma_{\mathbf{y}} + u_\infty \quad \forall\, \mathbf{x} \in \Omega^+. \tag{6}$$

The first integral in (6) stands for the double layer potential with density u^+ whereas the second integral represents the single layer potential with density $\frac{\partial u}{\partial \mathbf{n}}^+$. The constant u_∞ accounts for the asymptotic behavior of u at infinity and can be computed from the trace of u and its normal derivative on Γ (see [64]).

We remark that, once we know u^+ and $\frac{\partial u}{\partial \mathbf{n}}^+$, the solution to problem (3) can be computed in any point of Ω^+ using (6). Therefore, taking into account the transmission conditions (4), to solve problem (1) it is enough to determine the solution in the bounded domain Ω^- and its normal derivative $\frac{\partial u}{\partial \mathbf{n}}^-$. We need an additional condition relating these two unknowns. A usual choice is to consider the integral equation derived from (6) as $\mathbf{x}$ tends to Γ (cf. [19]):

$$\frac{1}{2}\, u = \widehat{\mathcal{K}} u - \widehat{\mathcal{V}} \frac{\partial u}{\partial \mathbf{n}} + u_\infty \qquad \text{on } \Gamma, \tag{7}$$

where $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{V}}$ denote, respectively, the double layer and the single layer operators, formally defined as the corresponding potentials. On the other hand, the asymptotic behavior of u at infinity and (6) imply that $\int_\Gamma \frac{\partial u}{\partial \mathbf{n}}\, d\sigma = 0$. Then, the integral equation (7) can be tested with functions of zero mean on Γ, so that the constant u_∞ disappears (it can be recovered testing (7) with a constant function).

Equations (5) and (7) are the basis of the BEM-FEM method analyzed in [48] by C. Johnson and J.-C. Nédélec. Until very recently (see [65]), it was thought that this method only worked well if the double layer operator $\widehat{\mathcal{K}}$ was compact. As a result, the choice of a smooth coupling boundary Γ was mandatory for the Laplace equation. In other applications, such as linear elasticity (where the double layer operator is never compact), this formulation is not used. In addition, this procedure leads to nonsymmetric systems of linear equations. For these reasons, M. Costabel [21] and H.D. Han [46] introduced the *symmetric method* of coupling boundary elements and finite elements. This method is based on adding a second integral equation on the coupling boundary:

$$\frac{1}{2}\frac{\partial u}{\partial \mathbf{n}} = -\widehat{\mathcal{W}}u - \widehat{\mathcal{K}}^*\frac{\partial u}{\partial \mathbf{n}} \qquad \text{on } \Gamma,$$

where $\widehat{\mathcal{W}}$ is the *hypersingular* operator and $\widehat{\mathcal{K}}^*$ is the adjoint of the double layer operator. We recall from [49] that, for the Laplacian, the hypersingular operator can be expressed in terms of the single layer operator as follows:

$$\widehat{\mathcal{W}} = -\frac{\partial}{\partial \mathbf{t}}\widehat{\mathcal{V}}\frac{\partial}{\partial \mathbf{t}}\,.$$

In the symmetric method, the compactness of the double layer operator does not play any role. Then, it is possible to choose a polygonal curve as coupling boundary and this is, in fact, what all the authors do. However, this choice leads to additional difficulties in the approximation of the boundary terms, that include integrals with singular kernels. Moreover, in this case we do not know how to analyze the effect of numerical quadrature on convergence.

In [43], we followed [56] and choose a smooth coupling boundary Γ. Using a parametrization of Γ, we obtain a new version of the symmetric method for exterior boundary value problems in the plane. The new formulation is equivalent to the standard symmetric method introduced in [21, 46], but allows to approximate the singular integrals from the boundary element method using only low order quadrature formulas. In addition, with this approach it is possible to analyze the effect of numerical quadrature on convergence, which is in fact the main contribution in [43].

In what follows, we assume that Γ is of class $\mathcal{C}^\infty$ and let $\boldsymbol{x}\colon \mathbb{R} \to \mathbb{R}^2$ be a 1-periodic parametrization of Γ. We consider the 1-periodic Sobolev space of index 1/2, defined by

$$H^{1/2} := \Big\{ \phi \in L^2[0,1] \,:\, \sum_{m=-\infty}^{\infty} (1+m^2)^{1/2}\, |\hat{\phi}(m)|^2 < \infty \Big\},$$

where $\hat{\phi}(m) := \int_0^1 \phi(s)\, e^{-2\pi i m s}\, ds$, for $m \in \mathbb{Z}$, are the Fourier coefficients of ϕ. We denote by $H^{-1/2}$ the dual space of $H^{1/2}$, and by $(\cdot,\cdot)$ the duality product between $H^{-1/2}$ and $H^{1/2}$. We consider parametrized versions, $\mathcal{V}$ and $\mathcal{K}$, of the single and double layer operators (cf. [43, Section 1.2]), that inherit the properties of the standard ones. Indeed, $\mathcal{V}\colon H^{-1/2} \to H^{1/2}$ is continuous and elliptic on $H_0^{-1/2} := \{\eta \in H^{-1/2} : (\eta, 1) = 0\}$, and $\mathcal{K}\colon H^{1/2} \to H^{1/2}$ is compact.

Then, introducing the parametrization $\boldsymbol{x}$ in the integrals over Γ that appear in the symmetric method, we derive a new continuous BEM-FEM formulation, equivalent to the one introduced in [21, 46]:

$$\left\{\begin{array}{rcll} \multicolumn{4}{l}{\text{find } u \in V \text{ and } \xi \in H_0^{-1/2} \text{ such that}} \\ a(u,v) + d(u,v) - c(v,\xi) & = & (f,v)_{L^2(\Omega^-)} & \forall\, v \in V \\ c(u,\eta) + b(\xi,\eta) & = & 0 & \forall\, \eta \in H_0^{-1/2} \end{array}\right. \tag{8}$$

where, for simplicity, we substitute the unknown $\frac{\partial u}{\partial \mathbf{n}}^{+}$ by $\xi := |\boldsymbol{x}'| \left(\frac{\partial u}{\partial \mathbf{n}}^{+} \circ \boldsymbol{x} \right)$, and for any $\xi, \eta \in H^{-1/2}$ and $u, v \in H^1(\Omega^-)$, define $b(\xi,\eta) := (\eta, \mathcal{V}\xi)$, $d(u,v) := b(\gamma(u)', \gamma(v)')$ and

$$c(v,\eta) := \left(\eta, \left(\frac{1}{2}\mathcal{I} - \mathcal{K} \right) \gamma(v) \right),$$

where $\gamma : H^1(\Omega^-) \to H^{1/2}$ is the parametrized trace, that extends the map $u \mapsto \gamma(u) := u|_\Gamma \circ \boldsymbol{x}$. Existence and uniqueness of a solution to (8) are a consequence of Lax–Milgram's Lemma and the properties of the single and double layer operators.

In [43] the discrete problem is defined using a regular family of exact triangulations $\{\mathcal{T}_h^-\}_h$ of the bounded domain $\overline{\Omega^-}$ (cf. [73]), that contains straight triangles and triangles with exactly one curved side (the one that fits the coupling boundary). The corresponding finite element subspaces, $V_h \subset V$, are defined using Lagrange curved finite elements of order one on the curved triangles combined with Lagrange finite elements of the same order over the straight triangles, so that global finite element functions are continuous in $\overline{\Omega^-}$. To approximate the unknown ξ, a family of subspaces $H_h \subset H_0^{-1/2}$ consisting of 1-periodic splines of order one over a uniform partition of the real line, $s_i := ih$, $i \in \mathbb{Z}$, are used. Using interpolation error bounds on curved triangles and an approximation result from [68] we derive optimal error estimates. We remark that this method can be generalized without any difficulty to higher order approximations. It is also possible to consider different approximating functions on an independent mesh of the boundary; for instance, we could use trigonometric polynomials (see [59, 60]). On the other hand, the use of ideal triangles could be avoided (see [64]).

In practice, it is not possible to compute exactly some coefficients of the linear system obtained from the discretization process and it is necessary the use of quadrature formulas. We describe next the quadrature formulas used to approximate the integrals of the discrete problem:

- *Integrals over a triangle* $T \in \mathcal{T}_h^-$. We consider a quadrature formula of order zero over a reference (straight) triangle and define the corresponding formula over a triangle $T \in \mathcal{T}_h^-$ through a change of variables. In this way, we define approximations $a_h(\cdot,\cdot)$ and $l_h(\cdot)$ of the bilinear form $a(\cdot,\cdot)$ and the linear form $l(\cdot) := (f,\cdot)_{L^2(\Omega^-)}$, respectively.

- *Approximation of the bilinear form* $b(\cdot,\cdot)$. The single layer operator, $\mathcal{V}$, shows a singularity of logarithmic type. Then, to approximate the associated bilinear form, $b(\cdot,\cdot)$, we decompose the integrand as in G.C. Hsiao et al. [47], so that it can be written as a sum of a $\mathcal{C}^\infty$ function,

$$F(s,t) := \begin{cases} \log \dfrac{|\boldsymbol{x}(s) - \boldsymbol{x}(t)|^2}{(s-t)^2} & \text{if } s \neq t, \\ \log |\boldsymbol{x}'(s)|^2 & \text{if } s = t, \end{cases}$$

and a term that can be computed exactly. Let $\hat{\ell}_2$ be a quadrature formula of order one on the unit square. Then, the bilinear form $b_h \colon H_h \times H_h \to \mathbb{R}$ is defined by

$$b_h(\xi_h, \eta_h) := \sum_{i,j=1}^{N} \xi_h|_{(s_{i-1},s_i)} \eta_h|_{(s_{i-1},s_i)} \tilde{b}_{i,j} \quad \forall\, \xi_h, \eta_h \in H_h,$$

where, for $i,j = 1,\ldots,N$,

$$\tilde{b}_{i,j} := -\frac{1}{4\pi}\, h^2 \left(\hat{\ell}_2(F(s_{\underline{i}-1} + h\,\cdot, s_{\underline{j}-1} + h\,\cdot)) + \log h^2 + B_{\underline{i}-\underline{j}}\right),$$

with

$$(\underline{i},\underline{j}) := \begin{cases} (i,j) & \text{if } |i-j| \leq N/2, \\ (i, j+N) & \text{if } i-j > N/2, \\ (i, j-N) & \text{if } j-i > N/2, \end{cases}$$

and $B_0 = -3$, $B_1 = 4\log(2) - 3$ and for $k \geq 2$,

$$B_k = 2\log(k) - \sum_{n=1}^{\infty} \frac{1}{n(n+1)(2n+1)} \frac{1}{k^{2n}}\,.$$

- *Approximation of the bilinear form* $d(\cdot,\cdot)$. We define the bilinear form $d_h \colon V_h \times V_h \to \mathbb{R}$ by

$$d_h(u_h, v_h) := \sum_{i,j=1}^{N} \gamma u(s_i)\, \gamma v(s_j)\, \tilde{d}_{i,j} \quad \forall\, u_h, v_h \in V_h,$$

where, for $i,j = 1,\ldots,N$, $\tilde{d}_{i,j} := \big(\tilde{b}_{i,j} - \tilde{b}_{i,j+1} - \tilde{b}_{i+1,j} + \tilde{b}_{i+1,j+1}\big)/h^2$.

- *Approximation of the bilinear form $c(\cdot,\cdot)$.* Let $v_h \in V_h$ and $\eta_h \in H_h$. We remark that $\gamma(v_h) \in T_h$, where

$$T_h := \{\eta_h \in \mathcal{C}(\mathbb{R})\ ;\ \eta_h \text{ 1-periodic and } \quad \eta_h|_{(s_{i-1},s_i)} \in \mathcal{P}_1 \quad \forall\, i \in \mathbb{Z}\}.$$

Let $\{l_i\}_{i=1}^N$ be the nodal basis of T_h. We compute $(\eta_h, \gamma v_h)$ exactly, so that it only remains to approximate the coefficients

$$c_{i,j} := \int_{s_{j-1}}^{s_j} \left(\int_{s_{i-1}}^{s_{i+1}} K(s,t) l_i(t)\, dt \right) ds \qquad i,j = 1,\ldots,N,$$

where $K(\cdot,\cdot)$ is the kernel of the double layer operator $\mathcal{K}$. Since $K(\cdot,\cdot)$ is a function of class $\mathcal{C}^\infty$, we can use $\hat{\ell}_2$ to define the approximations

$$\tilde{c}_{i,j} := h^2 \hat{\ell}_2\big(K(s_{j-1}+h\cdot, s_{i-1}+h\cdot) l_i(s_{i-1}+h\cdot) + K(s_{j-1}+h\cdot, s_i+h\cdot) l_i(s_i+h\cdot)\big),$$

Then, we define the bilinear form $c_h \colon V_h \times H_h \to \mathbb{R}$ by

$$c_h(v_h,\eta_h) := \frac{h}{4} \sum_{j=1}^N \eta_j \big(\gamma v(s_{j-1}) + \gamma v(s_j)\big) - \sum_{i,j=1}^N \eta_j \gamma v(s_i) \tilde{c}_{i,j}\,.$$

We proved in [43] that the fully discrete scheme based on these approximations:

$$\left\{ \begin{array}{rcll} \multicolumn{4}{l}{\text{find } u_h^* \in V_h \text{ and } \xi_h^* \in H_h \text{ such that}} \\ a_h(u_h^*, v_h) + d_h(u_h^*, v_h) - c_h(v_h, \xi_h^*) &=& l_h(v_h) & \forall\, v_h \in V_h \\ c_h(u_h^*, \eta_h) + b_h(\xi_h^*, \eta_h) &=& 0 & \forall\, \eta_h \in H_h \end{array} \right. \tag{9}$$

is well posed for h sufficiently small and derived optimal error estimates. More precisely, if $f \in W^{1,\infty}(\Omega^-)$ and $u \in H^2(\Omega^-)$, then there exists a constant $C > 0$, independent of h, such that

$$\|u - u_h^*\|_{H^1(\Omega^-)} + \|\xi - \xi_h^*\|_{H^{-1/2}} \le Ch \left(\|u\|_{H^2(\Omega^-)} + \|f\|_{W^{1,\infty}(\Omega^-)} \right).$$

The fully discrete scheme (9) can be implemented in the computer directly. However, this method leads to a system of equations of the form:

$$\begin{pmatrix} A + D & C^t \\ C & -B \end{pmatrix} \begin{pmatrix} \mathbf{u}_h^* \\ \boldsymbol{\xi}_h^* \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix}, \tag{10}$$

where A and B are symmetric and positive definite matrices, and D is symmetric and semidefinite; moreover, A is a sparse matrix whereas B, C and D are dense matrices. Therefore, one has to solve a symmetric indefinite linear system of equations which, in addition, is ill-conditioned and badly structured. The use of an efficient iterative solver is then necessary.

