

# Applications of Operator-Splitting Methods to the Direct Numerical Simulation of Particulate and Free-Surface Flows and to the Numerical Solution of the Two-Dimensional Elliptic Monge–Ampère Equation

Dedicated to J. Douglas, G.I. Marchuk, D.H. Peaceman and H.H. Rachford

R. GLOWINSKI\*, E.J. DEAN<sup>†</sup>, G. GUIDOBONI<sup>‡</sup>, L.H. JUÁREZ<sup>§</sup> and T.-W. PAN<sup>¶</sup>

\**Department of Mathematics, University of Houston  
Houston, Texas 77204, USA and  
Laboratoire Jacques-Louis Lions  
University of Pierre et Marie Curie, Paris, France  
E-mail: roland@math.uh.edu*

<sup>†</sup>*Department of Mathematics, University of Houston  
Houston, Texas 77204, USA*

<sup>‡</sup>*Department of Mathematics, University of Houston  
Houston, Texas 77204, USA*

<sup>§</sup>*Departamento de Matemáticas  
Universidad Autónoma Metropolitana-Iztapalapa  
Iztapalapa, D.F. 09340, Mexico*

<sup>¶</sup>*Department of Mathematics, University of Houston  
Houston, Texas 77204, USA*

Received September 14, 2005

Revised August 3, 2007

The main goal of this article is to review some recent applications of operator-splitting methods. We will show that these methods are well-suited to the numerical solution of outstanding problems from various areas in Mechanics, Physics and Differential Geometry, such as the direct numerical simulation of particulate flow, free boundary problems with surface tension for incompressible viscous fluids, and the elliptic real Monge–Ampère equation. The results of numerical experiments will illustrate the capabilities of these methods.

*Key words:* operator-splitting, particulate flow, free surface flow, Monge–Ampère equation, fictitious domain

## 1. Introduction

During the year 2005, the Scientific Community celebrated “50 Years of Alternating Directions,” a tribute to the well-known article by D. Peaceman and H. Rachford (ref. [56]) and many contributions on related topics which followed its publication (ref. [23], in particular). Actually, Alternating Direction Methods are particular and important cases of a more general class of methods, namely those based on the concept of Operator-Splitting. To the best of our knowledge, the most complete text on operator-splitting methods (including alternating direction ones) is the review article by G.I. Marchuk (ref. [43]); it provides a thorough discussion of these methods and describes many applications (our main regret concerning ref. [43] is that it never appeared as a single book). Our main goal with this article is, in some sense, to go beyond [43] by discussing the operator-splitting solution of

a variety of problems that the authors have investigated together these past years, the idea being here to convey the versatility and flexibility of operator-splitting methods. Indeed, it has been known for decades that operator-splitting methods provide efficient tools for the numerical solution of complicated problems from various areas in Science and Engineering. New applications appear almost daily and we know of many instances where the *only available* practical solution methods are of the operator-splitting type. Our goal, here, is to show that operator-splitting makes relatively simple the numerical solution of outstanding problems such as:

- (i) The simulation of *particulate flow* when the number of particles exceeds 100.
- (ii) *Free boundary problems* with *surface tension* for incompressible viscous fluids.
- (iii) *Fully nonlinear elliptic equations* of the real *Monge–Ampère* type.

The content of this article is as follows: In Section 2 we will discuss briefly the *time-discretization of initial value problems* by various operator-splitting methods and show that well-known iterative methods are indeed disguised operator-splitting schemes. In Section 3 we will address the *direct numerical simulation of particulate flow* via a methodology combining operator-splitting and *fictitious domain* methods. Then, in Section 4 we will address the solution of *free boundary problems* with surface tension for incompressible viscous fluids. Finally, in Section 5, we will discuss the numerical solution of the two-dimensional Dirichlet problem for the *elliptic Monge–Ampère equation*. The results of numerical experiments will be given in Sections 3, 4 and 5; they will confirm the capabilities of operator-splitting methods concerning the solution of problems still considered complicated by today standards.

Operator-splitting methods have generated a huge literature; let us mention, among many others, refs. [43], [13], [31], [28, Chapters 2 and 6], and [68] (see also the references therein).

## 2. Operator-splitting schemes for the time-discretization of initial value problems

### 2.1. Generalities

Let us consider the following autonomous initial value problem:

$$\begin{cases} \frac{d\phi}{dt} + A(\phi) = 0 & \text{on } (0, T) \quad (\text{with } 0 < T \leq +\infty), \\ \phi(0) = \phi_0. \end{cases} \quad (\text{IVP})$$

Operator  $A$  maps the vector space  $V$  into itself and we suppose that  $\phi_0 \in V$ . We suppose also that  $A$  has a *non-trivial decomposition* such as

$$A = \sum_{j=1}^J A_j, \quad (2.1)$$

with  $J \geq 2$  (by *non-trivial* we mean that the operators  $A_j$  are individually simpler than  $A$ ).

A question which arises naturally is clearly:

*Can we take advantage of decomposition (2.1) for the solution of (IVP)?*

It has been known for a long time that the answer to the above question is definitely *yes*.

Many schemes have been designed to take advantage of the decomposition (2.1) when solving (IVP); two of them will be briefly discussed in the following paragraphs, namely the *Lie's scheme* and the *Strang's scheme*.

## 2.2. Time-discretization of (IVP) by the Lie's scheme

Let  $\tau (> 0)$  be a *time-discretization step* (we suppose  $\tau$  uniform, for simplicity); we denote  $n\tau$  by  $t^n$ . With  $\phi^n$  denoting an approximation of  $\phi(t^n)$ , the *Lie's scheme* reads as follows (for its derivation see, e.g., [28, Chapter 6]):

$$\phi^0 = \phi_0; \tag{2.2}$$

then, for  $n \geq 0$ , assuming that  $\phi^n$  is known, compute  $\phi^{n+1}$  via

$$\begin{cases} \frac{d\phi_j}{dt} + A_j(\phi_j) = 0 & \text{on } (t^n, t^{n+1}), \\ \phi_j(t^n) = \phi^{n+(j-1)/J}; & \phi^{n+j/J} = \phi_j(t^{n+1}), \end{cases} \tag{2.3}$$

for  $j = 1, \dots, J$ .

If (IVP) is taking place in a finite dimensional space and if the operators  $A_j$  are smooth enough, then  $\|\phi(t^n) - \phi^n\| = O(\tau)$ , function  $\phi$  being the solution of (IVP).

REMARK 2.1. The above scheme applies also for *multivalued* operators (such as the *sub-gradient* of proper *l.s.c.* convex functionals) but in such a case first order accuracy is not guaranteed anymore. A related application will be given in Section 2.4.

REMARK 2.2. The above scheme is easy to generalize to *non-autonomous problems* by observing that

$$\begin{cases} \frac{d\phi}{dt} + A(\phi, t) = 0, \\ \phi(0) = \phi_0 \end{cases} \iff \begin{cases} \frac{d\phi}{dt} + A(\phi, \theta) = 0, \\ \frac{d\theta}{dt} - 1 = 0, \\ \phi(0) = \phi_0, \quad \theta(0) = 0. \end{cases}$$

REMARK 2.3. Scheme (2.2), (2.3) is semi-constructive in the sense that we still have to solve the sub-initial value problems in (2.3) for each  $j$ . Suppose that we discretize these sub-problems using just *one step of the backward Euler scheme*. The resulting scheme reads as follows:

$$\phi^0 = \phi_0; \tag{2.4}$$

then, for  $n \geq 0$ , assuming that  $\phi^{n+(j-1)/J}$  is known, compute  $\phi^{n+j/J}$  via

$$\frac{\phi^{n+j/J} - \phi^{n+(j-1)/J}}{\tau} + A_j(\phi^{n+j/J}) = 0, \quad (2.5)$$

for  $j = 1, \dots, J$ .

Scheme (2.4), (2.5) is known as the *Marchuk-Yanenko scheme* (see, e.g., refs. [43] and [28, Chapter 6] for more details).

### 2.3. Time-discretization of (IVP) by the Strang's scheme

In order to improve the accuracy of the Lie's scheme, G. Strang suggested a *symmetrized* variant of scheme (2.2), (2.3) (ref. [68]). When applied to non-autonomous problems, in the case where  $J = 2$ , we obtain (with  $t^{n+1/2} = (n + 1/2)\tau$ ):

$$\phi^0 = \phi_0; \quad (2.6)$$

then, for  $n \geq 0$ , assuming that  $\phi^n$  is known, compute  $\phi^{n+1}$  via

$$\begin{cases} \frac{d\phi_1}{dt} + A_1(\phi_1, t) = 0 & \text{on } (t^n, t^{n+1/2}), \\ \phi_1(t^n) = \phi^n; \quad \phi^{n+1/2} = \phi_1(t^{n+1/2}), \end{cases} \quad (2.7)$$

$$\begin{cases} \frac{d\phi_2}{dt} + A_2(\phi_2, t^{n+1/2}) = 0 & \text{on } (0, \tau), \\ \phi_2(0) = \phi^{n+1/2}; \quad \hat{\phi}^{n+1/2} = \phi_2(\tau), \end{cases} \quad (2.8)$$

$$\begin{cases} \frac{d\phi_1}{dt} + A_1(\phi_1, t) = 0 & \text{on } (t^{n+1/2}, t^{n+1}), \\ \phi_1(t^{n+1/2}) = \hat{\phi}^{n+1/2}; \quad \phi^{n+1} = \phi_1(t^{n+1}). \end{cases} \quad (2.9)$$

If (IVP) is taking place in a finite dimensional space and if operators  $A_1$  and  $A_2$  are smooth enough, then  $\|\phi(t^n) - \phi^n\| = O(\tau^2)$ , function  $\phi$  being the solution of (IVP).

REMARK 2.4. In order to preserve the second order accuracy of scheme (2.6)–(2.9) (assuming that such a property holds) we still have to discretize the initial value problems in (2.7), (2.8) and (2.9) by schemes which are themselves second order accurate (at least); examples of such schemes can be found in, e.g., [28, Chapter 6] (which contains also a discussion of the case  $J > 2$ ).

### 2.4. Application

It is not an exaggeration to say that applications of operator-splitting are everywhere; indeed, some well-known methods or algorithms are *disguised* operator-splitting schemes. Our favorite example in that direction is the following:

Suppose that  $\mathbf{A}$  is a real  $d \times d$  matrix, symmetric and positive definite. Ordering the eigenvalues of  $\mathbf{A}$  as follows:  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ , our goal is to compute  $\lambda_1$ . We have (with obvious notation)

$$\lambda_1 = \min_{\mathbf{v} \in S} \mathbf{v}^t \mathbf{A} \mathbf{v}, \quad \text{with } S = \{\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^d, \|\mathbf{v}\| = 1\}, \quad (2.10)$$

the norm in (2.10) being the canonical Euclidean one. The minimization problem in (2.10) is *equivalent* to

$$\min_{\mathbf{v} \in \mathbb{R}^d} \left\{ \frac{1}{2} \mathbf{v}^t \mathbf{A} \mathbf{v} + I_S(\mathbf{v}) \right\}, \quad (2.11)$$

where, in (2.11), the functional  $I_S : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined as follows

$$I_S(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} \in S, \\ +\infty & \text{otherwise,} \end{cases}$$

implying that  $I_S$  is the *indicator functional* of the sphere  $S$ . Suppose that  $\mathbf{u}$  is a solution of problem (2.11); we have then

$$\mathbf{A} \mathbf{u} + \partial I_S(\mathbf{u}) \ni \mathbf{0}, \quad (2.12)$$

$\partial I_S(\mathbf{u})$  being in (2.12) a (kind of) generalized gradient of functional  $I_S$  at  $\mathbf{u}$  (indeed,  $\partial I_S(\cdot)$  is a multivalued operator). Next, we associate with the (necessary) optimality system (2.12) the following initial value problem (*flow* in the *Dynamical System* terminology):

$$\begin{cases} \frac{d\mathbf{u}}{dt} + \mathbf{A} \mathbf{u} + \partial I_S(\mathbf{u}) \ni \mathbf{0} & \text{on } (0, +\infty), \\ \mathbf{u}(0) = \mathbf{u}_0. \end{cases} \quad (2.13)$$