We proposed an algorithm based on a preconditioning technique due to J.H. Bramble and J.E. Pasciak [7]. The idea is to transform the original system (10) into an equivalent one, with a symmetric and positive definite matrix in a certain inner product. Indeed, let R be a preconditioner of matrix A. Then, the system of linear equations

$$\begin{pmatrix} R^{-1}(A+D) & R^{-1}C^t \\ CR^{-1}(A+D)-C & B+CR^{-1}C^t \end{pmatrix} \begin{pmatrix} \mathbf{u}_h^* \\ \boldsymbol{\xi}_h^* \end{pmatrix} = \begin{pmatrix} R^{-1}\mathbf{f} \\ CR^{-1}\mathbf{f} \end{pmatrix}, \tag{11}$$

is equivalent to system (10). We proved that the matrix of the linear system (11) is symmetric and positive definite in the inner product

$$\left[\begin{pmatrix} u_h \\ \xi_h \end{pmatrix}, \begin{pmatrix} v_h \\ \eta_h \end{pmatrix} \right] := ((A+D-R)u_h, v_h)_{L^2(\Omega^-)} + (\xi_h, \eta_h).$$

Therefore, (11) can be solved by a preconditioned conjugate gradient method in the inner product $[\cdot,\cdot]$. Moreover, using Theorem 1 in [7], we showed that (11) can be preconditioned easily using only a preconditioner P of B (cf. [43] for more details).

This technique allows to *uncouple* the problem, since in each iteration we solve independently a problem using BEM and another problem using FEM. The method requires two preconditioners, one for the FEM stiffness matrix A and another one for the matrix associated with the single layer operator, B. Numerical experiments confirm the theoretical results and show that the algorithm is optimal in the sense that the number of iterations is independent of the discretization parameter.

Extension to nonlinear elliptic problems. The standard symmetric method was applied successfully to nonlinear boundary value problems that become homogeneous and linear with constant coefficients outside a bounded region (cf. [23, 37]). In these extensions, the error analysis is done assuming that the nonlinear operator is strongly monotone and Lipschitz-continuous, since in this case a Céa-type estimate is available. The parametrized version of the symmetric method can be extended fairly straightforward to this case (cf. [43]). The analysis of the continuous problem follows [37] and is based on the theory of monotone operators and Banach Fixed Point Theorem. The analysis of the discrete problem, based on a Céa type estimate, required to prove a technical result on the approximation in Sobolev spaces of non-integer index (which is a generalization of a result given in [70]). In addition, we analyzed a fully discrete scheme defined using only low order quadrature formulas. The analysis relies on a Strang-type inequality and on the analysis of the effect of numerical quadratures in the FEM for a nonlinear equation (cf. [25]).

On the other hand, J. Xu [71] introduced a technique to carry out the numerical analysis of nonlinear problems in bounded domains without using a Céa estimate. The idea is to linearize the nonlinear partial differential equation around an isolated solution and consider the finite element discretization of the linearized problem. In [58] we extended this technique to analyze exterior nonlinear problems without using discrete Green functions, at the expense of certain restrictions in the type of nonlinearity. We cannot deal with the general case because we do not know bounds for discrete Green functions associated with BEM–FEM formulations. Existence of a solution to the discrete problem and error estimates are derived applying Brouwer's Fixed Point Theorem; local uniqueness is also proved. The main contribution is the analysis of a fully discrete nonlinear BEM–FEM formulation without using Strang's lemma. We proved that the method described in [71] can be completed in this case to study the effect of numerical quadratures on convergence. This question remained open even in the bounded case and still is for a general nonlinear equation.

Nonlinear parabolic-elliptic problems. In [22], M. Costabel et al. applied the symmetric method to an exterior linear parabolic-elliptic problem. They used the Crank–Nicolson method for the time discretization and proved convergence of the solutions to the discrete schemes and theoretical error estimates. In [44], we applied the parametrized version of the symmetric method to a nonlinear parabolic-elliptic problem in the plane. This kind of problems appears in the modeling of quasi-stationary electromagnetic fields. The discrete problem is defined using the backward Euler method for the time discretization and an exact triangulation of the bounded domain. The analysis follows essentially [72]: existence, uniqueness, convergence of the discrete solutions and optimal error estimates are derived assuming that the nonlinear operator is strongly monotone and Lipschitz-continuous. In addition, we proposed a fully discrete scheme using only quadrature formulas of low order and, under some additional conditions on the nonlinear operator, proved that the order of convergence is optimal.

Extensions to elasticity. In [23, 36], the symmetric method was applied to a problem of three-dimensional elasticity theory, where an elastoplastic material is embedded into a linear elastic material. In two dimensions, this problem was analyzed in [16]. In [57] we generalized the parametrized version of the symmetric method to study an homogeneous isotropic linear elastic material in an exterior domain of the plane. In this case, the solution in the exterior region is given by Betti-Somigliana's formula (cf. [19]), and can be computed once we know the values of the solution and its traction on the coupling boundary. We solved the difficulties that result from the singularities of the integral operators and proposed a fully discrete scheme based on the use of quadrature formulas of low order. Optimal error estimates are derived and a preconditioning technique based on that of [7] is suggested to solve the corresponding linear systems. In [43] we showed that this technique can also be applied to the problem considered in [16] and proposed a fully discrete scheme that entails a great computational saving.

3 Dual-dual mixed methods in fluid mechanics

In this section we recall a dual-dual mixed finite element method to solve a class of quasi-Newtonian Stokes flows and discuss its application to the generalized Stokes problem. Mixed finite element methods are widely used to solve boundary value problems because they allow to approximate unknowns of physical interest directly. The standard mixed finite element method introduces the flux as an additional unknown; then, the gradient is expressed in terms of the flux and an integration by parts is done (cf. for instance [8, 42]).

When the constitutive law cannot be inverted explicitly, two basic strategies are available to obtain a mixed formulation. One possibility consists in inverting the constitutive law using the implicit function theorem (cf. [61, 62, 51]). The other strategy is based on introducing additional unknowns (preferably of physical interest) and rewriting the problem as a twofold saddle point operator equation, that is, the left-hand-side of the operator equation shows the following structure:

$$\begin{pmatrix} A & B^* \\ B & O \end{pmatrix} \qquad \text{with} \qquad A = \begin{pmatrix} A_1 & B_1^* \\ B_1 & O \end{pmatrix}, \tag{12}$$

where B and B_1 are linear bounded operators and A_1 is a nonlinear operator. This kind of formulations are called *dual–dual* formulations. It is important to emphasize that no inversion process is required in their derivation, which constitutes one of their main advantages.

Although the structure of (12) is very similar to the standard one, results from [52, 53, 66] cannot be applied. Fortunately, the standard theory of Babuška–Brezzi was generalized in [26] to deal with this kind of problems (see also [35]). On the other hand, this type of formulations leads to the solution of linear systems with a twofold saddle point structure, which are symmetric, indefinite and ill-conditioned. Efficient iterative solvers are already available (see [34, 32, 33]).

Dual-dual formulations were introduced in the context of coupling mixed finite elements and boundary elements (cf. [31, 39, 33, 26]). The first dual-dual method for a finite element discretization was analyzed in [32], where a linear second-order elliptic equation in divergence form is considered and, besides the scalar unknown u and the flux, the gradient ∇u is introduced as a third explicit unknown[1]. Then, this technique was applied in [35] to obtain a fully discrete dual-dual formulation of a nonlinear elliptic problem in divergence form. Dual-dual formulations were also used in nonlinear elasticity, where the strain tensor is introduced as additional unknown (cf. [3] for a dual-dual formulation of a hyperelastic material and [40] for the incompressible case).

Concerning the derivation of dual-dual formulations in fluid mechanics, we considered in [29] a class of quasi-Newtonian Stokes flows, and extended the

[1] The idea of introducing the gradient as an additional unknown was suggested by G.N. Gatica and W.L. Wendland [41] in the context of coupling mixed finite elements and boundary elements. It was also used in T. Arbogast et al. [1] and in Z. Chen [17, 18], where it was called *expanded mixed finite element method*. Additional variables are also used in least-squares finite element methods (see, e.g. [13]).

procedure in [11] to the generalized Stokes problem. The mixed finite element methods proposed in [29, 11] simply rely on the introduction of the flux and the tensor gradient of the velocity as additional unknowns. Then, the variational formulation is written as a twofold saddle point operator equation, so that the abstract theory developed in [26] can be applied to prove that the continuous and discrete schemes are well posed. In particular, we showed that the stability of the Galerkin schemes can be ensured using only low-order finite element subspaces. We remark that the usual Stokes equations are included in the class of problems considered in [29], so that we obtained, as a by-product, a new mixed finite element method for the Stokes problem.

Next we describe the derivation of a low-order mixed FEM based on a dual-dual formulation for quasi-Newtonian Stokes flows. We let Ω be a bounded domain in $\mathbb{R}^2$ with a Lipschitz-continuous boundary Γ, and consider a nonlinear Stokes fluid occupying the region Ω under the action of an external force. Let $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ be the nonlinear kinematic viscosity function of the fluid. Given $\mathbf{f} \in [L^2(\Omega)]^2$ and $\mathbf{g} \in [H^{1/2}(\Gamma)]^2$, we look for the velocity $\mathbf{u} := (u_1, u_2)^{\mathtt{t}}$ and the pressure p such that:

$$\left\{ \begin{array}{rcll} -\,\mathbf{div}\,(\,\psi(|\nabla \mathbf{u}|)\,\nabla \mathbf{u} \,-\, p\,\mathbf{I}\,) & = & \mathbf{f} & \text{in } \Omega, \\ \operatorname{div}(\mathbf{u}) & = & 0 & \text{in } \Omega, \\ \mathbf{u} & = & \mathbf{g} & \text{on } \Gamma. \end{array} \right. \tag{13}$$

We recall that the Dirichlet datum $\mathbf{g}$ must satisfy the compatibility condition

$$\int_\Gamma \mathbf{g} \cdot \mathbf{n}\, ds = 0, \tag{14}$$

where $\mathbf{n}$ is the unit outward normal to Γ.

This kind of nonlinear Stokes problem arises in the modeling of a large class of non-Newtonian fluids (biological fluids, lubricants, paints and polymeric fluids among others). In particular, the Ladyzhenskaya law for fluids with large stresses (see [50]), the power law used to model many polimeric solutions and melts (see [45]), and the Carreau law, used to model viscoplastic flows and creeping flow of metals (see, e.g. [54, 67]), are included in this framework. For the nonlinear model satisfying the power law, a dual-mixed variational formulation based on inverting the relation $\tilde{\boldsymbol{\sigma}} = \psi(|\nabla \mathbf{u}|)\nabla \mathbf{u}$ to obtain $\nabla \mathbf{u}$ as an explicit function of $\tilde{\boldsymbol{\sigma}}$ was studied in [55]. However, this procedure cannot be applied to the Carreau law since in this case such explicit inversion formula is not available.

To derive a dual-dual mixed variational formulation for the boundary value problem (13), we introduce the flux, $\boldsymbol{\sigma} := \psi(|\nabla \mathbf{u}|)\,\nabla \mathbf{u} - p\,\mathbf{I}$, and the tensor gradient of the velocity, $\mathbf{t} := \nabla \mathbf{u}$, as additional unknowns. Let us denote by $\boldsymbol{\psi}(\mathbf{r}) := (\psi(|\mathbf{r}|)\, r_{ij})$, for all $\mathbf{r} \in \mathbb{R}^{2\times 2}$. Then, the nonlinear constitutive law and the equilibrium equation become, respectively,

$$\boldsymbol{\sigma} \;=\; \boldsymbol{\psi}(\mathbf{t}) \,-\, p\,\mathbf{I} \quad \text{and} \quad -\,\mathbf{div}(\boldsymbol{\sigma}) \;=\; \mathbf{f} \qquad \text{in } \Omega. \tag{15}$$

In addition, since $\operatorname{div}(\mathbf{u}) = \operatorname{tr}(\nabla \mathbf{u})$, the incompressibility condition can be rewritten as $\operatorname{tr}(\mathbf{t}) = 0$ in Ω. Multiplying the relation $\mathbf{t} = \nabla \mathbf{u}$ by a tensor

$\boldsymbol{\tau}$, integrating by parts, using that $\mathbf{u} = \mathbf{g}$ on Γ and testing appropriately the equations in (15) and the incompressibility condition, we obtain the following mixed variational formulation of (13): find $(\mathbf{t}, \boldsymbol{\sigma}, p, \mathbf{u}, \xi) \in [L^2(\Omega)]^{2\times 2} \times H(\mathbf{div};\Omega) \times L^2(\Omega) \times [L^2(\Omega)]^2 \times \mathbb{R}$ such that

$$\begin{aligned}
\int_\Omega \boldsymbol{\psi}(\mathbf{t}) : \mathbf{s} - \int_\Omega \boldsymbol{\sigma} : \mathbf{s} - \int_\Omega p\,\mathrm{tr}(\mathbf{s}) \quad &= \quad 0,\\
-\int_\Omega \boldsymbol{\tau} : \mathbf{t} - \int_\Omega q\,\mathrm{tr}(\mathbf{t}) - \int_\Omega \mathbf{u}\cdot\mathbf{div}(\boldsymbol{\tau}) + \xi\int_\Omega \mathrm{tr}(\boldsymbol{\tau}) \quad &= \quad -\langle \boldsymbol{\tau}\mathbf{n}, \mathbf{g}\rangle_\Gamma, \\
-\int_\Omega \mathbf{v}\cdot\mathbf{div}(\boldsymbol{\sigma}) + \eta\int_\Omega \mathrm{tr}(\boldsymbol{\sigma}) \quad &= \quad \int_\Omega \mathbf{f}\cdot\mathbf{v},
\end{aligned} \tag{16}$$

for all $(\mathbf{s}, \boldsymbol{\tau}, q, \mathbf{v}, \eta) \in [L^2(\Omega)]^{2\times 2} \times H(\mathbf{div};\Omega) \times L^2(\Omega) \times [L^2(\Omega)]^2 \times \mathbb{R}$. We introduce in (16) the additional unknown ξ, which is a Lagrange multiplier associated with the restriction $\int_\Omega \mathrm{tr}(\boldsymbol{\sigma}) = 0$, added to ensure uniqueness (see [8]). Actually, we know in advance that $\xi = 0$, but we keep this artificial unknown to ensure the symmetry of the whole formulation.