If we apply the Marchuk–Yanenko scheme (2.4), (2.5) to the solution of problem (2.13) we obtain

$$\mathbf{u}^0 = \mathbf{u}_0, \quad (2.14)$$

and for  $n \geq 0$ ,  $\mathbf{u}^n$  being known,

$$\frac{\mathbf{u}^{n+1/2} - \mathbf{u}^n}{\tau} + \mathbf{A} \mathbf{u}^{n+1/2} = \mathbf{0}, \quad (2.15)$$

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^{n+1/2}}{\tau} + \partial I_S(\mathbf{u}^{n+1}) \ni \mathbf{0}. \quad (2.16)$$

Relation (2.15) implies

$$\mathbf{u}^{n+1/2} = (\mathbf{I} + \tau \mathbf{A})^{-1} \mathbf{u}^n. \quad (2.17)$$

On the other hand, relation (2.16) can be interpreted as a necessary optimality condition for the following minimization problem

$$\min_{\mathbf{v} \in S} \left\{ \frac{1}{2} \|\mathbf{v}\|^2 - \mathbf{v}^t \mathbf{u}^{n+1/2} \right\}. \quad (2.18)$$

Since  $\|\mathbf{v}\| = 1$  over  $S$ , the solution of problem (2.18) is given by

$$\mathbf{u}^{n+1} = \frac{\mathbf{u}^{n+1/2}}{\|\mathbf{u}^{n+1/2}\|}. \quad (2.19)$$

Algorithm (2.14)–(2.16) reduces then to (2.14), (2.17), (2.19) which is nothing but the *inverse power method with shift*, a well-known algorithm from *Numerical Linear Algebra*. Clearly, numerical analysts have not been waiting for operator-splitting to compute matrix eigenvalues and eigenvectors; on the other hand, operator-splitting has provided efficient algorithms for the solution of complicated problems from Differential Geometry, Mechanics, Physics, Physico-Chemistry, etc., including some nonlinear eigenvalue problems (as shown in, e.g., [25]). Some of these applications will be discussed in Sections 3, 4 and 5.

### 3. Operator-splitting methods for the direct numerical simulation of particulate flow

#### 3.1. Generalities: Problem formulation

It is the (very likely biased) opinion of these authors that the *direct numerical simulation of particulate flow* has been one of the success stories of operator-splitting methods. Albeit this “story” has been told in several publications (see, e.g., [28, Chapters 8 and 9] and [33], and the references therein), owing to its importance we decided to return to it again. For simplicity, we consider only the *one-particle* case (see the two above references for the *multi-particle case*).

Let  $\Omega$  be a bounded, connected and open region of  $\mathbb{R}^d$  ( $d = 2$  or  $3$  in applications); the boundary of  $\Omega$  is denoted by  $\Gamma$ . We suppose that  $\Omega$  contains:

- (i) A Newtonian incompressible viscous fluid of density  $\rho_f$  and viscosity  $\mu_f$ ;  $\rho_f$  and  $\mu_f$  are both positive constants.
- (ii) A rigid body  $B$  of boundary  $\partial B$ , mass  $M$ , center of mass  $G$ , and inertia  $\mathbf{I}$  at the center of mass (see Fig. 3.1, for additional details).

The fluid occupies the region  $\Omega \setminus \bar{B}$  and we suppose that *distance*  $(\partial B(0), \Gamma) > 0$ . From now on,  $\mathbf{x} = \{x_i\}_{i=1}^d$  will denote the generic point of  $\mathbb{R}^d$ ,  $d\mathbf{x} = dx_1 \cdots dx_d$ , while  $\phi(t)$  will denote the function  $\mathbf{x} \rightarrow \phi(\mathbf{x}, t)$ . Assuming that the only external force is *gravity*, the *fluid flow-rigid body motion* coupling is modeled by

$$\rho_f \left( \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right) - \mu_f \Delta \mathbf{u} + \nabla p = \rho_f \mathbf{g} \quad \text{in } \Omega \setminus \bar{B}(t), \quad \forall t \in (0, T), \quad (3.1)$$

$$\nabla \cdot \mathbf{u}(t) = 0 \quad \text{in } \Omega \setminus \bar{B}(t), \quad \forall t \in (0, T), \quad (3.2)$$

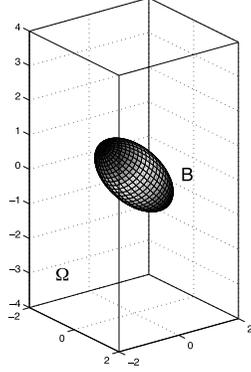


Fig. 3.1. Visualization of the flow region and of the rigid body.

$$\mathbf{u}(t) = \mathbf{u}_\Gamma(t) \quad \text{on } \Gamma, \quad \forall t \in (0, T), \quad \text{with } \int_\Gamma \mathbf{u}_\Gamma(t) \cdot \mathbf{n} \, d\Gamma = 0, \quad (3.3)$$

$$\mathbf{u}(0) = \mathbf{u}_0 \quad \text{in } \Omega \setminus \bar{B}(0) \quad \text{with } \nabla \cdot \mathbf{u}_0 = 0, \quad (3.4)$$

and

$$\frac{dG}{dt} = \mathbf{V}, \quad (3.5)$$

$$M \frac{d\mathbf{V}}{dt} = M\mathbf{g} + \mathbf{R}_H, \quad (3.6)$$

$$\frac{d(\mathbf{I}\boldsymbol{\omega})}{dt} = \mathbf{T}_H, \quad (3.7)$$

$$G(0) = G_0, \quad \mathbf{V}(0) = \mathbf{V}_0, \quad \boldsymbol{\omega}(0) = \boldsymbol{\omega}_0, \quad B(0) = B_0. \quad (3.8)$$

In relations (3.1)–(3.8):

- Vector  $\mathbf{u} = \{u_i\}_{i=1}^d$  is the *fluid (flow) velocity* and  $p$  is the *pressure*.
- $\mathbf{u}_0$  and  $\mathbf{u}_\Gamma$  are given functions.
- $\mathbf{V}$  is the velocity of the center of mass of body  $B$ , while  $\boldsymbol{\omega}$  is the angular velocity.
- $\mathbf{R}_H$  and  $\mathbf{T}_H$  denote, respectively, the *resultant* and the *torque* of the *hydrodynamical forces*, namely the forces that the fluid exerts on  $B$ ; we have, actually,

$$\mathbf{R}_H = \int_{\partial B} \boldsymbol{\sigma} \mathbf{n} \, d\gamma \quad \text{and} \quad \mathbf{T}_H = \int_{\partial B} \overrightarrow{G\mathbf{x}} \times \boldsymbol{\sigma} \mathbf{n} \, d\gamma. \quad (3.9)$$

In (3.9) the *stress-tensor*  $\boldsymbol{\sigma}$  is defined by  $\boldsymbol{\sigma} = 2\mu_f D(\mathbf{u}) - p\mathbf{I}_d$ , with  $D(\mathbf{v}) = \frac{1}{2}(\nabla \mathbf{v} + (\nabla \mathbf{v})^t)$ , while  $\mathbf{n}$  is a unit normal vector at  $\partial B$  and  $\mathbf{I}_d$  is the *identity tensor*.

Concerning the compatibility conditions on  $\partial B$  we have: (i) the forces exerted by the fluid on the solid body *balance* those exerted by the solid body on the fluid, and we shall assume that: (ii) on  $\partial B$  the *no-slip boundary condition* holds, namely

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{V}(t) + \boldsymbol{\omega}(t) \times \overrightarrow{G(t)\mathbf{x}}, \quad \forall \mathbf{x} \in \partial B(t). \quad (3.10)$$

REMARK 3.1. System (3.1)–(3.4) (resp., (3.5)–(3.8)) is of the *incompressible Navier–Stokes* (resp., *Euler–Newton*) type. Also, the above model can be generalized to multiple-particles situations and/or non-Newtonian incompressible viscous fluids.

The (local in time) *existence of weak solutions* for problems such as (3.1)–(3.8) has been proved in ref. [21], assuming that, at  $t = 0$ , the particles do not touch each other and do not touch  $\Gamma$  (see also [34], [64]). Concerning the numerical solution of (3.1)–(3.4) and (3.5)–(3.8) completed by the above interface conditions we can divide them, roughly, in two classes, namely: (i) The *Arbitrary Lagrange–Euler (ALE)* methods; these methods, which rely on *moving meshes*, are discussed in, e.g., refs. [37], [48], [35]. (ii) The non-boundary fitted *fictitious domain* methods; these methods rely on fixed meshes and are discussed in, e.g., [28, Chapter 8] and [33] (see also the references therein). These methods seem to enjoy a growing popularity, justifying thus the (brief) discussion hereafter.

REMARK 3.2. Even if theory suggests that collisions may never take place in finite time (if we assume that the flow is still modeled by the Navier–Stokes equations as long as the particles do not touch each others, or the boundary), near-collisions take place, and, after discretization, “real” collisions may occur. These phenomena can be avoided by introducing well-chosen short range repulsion potentials reminiscent of those encountered in *Molecular Dynamics* (see [28, Chapter 8] and [33] for details). For more information on the numerical treatment of particle collisions in flow, see, e.g., [66] (and the references therein).

### 3.2. A fictitious domain formulation

Considering the fluid-rigid body mixture as a unique medium we are going to derive a *fictitious domain based variational formulation*. The principle of this derivation is pretty simple; it relies on the following steps (see, e.g., [28] and [33] for details):

a. Start from the following *global weak* formulation (of the *virtual power* type):

$$\left\{ \begin{array}{l} \rho_f \int_{\Omega \setminus \bar{B}(t)} \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] \cdot \mathbf{v} \, d\mathbf{x} + 2\mu_f \int_{\Omega \setminus \bar{B}(t)} \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) \, d\mathbf{x} \\ - \int_{\Omega \setminus \bar{B}(t)} p \nabla \cdot \mathbf{v} \, d\mathbf{x} + M \frac{d\mathbf{V}}{dt} \cdot \mathbf{Y} + \frac{d(\mathbf{I}\boldsymbol{\omega})}{dt} \cdot \boldsymbol{\theta} \\ = \rho_f \int_{\Omega \setminus \bar{B}(t)} \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} + M \mathbf{g} \cdot \mathbf{Y}, \end{array} \right. \quad (3.11)$$

$$\left\{ \begin{array}{l} \forall \{\mathbf{v}, \mathbf{Y}, \boldsymbol{\theta}\} \in (H^1(\Omega \setminus \bar{B}(t)))^d \times \mathbb{R}^d \times \boldsymbol{\Theta} \text{ and verifying} \\ \mathbf{v} = 0 \text{ on } \Gamma, \quad \mathbf{v}(\mathbf{x}) = \mathbf{Y} + \boldsymbol{\theta} \times \overrightarrow{G(t)\mathbf{x}}, \quad \forall \mathbf{x} \in \partial B(t), \quad t \in (0, T), \\ \text{with } \boldsymbol{\Theta} = \mathbb{R}^3 \text{ if } d = 3, \quad \boldsymbol{\Theta} = \{(0, 0, \theta) \mid \theta \in \mathbb{R}\} \text{ if } d = 2, \end{array} \right.$$

$$\int_{\Omega \setminus \bar{B}(t)} q \nabla \cdot \mathbf{u}(t) \, d\mathbf{x} = 0, \quad \forall q \in L^2(\Omega \setminus \bar{B}(t)), \quad t \in (0, T), \quad (3.12)$$

$$\mathbf{u}(t) = \mathbf{u}_\Gamma(t) \text{ on } \Gamma, \quad t \in (0, T), \quad (3.13)$$

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{V}(t) + \boldsymbol{\omega}(t) \times \overrightarrow{G(t)\mathbf{x}}, \quad \forall \mathbf{x} \in \partial B(t), \quad t \in (0, T), \quad (3.14)$$

$$\frac{dG}{dt} = \mathbf{V}, \quad (3.15)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(x), \quad \forall \mathbf{x} \in \Omega \setminus \bar{B}(0), \quad (3.16)$$

$$G(0) = G_0, \quad \mathbf{V}(0) = \mathbf{V}_0, \quad \boldsymbol{\omega}(0) = \boldsymbol{\omega}_0, \quad B(0) = B_0. \quad (3.17)$$

- b. Fill  $B$  with the surrounding fluid.
- c. Impose a rigid body motion to the fluid inside  $B$ .
- d. Modify the global weak formulation (3.11)–(3.17) accordingly, taking advantage of the fact that if  $\mathbf{v}$  is a rigid body motion velocity field, then  $\nabla \cdot \mathbf{v} = 0$  and  $\mathbf{D}(\mathbf{v}) = \mathbf{0}$ .
- e. Use a *Lagrange multiplier* defined over  $B$  to force the rigid body motion inside  $B$ .

Assuming that  $B$  is made of a homogeneous material of density  $\rho_s$ , the above “program” leads to:

$$\left\{ \begin{aligned} & \rho_f \int_{\Omega} \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] \cdot \mathbf{v} \, d\mathbf{x} + 2\mu_f \int_{\Omega} \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ & + (1 - \rho_f/\rho_s) \left[ M \frac{d\mathbf{V}}{dt} \cdot \mathbf{Y} + \frac{d(\mathbf{I}\boldsymbol{\omega})}{dt} \cdot \boldsymbol{\theta} \right] + \left\langle \boldsymbol{\lambda}, \mathbf{v} - \mathbf{Y} - \boldsymbol{\theta} \times \overrightarrow{G(t)\mathbf{x}} \right\rangle_{B(t)} \\ & = \rho_f \int_{\Omega} \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} + (1 - \rho_f/\rho_s) M \mathbf{g} \cdot \mathbf{Y}, \\ & \forall \{\mathbf{v}, \mathbf{Y}, \boldsymbol{\theta}\} \in (H^1(\Omega))^d \times \mathbb{R}^d \times \boldsymbol{\Theta}, \quad t \in (0, T), \\ & \text{with } \boldsymbol{\Theta} = \mathbb{R}^3 \quad \text{if } d = 3, \quad \boldsymbol{\Theta} = \{(0, 0, \theta) \mid \theta \in \mathbb{R}\} \quad \text{if } d = 2, \end{aligned} \right. \quad (3.18)$$

$$\int_{\Omega} q \nabla \cdot \mathbf{u}(t) \, d\mathbf{x} = 0, \quad \forall q \in L^2(\Omega), \quad t \in (0, T), \quad (3.19)$$

$$\mathbf{u}(t) = \mathbf{u}_\Gamma(t) \quad \text{on } \Gamma, \quad t \in (0, T), \quad (3.20)$$

$$\left\{ \begin{aligned} & \left\langle \boldsymbol{\mu}, \mathbf{u}(\mathbf{x}, t) - \mathbf{V}(t) - \boldsymbol{\omega}(t) \times \overrightarrow{G(t)\mathbf{x}} \right\rangle_{B(t)} = 0, \\ & \forall \boldsymbol{\mu} \in \boldsymbol{\Lambda}(t) \quad (= (H^1(B(t)))^d), \quad t \in (0, T), \end{aligned} \right. \quad (3.21)$$

$$\frac{dG}{dt} = \mathbf{V}, \quad (3.22)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(x), \quad \forall \mathbf{x} \in \Omega \setminus \bar{B}_0, \quad (3.23)$$

$$\left\{ \begin{aligned} & G(0) = G_0, \quad \mathbf{V}(0) = \mathbf{V}_0, \quad \boldsymbol{\omega}(0) = \boldsymbol{\omega}_0, \quad B(0) = B_0, \\ & \mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega \setminus \bar{B}_0, \quad \mathbf{u}(x, 0) = \mathbf{V}_0 + \boldsymbol{\omega}_0 \times \overrightarrow{G_0\mathbf{x}}, \quad \forall \mathbf{x} \in \bar{B}_0. \end{aligned} \right. \quad (3.24)$$

From a theoretical point of view, a natural choice for  $\langle \cdot, \cdot \rangle_{B(t)}$  is provided by, e.g.,

$$\langle \boldsymbol{\mu}, \mathbf{v} \rangle_{B(t)} = \int_{B(t)} [\boldsymbol{\mu} \cdot \mathbf{v} + l^2 \mathbf{D}(\boldsymbol{\mu}) : \mathbf{D}(\mathbf{v})] \, d\mathbf{x}; \quad (3.25)$$

in (3.25),  $l$  is a characteristic length, the diameter of  $B$ , for example. From a practical point of view, when it come to space discretization, a simple and efficient strategy is the following one (cf. [28, Chapter 8] and [33]): “approximate”  $\Lambda(t)$  by

$$\Lambda_h(t) = \left\{ \boldsymbol{\mu} \left| \boldsymbol{\mu} = \sum_{j=1}^N \boldsymbol{\mu}_j \delta(\mathbf{x} - \mathbf{x}_j), \text{ with } \boldsymbol{\mu}_j \in \mathbb{R}^d, \forall j = 1, \dots, N \right. \right\}, \quad (3.26)$$

and the above pairing by

$$\langle \boldsymbol{\mu}, \mathbf{v} \rangle_{(B(t), h)} = \sum_{j=1}^N \boldsymbol{\mu}_j \cdot \mathbf{v}(\mathbf{x}_j). \quad (3.27)$$

In (3.26), (3.27),  $\mathbf{x} \rightarrow \delta(\mathbf{x} - \mathbf{x}_j)$  is the Dirac measure at  $\mathbf{x}_j$ , and the set  $\{\mathbf{x}_j\}_{j=1}^N$  is the union of two subsets, namely: (i) The set of the points of the velocity grid contained in  $B(t)$  and whose distance at  $\partial B(t)$  is  $\geq ch$ ,  $h$  being a space discretization step and  $c$  a constant  $\approx 1$ . (ii) A set of control points located on  $\partial B(t)$  and forming a mesh whose step size is of the order of  $h$ . It is clear that, using the approach above, one forces the rigid body motion inside the particle by *collocation*.

### 3.3. Solving problem (3.18)–(3.24) by operator-splitting

We do not consider *collisions*; after (formal) elimination of  $p$  and  $\boldsymbol{\lambda}$ , problem (3.18)–(3.24) reduces to a *dynamical system* of the following form

$$\frac{d\mathbf{X}}{dt} + \sum_{j=1}^J A_j(\mathbf{X}, t) = \mathbf{0} \quad \text{on } (0, T), \quad \mathbf{X}(0) = \mathbf{X}_0, \quad (3.28)$$

where  $\mathbf{X} = \{\mathbf{u}, \mathbf{V}, \boldsymbol{\omega}, G\}$  (or  $\{\mathbf{u}, \mathbf{V}, \mathbf{I}\boldsymbol{\omega}, G\}$ ). A typical situation will be the one where, with  $J = 4$ , operator  $A_1$  will be associated to incompressibility,  $A_2$  to advection,  $A_3$  to diffusion,  $A_4$  to fictitious domain and body motion; other decompositions are possible as shown in, e.g., [28, Chapter 8] and [33]. The Lie’s scheme (2.2), (2.3) applies “beautifully” to the solution of the formulation (3.28) of problem (3.18)–(3.24). The resulting method is quite modular implying that different space and time approximations can be used to treat the various steps; the only constraint is that two successive steps have to communicate (by projection in general).

## 3.4. Numerical experiments

### 3.4.1. Generalities

The methods described (quite briefly) in the above paragraphs have been validated by numerous experiments (see, e.g., [28, Chapters 8 and 9], [33], and [55]). In this article, we will consider four test problems, the first, second and third ones involving one, two and three particles, respectively, while the fourth test problem concerns a channel flow with 300 particles. “Our” fictitious domain/operator-splitting approach has made the solution of these problems (almost) routine, but no

latter than the mid-nineties, solving such problems was considered a *Grand Challenge* (actually, our first systematic attack of such problems was part of a *National Science Foundation* supported *Grand Challenge* project). All the flow computations have been done using the *Bercovier–Pironneau finite element approximation*; namely (see [28, Chapters 5, 8 and 9] for details) we used a globally continuous piecewise affine approximation of the velocity (resp., the pressure) associated to a triangulation (in 2-D) or tetrahedral partition (in 3-D)  $\mathcal{T}_h$  (resp.,  $\mathcal{T}_{2h}$ ) of  $\Omega$ ,  $h$  being a space discretization step. The pressure mesh is thus *twice coarser* than the velocity one. All our calculations have been done using *uniform partitions*  $\mathcal{T}_h$  and  $\mathcal{T}_{2h}$ .

### 3.4.2. First test problem: Lifting of a ball by a pressure driven flow

Let us denote by  $\Omega$  a truncated circular cylinder of length  $L$  and diameter  $D$ ; we denote by  $\Gamma$  the boundary of  $\Omega$  and we suppose that: (a) The axis of the cylinder is parallel to the horizontal axis  $Ox_2$ . (b) The cylinder has been truncated by the vertical planes  $x_2 = 0$  and  $x_2 = L$ ; we denote by  $\Gamma_1$  and  $\Gamma_2$  the two vertical disks limiting  $\Omega$ . (c) The cylinder is filled with an incompressible Newtonian viscous fluid of density  $\rho_f$  and viscosity  $\mu_f$ ; it contains also a solid ball  $B$  of diameter  $d$  and density  $\rho_s$ . (d) The only external force is gravity. (e) At  $t = 0$ , the fluid and the solid are at rest, the ball lying on the bottom of the cylinder. (f) There exists a pressure drop  $\Delta P$  between  $\Gamma_1$  and  $\Gamma_2$ . (g) The following boundary conditions prevail on  $\Gamma$ :  $\mathbf{u} = \mathbf{0}$  on  $\Gamma \setminus (\Gamma_1 \cup \Gamma_2)$ , and  $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}(\mathbf{x} + Le_2, t)$ ,  $\forall \mathbf{x} \in \Gamma_1$ ,  $e_2$  being the unit vector of axis  $Ox_2$  (*space-periodicity* in the  $Ox_2$  direction). We suppose also that when the ball exits the cylinder on the right (i.e., through  $\Gamma_2$ ), an identical ball enters the cylinder through  $\Gamma_1$  with identical translation and angular velocities, the centers of mass of the two balls being located on a same line parallel to the axis  $Ox_2$ .

If  $\Delta P$  is large enough, the *lift* acting on  $B$  will be sufficient for the ball to take off and possibly reach an equilibrium height. This is indeed the phenomenon we intend to simulate. In order to do so we are going to apply the *fictional domain methodology* discussed in Sections 3.2 and 3.3. Actually the above methodology is used twice: (1) To treat the fluid-rigid ball coupling as done in Section 3.2. (2) To solve all the partial differential equations in a “box”  $\tilde{\Omega}$  containing  $\Omega$ ; we impose  $\mathbf{u} = 0$  in  $\tilde{\Omega} \setminus \Omega$  via a distributed Lagrange multiplier method very close to the one discussed in [28, Section 41].