Next, we remark that (16) has a twofold saddle point structure. Indeed, let us introduce the spaces $X_1 := [L^2(\Omega)]^{2\times 2}$, $M_1 := H(\mathbf{div};\Omega) \times L^2(\Omega)$ and $M := [L^2(\Omega)]^2 \times \mathbb{R}$, and define the operators $A_1 : X_1 \to X_1'$, $B_1 : X_1 \to M_1'$ and $B : M_1 \to M'$ as follows:

$$[A_1(\mathbf{r}), \mathbf{s}] := \int_\Omega \boldsymbol{\psi}(\mathbf{r}) : \mathbf{s}\,, \quad [B_1(\mathbf{r}), (\boldsymbol{\tau}, q)] := -\int_\Omega \boldsymbol{\tau} : \mathbf{r} \,-\, \int_\Omega q\,\mathrm{tr}(\mathbf{r}),$$

$$[B(\boldsymbol{\tau}, q), (\mathbf{v}, \eta)] := -\int_\Omega \mathbf{v}\cdot\mathbf{div}(\boldsymbol{\tau}) \,+\, \eta\int_\Omega \mathrm{tr}(\boldsymbol{\tau}),$$

for all $\mathbf{r}, \mathbf{s} \in X_1$, $(\boldsymbol{\tau}, q) \in M_1$ and $(\mathbf{v}, \eta) \in M$, where $[\cdot,\cdot]$ stands for the duality pairing induced by the corresponding operators. Then, with the previous definitions for A_1, B_1 and B, (16) can be written as an operator equation with a matrix operator of the form (12). Under suitable assumptions on the nonlinear kinematic viscosity function ψ (see equations (1.2) and (1.3) in [29]), we proved that the continuous formulation (16) is well posed. The proof reduces to show that the hypotheses of Theorem 2.4 in [26] are satisfied.

In order to define the corresponding mixed finite element scheme, we assume for simplicity that Γ is a polygonal curve, and let $\{\mathcal{T}_h\}_{h>0}$ be a regular family of triangulations of $\overline{\Omega}$ by triangles T of diameter h_T such that $h := \max\{\, h_T : T \in \mathcal{T}_h \,\}$ and $\overline{\Omega} = \cup\{\, T : T \in \mathcal{T}_h \,\}$. For each $T \in \mathcal{T}_h$, we let $\mathcal{RT}_0(T)$ be the local Raviart-Thomas space of lowest order and, for any non-negative integer k, we denote by $\mathcal{P}_k(T)$ the space of polynomials defined on T of degree $\leq k$. Then, we introduce the following finite element subspaces:

$$X_{1,h} := \left\{ \mathbf{s} \in [L^2(\Omega)]^{2\times 2} \,:\, \mathbf{s}|_T \in [\mathcal{P}_0(T)]^{2\times 2} \quad \forall\, T \in \mathcal{T}_h \right\},$$

$$M_{1,h}^{\boldsymbol{\sigma}} := \{ \boldsymbol{\tau} := (\tau_{ij}) \in H(\mathbf{div};\Omega) \,:\, (\tau_{i1}\ \tau_{i2})^{\mathrm{t}}|_T \in \mathcal{RT}_0(T)\ i = 1,2\,,\ \forall\, T \in \mathcal{T}_h \}\,,$$

$$M_{1,h}^{p} := \{\, q \in L^2(\Omega) \,:\, q|_T \in \mathcal{P}_0(T) \quad \forall\, T \in \mathcal{T}_h \,\},$$

$$M_h^{\mathbf{u}} := \{ \mathbf{v} \in [L^2(\Omega)]^2 \; : \; \mathbf{v}|_T \in [\mathcal{P}_0(T)]^2 \quad \forall T \in \mathcal{T}_h \}.$$

We showed that the corresponding Galerkin scheme has a unique solution $(\mathbf{t}_h, \boldsymbol{\sigma}_h, p_h, \mathbf{u}_h, \xi_h) \in X_{1,h} \times M_{1,h}^{\boldsymbol{\sigma}} \times M_{1,h}^{p} \times M_h^{\mathbf{u}} \times \mathbb{R}$. Furthermore, using a Céa estimate and the approximation properties of the subspaces $X_{1,h}$, $M_{1,h}^{\boldsymbol{\sigma}}$, $M_{1,h}^{p}$ and $M_h^{\mathbf{u}}$, that follow from classical error estimates for projection and equilibrium interpolation operators (see e.g. [63]), we obtained the following rate of convergence. If $\mathbf{t} \in [H^1(\Omega)]^{2\times 2}$, $\boldsymbol{\sigma} \in [H^1(\Omega)]^{2\times 2}$, $\mathbf{div}(\boldsymbol{\sigma}) \in [H^1(\Omega)]^2$, $p \in H^1(\Omega)$ and $\mathbf{u} \in [H^1(\Omega)]^2$, then there exists $C > 0$, independent of h, such that

$$\begin{aligned} \|(\mathbf{t}, \boldsymbol{\sigma}, p, \mathbf{u}, \xi) - (\mathbf{t}_h, \boldsymbol{\sigma}_h, p_h, \mathbf{u}_h, \xi_h)\| \, \leq \, C\, h \Big(& \|\mathbf{t}\|_{[H^1(\Omega)]^{2\times 2}} + \\ & + \|\boldsymbol{\sigma}\|_{[H^1(\Omega)]^{2\times 2}} + \|\mathbf{div}(\boldsymbol{\sigma})\|_{[H^1(\Omega)]^2} + \|p\|_{H^1(\Omega)} + \|\mathbf{u}\|_{[H^1(\Omega)]^2} \Big). \end{aligned}$$

Recently, V.J. Ervin et al. [24] recasted the formulation introduced in [29] in appropriate Sobolev spaces, providing tighter error estimates for the approximate solution and showing that higher-order approximating spaces can be used.

On the other hand, the application of adaptive algorithms based on a posteriori error estimates usually guarantees the quasi-optimal rate of convergence of the finite element solution to a boundary value problem. These techniques are specially useful for nonlinear problems, where no a priori hints on how to build suitable meshes are available. As shown in [3, 40], the combination of the usual Bank-Weiser approach from [2] with the analysis from [9] and [10] allows to derive fully explicit and reliable a posteriori error estimates for dual-dual variational formulations. In [30] we followed [3] and obtained reliable and quasi-efficient a posteriori error estimators for the nonlinear Stokes problems analyzed in [29]. Numerical experiments illustrate the performance of the mixed finite element scheme and confirm the reliability and quasi-efficiency of the a posteriori error estimators. They also show that the associated adaptive algorithm is much more efficient than a uniform refinement procedure.

As we mentioned before, in [11] we applied the approach from [29] to derive a low-order mixed FEM for the generalized Stokes problem. The generalized Stokes problem is a Stokes-like linear system with a dominating zeroth order term. This problem arises naturally in the time discretization of the corresponding non-steady equations and hence, plays a fundamental role in the numerical simulation of viscous incompressible flows (laminar and turbulent). Indeed, the most expensive part of the solution procedure for the time-dependent Navier-Stokes equations reduces to solve the generalized Stokes problem at each nonlinear iteration. Given $\mathbf{f} \in [L^2(\Omega)]^2$ and $\mathbf{g} \in [H^{1/2}(\Gamma)]^2$, we look for the velocity $\mathbf{u} := (u_1, u_2)^{\mathtt{t}}$ and the pressure p of a fluid occupying the region Ω, and such that

$$\left\{ \begin{array}{rcll} \alpha\, \mathbf{u} \, - \, \nu\, \Delta \mathbf{u} \, + \, \nabla p & = & \mathbf{f} & \text{in } \Omega, \\ \mathrm{div}(\mathbf{u}) & = & 0 & \text{in } \Omega, \\ \mathbf{u} & = & \mathbf{g} & \text{on } \Gamma. \end{array} \right. \tag{17}$$

where $\nu > 0$ is the kinematic viscosity of the fluid, that we assume constant, and α is a positive parameter proportional to the inverse of the time-step (we may assume, without loss of generality, that $\alpha \geq \nu$). We recall that, due to the incompressibility of the fluid, the Dirichlet datum $\mathbf{g}$ must satisfy the compatibility condition (14).

Now, we proceed as in [29] and introduce the tensor gradient of the velocity $\mathbf{t} := \nabla \mathbf{u}$ and the flux $\boldsymbol{\sigma} := \nu \nabla \mathbf{u} - p\,\mathbf{I}$ as additional unknowns in Ω. In this way, we obtain a mixed variational formulation of problem (17) that shows a twofold saddle point structure (see (12)). Indeed, let us define the spaces $X_1 := [L^2(\Omega)]^{2\times 2} \times [L^2(\Omega)]^2$, $M_1 := H(\mathbf{div};\Omega)$, $X := X_1 \times M_1$ and $M := L^2(\Omega) \times \mathbb{R}$, and the operators $A_1 : X_1 \to X_1'$, $B_1 : X_1 \to M_1'$ and $B : X \to M$ as follows:

$$\begin{aligned}
[A_1(\mathbf{s},\mathbf{v}),(\mathbf{r},\mathbf{w})] &:= \nu \int_\Omega \mathbf{s} : \mathbf{r} + \alpha \int_\Omega \mathbf{v}\cdot\mathbf{w},\\
[B_1(\mathbf{s},\mathbf{v}),\boldsymbol{\tau}] &:= -\int_\Omega \boldsymbol{\tau} : \mathbf{s} - \int_\Omega \mathbf{div}(\boldsymbol{\tau})\cdot\mathbf{v},\\
[B(\mathbf{s},\mathbf{v},\boldsymbol{\tau}),(q,\eta)] &:= -\int_\Omega q\,\mathrm{tr}(\mathbf{s}) + \eta \int_\Omega \mathrm{tr}(\boldsymbol{\tau}),
\end{aligned}$$

for all $(\mathbf{s},\mathbf{v}),(\mathbf{r},\mathbf{w}) \in X_1$, $\boldsymbol{\tau} \in M_1$ and $(q,\eta) \in M$. Then, the variational formulation of (17) can be set equivalently as: find $((\mathbf{t},\mathbf{u},\boldsymbol{\sigma}),(p,\xi)) \in X \times M$ such that

$$\begin{pmatrix} A & B^* \\ B & O \end{pmatrix} \begin{pmatrix} (\mathbf{t},\mathbf{u},\boldsymbol{\sigma}) \\ (p,\xi) \end{pmatrix} = \begin{pmatrix} F \\ O \end{pmatrix}, \tag{18}$$

where A is defined as in (12) and $[F,(\mathbf{s},\mathbf{v},\boldsymbol{\tau})] := \int_\Omega \mathbf{f}\cdot\mathbf{v} - \langle \boldsymbol{\tau}\mathbf{n},\mathbf{g}\rangle_\Gamma$. Using the abstract theory from [26], we proved that problem (18) has a unique solution $((\mathbf{t},\mathbf{u},\boldsymbol{\sigma}),(p,\xi)) \in X \times M$, and that there exists a positive constant $C(\alpha,\nu) = \mathcal{O}(\frac{\alpha^3}{\nu})$ such that

$$\|((\mathbf{t},\mathbf{u},\boldsymbol{\sigma}),(p,\xi))\|_{X\times M} \;\leq\; C(\alpha,\nu)\;\left(\|\mathbf{f}\|_{L^2(\Omega)} + \|\mathbf{g}\|_{H^{1/2}(\Gamma)}\right). \tag{19}$$

The corresponding Galerkin scheme is defined using the same finite element subspaces as in [29] for the unknowns $\mathbf{u}$, $\boldsymbol{\sigma}$ and p. However, to guarantee stability, the approximating space of the tensor gradient of the velocity, $\mathbf{t}$, has to be suitably enriched. Indeed, we have to include the deviator of the vector Raviart-Thomas space of lowest order, that is, we define

$$X_{1,h}^{\mathbf{t}} := \{\mathbf{s} \in [L^2(\Omega)]^{2\times 2} \;:\; \mathbf{s}|_T \in \mathcal{A}_0(T) \quad \forall\, T \in \mathcal{T}_h\},$$

where

$$\mathcal{A}_0(T) \;:=\; [\mathcal{P}_0(T)]^{2\times 2} \oplus \left\langle \left\{ \begin{pmatrix} x_1 & 2x_2 \\ 0 & -x_1 \end{pmatrix}, \begin{pmatrix} -x_2 & 0 \\ 2x_1 & x_2 \end{pmatrix} \right\} \right\rangle.$$

We proved that the discrete scheme is well-posed and derived the corresponding rate of convergence, in which a constant $\bar{C}(\alpha,\nu) = \mathcal{O}(\frac{\alpha^3}{\nu})$ is involved.

Theoretical results suggest that the rate of convergence is affected by large values of α, which, nevertheless, is not too severe in the numerical experiments (see [11]). Further, since α is proportional to the inverse of the time-step, Δt, the estimates also lead us to think that the convergence of time-dependent solutions should deteriorate as Δt decreases.

We followed [3, 30] and developed an a posteriori error analysis based on local problems. In this way, we obtained reliable and quasi-efficient a posteriori error estimators, that depend on the choice of two auxiliary functions. Numerical experiments confirm the reliability and quasi-efficiency of the estimators and illustrate the ability of the associated adaptive algorithm to localize the boundary layers, inner layers and singularities of the solution. Moreover, according to the theory, one would expect effectivity indexes between $\mathcal{O}(\alpha^{-1})$ and $\mathcal{O}(\frac{\alpha^3}{\nu})$. However, we observe in practice that they all lie on ranges much tighter than that, they do not deteriorate as the number of degrees of freedom increases and, in addition, they improve from uniform to adaptive refinements. The above observations yield the conjecture that these constants are overestimated. To conclude, this mixed method is perhaps not so competitive for extremely large values of α, but constitutes a good alternative for moderately large values of this parameter. Numerical experiments show that the adaptive algorithm is much more efficient than a uniform refinement when solving the discrete scheme.

Most approaches to the problems considered in [29, 11] deal with the usual pressure-velocity formulation, in which the velocity lives in $[H^1(\Omega)]^2$. This means, in particular, that the finite element subspace for the velocity needs to be a subset of the continuous functions. In addition, the Dirichlet boundary condition, being essential and non-homogeneous, cannot be incorporated either in the continuous and discrete formulations or in the definitions of the spaces involved, and therefore one is necessarily led to a non-conforming Galerkin scheme (certainly, we refer to the theoretical analysis of the method, since the interpolation of essential boundary conditions does not cause any difficulty in practice). In turn, in a dual-mixed setting the velocity becomes an unknown in $[L^2(\Omega)]^2$, which gives more flexibility to choose the associated finite element subspace (for instance, piecewise constant functions are a feasible choice). Furthermore, the Dirichlet boundary condition, being now natural, is incorporated directly into the right hand sides of the continuous and discrete formulations and hence, we avoid the error analysis of a non-conforming scheme.