For the particular problem that we are considering:  $L = 2$ ,  $D = 1$ ,  $d = 0.2$ ,  $\rho_f = 1$ ,  $\mu_f = 1$ ,  $\rho_s = 1.001$ ,  $\tilde{\Omega} = (0, 1) \times (0, 2) \times (0, 1)$  is approximated by  $\tilde{\Omega}_h = (0, 1 + 4h) \times (0, 2) \times (0, 1 + 4h)$  ( $h$  being a uniform space-discretization step), the cylinder axis is on the line  $x_1 = 1/2 + 2h$ ,  $x_3 = 1/2 + 2h$ , and  $\Delta P = 160$ , so that, without the ball, the maximal steady flow speed would have been 10. At  $t = 0$ , the ball mass center is located right under the cylinder axis at a distance of  $0.375D$  (the height of the ball center is 0.125 from the bottom of the cylinder).

The mesh size used to compute the velocity field (resp., the pressure) is  $h_v = h = 1/96$  (resp.,  $h_p = 2h = 1/48$ ), while we took  $1/1000$  for the time-discretization step; with these values the velocity mesh contains approximately  $2 \times 10^6$  grid-points.

The lifting phenomenon has been simulated for  $t \in [0, 40]$ , leading to the following results:

- The translational speed of the ball averaged during the last five time units is 6.126.
- The particle Reynolds number averaged during the last five time units is 1.2252.
- The distance of the ball center to the cylinder axis, also averaged during the last five time units, is 0.30129.

The followings have been visualized on Figs. 3.2 and 3.3, respectively:

- The position of the ball at time  $t = 0$ .
- The position of the ball at  $t = 40$ ; at that time it has, essentially, reached its equilibrium height (the height of the ball center is about 0.2 from the bottom of the cylinder). Here we follow the mass center of the ball to plot the position related to the cylinder, so it seems the ball does not move.
- The variation versus time of the ratio of the distance of the ball center to the cylinder axis and the radius of the cylinder; the *equilibrium height* has been reached around  $t = 27.5$ .

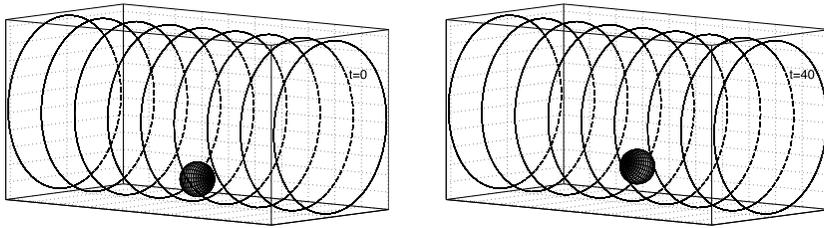


Fig. 3.2. Position of the ball at  $t = 0$  (left) and  $t = 40$  (right).

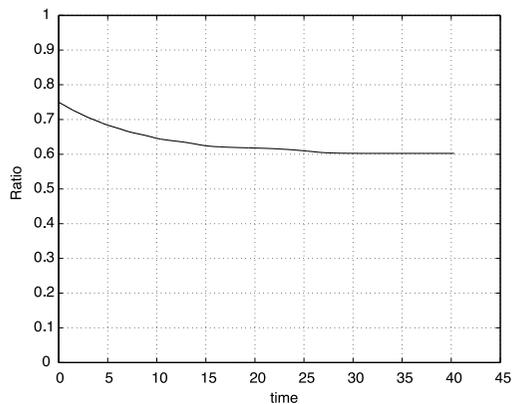


Fig. 3.3. Time variation of the ratio of the distance of the ball center to the cylinder axis and the radius of the cylinder.

### 3.4.3. Settling of two balls and three balls in a vertical narrow tube

The main goal of this paragraph is to discuss the interaction of two balls and then three balls settling in a narrow tube of rectangular cross section and containing an incompressible Newtonian viscous fluid. Theoretically, the tube should be infinitely long, but for practicality we first consider the settling of the balls in a truncated cylinder of length 4 whose cross-section is the rectangle  $(0, 1) \times (0, 1/4)$  for the interaction of two balls; this cylinder is moving with the balls in such a way that the center of the lower ball is in the horizontal symmetry plane (a possible alternative would have been to specify periodicity in the vertical direction). At time  $t = 0$ , we suppose that the truncated cylinder coincides with the “box”  $\Omega = (0, 1) \times (0, 1/4) \times (0, 4)$  and that the centers of the balls are on the vertical axis of the cylinder at the points  $x_1 = 1/2$ ,  $x_2 = 1/8$ ,  $x_3 = 0.9$  and  $1.25$ . The parameters chosen in this case are  $d = 0.2$ ,  $\rho_s = 1.1$ ,  $\rho_f = 1$ ,  $\mu_f = 0.01$ . The mesh size used to compute the velocity field (resp., the pressure) is  $h_v = h = 1/96$  (resp.,  $h_p = 2h = 1/48$ ), while we took  $1/1000$  for the time-discretization step; the initial velocity of the flow is also zero while the two balls are released from rest. The velocity on the cylinder wall is always zero. Over 10 time units, we have observed drafting, kissing, and tumbling twice and then the two balls settle in a stable configuration as shown in Figs. 3.4 to 3.6. The averaged speed of the two balls for the last two time units is 2.4362 and the averaged particle Reynolds number for the last two time units is 48.724.

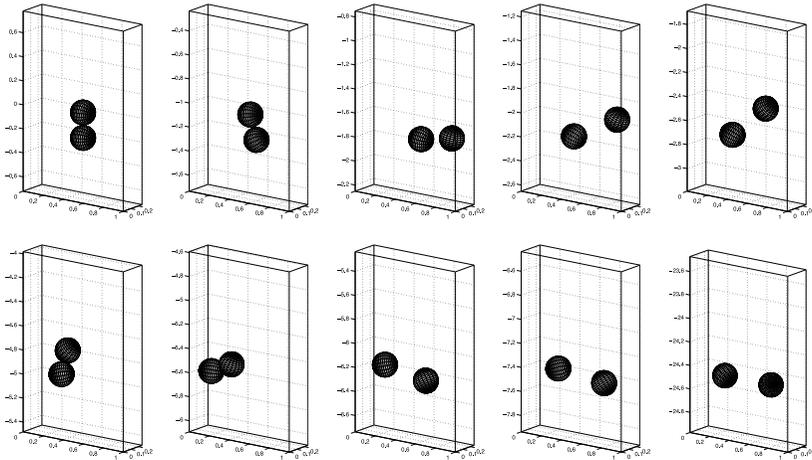


Fig. 3.4. Related position of the two balls at  $t = 0.4, 0.65, 0.85, 1, 1.2, 2, 2.2, 2.5, 3$  and  $10$  (from left to right and from top to bottom).

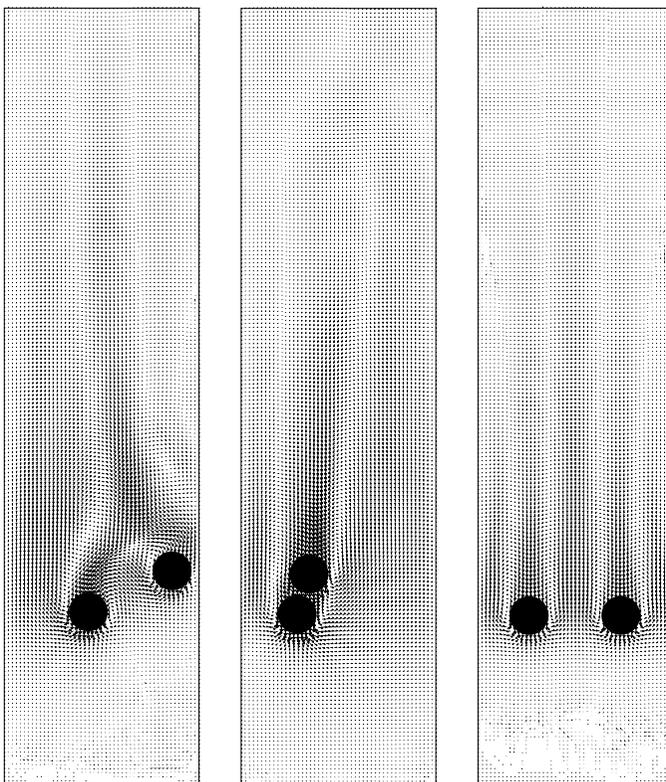


Fig. 3.5. Projection of the velocity field to the plane passing through the mass center of the left ball and parallel to the plane  $x_2 = 1/8$  at  $t = 1, 2$  and  $10$  (from left to right).

For the case of three ball interaction, the truncated cylinder coincides with the “box”  $\Omega = (0, 1.5) \times (0, 1/4) \times (0, 6)$  and the centers of the balls are on the vertical axis of the cylinder at the points  $x_1 = 3/4$ ,  $x_2 = 1/8$ ,  $x_3 = 1, 1.3$  and  $1.6$ . The other parameters chosen in this case are the same as in the above two ball case. The three balls are released from rest. Over 15 time units, we have observed drafting, kissing, and tumbling several times and then a stable configuration is taking place as shown in Figs. 3.6, 3.7 and 3.8. The averaged speed of the three balls for the last two time units is 2.4653 and the averaged particle Reynolds number for the last two time units is 49.304. Further study and analysis of the interaction between balls will be reported in a forthcoming paper.

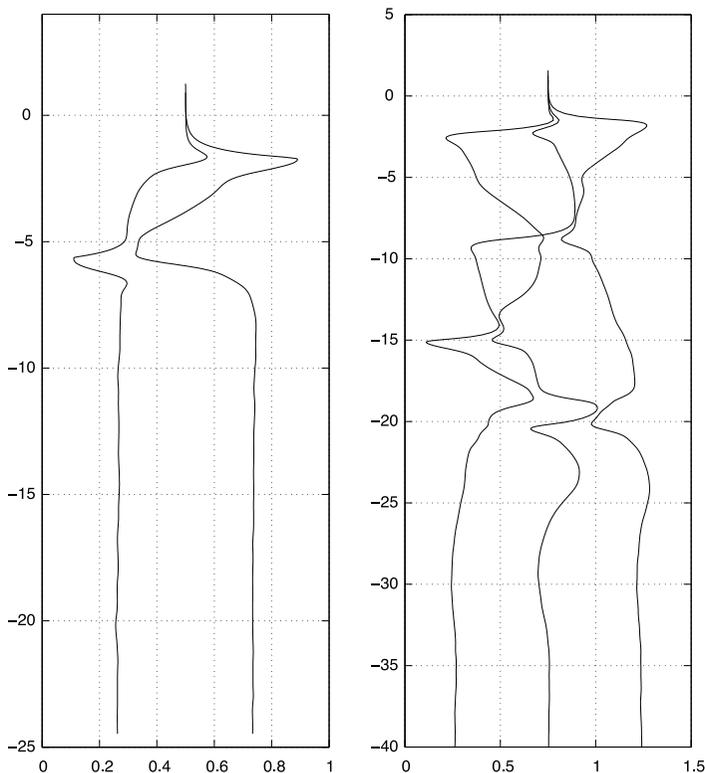


Fig. 3.6. The trajectories of the mass centers of the two balls (left) and three balls (right) projected on the  $x_1x_3$ -plane.