Another important advantage of using dual-mixed methods, already pointed out, is the possibility of introducing further unknowns of physical interest (like the flux). These unknowns are then approximated directly, avoiding any numerical postprocessing that could yield additional sources of error. Moreover, the conservativity properties are transferred to some of these unknowns (for instance, continuity of the normal components of the flux), which, as we have seen, can be approximated with finite elements of very low order as well. Finally, we recall that the derivation of finite element subspaces guaranteeing unique solvability and stability of the Galerkin schemes for dual-dual formulations in elasticity and fluid mechanics was unified in [12].

4 A posteriori error analysis of augmented mixed finite element methods in elasticity

In this section we consider the augmented mixed finite element method introduced in [27] for the linear elasticity system in the plane and outline a residual-based a posteriori error analysis developed in [4] in the case of pure homogeneous Dirichlet boundary conditions. The analysis in the case of mixed boundary conditions can be found in [5].

Let Ω be a simply connected domain in $\mathbb{R}^2$. We assume, for simplicity, that Ω has a polygonal boundary Γ. Given a volume force $\mathbf{f} \in [L^2(\Omega)]^2$, we consider the problem of computing the displacements $\mathbf{u}$ and the stress tensor $\boldsymbol{\sigma}$ of a linear elastic material occupying the region Ω and such that

$$\left\{ \begin{array}{rcll} \boldsymbol{\sigma} & = & \mathcal{C}\,\mathbf{e}(\mathbf{u}) & \text{in } \Omega, \\ -\,\mathbf{div}(\boldsymbol{\sigma}) & = & \mathbf{f} & \text{in } \Omega, \\ \mathbf{u} & = & \mathbf{0} & \text{on } \Gamma. \end{array} \right. \tag{20}$$

Hereafter, $\mathbf{e}(\mathbf{u}) := \frac{1}{2}\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^{\mathtt{t}}\right)$ is the strain tensor of small deformations and $\mathcal{C}$ is the elasticity tensor determined by Hooke's law:

$$\mathcal{C}\,\boldsymbol{\zeta} \;:=\; \lambda\,\mathrm{tr}(\boldsymbol{\zeta})\,\mathbf{I} \;+\; 2\,\mu\,\boldsymbol{\zeta} \qquad \forall\,\boldsymbol{\zeta} \,\in\, [L^2(\Omega)]^{2\times 2}\,, \tag{21}$$

where $\lambda,\ \mu\ >\ 0$ are the Lamé constants.

Recently, a new stabilized mixed finite element method for plane linear elasticity was presented and analyzed in [27]. The approach is based on the introduction of suitable Galerkin least-squares terms arising from the constitutive and equilibrium equations, and from the relation defining the rotation $\boldsymbol{\gamma}$ in terms of the displacement, $\boldsymbol{\gamma} \;:=\; \frac{1}{2}\left(\nabla\mathbf{u} \,-\, (\nabla\mathbf{u})^{\mathtt{t}}\right)$. In particular, given positive parameters, κ_1, κ_2 and κ_3, independent of λ, the following augmented variational formulation for problem (20) is proposed: find $(\boldsymbol{\sigma},\mathbf{u},\boldsymbol{\gamma}) \,\in\, \hat{H}_0 \,:=\, H_0 \times [H_0^1(\Omega)]^2 \times [L^2(\Omega)]^{2\times 2}_{\mathtt{skew}}$ such that

$$A((\boldsymbol{\sigma},\mathbf{u},\boldsymbol{\gamma}),(\boldsymbol{\tau},\mathbf{v},\boldsymbol{\eta})) \;=\; F(\boldsymbol{\tau},\mathbf{v},\boldsymbol{\eta}) \qquad \forall\,(\boldsymbol{\tau},\mathbf{v},\boldsymbol{\eta}) \in \hat{H}_0, \tag{22}$$

where $[L^2(\Omega)]^{2\times 2}_{\mathtt{skew}} := \left\{ \boldsymbol{\eta} \in [L^2(\Omega)]^{2\times 2} \,:\, \boldsymbol{\eta} + \boldsymbol{\eta}^{\mathtt{t}} \,=\, \mathbf{0} \right\}$,

$$H_0 := \left\{ \boldsymbol{\tau} \in H(\mathbf{div};\Omega) \,:\, \int_\Omega \mathrm{tr}(\boldsymbol{\tau}) \,=\, 0 \right\},$$

and the bilinear form $A : \hat{H}_0 \times \hat{H}_0 \to \mathbb{R}$ and the functional $F : \hat{H}_0 \to \mathbb{R}$ are defined by

$$
\begin{aligned}
&A((\boldsymbol{\sigma},\mathbf{u},\gamma),(\boldsymbol{\tau},\mathbf{v},\boldsymbol{\eta})) := \\
&\quad := \int_\Omega \mathcal{C}^{-1}\boldsymbol{\sigma} : \boldsymbol{\tau} + \int_\Omega \mathbf{u}\cdot\mathbf{div}(\boldsymbol{\tau}) + \int_\Omega \gamma : \boldsymbol{\tau} - \int_\Omega \mathbf{v}\cdot\mathbf{div}(\boldsymbol{\sigma}) - \int_\Omega \boldsymbol{\eta} : \boldsymbol{\sigma} \\
&\quad + \kappa_1 \int_\Omega \left(\mathbf{e}(\mathbf{u}) - \mathcal{C}^{-1}\boldsymbol{\sigma}\right) : \left(\mathbf{e}(\mathbf{v}) + \mathcal{C}^{-1}\boldsymbol{\tau}\right) \;+\; \kappa_2 \int_\Omega \mathbf{div}(\boldsymbol{\sigma})\cdot\mathbf{div}(\boldsymbol{\tau}) \\
&\quad + \kappa_3 \int_\Omega \left(\gamma - \frac{1}{2}(\nabla\mathbf{u} - (\nabla\mathbf{u})^{\mathtt{t}})\right) : \left(\boldsymbol{\eta} + \frac{1}{2}(\nabla\mathbf{v} - (\nabla\mathbf{v})^{\mathtt{t}})\right),
\end{aligned}
$$

and

$$F(\boldsymbol{\tau},\mathbf{v},\boldsymbol{\eta}) \;:=\; \int_\Omega \mathbf{f}\cdot(\,\mathbf{v} \;-\; \kappa_2\,\mathbf{div}(\boldsymbol{\tau})\,)\;.$$

We recall that it is easy to see from (21) that the inverse tensor $\mathcal{C}^{-1}$ reduces to

$$\mathcal{C}^{-1}\,\boldsymbol{\zeta} \;:=\; \frac{1}{2\,\mu}\,\boldsymbol{\zeta} \;-\; \frac{\lambda}{4\,\mu\,(\lambda+\mu)}\,\mathrm{tr}(\boldsymbol{\zeta})\,\mathbf{I} \qquad \forall\,\boldsymbol{\zeta} \,\in\, [L^2(\Omega)]^{2\times 2}.$$

Assume that $(\kappa_1,\kappa_2,\kappa_3)$ is independent of λ and such that

$$0 < \kappa_3 < \kappa_1 < 2\,\mu \qquad \text{and} \qquad 0 < \kappa_2. \tag{23}$$

Then, the bilinear form $A(\cdot,\cdot)$ is strongly coercive and continuous, and therefore, problem (22) is well-posed (see Theorems 3.1 and 3.2 in [27]). In particular, if we take $\kappa_2 = \frac{1}{\mu}\left(1 - \frac{\kappa_1}{2\,\mu}\right)$, then the stability constant depends only on μ, $\dfrac{1}{\mu}$ and Ω.

The augmented variational formulation (22), being strongly coercive, allows to use arbitrary finite element subspaces to define the corresponding discrete scheme. This constitutes one of its main advantages, as compared with the traditional mixed finite element schemes for the linear elasticity problem (see [8]). Indeed, given a finite element subspace $\hat{H}_{0,h} := H_{0,h}^{\boldsymbol{\sigma}} \times H_{0,h}^{\mathbf{u}} \times H_{0,h}^{\boldsymbol{\gamma}} \subseteq \hat{H}_0$, the Galerkin scheme associated to (22) reads: find $(\boldsymbol{\sigma}_h,\mathbf{u}_h,\gamma_h) \in \hat{H}_{0,h}$ such that

$$A((\boldsymbol{\sigma}_h,\mathbf{u}_h,\gamma_h),(\boldsymbol{\tau}_h,\mathbf{v}_h,\boldsymbol{\eta}_h)) = F(\boldsymbol{\tau}_h,\mathbf{v}_h,\boldsymbol{\eta}_h) \qquad \forall\,(\boldsymbol{\tau}_h,\mathbf{v}_h,\boldsymbol{\eta}_h) \in \hat{H}_{0,h}. \tag{24}$$

If the parameters κ_1, κ_2 and κ_3 satisfy (23), then the discrete problem (24) is well-posed for any arbitrary choice of the subspace $\hat{H}_{0,h}$. In particular, it is possible to use Raviart-Thomas spaces of lowest order to approximate the stress tensor $\boldsymbol{\sigma}$, piecewise linear elements for the displacement $\mathbf{u}$, and piecewise constants for the rotation $\boldsymbol{\gamma}$. The rate of convergence of (24) for this specific finite element subspace is given in Theorem 4.2 in [27].

As compared with more traditional mixed methods, such as PEERS and BDM, and besides the fact of being able to choose any finite element subspace, the augmented approach presents other important advantages. Indeed, it becomes a much cheaper alternative since the global number of degrees of freedom in terms of the number of triangles is much smaller (see [4, Section 5]). In addition, if we choose the finite element subspace of the lowest order, the augmented scheme (24) yields simpler computations.

The competitive character of the augmented mixed finite element method (24) motivated the derivation of a posteriori error estimators for this scheme. We need to introduce some notations. Let $\{\mathcal{T}_h\}_{h>0}$ be a regular family of triangulations of $\overline{\Omega}$ by triangles T of diameter h_T such that $h := \max\{\, h_T : T \in \mathcal{T}_h \,\}$ and $\overline{\Omega} = \cup\{\, T : T \in \mathcal{T}_h \,\}$. Given $T \in \mathcal{T}_h$, we let $E(T)$ be the set of its edges and let $E_h(\Omega)$ be the set of all interior edges of the triangulation $\mathcal{T}_h$. In what follows, h_e stands for the length of edge e. Further, given $\boldsymbol{\tau} \in [L^2(\Omega)]^{2\times 2}$ (such that $\boldsymbol{\tau}|_T \in \mathcal{C}(T)$ on each $T \in \mathcal{T}_h$), an edge $e \in E(T) \cap E_h(\Omega)$ and the unit tangential vector $\mathbf{t}$ along e, we denote by $J[\boldsymbol{\tau}\mathbf{t}]$ the tangential jump across e, that is, $J[\boldsymbol{\tau}\mathbf{t}] := (\boldsymbol{\tau}|_T - \boldsymbol{\tau}|_{T'})|_e \mathbf{t}$, where $T' \in \mathcal{T}_h$ is such that $T \cap T' = e$. We recall that $\mathbf{t} := (-n_2, n_1)^{\mathtt{t}}$, where $\mathbf{n} := (n_1, n_2)^{\mathtt{t}}$ is the unit outward normal to ∂T. The normal jumps $J[\boldsymbol{\tau}\mathbf{n}]$ are defined analogously.

Then, if $(\boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\gamma}) \in \hat{H}_0$ and $(\boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\gamma}_h) \in \hat{H}_{0,h}$ are, respectively, the solutions to the continuous and discrete formulations, (22) and (24), we define the error indicator θ_T, for any $T \in \mathcal{T}_h$, as follows:

$$
\begin{aligned}
\theta_T^2 :=\; & \|\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)\|^2_{[L^2(T)]^2} + \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^{\mathtt{t}}\|^2_{[L^2(T)]^{2\times 2}} \\
& + \|\boldsymbol{\gamma}_h - \tfrac{1}{2}(\nabla \mathbf{u}_h - (\nabla \mathbf{u}_h)^{\mathtt{t}})\|^2_{[L^2(T)]^{2\times 2}} + h_T^2 \,\|\,\mathbf{curl}(\mathcal{C}^{-1}\boldsymbol{\sigma}_h + \boldsymbol{\gamma}_h)\|^2_{[L^2(T)]^2} \\
& + \sum_{e \in E(T)} h_e \,\|J[(\mathcal{C}^{-1}\boldsymbol{\sigma}_h - \nabla \mathbf{u}_h + \boldsymbol{\gamma}_h)\mathbf{t}]\|^2_{[L^2(e)]^2} \\
& + h_T^2 \,\|\,\mathbf{curl}(\mathcal{C}^{-1}(\mathbf{e}(\mathbf{u}_h) - \mathcal{C}^{-1}\boldsymbol{\sigma}_h))\|^2_{[L^2(T)]^2} \\
& + \sum_{e \in E(T)} h_e \,\|J[(\mathcal{C}^{-1}(\mathbf{e}(\mathbf{u}_h) - \mathcal{C}^{-1}\boldsymbol{\sigma}_h))\mathbf{t}]\|^2_{[L^2(e)]^2} \\
& + h_T^2 \,\|\mathbf{div}(\mathbf{e}(\mathbf{u}_h) - \tfrac{1}{2}(\mathcal{C}^{-1}\boldsymbol{\sigma}_h + (\mathcal{C}^{-1}\boldsymbol{\sigma}_h)^{\mathtt{t}}))\|^2_{[L^2(T)]^2} \\
& + \sum_{e \in E(T) \cap E_h(\Omega)} h_e \,\|J[(\mathbf{e}(\mathbf{u}_h) - \frac{1}{2}(\mathcal{C}^{-1}\boldsymbol{\sigma}_h + (\mathcal{C}^{-1}\boldsymbol{\sigma}_h)^{\mathtt{t}}))\mathbf{n}]\|^2_{[L^2(e)]^2} \\
& + h_T^2 \,\|\mathbf{div}(\boldsymbol{\gamma}_h - \tfrac{1}{2}(\nabla \mathbf{u}_h - (\nabla \mathbf{u}_h)^{\mathtt{t}}))\|^2_{[L^2(T)]^2} \\
& + \sum_{e \in E(T) \cap E_h(\Omega)} h_e \,\|J[(\boldsymbol{\gamma}_h - \frac{1}{2}(\nabla \mathbf{u}_h - (\nabla \mathbf{u}_h)^{\mathtt{t}}))\mathbf{n}]\|^2_{[L^2(e)]^2}\,.
\end{aligned}
$$

The residual character of each term involved in the definition of θ_T is quite clear. In addition, we observe that some of these terms are known from residual estimators for the usual (non-augmented) mixed finite element method in linear elasticity (see, e.g. [15]). However, most of them are new since they arise from the new Galerkin least-squares terms introduced in (22). Finally, we remark that when $\boldsymbol{\sigma}_h|_T \in [\mathcal{RT}_0(T)^{\mathtt{t}}]^2$, $\mathbf{u}_h|_T \in [\mathcal{P}_1(T)]^2$ and $\boldsymbol{\gamma}_h|_T \in [\mathcal{P}_0(T)]^{2\times 2}$, some of the terms in the definition of θ_T vanish.

As usual, $\theta := \left(\sum_{T\in\mathcal{T}_h} \theta_T^2 \right)^{1/2}$ is used as the global residual error estimator. We proved in [4] that the a posteriori error estimator θ is reliable and efficient, that is, there exist $C_{\texttt{eff}}, C_{\texttt{rel}} > 0$, independent of h and λ, such that

$$C_{\texttt{eff}}\, \theta \ \leq\ \|(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h)\|_{\hat{H}_0} \ \leq\ C_{\texttt{rel}}\, \theta\,. \tag{25}$$

Reliability (upper bound in (25)) ensures that we obtain a numerical solution with an accuracy below a prescribed tolerance. Local lower bounds are necessary to ensure that the mesh is correctly refined so that one obtains a numerical solution with a prescribed tolerance using a (nearly) minimal number of nodes.

To prove that θ is reliable, we combined a technique used in mixed finite element schemes (see, e.g. [14, 15]) with the usual procedure applied to primal finite element methods (see [69]). It is important to remark that just one of these approaches by itself would not be enough in this case. Up to our knowledge, this combined analysis seems to be applied in [4] for the first time. We provide next a sketch of the proof.

We consider the following auxiliary problem: find $\mathbf{z} \in [H_0^1(\Omega)]^2$ such that

$$\left\{ \begin{array}{rcll} -\mathbf{div}(\mathbf{e}(\mathbf{z})) & = & \mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h) & \text{in } \Omega, \\ \mathbf{z} & = & \mathbf{0} & \text{on } \Gamma, \end{array} \right. \tag{26}$$

and define $\boldsymbol{\sigma}^* := \mathbf{e}(\mathbf{z})$, where $\mathbf{z}$ is the unique solution to problem (26). It follows that $\boldsymbol{\sigma}^* \in H_0$ and, because of the continuous dependence result, there exists $c > 0$ such that

$$\|\boldsymbol{\sigma}^*\|_{H(\mathbf{div};\Omega)} \ \leq\ c\, \|\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)\|_{[L^2(\Omega)]^2}. \tag{27}$$

In addition, it is easy to see that $\mathbf{div}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \boldsymbol{\sigma}^*) = \mathbf{0}$ in Ω. Then, using the triangle inequality, that A is coercive and bounded, and (27), we obtain that there exists $C > 0$, independent of h and λ, such that

$$\begin{aligned} C\, \|(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h)\|_{\hat{H}_0} &\leq \|\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)\|_{[L^2(\Omega)]^2} + \\ &+ \sup_{\substack{\mathbf{0} \neq (\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \hat{H}_0 \\ \mathbf{div}(\boldsymbol{\tau}) = \mathbf{0}}} \frac{A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}))}{\|(\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta})\|_{\hat{H}_0}}\,. \end{aligned} \tag{28}$$

It remains to bound the second term on the right hand side of (28). To this end, we make use of the well-known Clément interpolation operator, $I_h : H^1(\Omega) \to X_h$, where X_h is the space of continuous piecewise linear functions on $\mathcal{T}_h$, which satisfies the standard local approximation properties stated in [20]. We recall that $I_h(v) \in X_h \cap H_0^1(\Omega)$ for all $v \in H_0^1(\Omega)$.

Now, let $(\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \hat{H}_0$, $(\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \neq \mathbf{0}$, be such that $\mathbf{div}(\boldsymbol{\tau}) = \mathbf{0}$ in Ω. Since we assume the domain Ω to be simply connected, there exists a stream function $\boldsymbol{\varphi} := (\varphi_1, \varphi_2) \in [H^1(\Omega)]^2$ such that $\int_\Omega \varphi_i = 0$, for $i = 1, 2$, and $\boldsymbol{\tau} = \underline{\mathbf{curl}}(\boldsymbol{\varphi})$. Then, we define $\boldsymbol{\varphi}_h := (\varphi_{1,h}, \varphi_{2,h})$, with $\varphi_{i,h} := I_h(\varphi_i)$ for

$i = 1, 2$, and $\boldsymbol{\tau}_h := \underline{\mathbf{curl}}(\boldsymbol{\varphi}_h)$. Note that we can write $\boldsymbol{\tau}_h = \boldsymbol{\tau}_{h,0} + d_h\,\mathbf{I}$, where $\boldsymbol{\tau}_{h,0} \in H^{\boldsymbol{\sigma}}_{0,h}$ and $d_h = \frac{\int_\Omega \mathrm{tr}(\boldsymbol{\tau}_h)}{2|\Omega|} \in \mathbb{R}$.

On the other hand, an immediate consequence of (22) and (24) is the Galerkin orthogonality:

$$A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau}_h, \mathbf{v}_h, \boldsymbol{\eta}_h)) = 0 \quad \forall\, (\boldsymbol{\tau}_h, \mathbf{v}_h, \boldsymbol{\eta}_h) \in \hat{H}_{0,h}\,. \tag{29}$$

Let $\mathbf{v}_h := (I_h(v_1), I_h(v_2)) \in H^{\mathbf{u}}_{0,h}$ be the vector Clément interpolant of $\mathbf{v} := (v_1, v_2) \in [H^1_0(\Omega)]^2$. Then, it follows from (29) that

$$\begin{aligned} &A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta})) = \\ &\qquad\qquad = A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau} - \boldsymbol{\tau}_{h,0}, \mathbf{v} - \mathbf{v}_h, \boldsymbol{\eta})). \end{aligned} \tag{30}$$

Since $\int_\Omega \mathrm{tr}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) = 0$ and $\mathbf{u} - \mathbf{u}_h = \mathbf{0}$ on Γ, using the orthogonality between symmetric and skew-symmetric tensors, we obtain that

$$A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (d_h\mathbf{I}, \mathbf{0}, \mathbf{0})) = 0.$$

Hence, from (30) and (22) we deduce that

$$\begin{aligned} &A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta})) = \\ &\quad = F(\boldsymbol{\tau} - \boldsymbol{\tau}_h, \mathbf{v} - \mathbf{v}_h, \boldsymbol{\eta}) - A((\boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\gamma}_h), (\boldsymbol{\tau} - \boldsymbol{\tau}_h, \mathbf{v} - \mathbf{v}_h, \boldsymbol{\eta}))\,. \end{aligned}$$

According to the definitions of the forms $A(\cdot,\cdot)$ and $F(\cdot)$, taking into account that $\mathbf{div}(\boldsymbol{\tau} - \boldsymbol{\tau}_h) = \mathbf{div}(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h)) = \mathbf{0}$ and using again (29), we find (after some algebraic manipulations) that

$$\begin{aligned} A((\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h, \boldsymbol{\gamma} - \boldsymbol{\gamma}_h), (\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta})) &= \int_\Omega (\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)) \cdot (\mathbf{v} - \mathbf{v}_h) \\ &+ \frac{1}{2} \int_\Omega (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^{\mathtt{t}}) : \boldsymbol{\eta} - \kappa_3 \int_\Omega \left(\boldsymbol{\gamma}_h - \frac{1}{2}(\nabla\mathbf{u}_h - (\nabla\mathbf{u}_h)^{\mathtt{t}}) \right) : \boldsymbol{\eta} \\ &- \int_\Omega \left((\mathcal{C}^{-1}\boldsymbol{\sigma}_h - \nabla\mathbf{u}_h + \boldsymbol{\gamma}_h) + \kappa_1\, \mathcal{C}^{-1}(\mathbf{e}(\mathbf{u}_h) - \mathcal{C}^{-1}\boldsymbol{\sigma}_h) \right) : (\boldsymbol{\tau} - \boldsymbol{\tau}_h) \\ &- \kappa_1 \int_\Omega \left(\mathbf{e}(\mathbf{u}_h) - \frac{1}{2}(\mathcal{C}^{-1}\boldsymbol{\sigma}_h + (\mathcal{C}^{-1}\boldsymbol{\sigma}_h)^{\mathtt{t}}) \right) : \nabla(\mathbf{v} - \mathbf{v}_h) \\ &+ \kappa_3 \int_\Omega \left(\boldsymbol{\gamma}_h - \frac{1}{2}(\nabla\mathbf{u}_h - (\nabla\mathbf{u}_h)^{\mathtt{t}}) \right) : \nabla(\mathbf{v} - \mathbf{v}_h). \end{aligned} \tag{31}$$

The rest of the proof of reliability consists in deriving suitable upper bounds for each one of the terms appearing on the right hand side of (31); we omit the details and refer the reader to Section 3 in [4].

On the other hand, to show that the a posteriori error estimator θ is efficient (lower bound in (25)), we proceed as in [14] and [15] and apply inverse inequalities and the localization technique introduced in [69], which is based on triangle-bubble and edge-bubble functions (see Section 4 in [4] for more details). We remark that, because of the new terms in the definition of θ (those involving the **curl** and **div** operators and the normal and tangential jumps across the edges of the triangulation), we needed to establish more general versions of some technical lemmas concerning inverse estimates and piecewise polynomials (see Lemmas 4.3-4.6 in [4]). The generality of these results allows to eventually apply them not only in the present context, but also in the a posteriori error analysis of other primal and mixed finite element methods.

We proposed the following adaptive algorithm, based on the a posteriori error estimator θ, to compute the solutions of (24) (cf. [69]):

1. Start with a coarse mesh $\mathcal{T}_h$.
2. Solve the Galerkin scheme (24) for the current mesh $\mathcal{T}_h$.
3. Compute θ_T for each triangle $T \in \mathcal{T}_h$.
4. Consider stopping criterion and decide to finish or go to next step.
5. Use *blue-green* procedure to refine each element $T' \in \mathcal{T}_h$ such that
$$\theta'_T \geq \frac{1}{2} \max_{T \in \mathcal{T}_h} \theta_T.$$
6. Define resulting mesh as the new $\mathcal{T}_h$ and go to step 2.

Numerical experiments underline the reliability and efficiency of the a posteriori error estimator θ and strongly demonstrate that the associated adaptive algorithm is much more suitable than a uniform discretization procedure when solving problems with non-smooth solutions. The robustness of θ with respect to the Poisson ratio and the ability of the adaptive algorithm to localize the singularities and large stress regions of the solution are also illustrated.

We recognize that θ is certainly more expensive than, for instance, the error indicator introduced in [6]. However, it is clear that the reliability and efficiency of θ become more advantageous features than the sole reliability of the estimator from [6]. Finally, in connection with the residual-based a posteriori error estimator developed in [15] for PEERS and BDM, which is also reliable and efficient, we point out that the advantage of θ, though a bit more expensive, is still the freedom to choose the finite element subspaces in the augmented scheme (24).

Finally, we mention that we have introduced an augmented primal-mixed method for the linear elasticity problem in the plane (see [28]) that involves four unknowns, namely: the displacement, the stress tensor, the strain tensor of small deformations and the pressure. This new variational formulation relies on the Hu-Washizu principle and was obtained by adding a least-squares term that involves the strain tensor of small deformations. We established sufficient conditions for the well-posedness of the corresponding Galerkin scheme and described a way to obtain stable finite element subspaces from any stable pair for the Stokes problem. Error estimates are also provided.

Acknowledgments

I would like to take this opportunity to express my gratitude to my thesis advisor, Salim Meddahi Bouras, and to Gabriel N. Gatica, for their generosity and support during these years. I am also very grateful to the Society for honoring me with the XI SEMA Prize to Young Researchers. This work is dedicated to my family and friends.

References

[1] T. Arbogast, M.F. Wheeler and I. Yotov. *Mixed finite elements for elliptic problems with tensor coefficients as cell–centered finite differences.* SIAM J. Numer. Anal. 34, no. 2, 828–852 (1997).

[2] R.E. Bank and A. Weiser. *Some a posteriori error estimators for elliptic partial differential equations.* Math. Comp. 44, no. 170, 283-301 (1985).

[3] M.A. Barrientos, G.N. Gatica and E.P. Stephan. *A mixed finite element method for nonlinear elasticity: two-fold saddle point approach and a-posteriori error estimate.* Numer. Math. 91, no. 2, 197-222 (2002).

[4] T.P. Barrios, G.N. Gatica, M. González and N. Heuer. *A residual based a posteriori error estimator for an augmented mixed finite element method in linear elasticity.* M2AN Math. Model. Numer. Anal. 40, no. 5, 843-869 (2006).

[5] T.P. Barrios, E.M. Behrens and M. González. *Residual based a posteriori error estimators for an augmented mixed finite element method in linear elasticity with mixed nonhomogeneous boundary conditions.* Pre-print (2009).

[6] D. Braess, O. Klaas, R. Niekamp, E. Stein and F. Wobschal. *Error indicators for mixed finite elements in 2-dimensional linear elasticity.* Comput. Methods Appl. Mech. Engrg. 127, no. 1-4, 345-356 (1995).

[7] J.H. Bramble and J.E. Pasciak. *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems.* Math. Comp. 50, no. 181, 1-17 (1988). Corrigenda: Math. Comp. 51, no. 183, 387-388 (1988).

[8] F. Brezzi and M. Fortin. Mixed and hybrid finite element methods. Springer Series in Computational Mathematics, 15. Springer-Verlag, New York, 1991.

[9] U. Brink and E. Stein. *A posteriori error estimation in large-strain elasticity using equilibrated local Neumann problems.* Comput. Methods Appl. Mech. Engrg. 161, no. 1-2, 77-101 (1998).

[10] U. Brink and E.P. Stephan. *Adaptive coupling of boundary elements and mixed finite elements for incompressible elasticity.* Numer. Methods Partial Differential Equations, 17, no. 1, 79-92 (2001).

[11] R. Bustinza, G.N. Gatica and M. González. *A mixed finite element method for the generalized Stokes problem.* Internat. J. Numer. Methods Fluids, 49, no. 8, 877-903 (2005).

[12] R. Bustinza, G.N. Gatica, M. González, S. Meddahi and E.P. Stephan. *Enriched finite element subspaces for dual-dual mixed formulations in fluid mechanics and elasticity.* Comput. Methods Appl. Mech. Engng. 194, no. 2-5, 427-439 (2005).

[13] Z. Cai, T.A. Manteuffel and S.F. McCormick. *First-order system least squares for the Stokes equations, with application to linear elasticity.* SIAM J. Numer. Anal. 34, no. 5, 1727-1741 (1997).