#### 3.4.4. Motion of 300 neutrally buoyant disks in a two-dimensional horizontal channel

The fourth test problem involving 300 particles and a *solid volume/fluid volume* ratio of the order of 0.3786, collisions (or near-collisions) have to be accounted for in the simulations; to do so we have used the methods discussed in, e.g., refs. [28, Chapter 8] and [33]. Another peculiarity of this fourth test problem is that  $\rho_s = \rho_f$  for all particles (a *neutrally buoyant* situation). Indeed, neutrally buoyant models are more delicate to handle than those in the general case, since  $1 - \rho_f/\rho_s = 0$  in (3.18); however, this difficulty can be overcome, as shown in ref. [55]. For this test problem, we have: (a)  $\Omega = (0, 42) \times (0, 12)$ . (b)  $\Omega$  contains the mixture of a Newtonian incompressible viscous fluid of density 1 and viscosity 1 with 300 rigid solid disks of density 1 and radius 0.45. (c) At time  $t = 0$ , fluid and particles are at rest, the particle centers being located at the points of a regular lattice. (d) The mixture is put into motion by a uniform pressure drop of 20/18 per unit length (without the particles the steady flow would have been of the Poiseuille's type with 20 as maximal flow speed). (e) The boundary conditions are given by  $\mathbf{u}(x_1, x_2, t) = \mathbf{0}$  if  $0 \leq x_1 \leq 42$ ,  $x_2 = 0$  and 12, and  $0 \leq t \leq 400$  (no-slip boundary condition on the horizontal parts

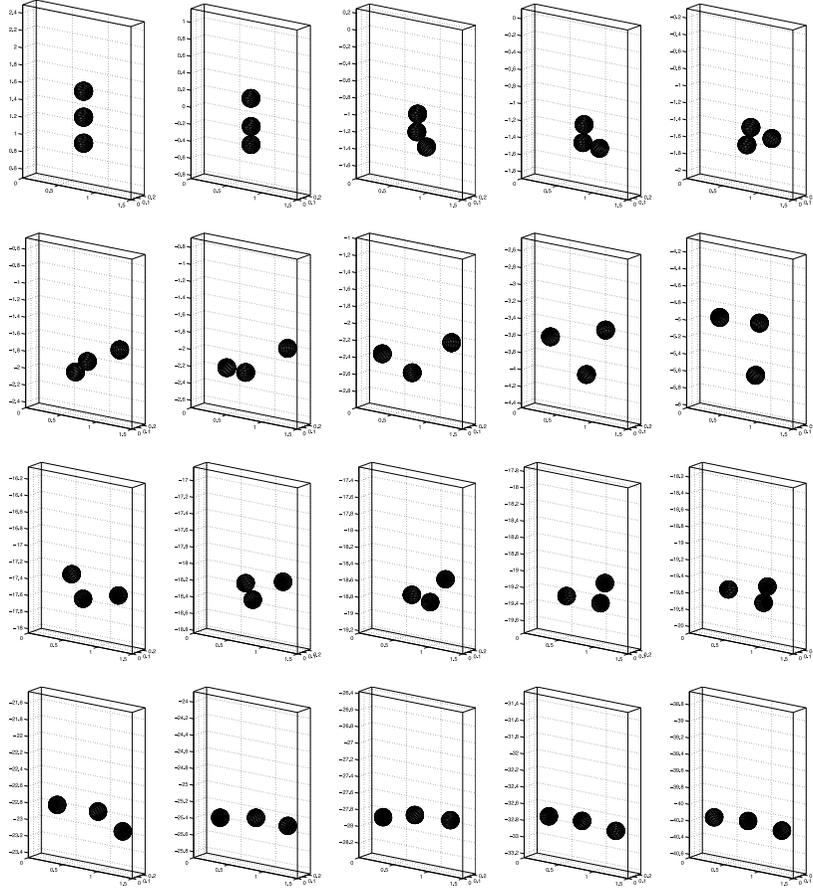


Fig. 3.7. Related position of the three balls at  $t = 0, 0.4, 0.6, 0.65, 0.7, 0.8, 0.9, 1, 1.5, 2, 6, 6.25, 6.4, 6.6, 6.7, 8, 9, 10, 12$  and  $15$  (from left to right and from top to bottom).

of the boundary), and then  $\mathbf{u}(0, x_2, t) = \mathbf{u}(42, x_2, t)$ ,  $0 < x_2 < 12$ ,  $0 \leq t \leq 400$  (space-periodic boundary conditions in the  $Ox_1$  direction). (f)  $h_v = h = 1/10$ ,  $h_p = 2h = 1/5$ , the time-discretization step being  $1/1000$ .

The particle distribution at  $t = 100, 107.8, 114, 200$  and  $400$  has been visualized on Fig. 3.9. These figures show that, initially, we have the sliding motion of horizontal particle layers, then after some time a chaotic flow-motion takes place in very few time-units, the higher particle concentration being along the axis of the channel (actually, a careful inspection of the results shows that the transition to chaos takes place just after  $t = 107.8$ ). The maximal speed at  $t = 400$  is  $7.9$ , implying that the corresponding particle Reynolds number is very close to  $7.1$ .

For more details, and further comments and results on pressure driven neutrally buoyant particulate flow in two-dimensional channels (including simulations with much larger numbers of particles, the largest one being  $1,200$ ), see [28, Chapter 9]

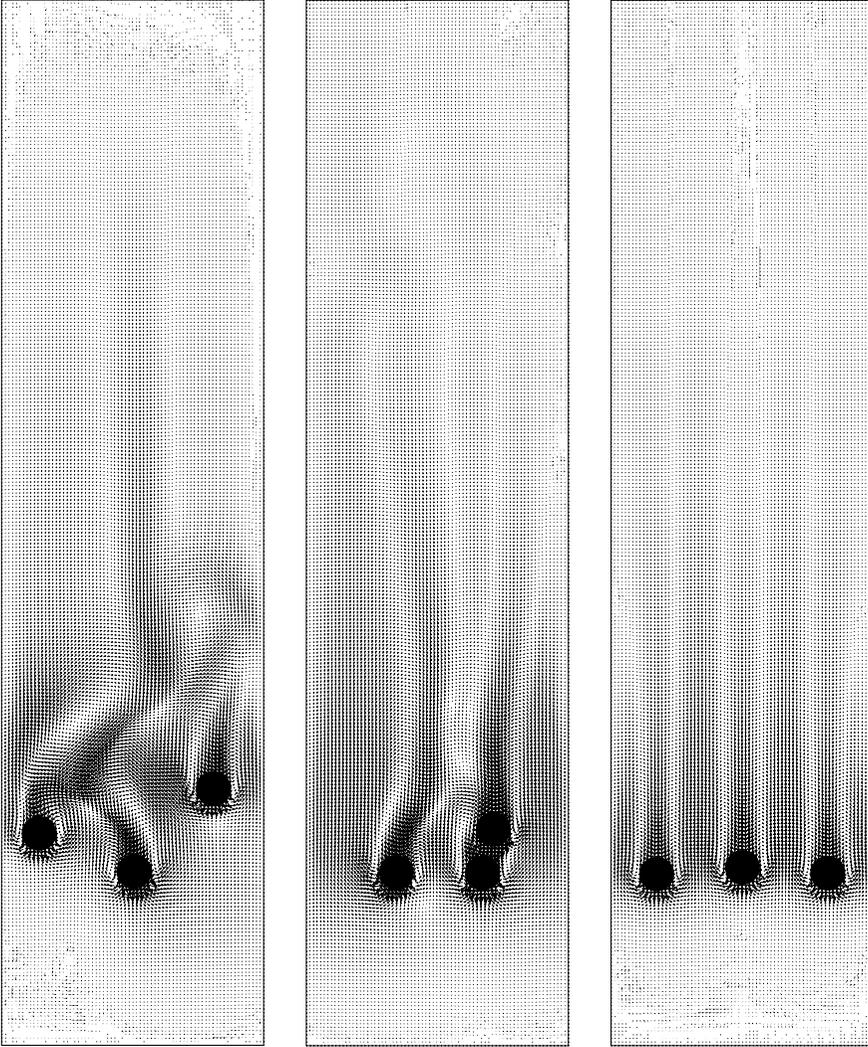


Fig. 3.8. Projection of the velocity field to a plane passing through the mass center of one of the three balls and parallel to the plane  $x_2 = 1/8$  at  $t = 1.1$ ,  $6.6$  and  $15$  (from left to right).

and [55]; among these results, we chose to report the following two: (i) Fig. 3.10 shows the averaged solid fraction as a function of the coordinate  $x_2$ , the averaging set being  $\{(x_1, t) \mid 0 \leq x_1 < 42, 380 \leq t \leq 400\}$ . We observe that particles tend to aggregate along the axis of the channel, since the solid fraction is close to  $0.58$  there, while the global solid fraction is  $0.3786$  (vertical line in the figure). (ii) Fig. 3.11 shows the  $x_1$ -averaged horizontal component of the mixture velocity at  $t = 400$ , as a function of  $x_2$ . The dashed line corresponds to the horizontal velocity distribution of the flow of the same fluid, with no particle in the channel, for the same

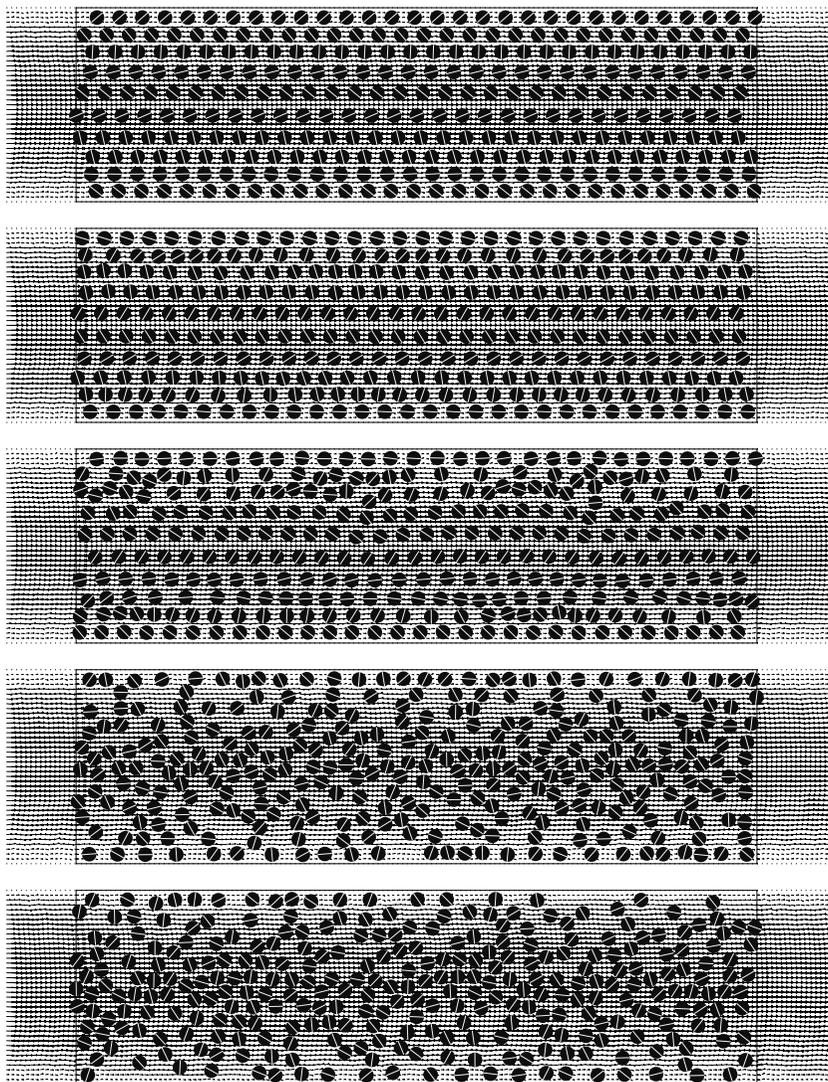


Fig. 3.9. The 300 particle positions  $t = 100, 107.8, 114, 200$  and  $400$  (from top to bottom).

pressure drop; the corresponding velocity profile is (of course) of the Poiseuille's type and shows that the mixture behaves like a viscous fluid whose viscosity is (approximately) 2.5 larger than  $\mu_f$ . Actually, a closer inspection (see ref. [55] for details) shows that the mixture behaves like a non-Newtonian incompressible viscous fluid of the power law type, for an exponent  $s = 1.7093$  ( $s = 2$  corresponding to a Newtonian fluid and  $s = 1$  to a visco-plastic material). The two above figures show also that, as well known, some order is often found in chaos.