[14] C. Carstensen. *A posteriori error estimate for the mixed finite element method.* Math. Comp. 66, no. 218, 465-476 (1997).

[15] C. Carstensen and G. Dolzmann. *A posteriori error estimates for mixed FEM in elasticity.* Numer. Math. 81, no. 2, 187-209 (1998).

[16] C. Carstensen, S.A. Funken and E.P. Stephan. *On the adaptive coupling of FEM and BEM in 2-d-elasticity.* Numer. Math. 77, no. 2, 187-221 (1997).

[17] Z. Chen. *Expanded mixed finite element methods for linear second-order elliptic problems. I.* RAIRO Modél. Math. Anal. Numér. 32, no. 4, 479-499 (1998).

[18] Z. Chen. *Expanded mixed finite element methods for quasilinear second order elliptic problems. II.* RAIRO Modél. Math. Anal. Numér. 32, no. 4, 501-520 (1998).

[19] G. Chen and J. Zhou. Boundary element methods. Computational Mathematics and Applications. Academic Press, Ltd., London, 1992.

[20] Ph. Clément. *Approximation by finite element functions using local regularization.* RAIRO Analyse Numérique 9, no. R-2, 77-84 (1975).

[21] M. Costabel. *Symmetric methods for the coupling of finite elements and boundary elements.* Boundary elements IX, Vol. 1 (Stuttgart, 1987), 411–420, Comput. Mech., Southampton, 1987.

[22] M. Costabel, V.J. Ervin and E.P. Stephan. *Symmetric coupling of finite elements and boundary elements for a parabolic-elliptic interface problem.* Quart. Appl. Math. 48, no. 2, 265-279 (1990).

[23] M. Costabel and E.P. Stephan. *Coupling of finite and boundary element methods for an elastoplastic interface problem*. SIAM J. Numer. Anal. 27, no. 5, 1212-1226 (1990).

[24] V.J. Ervin, J.S. Howell and I. Stanculescu. *A dual-mixed approximation method for a three-field model of a nonlinear generalized Stokes problem.* Comput. Methods Appl. Mech. Engrg. 197, no. 33-40, 2886-2900 (2008).

[25] M. Feistauer. *On the finite element approximation of a cascade flow problem.* Numer. Math. 50, no. 6, 655-684 (1987).

[26] G.N. Gatica. *Solvability and Galerkin approximations of a class of nonlinear operator equations*. Z. Anal. Anwendungen, 21, no. 3, 761-781 (2002).

[27] G.N. Gatica. *Analysis of a new augmented mixed finite element method for linear elasticity allowing* $\mathbb{RT}_0 - \mathbb{P}_1 - \mathbb{P}_0$ *approximations.* M2AN Math. Model. Numer. Anal. 40, no. 1, 1-28 (2006).

[28] G.N. Gatica, L.F. Gatica and M. González. *A new augmented primal-mixed finite element method in elasticity.* Pre-print (2009).

[29] G.N. Gatica, M. González and S. Meddahi. *A low-order mixed finite element method for a class of quasi-Newtonian Stokes flows. I. A priori error analysis.* Comput. Methods Appl. Mech. Engrg. 193, no. 9-11, 881-892 (2004).

[30] G.N. Gatica, M. González and S. Meddahi. *A low-order mixed finite element method for a class of quasi-Newtonian Stokes flows. II. A posteriori error analysis.* Comput. Methods Appl. Mech. Engrg. 193, no. 9-11, 893-911 (2004).

[31] G.N. Gatica and N. Heuer. *A dual-dual formulation for the coupling of mixed-FEM and BEM in hyperelasticity.* SIAM J. Numer. Anal. 38, no. 2, 380-400 (2000).

[32] G.N. Gatica and N. Heuer. *An expanded mixed finite element approach via a dual-dual formulation and the minimum residual method.* J. Comput. Appl. Math. 132, no. 2, 371-385 (2001).

[33] G.N. Gatica and N. Heuer. *Minimum residual iteration for a dual-dual mixed formulation of exterior transmission problems.* Numer. Linear Algebra Appl. 8, no. 3, 147-164 (2001).

[34] G.N. Gatica and N. Heuer. *Conjugate gradient method for dual-dual mixed formulations.* Math. Comp., 71, no. 240, 1455-1472 (2002).

[35] G.N. Gatica, N. Heuer and S. Meddahi. *On the numerical analysis of nonlinear twofold saddle point problems.* IMA J. Numer. Anal. 23, no. 2, 301-330 (2003).

[36] G.N. Gatica and G.C. Hsiao. *On a class of variational formulations for some nonlinear interface problems*. Rend. Mat. Appl. 10, no. 4, 681-715 (1991).

[37] G.N. Gatica and G.C. Hsiao. *On the coupled BEM and FEM for a nonlinear exterior Dirichlet problem in* $\mathbb{R}^2$. Numer. Math. 61, no. 2, 171-214 (1992).

[38] G.N. Gatica and G.C. Hsiao. *Boundary-field equation methods for a class of nonlinear problems*. Pitman Research Notes in Mathematics Series, vol. 331, 1995.

[39] G.N. Gatica and S. Meddahi. *A dual-dual mixed formulation for nonlinear exterior transmission problems*. Math. Comp. 70, no. 236, 1461-1480 (2001).

[40] G.N. Gatica and E.P. Stephan. *A mixed-FEM formulation for nonlinear incompressible elasticity in the plane*. Numer. Methods Partial Differential Equations 18, no. 1, 105-128 (2002).

[41] G.N. Gatica and W.L. Wendland. *Coupling of mixed finite elements and boundary elements for linear and nonlinear elliptic problems*. Appl. Anal. 63, no. 1-2, 39-75 (1996).

[42] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations. Theory and algorithms*. Springer Series in Computational Mathematics, 5. Springer-Verlag, Berlin, 1986.

[43] M. González. *Análisis numérico de problemas de contorno en dominios no acotados del plano*. Ph.D. Thesis. Universidad de Oviedo, 2000.

[44] M. González. *Fully discrete FEM-BEM method for a class of exterior nonlinear parabolic-elliptic problems in 2D*. Appl. Numer. Math. 56, no. 10-11, 1340-1355 (2006).

[45] C.D. Han. Multiphase flow in polymer processing. Academic Press, New York, 1981.

[46] H.D. Han. *A new class of variational formulations for the coupling of finite and boundary element methods*. J. Comput. Math. 8, no. 3, 223-232 (1990).

[47] G.C. Hsiao, P. Kopp and W.L. Wendland. *A Galerkin collocation method for some integral equations of the first kind*. Computing 25, no. 2, 89-130 (1980).

[48] C. Johnson and J.-C. Nédélec. *On the coupling of boundary integral and finite element methods*. Math. Comp. 35, no. 152, 1063-1079 (1980).

[49] R. Kress. Linear integral equations. Second edition. Applied Mathematical Sciences, 82. Springer-Verlag, New York, 1999.

[50] O. A. Ladyzhenskaya. *New equations for the description of the viscous incompressible fluids and solvability in the large for the boundary value problems of them.* In: Boundary value problems of mathematical physics. V. Edited by O. A. Ladyzhenskaya. American Mathematical Society, Providence, R.I. 1970.

[51] M. Lee and F.A. Milner. *Mixed finite element methods for nonlinear elliptic problems: the p-version.* Numer. Methods Partial Differential Equations, 12, no. 6, 729-741 (1996).

[52] P. Le Tallec. *Existence and approximation results for nonlinear mixed problems: application to incompressible finite elasticity.* Numer. Math. 38, no. 3, 365-382 (1981/82).

[53] P. Le Tallec and V. Rúas. *On the convergence of the bilinear-velocity constant-pressure finite method in viscous flow.* Comput. Methods Appl. Mech. Engrg. 54, no. 2, 235-243 (1986).

[54] A.F.D. Loula and J.N.C. Guerreiro. *Finite element analysis of nonlinear creeping flows.* Comput. Methods Appl. Mech. Engrg. 79, no. 1, 87-109 (1990).

[55] H. Manouzi and M. Farhloul. *Mixed finite element analysis of a non-linear three-fields Stokes model*. IMA J. Numer. Anal. 21, no. 1, 143-164 (2001).

[56] S. Meddahi. *An optimal iterative process for the Johnson-Nedelec method of coupling boundary and finite elements.* SIAM J. Numer. Anal. 35, no. 4, 1393-1415 (1998).

[57] S. Meddahi and M. González. *A fully discrete BEM-FEM method for an exterior elasticity system in the plane.* J. Comput. Appl. Math. 134, no. 1-2, 127-141 (2001).

[58] S. Meddahi, M. González and P. Pérez. *On a FEM-BEM formulation for an exterior quasilinear problem in the plane.* SIAM J. Numer. Anal. 37, no. 6, 1820-1837 (2000).

[59] S. Meddahi and A. Márquez. *A combination of spectral and finite elements methods for an exterior problem in the plane.* Appl. Numer. Math. 43, 275-295 (2002).

[60] S. Meddahi, A. Márquez and V. Selgas. *Computing acoustic waves in an inhomogeneous medium of the plane by a coupling of spectral and finite elements.* SIAM J. Numer. Anal. 41, no. 5, 1729-1750 (2003).

[61] F.A. Milner and E.-J. Park. *A mixed finite element method for a strongly nonlinear second-order elliptic problem.* Math. Comp. 64, no. 211, 973-988 (1995).

[62] E.-J. Park. *Mixed finite element methods for nonlinear second-order elliptic problems.* SIAM. J. Numer. Anal. 32, no. 3, 865-885 (1995).

[63] P.-A. Raviart and J.-M. Thomas. Introduction à l'analyse numérique des équations aux dérivées partielles. (In French) Masson, Paris, 1991.

[64] F.-J. Sayas. *A nodal coupling of finite and boundary elements.* Numer. Methods Partial Differential Equations, 19, no. 5, 555-570 (2003).

[65] F.J. Sayas. *The validity of Johnson-Nédélec's BEM-FEM coupling on polygonal interfaces.* SIAM J. Numer. Anal. (in revision).

[66] B. Scheurer. *Existence et approximation de point selles pour certains problèmes non linéaires.* RAIRO Anal. Numér. 11, no. 4, 369-400 (1977).

[67] D. Sandri. *Sur l'approximation numérique des écoulements quasi-newtoniens dont la viscosité suit la loi puissance ou la loi de Carreau.* (In French) RAIRO Modél. Math. Anal. Numér. 27, no. 2, 131-155 (1993).

[68] I.H. Sloan. *Error analysis of boundary integral methods.* Acta numerica, 1992, 287-339, Acta Numer., Cambridge Univ. Press, Cambridge, 1992.

[69] R. Verfürth. A review of a posteriori error estimation and adaptive mesh-refinement techniques. Wiley-Teubner, Chichester, 1996.

[70] J. Xu. *Theory of multilevel methods.* Ph.D. Thesis. Cornell University, 1989.

[71] J. Xu. *Two-grid discretization techniques for linear and nonlinear PDEs.* SIAM J. Numer. Anal. 33, no. 5, 1759–1777 (1996).

[72] A. Ženíšek. *Finite element variational crimes in parabolic-elliptic problems. I. Nonlinear schemes.* Numer. Math. 55, no. 3, 343-376 (1989).

[73] A. Ženíšek. *Nonlinear elliptic and evolution problems and their finite element approximations.* Computational Mathematics and Applications. Academic Press, Inc., London, 1990.

[74] O.C. Zienkiewicz, D.W. Kelly and P. Bettess. *The coupling of the finite element method and boundary solution procedures.* Internat. J. Numer. Methods Engrg. 11, no. 2, 355-375 (1977).

[75] O.C. Zienkiewicz, D.W. Kelly and P. Bettess. *Marriage à la mode: the best of both worlds (finite elements and boundary integrals).* Energy methods in finite element analysis, pp. 81–107, Wiley, Chichester, 1979.

María González Taboada es licenciada en Matemáticas en 1996 por la especialidad de Matemática Aplicada y Computación, con Premio Extraordinario de licenciatura de la Universidad de Oviedo y Premio Fin de Carrera Arthur Andersen. Es Doctora en Matemáticas en 2000 por la Universidad de Oviedo, con Premio Extraordinario de Doctorado. Actualmente es Profesora Titular del Área de Matemática Aplicada de la Universidad de A Coruña, desde septiembre de 2003.

Tras la presentación de su tesis doctoral, dirigida por el profesor Salim Meddahi de la Universidad de Oviedo, ha realizado varias estancias de investigación en el Departamento de Ingeniería Matemática de la Universidad de Concepción (Chile) y en la Universidad Católica de la Santísima Concepción (Chile).

También ha colaborado en proyectos de transferencia entre la universidad y la empresa, siendo responsable de un contrato con la empresa Advaced Dynamics S.A. y participando en otro con la empresa Navantia.

Título:	CONTRIBUCIÓN AL ESTUDIO DE MODELOS MATEMÁTICOS TERMOHIDRODINÁMICOS EN LUBRICACIÓN.
Doctorando:	José Pereira Pérez.
Director/es:	José Durany Castrillo, Fernando Varas Mérida.
Defensa:	26 de Marzo de 2005, Vigo.
Calificación:	Sobresaliente cum laude por unanimidad.

Resumen:

El objetivo científico de esta Tesis Doctoral ha sido el estudio matemático y numérico de problemas complejos que surgen de la termohidrodinámica de cojinentes sometidos a procesos de lubricación. Este tipo de dispositivo mecánico ha sido y continúa siendo utilizado en una gran variedad de maquinaria debido a su simplicidad de construcción y a las buenas propiedades de comportamiento que presenta.

En concreto, en el Capítulo 1 se realiza una exposición de los modelos matemáticos que serán empleados en el análisis y resolución numérica del problema termohidrodinámico en el par eje–cojinete, tanto para el caso estacionario como el transitorio. Se plantean las ecuaciones en derivadas parciales (EDP) que rigen los distintos fenómenos involucrados y sus condiciones de contorno. Se describen las variables, parámetros y coeficientes que intervienen en las mismas, y los acoplamientos entre las ecuaciones térmica e hidrodinámica. En este Capítulo se aporta un análisis riguroso de la acotación de los coeficientes viscosos en la ecuación de Reynolds. Esta acotación es fundamental para obtener el resultado de existencia de solución del problema hidrodinámico, estimaciones de la misma y su posterior resolución numérica. Asimismo, se obtienen estimaciones de la derivada de la presión en el caso unidimensional.

En el Capítulo 2 se expone detalladamente el método numérico utilizado en la resolución del problema hidrodinámico con el modelo de cavitación de Elrod-Adams. Este es un problema de frontera libre, para el que existen diferentes estrategias numéricas de resolución . La elección que se considera aquí utiliza métodos de dominio fijo conjuntamente con métodos upwind para los términos convectivos, y métodos de dualidad para tratar las nolinealidades. La combinación de métodos de características con métodos de dualidad se justifica por haber sido utilizados con gran éxito en problemas similares de flujo de gases, cambios de fase y, también, en lubricación . Aunque el esquema es ya conocido, en este Capítulo se analiza la convergencia de la presión en el modelo Elrod-Adams para la cavitación a partir de una solución exacta para el caso 1-D. Se

analizan también las influencias de la temperatura y condiciones de alimentación sobre la solución obtenida.

El Capítulo 3 se dedica al método de resolución del problema térmico en el fluido. La aplicación de un método cell-vertex para la resolución de la ecuación de la energía constituye una de las aportaciones más novedosas e interesantes de la Tesis. La aplicación del método a la resolución del problema térmico en el fluido se expone en detalle, comprobando que permite una sencilla construcción del sistema de ecuaciones a resolver; similar a los ensamblados de los métodos de elementos finitos. Se demuestra que presenta una convergencia de segundo orden, superior a la de otros esquemas habituales. Esto permite lograr mejores aproximaciones con mallas menos refinadas.

El Capítulo 4 está dedicado al problema acoplado termohidrodinámico (THD), en el que se deben resolver simultáneamente la ecuación de Reynolds con el modelo de cavitación de Elrod-Adams y el problema térmico en el fluido, el eje y el cojinete. En este Capítulo se propone una forma novedosa de abordar el problema térmico en el cojinete consistente en el empleo de elementos de contorno (MEC). Este procedimiento elimina la necesidad de crear una malla para el cojinete, resuelve la temperatura únicamente en las fronteras del mismo y con ello reduce el número de incógnitas y el coste computacional. El procedimiento de construcción del sistema de ecuaciones asociado al MEC se expone en detalle, pudiendo comprobarse que la simetría del problema facilita de forma importante los cálculos. También se realiza un estudio para comprobar que se trata de un método de orden dos. Dentro de este mismo Capítulo se presentan los resultados correspondientes a la resolución del problema THD con diferentes condiciones de trabajo y de frontera.

En el Capítulo 5 se realizan las extensiones de los métodos al caso termohidrodinámico transitorio. Es de especial interés el estudio realizado para extender el método de resolución del problema térmico en el cojiente mediante los elementos de contorno (MEC) al caso temporal, lo que supone otra aportación importante de la Tesis. En concreto, la utilización del método de reciprocidad dual (MRD), que obliga al cálculo de la temperatura en puntos internos al dominio y da lugar a una convergencia de orden menor que dos.

Finalmente, el Capítulo 6 está dedicado al estudio de la estabilidad dinámica del sistema eje-cojinete. En este Capítulo se presenta una novedosa solución analítica del problema 1-D con modelo de cavitación del Elrod-Adams y se estudia su rango de estabilidad, comparando los resultados con los de los modelos de cojinete corto e infinitamente largo, habituales en estudios de estabilidad de estos dispositivos. Además, se propone un nuevo procedimiento para el análisis de la estabilidad que consiste en resolver el acoplamiento de la dinámica del eje con la hidrodinámica del fluido lubricante mediante un método de Euler implícito. A cada paso temporal del método de Euler, se calcula la solución de un sistema de ecuaciones no lineales mediante un método de Broyden, empleando la técnica de Armijo-Goldstein para la elección del paso de descenso. Se presentan los resultados del análisis de estabilidad con este nuevo método para distintas presiones de alimentación y situaciones geométricas de la ranura de alimentación. Finalmente, se analiza también la influencia de los aspectos térmicos en las curvas de estabilidad del dispositivo.

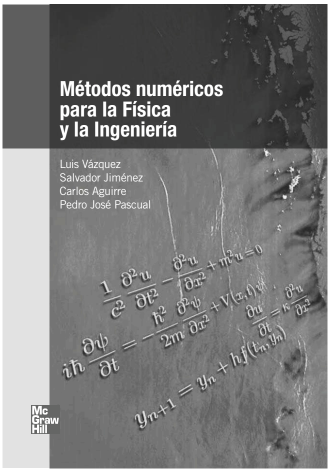
Título:	THEORETICAL AND NUMERICAL STUDY OF THE STABILITY OF A CONVECTION PROBLEM WITH VARIABLE VISCOSITY.
Doctorando:	Francisco Plá Martos.
Director/es:	Henar Herrero Sanz, Ana María Mancho Sánchez.
Defensa:	14 de abril de 2009, Universidad de Castilla-La Mancha..
Calificación:	Sobresaliente cum laude por unanimidad..

Resumen:

En el presente trabajo estudiamos el problema de convección de Rayleigh-Bénard con viscosidad variable, dependiente de la temperatura, como una primera aproximación a la convección en el manto terrestre o planetario. En los resultados numéricos la viscosidad tiene un perfil exponencial con la temperatura en el término de la divergencia de las ecuaciones del movimiento. El problema será resuelto en tres geometrías cartesianas distintas. En primer lugar consideramos un dominio tridimensional (3D) en el que el fluido se encuentra entre dos planos paralelos no acotados. El fluido es calentando uniformemente desde abajo y nos planteamos el problema de estabilidad lineal de la solución conductiva. Se demuestra que la solución conductiva pierde su estabilidad por medio de una bifurcación estacionaria. Se obtienen las curvas de estabilidad marginales para distintos perfiles de viscosidad y condiciones de contorno: rígidas en ambos planos y por otro lado, rígidas en el plano inferior y libres en el superior. Los números de Rayleigh críticos son calculados. En segundo lugar estudiamos la estabilidad de la solución estacionaria en un dominio bidimensional (2D) acotado. El fluido es calentado uniformemente desde la pared inferior y se consideran condiciones de contorno rígidas para la velocidad en el plano inferior y libres en el resto. Las respectivas curvas de estabilidad marginales están en función de la relación de aspecto de la celda 2D. A través de la teoría de bifurcación, encontramos distintas soluciones que pueden coexistir bajo un mismo régimen de parámetros. Los diagramas de bifurcación presentan diagramas saddle-node y subcríticos en el caso de viscosidad variable y de tipo doble-pitchfork en el caso de viscosidad constante. Esto puede sugerir que la evolución térmica en los planetas no solo depende de los componentes químicos o geofísicos sino de su tamaño. El método numérico utilizado está basado en un método de Chebyshev de colocación. Finalmente, se estudia la estabilidad lineal del estado básico en un dominio 3D no acotado según uno de los ejes y con gradiente horizontal de temperatura en el plano inferior. El movimiento del fluido tiende a localizarse en las paredes más calientes y se obtienen estructuras en la dirección no acotada del dominio mientras que con calentamiento uniforme las soluciones no suelen presentar una estructura 3D. En las tres geometrías se observa que grandes contrastes de viscosidad favorece la inestabilidad del sistema y el movimiento del fluido se concentra en el plano inferior.

Métodos numéricos para la Física y la Ingeniería
Luis Vázquez, Salvador Jiménez, Carlos Aguirre y Pedro José Pascual
Editorial McGraw-Hill
ISBN: 978-84-481-66021 (384 páginas) – 2008

Por Miguel A. F. Sanjuán



¿Otro libro más de métodos numéricos? Muy bien podría ser la pregunta que uno podría hacerse ante esta nueva aportación a la literatura de los profesores Luis Vázquez, Salvador Jiménez, Carlos Aguirre y Pedro José Pascual. La cuestión es que si bien se trata de otro libro de métodos numéricos, sus características son tales que lo hacen especial, especialmente debido a la formación de sus autores, ya que además de su experiencia docente de años en estas materias, a ello se une su extensa actividad de investigación en simulación numérica de problemas de ciencia y tecnología.

Los contenidos del libro abarcan desde temas sencillos a otros más avanzados y bien puede considerarse como un buen manual útil para la enseñanza de métodos numéricos a estudiantes de ciencias físicas o ingeniería. Los primeros capítulos contienen muchos aspectos prácticos que suelen ser de gran utilidad para el principiante a fin de hacerle comprender la naturaleza de los errores en el cálculo numérico, aportando numerosos ejemplos y ejercicios al final. Le siguen los métodos para calcular los ceros de una función, ilustrándose con algunos ejemplos de fácil comprensión. Posteriormente los métodos de integración y derivación numérica, para pasar a continuación a los métodos de aproximación de funciones, donde se incluyen los métodos de trazadores (splines). El capítulo que se dedica a los sistemas de ecuaciones lineales, autovalores y autovectores es muy extenso cubriendo los aspectos básicos y otros más avanzados. Un parte muy importante del libro es el dedicado a las ecuaciones diferenciales, tanto ordinarias como en derivadas parciales dada su importancia en las aplicaciones en la Física y la Ingeniería. Se describen numerosos ejemplos de aplicaciones con ejemplos de sistemas dinámicos no lineales.

Dado el carácter enormemente práctico del libro, se incluye un capítulo dedicado a ciertos tratamientos de problemas de Física como el dedicado al movimiento de una partícula clásica, donde se introducen nociones básicas de mecánica clásica y métodos de discretización de las ecuaciones del movimiento de una partícula, usando esquemas tanto conservativos, como disipativos e ilustrándose mediante el uso de modelos de osciladores no lineales como el oscilador de Duffing. En este contexto resulta muy interesante y además constituye un aspecto original de este libro el estudio que se presenta acerca de la partícula cargada en un campo electromagnético.

Por último se hace un estudio pormenorizado de la ecuación de ondas, introduciendo el método de las diferencias finitas y de las ecuaciones de ondas no lineales como la ecuación de Klein–Gordon, ecuaciones de ondas disipativas con soluciones tipo solitón. Las ecuaciones en derivadas parciales, tanto parabólicas, como la ecuación de difusión, tanto lineal como no lineal, como las elípticas, como las de Laplace y Poisson merecen un capítulo especial de especial interés por sus aplicaciones.

Para terminar se incluye un capítulo de especial importancia, que es el dedicado a las ecuaciones de Maxwell y a la ecuación de Schrödinger, con un tratamiento especial a la ecuación de Schrödinger no lineal, con soluciones tipo solitón. Aquí queda de manifiesto la experiencia en investigación de los autores, donde también se incluye una bibliografía muy extensa. En este contexto de aplicaciones de métodos numéricos a problemas de Física e Ingeniería, hubiera estado bien haber incluido un tratamiento de métodos numéricos aplicados a la Mecánica de Fluidos.

En conjunto se trata de una novedad que puede ser de gran utilidad para la enseñanza de métodos numéricos para estudiantes de Física e Ingeniería, y asimismo constituye una buena introducción para el aprendizaje de la fundamentación matemática de métodos numéricos que puede ser necesario para estudiante de doctorado.

Tipo de evento:	Simposio
Nombre:	Positive systems: Theory and Applications (POSTA09)
Lugar:	València
Fecha:	2–4 de septiembre de 2009
Organiza:	Universidad Politécnica de València
Información:	
E-mail:	`posta09@imm.upv.es`
WWW:	`http://posta09.webs.upv.es/`

Tipo de evento:	Workshop
Nombre:	International Workshop - Homogenization and Optimal Design
Lugar:	Sevilla
Fecha:	11–12 de septiembre de 2009
Organiza:	Lucio Boccardo (Università Roma 2, Italy), Carmen Calvo Jurado (Universidad de Extremadura, Spain), Juan Casado Díaz (Universidad de Sevilla, Spain), Julio Couce Calvo (Universidad de Sevilla, Spain), Faustino Maestre Caballero (Universidad de Sevilla, Spain), José D. Martín Gómez (Universidad de Sevilla, Spain), François Murat (Université Paris VI, France), Beatriz Ordóñez Flores (Universidad de Sevilla, Spain), Francisco J. Suárez Grau (Universidad de Sevilla, Spain)
Información:	
E-mail:	`mllaynez@us.es`
WWW:	`http://congreso.us.es/fqm309/`

Tipo de evento:	Congreso
Nombre:	ECCOMAS Thematic Conference: IV International Conference on Textile Composites and Inflatable Structures
Lugar:	Stuttgart, Germany
Fecha:	5–7 October 2009
Organiza:	E. Oñate, Universitat Politècnica de Catalunya; B. Kröplin, University of Stuttgart, Germany
Información:	
E-mail:	`membranes@cimne.upc.edu`
WWW:	`http://congress.cimne.upc.es/membranes09/frontal/`

Tipo de evento:	Congreso
Nombre:	SIAM Conference Mathematics for Industry: Challenges and Frontiers (MI09)
Lugar:	San Francisco, California
Fecha:	October 9–10, 2009
Organiza:	SIAM
Información:	
E-mail:	
WWW:	`http://www.siam.org/meetings/mi09/`

Tipo de evento:	Congreso
Nombre:	ECCOMAS Thematic Conference: International Conference on Particle–Based Methods (Particle 2009)
Lugar:	Barcelona
Fecha:	25–27 November 2009
Organiza:	D. R. J. Owen (Chairman), Swansea University, UK; E. Oñate (Co–Chairman), Univ. Politècnica de Catalunya, Spain; J. Bonet, Swansea University, UK; Y. T. Feng, Swansea University, UK; A. Huerta, Univ.Politècnica de Catalunya, Spain; S. Idelsohn, CIMNE, Barcelona, Spain; X. Oliver, Univ.Politècnica de Catalunya, Spain
Información:	
E-mail:	`particle-basedmethods@cimne.upc.edu`
WWW:	`http://congress.cimne.com/particles2009/frontal/`

Tipo de evento:	Congreso
Nombre:	SIAM Conference on Analysis of Partial Differential Equations
Lugar:	Miami, Florida
Fecha:	December 7–9, 2009
Organiza:	SIAM
Información:	
E-mail:	
WWW:	`http://www.siam.org/meetings/pd09/`

Tipo de evento:	Workshop
Nombre:	New Directions in Financial Mathematics
Lugar:	Institute for Pure and Applied Mathematics (IPAM), UCLA, Los Angeles, California
Fecha:	January 5–9, 2010
Organiza:	Rene Carmona (Princeton University, Mathematics); Jaska Cvitanic (California Institute of Technology); Nicole El Karoui (École Polytechnique); George Papanicolaou (Stanford University); Eduardo Schwartz (University of California, Los Angeles (UCLA), Anderson); Ronnie Sircar (Princeton University); Thaleia Zariphopoulou (University of Texas at Austin, Departments of Mathematics and IROM)
Información:	
E-mail:	`fin2010@ipam.ucla.edu`
WWW:	`http://www.ipam.ucla.edu/programs/fin2010/`

Tipo de evento:	Congreso
Nombre:	ACM–SIAM Symposium on Discrete Algorithms (SODA10)
Lugar:	Austin, Texas
Fecha:	January 17–19, 2010
Organiza:	SIAM
Información:	
E-mail:	
WWW:	`http://www.siam.org/meetings/da10/`