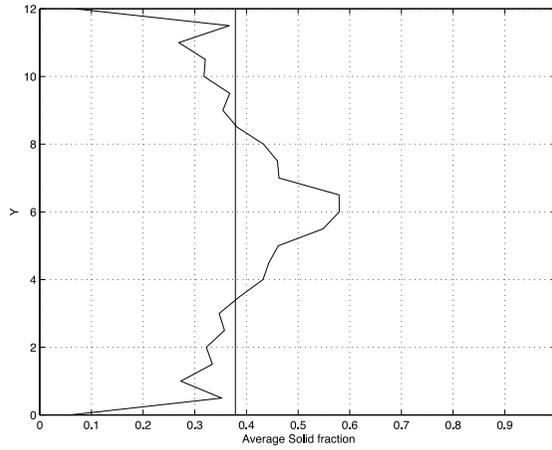
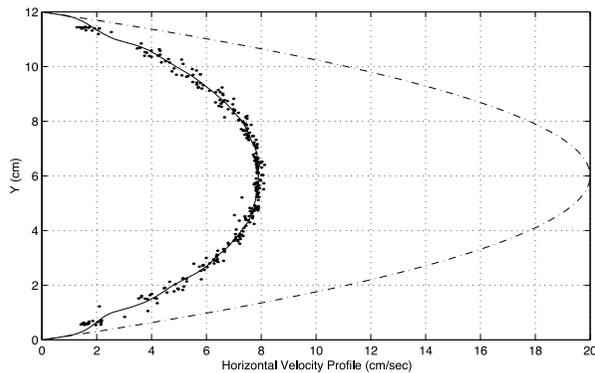


Fig. 3.10. Averaged solid fraction distribution.

Fig. 3.11. Horizontal velocity distribution at  $t = 400$ .

#### 4. Operator-splitting and finite element methods for a time-dependent incompressible viscous free-surface flow

##### 4.1. Generalities

Many flow problems in physics and applied sciences lead to the incompressible Navier–Stokes equations with a *free capillary surface*. These flows have many important industrial applications. Among these applications let us mention the so called “coating flows” (refs. [41], [58], [38], and [63]), flow in semi-conductor melts (refs. [3], [49], and [65]), and flow of melted aluminum ([7] and references therein). On a capillary free surface the normal stress of the flow field is balanced by *surface tension*, the basic assumption being that surface tension is proportional to the *mean curvature* of the free surface, implying in turn that surface tension plays a significant role in determining the shape of the free surface. Not surprisingly, this

type of flow with free surface has been studied extensively: experimentally ([59]), analytically ([1], [4], [53], [61], and [67]), asymptotically ([40]), and, of course, numerically ([63], [7], [2], [15], [39], [45], [46], [6], [72], [69], [70], [36], and [47], among many other references). In particular, a thorough discussion of a flow down an inclined perturbed plane can be found in ref. [60]; it includes a detailed comparison between experimental measurements, lubrication approximations, and the results of numerical simulation. Some numerical methods widely used today rely on *space approximation by Galerkin/finite element methods* themselves associated to a *variational formulation* of the flow problem; among the reasons for such a choice, let us mention the following one: formulated this way, the free surface problem takes an elegant and concise form with some of the boundary conditions integrated within the variational formulation in a straightforward manner; these methods are particularly well-suited to coping with the highly deformed free boundaries and complex shape flow regions occurring in these problems. Indeed, compared to flow problems in fixed regions, the discretization of the free-surface flow problems discussed here encounters additional difficulties due to the *free boundary curvature* related terms. Thus, many authors prefer to treat these terms in an implicit way, and, in general, a kind of iterative procedure is applied in order to locate the free surface which shifts from trial to trial and rarely coincides with any convenient co-ordinate surface. In addition, *isoparametric* quadratic finite element approximations are commonly used, combined with quadratic or cubic parametrization of the free-boundary, all this leading to further complications. The analysis of the stability and/or convergence properties of some computational methods can be found in e.g., refs. [2], and [50], [62].

The numerical solution of free-boundary problems is still an active field of research and novel efficient computational methods are needed in order to solve this kind of problems. Following A. Chorin (ref. [12]), most “modern” Navier–Stokes solvers are based on *operator-splitting* (see, e.g., [43], [28], [71], [44]). In this article (which follows very closely [30]), we consider the numerical solution of a time-dependent two-dimensional viscous free-surface flow with applications to lubrication and coating technology; the formulation of the above problem is given in Section 4.2. In Section 4.3, we take advantage of *operator-splitting* methods in order to avoid iterative procedures to locate the free-boundary; these iterative localization procedures are costly, in general. In Section 4.4, we introduce an *isoparametric* variant of the *Bercovier–Pironneau finite element method* ([28], [5], [26], [57]) to achieve space discretization, leading to the full (space-time) discretization described in Section 4.5. The resulting discrete equations have the same structure as in usual (i.e., with fixed boundaries) problems modeled by the Navier–Stokes equations, with, however, an additional step “taking care” of the relocation of the free-boundary. The results of two series of numerical experiments are presented in Section 4.6.

## 4.2. Formulation of the free-boundary problem

Let  $\Omega(t)$  be a 2-D space region, *periodic* in the  $Ox_1$  (i.e., *horizontal*) direction. We suppose that  $\Omega(t)$  contains an *incompressible Newtonian viscous fluid* of *density*  $\rho$  and *viscosity*  $\mu$  and that its top boundary is a *free capillary surface*  $\gamma(t)$ , as shown in Fig. 4.1. We shall assume that the free boundary  $\gamma(t)$  has the following representation:

$$\gamma(t) = \{\mathbf{x} \mid \mathbf{x} = \{x_1, x_2\}, 0 \leq x_1 \leq L, x_2 = \eta(x_1, t)\}, \quad (4.1)$$

the function  $\eta$  in (4.1) being  $x_1$ -periodic of period  $L$ .

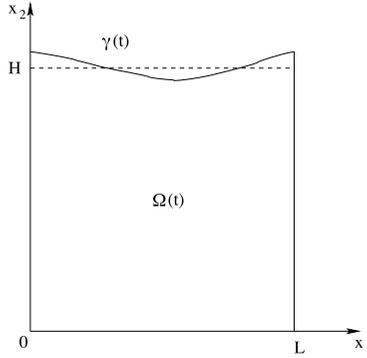


Fig. 4.1. Flow region with a free boundary.

Assuming that  $t \in [0, T]$ , and that the only external force acting on the fluid is *gravity*, to be denoted by  $\mathbf{g}$ , the fluid flow is modeled by the Navier–Stokes equations completed by appropriate boundary and initial conditions, namely:

$$\rho \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla \mathbf{u}) \right] = \rho \mathbf{g} + \nabla \cdot \boldsymbol{\sigma} \quad \text{in } \Omega(t), \quad 0 < t \leq T, \quad (4.2)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega(t), \quad 0 < t \leq T, \quad (4.3)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \quad \mathbf{x} \in \Omega(0), \quad (4.4)$$

$$\mathbf{u}(x_1, 0, t) = \mathbf{0}, \quad \forall \{x_1, t\} \in (0, L) \times (0, T), \quad (4.5)$$

$$\mathbf{u}(0, x_2, t) = \mathbf{u}(L, x_2, t), \quad \text{if } t \in (0, T) \text{ and } x_2 \in (0, \eta(0, t)) = (0, \eta(L, t)), \quad (4.6)$$

$$\boldsymbol{\sigma} \mathbf{n} = s \mathcal{H}(\eta) \mathbf{n} \quad \text{on } \gamma(t), \quad (\text{with } s \in \mathbb{R}), \quad (4.7)$$

$$\frac{\partial \eta}{\partial t} n_2 = \mathbf{u} \cdot \mathbf{n} \quad \text{on } \gamma(t), \quad \text{with } \eta(0, t) = \eta(L, t), \quad t \in (0, T). \quad (4.8)$$

In system (4.2)–(4.8):  $\mathbf{u}$  ( $= \{u_1, u_2\}$ ) denotes the *fluid velocity*;  $\boldsymbol{\sigma} = \mu[\nabla \mathbf{u} + (\nabla \mathbf{u})^t] - p \mathbf{I}$  is the *stress-tensor*,  $p$  being the *pressure*;  $\mathbf{n}$  ( $= \{n_1, n_2\}$ ) is the *unit outward normal vector* at the boundary of  $\Omega(t)$ ; the initial velocity  $\mathbf{u}_0(\mathbf{x})$  is periodic of period  $L$  in the  $Ox_1$  direction, and verifies  $\nabla \cdot \mathbf{u}_0 = 0$  in  $\Omega(0)$ ;  $s$  is the *surface tension coefficient*;  $\mathcal{H}(\eta)$  is the *curvature of the free boundary*  $\gamma(t)$  and is given by

$$\mathcal{H}(\eta) = \frac{\partial^2 \eta}{\partial x_1^2} \Big/ \left( 1 + \left| \frac{\partial \eta}{\partial x_1} \right|^2 \right)^{3/2}. \quad (4.9)$$

Assuming that the “volume” of fluid in the region  $\Omega(t)$  is  $HL$ , we have

$$\int_0^L \eta(x_1, t) dx_1 = HL, \quad \forall t \in (0, T), \quad (4.10)$$

which is consistent with

$$\frac{d}{dt} \int_0^L \eta dx_1 = \int_0^L \frac{\partial \eta}{\partial t} dx_1 = \int_{\gamma(t)} \frac{\partial \eta}{\partial t} n_2 d\gamma(t) = \int_{\gamma(t)} \mathbf{u} \cdot \mathbf{n} d\gamma(t) = 0,$$

(from (4.3), (4.5), (4.8) and the *divergence theorem*). On the free boundary  $\gamma(t)$  we have

$$\mathbf{n}(\mathbf{x}) = \{-\partial\eta/\partial x_1, 1\}/(1 + |\partial\eta/\partial x_1|^2)^{1/2}, \quad (4.11)$$

implying that equation (4.8) is equivalent to the (transport) equation

$$\frac{\partial \eta}{\partial t} + u_1 \frac{\partial \eta}{\partial x_1} = u_2 \quad \text{in } (0, L) \times (0, T), \quad \text{with } \eta(0, t) = \eta(L, t), \quad (4.12)$$

and

$$\eta(x_1, 0) = \eta_0(x_1), \quad x_1 \in (0, L). \quad (4.13)$$

In (4.12),  $u_i$  denotes,  $\forall i = 1, 2$ , the function  $\{x_1, t\} \rightarrow u_i(x_1, \eta(x_1, t), t)$ .

In order to obtain a *variational formulation* of problem (4.2)–(4.8) we introduce the following functional spaces:

$$\mathbf{V}_0(t) = \{\mathbf{v} \mid \mathbf{v} \in (H^1(\Omega(t)))^2, \mathbf{v}(x_1, 0) = \mathbf{0}, \forall x_1 \in (0, L), \\ \mathbf{v} \text{ is periodic of period } L \text{ in the } Ox_1 \text{ direction}\}, \quad (4.14)$$

$$L_0^2(\Omega(t)) = \left\{ q \mid q \in L^2(\Omega(t)), \int_{\Omega(t)} q dx = 0 \right\}, \quad (4.15)$$

$$S = \{\phi \mid \phi \in H^1(0, L), \phi \text{ is periodic of period } L\}. \quad (4.16)$$

Multiplying equations (4.2), (4.3) and (4.8) by the test functions  $\mathbf{v} \in \mathbf{V}_0(t)$ ,  $q \in L^2(\Omega(t))$  and  $\phi \in S$ , respectively, we obtain, after integration by parts over  $\Omega(t)$  and  $(0, L)$ , the following variational formulation of (4.2)–(4.8):

For a.e.  $t \in (0, T)$ , find  $\mathbf{u}(t) \in \mathbf{V}_0(t)$ ,  $p(t) \in L_0^2(\Omega(t))$  and  $\eta(t) \in S$ , such that

$$\begin{aligned} & \rho \int_{\Omega(t)} \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla \mathbf{u}) \right] \cdot \mathbf{v} dx \\ & + 2\mu \int_{\Omega(t)} \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) dx - \int_{\Omega(t)} p \nabla \cdot \mathbf{v} dx \\ & = \rho \int_{\Omega(t)} \mathbf{g} \cdot \mathbf{v} dx + s \int_{\gamma(t)} \mathcal{H}(\eta) \mathbf{n} \cdot \mathbf{v} d\gamma(t), \quad \forall \mathbf{v} \in \mathbf{V}_0(t), \end{aligned} \quad (4.17)$$

$$\int_{\Omega(t)} \nabla \cdot \mathbf{u}(t) q \, d\mathbf{x} = 0, \quad \forall q \in L^2(\Omega(t)), \quad (4.18)$$

$$\int_0^L \frac{\partial \eta}{\partial t} \phi \, dx_1 + \int_0^L \left[ u_1|_{\gamma(t)} \frac{\partial \eta}{\partial x_1} - u_2|_{\gamma(t)} \right] \phi \, dx_1 = 0, \quad \forall \phi \in S, \quad (4.19)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega(0), \quad \text{with } \nabla \cdot \mathbf{u}_0 = 0, \quad (4.20)$$

$$\eta(x_1, 0) = \eta_0(x_1), \quad \forall x_1 \in (0, L), \quad \text{with } \int_0^L \eta_0(x_1) \, dx_1 = HL; \quad (4.21)$$

in (4.17),  $\mathbf{D}(\mathbf{v})$  denotes the *deformation tensor*  $\frac{1}{2}[\nabla \mathbf{v} + (\nabla \mathbf{v})^t]$ .