Tipo de evento:	Congreso
Nombre:	METAMATERIALS: APPLICATIONS, ANALYSIS AND MODELING
Lugar:	Institute for Pure and Applied Mathematics (IPAM), UCLA, Los Angeles, California
Fecha:	January 25–29, 2010
Organiza:	Robert Kohn, Co–Chair (New York University, Courant Institute); Graeme Milton, Co–Chair (University of Utah, Mathematics); Susanne Brenner (Louisiana State University); Maria–Carme Calderer (University of Minnesota, Twin Cities); Tatsuo Itoh (University of California, Los Angeles (UCLA)); Jichun Li (University of Nevada, Las Vegas, Mathematical Sciences); Chi–Wang Shu (Brown University); Richard W. Ziolkowski (University of Arizona, Engineering)
Información:	
E-mail:	`meta2010@ipam.ucla.edu`
WWW:	`http://www.ipam.ucla.edu/programs/meta2010/`

Tipo de evento:	Congreso
Nombre:	THE INTERNATIONAL SYMPOSIUM ON STOCHASTIC MODELS IN RELIABILITY ENGINEERING, LIFE SCIENCES, AND OPERATIONS MANAGEMENT
Lugar:	Beer Sheva, Israel
Fecha:	February 8–10, 2010
Organiza:	Prof. N. Balakrishnan, McMaster University, Canada; Prof. Alan Hutson, State University of New York at Buffalo, USA; Prof. Zohar Laslo, Sami Shamoon College of Engineering, Israel
Información:	
E-mail:	`SMRLO10@sce.ac.il`
WWW:	`http://www.sce.ac.il/smrlo/`

Cañizo Rincón, José Alfredo
Investigador. *Líneas de investigación:* Teoría de existencia y comportamiento asintótico de ecuaciones cinéticas, especialmente modelos de coagulación-fragmentación – Univ. Autónoma de Barcelona – Fac. de Ciencias – Depto. de Matemáticas – 08193 - Bellaterra (Barcelona).
Tlf.: 935.813.104. *Fax:* 935.812.790.
e-mail: `canizo@mat.uab.cat`
`http://mat.uab.cat/~canizo`

Morales Rodrigo, Cristian
Líneas de investigación: – Univ. de Sevilla – Fac. de Matemáticas – Dpto. de Ecuaciones Diferenciales y Análisis Numérico – Campus Avda. Reina Mercedes. 41012 Sevilla.
Tlf.: 954.557.981. *Fax:* 954.552.898.
e-mail: `cristianm@us.es`

Direcciones útiles

Consejo Ejecutivo de S$\vec{e}$MA

Presidente:

Carlos Vázquez Cendón. (carlosv@udc.es).
Dpto. de Matemáticas. Facultad de Informática. Univ. de A Coruña. Campus de Elviña, s/n. 15071 A Coruña. *Tel:* 981 16 7000-1335.

Vicepresidente:

Rosa María Donat Beneito. (Rosa.M.Donat@uv.es)
Dpto. de Matemática Aplicada. Fac. de Matemàtiques. Univ. de Valencia. Dr. Moliner, 50. 46100 Burjassot (Valencia) *Tel:* 963 544 727.

Secretario:

Carlos Castro Barbero. (ccastro@caminos.upm.es).
Dpto. de Matemática e Informática. E.T.S.I. Caminos, Canales y Puertos. Univ. Politécnica de Madrid. Av. Aranguren s/n. 28040 Madrid. *Tel:* 91 336 6664.

Vocales:

Sergio Amat Plata. (sergio.amat@upct.es)
Dpto. de Matemática Aplicada y Estadística. Univ. Politécnica de Cartagena. Paseo de Alfonso XIII, 52. 30203 Cartagena (Murcia). *Tel:* 968 325 694.

Rafael Bru García. (rbru@mat.upv.es)
Dpto. de Matemática Aplicada. E.T.S.I. Agrónomos. Univ. Politécnica de Valencia. Camí de Vera, s/n. 46022 Valencia. *Tel:* 963 879 669.

José Antonio Carrillo de la Plata. (carrillo@mat.uab.es)
Dpto. de Matemáticas. Univ. Autónoma de Barcelona. Edifici C. 08193 Bellaterra (Barcelona). *Tel:* 935 812 413.

Inmaculada Higueras Sanz. (higueras@unavarra.es).
Dpto de Matemática e Informática Univ. Pública de Navarra. Campus de Arrosadía, s/n. *Tel:* 948 169 526. 31006 Pamplona.

Carlos Parés Madroñal. (carlos_pares@uma.es).
Dpto. de Análisis Matemático. Fac. de Ciencias. Univ. de Málaga. Campus de Teatinos, s/n. 29080 Málaga. *Tel:* 952 132 017.

Pablo Pedregal Tercero. (Pablo.Pedregal@uclm.es).
Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla-La Mancha. Avda. de Camilo José Cela, s/n. 13071 Ciudad Real. *Tel:* 926 295 436

Luis Vega González. (luis.vega@ehu.es).
Dpto. de Matemáticas. Fac. de Ciencias. Univ. del País Vasco. Aptdo. 644. 48080 Bilbao (Vizcaya). *Tel:* 944 647 700.

Tesorero:

Íñigo Arregui Álvarez. (arregui@udc.es).
Dpto. de Matemáticas. Fac. de Informática. Univ. de A Coruña. Campus de Elviña, s/n. 15071 A Coruña. *Tel:* 981 16 7000-1327.

Comité Científico del Boletín de S$\vec{e}$MA

Luis Vega González. (mtpvegol@lg.ehu.es).
Dpto. de Matemáticas. Fac. de Ciencias. Univ. del País Vasco. Aptdo. 644. 48080 Bilbao (Vizcaya). *Tel:* 944 647 700.

Chi-Wang Shu. (shu@dam.brown.edu).
Division of Applied Mathematics Box F. 182 George Street Brown University Providence RI 02912 *Tel:* (401) 863-2549

Enrique Zuazua Iriondo. (zuazua@bcamath.org).
Basque Center for Applied Mathematics Bizkaia Technology Park Building 208B 48170 - Zamudio (Vizcaya) *Tel:* 944 014 690

Grupo Editor del Boletín de S$\vec{e}$MA

Pablo Pedregal Tercero. (Pablo.Pedregal@uclm.es).
Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla-La Mancha. Avda. Camilo José Cela, s/n. 13071 Ciudad Real. *Tel:* 926 295 300 ext. 3809

Enrique Fernández Cara. (cara@us.es).
Dpto. de Ecuaciones Diferenciales y An. Numérico. Fac. de Matemáticas. Univ. de Sevilla. Tarfia, s/n. 41012 Sevilla. *Tel:* 954 557 992.

Ernesto Aranda Ortega. (Ernesto.Aranda@uclm.es).
Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla-La Mancha. Avda. Camilo José Cela, s/n. 13071 Ciudad Real. *Tel:* 926 295 300 ext. 3813

José Carlos Bellido Guerrero. (JoseCarlos.Bellido@uclm.es).
Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla-La Mancha. Avda. Camilo José Cela, s/n. 13071 Ciudad Real. *Tel:* 926 295 300 ext. 3859

Alberto Donoso Bellón. (Alberto.Donoso@uclm.es).
Dpto. de Matemáticas. E.T.S.I. Industriales. Univ. de Castilla-La Mancha. Avda. Camilo José Cela, s/n. 13071 Ciudad Real. *Tel:* 926 295 300 ext. 3859

Responsables de secciones del Boletín de S$\vec{e}$MA

Artículos:

Enrique Fernández Cara. (cara@us.es).
Dpto. de Ecuaciones Diferenciales y An. Numérico. Fac. de Matemáticas. Univ. de Sevilla. Tarfia, s/n. 41012 Sevilla. *Tel:* 954 557 992.

Matemáticas e Industria:

Mikel Lezaun Iturralde. (mepleitm@lg.ehu.es).
Dpto. de Matemática Aplicada, Estadística e I. O. Fac. de Ciencias. Univ. del País Vasco. Aptdo. 644. 48080 Bilbao (Vizcaya). *Tel:* 944 647 700.

Educación Matemática:

Roberto Rodríguez del Río. (rr_delrio@mat.ucm.es).
Dpto. de Matemática Aplicada. Fac. de Químicas. Univ. Compl. de Madrid. Ciudad Universitaria. 28040 Madrid. *Tel:* 913 944 102.

Resúmenes de libros:

Fco. Javier Sayas González. (jsayas@posta.unizar.es).
Dpto. de Matemática Aplicada. Centro Politécnico Superior . Universidad de Zaragoza. C/María de Luna, 3. 50015 Zaragoza. *Tel:* 976 762 148.

Noticias de S$\vec{e}$MA:

Carlos Castro Barbero. (`ccastro@caminos.upm.es`).
Dpto. de Matemática e Informática. E.T.S.I. Caminos, Canales y Puertos. Univ. Politécnica de Madrid. Av. Aranguren s/n. 28040 Madrid. *Tel:* 91 336 6664.

Anuncios:

Óscar López Pouso. (`oscarlp@usc.es`).
Dpto. de Matemática Aplicada. Fac. de Matemáticas. Univ. de Santiago de Compostela. Campus sur, s/n. 15782 Santiago de Compostela *Tel:* 981 563 100, ext. 13228.

Responsables de otras secciones de S$\vec{e}$MA

Gestión de Socios:

Íñigo Arregui Álvarez. (`arregui@udc.es`).
Dpto. de Matemáticas. Fac. de Informática. Univ. de A Coruña. Campus de Elviña, s/n. 15071 A Coruña. *Tel:* 981 16 7000-1327.

Página web: `www.sema.org.es/`**:**

Carlos Castro Barbero. (`ccastro@caminos.upm.es`).
Dpto. de Matemática e Informática. E.T.S.I. Caminos, Canales y Puertos. Univ. Politécnica de Madrid. Av. Aranguren s/n. 28040 Madrid. *Tel:* 91 336 6664.

1. Los artículos publicados en este Boletín podrán ser escritos en español o inglés y deberán ser enviados por correo certificado a

   Prof. E. FERNÁNDEZ CARA
   Presidente del Comité Científico, Boletín S$\vec{e}$MA
   Dpto. E.D.A.N., Facultad de Matemáticas
   Aptdo. 1160, 41080 SEVILLA

   También podrán ser enviados por correo electrónico a la dirección

   `boletin.sema@uclm.es`

   En ambos casos, el/los autor/es deberán enviar por correo certificado una carta a la dirección precedente mencionando explícitamente que el artículo es sometido a publicación e indicando el nombre y dirección del autor corresponsal. En esta carta, podrán sugerirse nombres de miembros del Comité Científico que, a juicio de los autores, sean especialmente adecuados para juzgar el trabajo.

   La decisión final sobre aceptación del trabajo será precedida de un procedimiento de revisión anónima.

2. Las contribuciones serán preferiblemente de una longitud inferior a 24 páginas y se deberán ajustar al formato indicado en los ficheros a tal efecto disponibles en la página web de la Sociedad (`http://www.sema.org.es/`).

3. El contenido de los artículos publicados corresponderá a un área de trabajo preferiblemente conectada a los objetivos propios de la Matemática Aplicada. En los trabajos podrá incluirse información sobre resultados conocidos y/o previamente publicados. Se anima especialmente a los autores a presentar sus propios resultados (y en su caso los de otros investigadores) con estilo y objetivos divulgativos.

Ficha de Inscripción Individual

Sociedad Española de Matemática Aplicada S$\vec{e}$MA

Remitir a: Iñigo Arregui, Dpto de Matemáticas, Fac. de Informática,
Universidad de A Coruña. Campus de Elviña, s/n. 15071 A Coruña.
CIF: G-80581911

Datos Personales

- Apellidos: ....
- Nombre: ....
- Domicilio: ....
- C.P.: .... Población: ....
- Teléfono: .... DNI/CIF: ....
- Fecha de inscripción: ....

Datos Profesionales

- Departamento: ....
- Facultad o Escuela: ....
- Universidad o Institución: ....
- Domicilio: ....
- C.P.: .... Población: ....
- Teléfono: .... Fax: ....
- Correo electrónico: ....
- Página web: `http://` ....
- Categoría Profesional: ....
- Líneas de Investigación: ....

  ....

Dirección para la correspondencia: ☐ **Profesional** ☐ **Personal**

Cuota anual para el año 2009

☐ Socio ordinario: 30€ ☐ Socio de reciprocidad con la RSME: 12€
☐ Socio estudiante: 15€

Datos bancarios

. . . de . . . . de 200. .

Muy Sres. Míos:

Ruego a Uds. que los recibos que emitan a mi cargo en concepto de cuotas de inscripción y posteriores cuotas anuales de S$\vec{e}$MA (Sociedad Española de Matemática Aplicada) sean pasados al cobro en la cuenta cuyos datos figuran a continuación

Entidad (4 dígitos)	Oficina (4 dígitos)	D.C. (2 dígitos)	Número de cuenta (10 dígitos)

- Entidad bancaria: . . . .
- Domicilio: . . . .
- C.P.: . . . . Población: . . . .

Con esta fecha, doy instrucciones a dicha entidad bancaria para que obren en consecuencia.

Atentamente,

Fdo. . . . .

Para remitir a la entidad bancaria

. . . de . . . . de 200. .

Muy Sres. Míos:

Ruego a Uds. que los recibos que emitan a mi cargo en concepto de cuotas de inscripción y posteriores cuotas anuales de S$\vec{e}$MA (Sociedad Española de Matemática Aplicada) sean cargados a mi cuenta corriente/libreta . . . . en esa Agencia Urbana y transferidas a

SEMA: 0128 - 0380 - 03 - 0100034244
Bankinter
C/ Hernán Cortés, 63
39003 Santander

Atentamente,

Fdo. . . . .

Ficha de Inscripción Institucional

Sociedad Española de Matemática Aplicada S$\vec{e}$MA

Remitir a: Iñigo Arregui, Dpto de Matemáticas, Fac. de Informática,
Universidad de A Coruña. Campus de Elviña, s/n. 15071 A Coruña.
CIF: G-80581911

Datos de la Institución

- Departamento: ....
- Facultad o Escuela: ....
- Universidad o Institución: ....
- Domicilio: ....
- C.P.: .... Población: ....
- Teléfono: .... DNI/CIF: ....
- Correo electrónico: ....
- Página web: `http://` ....
- Fecha de inscripción: ....

Forma de pago

La cuota anual para el año 2009 como Socio Institucional es de 150€.
El pago se realiza mediante transferencia bancaria a

SEMA: 0128 - 0380 - 03 - 0100034244
Bankinter
C/ Hernán Cortés, 63
39003 Santander