Problems (4.2)–(4.8), (4.17)–(4.21) is defined over  $\Omega(t)$ , a domain which changes with time and is not known a priori. We are going therefore to reformulate (4.2)–(4.8), (4.17)–(4.21) into a fixed reference domain.

Let  $(A_t)_{t \in (0, T)}$  be a family of mappings which for each  $t \in (0, T)$  map the current domain  $\Omega(t)$  onto the reference domain  $\hat{\Omega} = (0, L) \times (0, H)$  defined by

$$\begin{aligned} A_t: \Omega(t) &\rightarrow \hat{\Omega}, \\ \mathbf{x} = \{x_1, x_2\} &\rightarrow \boldsymbol{\xi} = \{\xi_1, \xi_2\} = A_t(\mathbf{x}) = \begin{cases} \xi_1 = x_1, \\ \xi_2 = \frac{H}{\eta(x_1, t)} x_2, \end{cases} \end{aligned} \quad (4.22)$$

as sketched in Fig. 4.2. We observe that the free surface  $\gamma(t)$  is mapped onto

$$\hat{\gamma} = \{\boldsymbol{\xi} \in \mathbb{R}^2 \mid \xi_1 \in (0, L), \xi_2 = H\}. \quad (4.23)$$

The Jacobian matrix  $\mathbf{J}$  of the above transformation, and its determinant  $J$  are given by

$$\mathbf{J} = \mathbf{J}(\eta) = \begin{bmatrix} 1 & 0 \\ -\frac{\xi_2}{\eta(\xi_1, t)} \frac{\partial \eta}{\partial \xi_1}(\xi_1, t) & \frac{H}{\eta(\xi_1, t)} \end{bmatrix}, \quad J = J(\eta) = \frac{H}{\eta(\xi_1, t)}. \quad (4.24)$$

It is clear that the above transformation is well defined as long as  $\eta(\xi_1, t) > 0$  or, in other words, as long as the free surface does not touch the bottom of the flow region. Let  $f = f(\mathbf{x}, t)$  be a function defined on  $\Omega(t)$  and  $\hat{f} = \hat{f}(\boldsymbol{\xi}, t) = f(A_t^{-1}(\boldsymbol{\xi}), t)$  be the corresponding function defined on  $\hat{\Omega}$ . Due to the change of variables, we have

$$\int_{\Omega(t)} f(\mathbf{x}, t) \, d\mathbf{x} = \int_{\hat{\Omega}} \hat{f}(\boldsymbol{\xi}, t) J \, d\boldsymbol{\xi}. \quad (4.25)$$

Moreover, it follows from the chain rule that

$$\frac{\partial f}{\partial t} = \frac{\partial \hat{f}}{\partial t} - \mathbf{w} \cdot \hat{\nabla} \hat{f}, \quad (4.26)$$

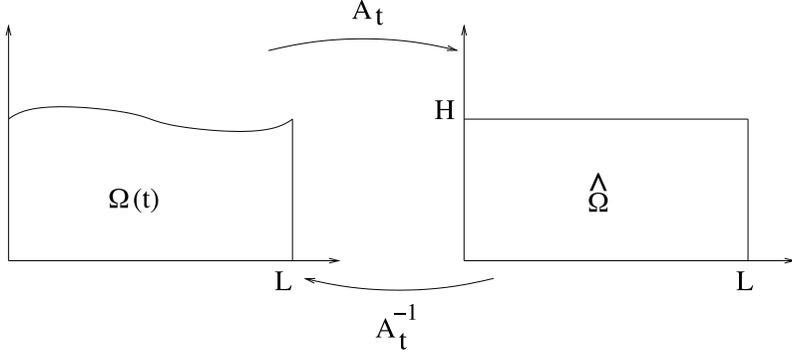


Fig. 4.2.  $A_t$  maps the current domain  $\Omega(t)$  onto the reference domain  $\hat{\Omega}$ .

where  $\hat{\nabla} = \mathbf{J}^{-t} \nabla$  and where  $\mathbf{w}$  is the *domain velocity*;  $\mathbf{w}$  is given by

$$\begin{aligned} \mathbf{w}(\boldsymbol{\xi}, t) &= -\frac{\partial \boldsymbol{\xi}}{\partial t} = \frac{\xi_2}{\eta} \frac{\partial \eta}{\partial t}(\xi_1, t) \mathbf{e}_2 \\ &= \frac{\xi_2}{\eta} \left[ u_2(\xi_1, \eta(\xi_1, t), t) - u_1(\xi_1, \eta(\xi_1, t), t) \frac{\partial \eta}{\partial x_1}(\xi_1, t) \right] \mathbf{e}_2. \end{aligned} \quad (4.27)$$

The last term is a consequence of the kinematic condition (4.12) while  $\mathbf{e}_2 = \{0, 1\}$ . Using (4.25) and (4.26), system (4.17)–(4.21) becomes (with obvious notation).

For a.e.  $t \in (0, T)$ , find  $\hat{\mathbf{u}}(t) \in \mathbf{V}_0(\hat{\Omega})$ ,  $\hat{p}(t) \in L_0^2(\hat{\Omega})$  and  $\eta(t) \in S$ , such that

$$\begin{aligned} &\rho \int_{\hat{\Omega}} \frac{\partial \hat{\mathbf{u}}}{\partial t} \cdot \hat{\mathbf{v}} J(\eta) \, d\xi \\ &- \rho \int_{\hat{\Omega}} (\mathbf{w} \cdot \hat{\nabla} \hat{\mathbf{u}}) \cdot \hat{\mathbf{v}} J(\eta) \, d\xi + \rho \int_{\hat{\Omega}} (\hat{\mathbf{u}} \cdot \hat{\nabla} \hat{\mathbf{u}}) \cdot \hat{\mathbf{v}} J(\eta) \, d\xi \\ &+ 2\mu \int_{\hat{\Omega}} \hat{\mathbf{D}}(\hat{\mathbf{u}}) : \hat{\mathbf{D}}(\hat{\mathbf{v}}) J(\eta) \, d\xi - \int_{\hat{\Omega}} \hat{p} \hat{\nabla} \cdot \hat{\mathbf{v}} J(\eta) \, d\xi \\ &= \rho \int_{\hat{\Omega}} \hat{\mathbf{g}} \cdot \hat{\mathbf{v}} J(\eta) \, d\xi \\ &\quad + s \int_0^L \mathcal{H}(\eta) \left[ \hat{v}_2(\xi_1, H) - \hat{v}_1(\xi_1, H) \frac{\partial \eta}{\partial \xi_1} \right] d\xi_1, \quad \forall \hat{\mathbf{v}} \in \mathbf{V}_0(\hat{\Omega}), \end{aligned} \quad (4.28)$$

$$\int_{\hat{\Omega}} \hat{q} \hat{\nabla} \cdot \hat{\mathbf{u}} J(\eta) \, d\xi = 0, \quad \forall \hat{q} \in L^2(\hat{\Omega}), \quad (4.29)$$

$$\int_0^L \frac{\partial \eta}{\partial t} \phi \, d\xi_1 + \int_0^L \left[ \hat{u}_1(\xi_1, H, t) \frac{\partial \eta}{\partial \xi_1} - \hat{u}_2(\xi_1, H, t) \right] \phi \, d\xi_1 = 0, \quad \forall \phi \in S, \quad (4.30)$$

$$\hat{\mathbf{u}}(0) = \hat{\mathbf{u}}_0, \quad \text{with } \hat{\nabla} \cdot \hat{\mathbf{u}}_0 = 0, \quad \text{and}$$

$$\eta(\xi_1, 0) = \eta_0(\xi_1), \quad \text{with } \int_0^L \eta_0(\xi_1) \, d\xi_1 = HL. \quad (4.31)$$

The notation  $\hat{\mathbf{D}}(\hat{\mathbf{v}}) = \frac{1}{2}[\hat{\nabla} \hat{\mathbf{v}} + (\hat{\nabla} \hat{\mathbf{v}})^t]$  has been used in (4.28).

### 4.3. Time-discretization by operator-splitting

We encountered in Section 3 particulate flows, which are also flows with moving boundaries. To simulate these flows, we employed a fractional-step approach, one of the steps being dedicated to the motion of the particles, i.e., to updating the location of the moving boundaries. A similar approach can be used here, the particle motion dedicated step being replaced by a step dedicated to the relocation of the free-surface, i.e., from a practical point of view, to the approximate solution of the system (4.12), (4.13), assuming that  $\mathbf{u}$  is already known (approximately, of course), from a previous fractional step. Indeed, the numerical solution of problem (4.28)–(4.31) contains three main difficulties: (a) the incompressibility condition and the related unknown pressure, (b) an advection term associated to the velocity of the fluid, (c) an advection term associated to the velocity of the physical domain. The difficulty associated with the unknown location of the free boundary has been removed through the transformation defined by (4.22). As in Section 3, a time-discretization by operator-splitting will allow to decouple the difficulties of the problem and to treat each of them with an appropriate method. As mentioned in Section 4.1, this approach makes possible to update the location of the free boundary as a sub-step, avoiding thus costly sub-iterations between the location of the free surface and the solution of the Navier–Stokes equations. Among the many possible operator-splitting schemes, the one we advocate is (partially) a Marchuk–Yanenko implementation of the Lie’s scheme (see Section 2.2) and is first order accurate at best. Its low order of accuracy is compensated by its simplicity, making it (relatively) easy to implement, and by its robustness. The scheme reads as follows (with  $\Delta t$  ( $> 0$ ) a time discretization step,  $t^n = n \Delta t$ ,  $\hat{\mathbf{u}}^n = \hat{\mathbf{u}}(t^n)$ ,  $\hat{p}^n = \hat{p}(t^n)$ , and  $\eta^n = \eta(t^n)$ ):

$$\hat{\mathbf{u}}^0 = \hat{\mathbf{u}}_0, \quad \text{and} \quad \eta^0 = \eta(0); \quad (4.32)$$

then, for  $n \geq 0$ ,  $\hat{\mathbf{u}}^n$  and  $\eta^n$  being known, solve the following sub-problems:

- (i) Find  $\hat{\mathbf{u}}^{n+1/5} \in \mathbf{V}_0(\hat{\Omega})$  and  $\hat{p}^{n+1/5} \in L_0^2(\hat{\Omega})$  such that

$$\left\{ \begin{array}{l} \rho \int_{\hat{\Omega}} \frac{\hat{\mathbf{u}}^{n+1/5} - \hat{\mathbf{u}}^n}{\Delta t} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi \\ + \mu \int_{\hat{\Omega}} \hat{\mathbf{D}}(\hat{\mathbf{u}}^{n+1/5}) : \hat{\mathbf{D}}(\hat{\mathbf{v}}) J(\eta^n) d\xi - \int_{\hat{\Omega}} \hat{p}^{n+1/5} \hat{\nabla} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi \\ = \rho \int_{\hat{\Omega}} \hat{\mathbf{g}} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi \\ + s \int_0^L \mathcal{H}(\eta^n) \left[ \hat{v}_2(\xi_1, H) - \hat{v}_1(\xi_1, H) \frac{\partial \eta^n}{\partial \xi_1} \right] d\xi_1, \quad \forall \hat{\mathbf{v}} \in \mathbf{V}_0(\hat{\Omega}), \\ \int_{\hat{\Omega}} \hat{q} \hat{\nabla} \cdot \hat{\mathbf{u}}^{n+1/5} J(\eta^n) d\xi = 0, \quad \forall \hat{q} \in L^2(\hat{\Omega}). \end{array} \right. \quad (4.33)$$

- (ii) Find  $\hat{\mathbf{u}}^{n+2/5} \in \mathbf{V}(\hat{\Omega})$  via the solution of the following pure advection problem on  $\hat{\Omega} \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_{\hat{\Omega}} \frac{\partial \hat{\mathbf{u}}}{\partial t} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi + \int_{\hat{\Omega}} (\hat{\mathbf{u}}^{n+1/5} \cdot \hat{\nabla}) \hat{\mathbf{u}} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi = 0, & \forall \hat{\mathbf{v}} \in \hat{\mathbf{v}}_-^{n+1/5}, \\ \hat{\mathbf{u}}(t^n) = \hat{\mathbf{u}}^{n+1/5}, \\ \hat{\mathbf{u}}(t) = \hat{\mathbf{u}}^{n+1/5} \quad \text{on } \hat{\Gamma}_-^{n+1/5} \times (t^n, t^{n+1}), \end{cases} \quad (4.34)$$

and then setting  $\hat{\mathbf{u}}^{n+2/5} = \hat{\mathbf{u}}(t^{n+1})$ . The functional spaces and set introduced just above are defined as follows:

$$\begin{aligned} \mathbf{V}(\hat{\Omega}) &= \{ \mathbf{v} \mid \mathbf{v} \in (H^1(\hat{\Omega}))^2, \\ &\quad \mathbf{v} \text{ is periodic of period } L \text{ in the } O\xi_1 \text{ direction} \}, \\ \hat{\mathbf{v}}_-^{n+1/5} &= \{ \mathbf{v} \mid \mathbf{v} \in \mathbf{V}(\hat{\Omega}), \mathbf{v} = \mathbf{0} \text{ on } \hat{\Gamma}_-^{n+1/5} \}, \\ \hat{\Gamma}_-^{n+1/5} &= \{ \xi \mid \xi \in \hat{\gamma}, \hat{\mathbf{u}}^{n+1/5}(\xi) \cdot \mathbf{e}_2 < 0 \}. \end{aligned}$$

- (iii) Find  $\hat{\mathbf{u}}^{n+3/5} \in \mathbf{V}_0(\hat{\Omega})$  and  $\hat{p}^{n+3/5} \in L_0^2(\hat{\Omega})$  such that

$$\begin{cases} \rho \int_{\hat{\Omega}} \frac{\hat{\mathbf{u}}^{n+3/5} - \hat{\mathbf{u}}^{n+2/5}}{\Delta t} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi \\ + \mu \int_{\hat{\Omega}} \hat{\mathbf{D}}(\hat{\mathbf{u}}^{n+3/5}) : \hat{\mathbf{D}}(\hat{\mathbf{v}}) J(\eta^n) d\xi \\ - \int_{\hat{\Omega}} \hat{p}^{n+3/5} \hat{\nabla} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi = 0, & \forall \hat{\mathbf{v}} \in \mathbf{V}_0(\hat{\Omega}), \\ \int_{\hat{\Omega}} \hat{q} \hat{\nabla} \cdot \hat{\mathbf{u}}^{n+3/5} J(\eta^n) d\xi = 0, & \forall \hat{q} \in L^2(\hat{\Omega}). \end{cases} \quad (4.35)$$

- (iv) Find  $\hat{\mathbf{u}}^{n+4/5} \in \mathbf{V}(\hat{\Omega})$  via the solution of the following pure advection problem on  $\hat{\Omega} \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_{\hat{\Omega}} \frac{\partial \hat{\mathbf{u}}}{\partial t} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi - \int_{\hat{\Omega}} (\hat{\mathbf{w}}^{n+3/5} \cdot \hat{\nabla}) \hat{\mathbf{u}} \cdot \hat{\mathbf{v}} J(\eta^n) d\xi = 0, & \forall \hat{\mathbf{v}} \in \hat{\mathbf{W}}_+^{n+3/5}, \\ \hat{\mathbf{u}}(t^n) = \hat{\mathbf{u}}^{n+3/5}, \\ \hat{\mathbf{u}}(t) = \hat{\mathbf{u}}^{n+3/5} \quad \text{on } \hat{\Gamma}_+^{n+3/5} \times (t^n, t^{n+1}). \end{cases} \quad (4.36)$$

Set then  $\hat{\mathbf{u}}^{n+4/5} = \hat{\mathbf{u}}(t^{n+1})$  and  $\hat{p}^{n+4/5} = \hat{p}^{n+1/5} + \hat{p}^{n+3/5}$ . Here

$$\mathbf{w}^{n+3/5}(\xi) = \frac{\xi_2}{\eta} \left[ \hat{u}_2^{n+3/5}(\xi_1, H) - \hat{u}_1^{n+3/5}(\xi_1, H) \frac{\partial \eta^n}{\partial \xi_1}(\xi_1) \right] \mathbf{e}_2, \quad (4.37)$$

and

$$\begin{aligned} \hat{\mathbf{W}}_+^{n+3/5} &= \{ \mathbf{v} \mid \mathbf{v} \in \mathbf{V}(\hat{\Omega}), \mathbf{v} = \mathbf{0} \text{ on } \hat{\Gamma}_+^{n+3/5} \}, \\ \hat{\Gamma}_+^{n+3/5} &= \{ \xi \mid \xi \in \hat{\gamma}, \mathbf{w}^{n+3/5}(\xi) \cdot \mathbf{e}_2 > 0 \}. \end{aligned}$$

- (v) Find the new position of the free surface  $\eta^{n+1} \in S$  by solving the following problem in  $(0, L) \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_0^L \frac{\partial \eta}{\partial t} \phi d\xi_1 + \int_0^L \left[ \hat{u}_1^{n+4/5}(\xi_1, H) \frac{\partial \eta}{\partial \xi_1} - \hat{u}_2^{n+4/5}(\xi_1, H) \right] \phi d\xi_1 = 0, \\ \forall \phi \in S, \\ \eta(t^n) = \eta^n; \end{cases} \quad (4.38)$$

set then  $\eta^{n+1} = \eta(t^{n+1})$ .

- (vi) The new physical domain  $\Omega^{n+1}$  is obtained from the reference domain through the mapping  $A_{t^{n+1}}^{-1}$  defined as

$$\begin{aligned} A_{t^{n+1}}^{-1} : \hat{\Omega} \subset \mathbb{R}^2 &\rightarrow \Omega^{n+1} \subset \mathbb{R}^2, \\ \boldsymbol{\xi} = \{\xi_1, \xi_2\} \rightarrow \mathbf{x} = \{x_1, x_2\} &= \begin{cases} x_1 = \xi_1, \\ x_2 = \frac{\eta^{n+1}(\xi_1)}{H} \xi_2. \end{cases} \end{aligned} \quad (4.39)$$

The velocity  $\mathbf{u}^{n+1}$  and the pressure  $p^{n+1}$  defined on the new domain  $\Omega^{n+1}$  are obtained as follows:

$$\begin{aligned} \mathbf{u}^{n+1}(x_1, x_2) &= \hat{\mathbf{u}}^{n+4/5}(x_1, Hx_2/\eta^{n+1}(x_1)), \\ p^{n+1}(x_1, x_2) &= \hat{p}^{n+4/5}(x_1, Hx_2/\eta^{n+1}(x_1)), \end{aligned} \quad (4.40)$$

with  $\{x_1, x_2\} \in \Omega^{n+1}$ .

We observe that  $\hat{\nabla} = \mathbf{J}^{-t}(\eta^n) \nabla$  in (4.33)–(4.36) and that, indeed, it is much easier to solve the sub-problems (4.33)–(4.35) using their equivalent formulation in the physical domain, whose fixed “value” is  $\Omega^n$  on the time interval  $(t^n, t^{n+1})$  (this is an important consequence from the way the free surface problem was split). Moreover, problems (4.33) and (4.35) can be significantly simplified by noticing that

$$\int_{\Omega(t)} \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) d\mathbf{x} = \frac{1}{2} \int_{\Omega(t)} \nabla \mathbf{u} : \nabla \mathbf{v} d\mathbf{x} + \frac{1}{2} \int_{\Omega(t)} (\nabla \mathbf{u})^t : \nabla \mathbf{v} d\mathbf{x}. \quad (4.41)$$

Taking (4.41) and the above observation into account leads us to replace (4.33)–(4.35) by, respectively:

(i) Find  $\mathbf{u}^{n+1/5} \in \mathbf{V}_0(\Omega^n)$  and  $p^{n+1/5} \in L_0^2(\Omega^n)$  such that

$$\left\{ \begin{array}{l} \rho \int_{\Omega^n} \frac{\mathbf{u}^{n+1/5} - \mathbf{u}^n}{\Delta t} \cdot \mathbf{v} \, d\mathbf{x} \\ + \frac{\mu}{2} \int_{\Omega^n} \nabla \mathbf{u}^{n+1/5} : \nabla \mathbf{v} \, d\mathbf{x} - \int_{\Omega^n} p^{n+1/5} \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ = \rho \int_{\Omega^n} \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} + s \int_{\gamma^n} \mathcal{H}(\eta^n) \mathbf{n} \cdot \mathbf{v} \, d\gamma^n \\ - \frac{\mu}{2} \int_{\Omega^n} (\nabla \mathbf{u}^n)^t : \nabla \mathbf{v} \, d\mathbf{x}, \quad \forall \mathbf{v} \in \mathbf{V}_0(\Omega^n), \\ \int_{\Omega^n} q \nabla \cdot \mathbf{u}^{n+1/5} \, d\mathbf{x} = 0, \quad \forall q \in L^2(\Omega^n). \end{array} \right. \quad (4.42)$$

(ii) Find  $\mathbf{u}^{n+2/5} \in \mathbf{V}_0(\Omega^n)$  via the solution of the following pure advection problem on  $\Omega^n \times (t^n, t^{n+1})$ :

$$\left\{ \begin{array}{l} \int_{\Omega^n} \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega^n} (\mathbf{u}^{n+1/5} \cdot \nabla) \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} = 0, \quad \forall \mathbf{v} \in \mathbf{V}_-^{n+1/5}, \\ \mathbf{u}(t^n) = \mathbf{u}^{n+1/5}, \\ \mathbf{u}(t) = \mathbf{u}^{n+1/5} \quad \text{on } \Gamma_-^{n+1/5} \times (t^n, t^{n+1}), \end{array} \right. \quad (4.43)$$

and then setting  $\mathbf{u}^{n+2/5} = \mathbf{u}(t^{n+1})$ . The functional spaces and set used just above are defined by

$$\begin{aligned} \mathbf{V}(\Omega^n) &= \{ \mathbf{v} \mid \mathbf{v} \in (H^1(\Omega^n))^2, \\ &\quad \mathbf{v} \text{ is periodic of period } L \text{ in the } O x_1 \text{ direction} \}, \\ \mathbf{V}_-^{n+1/5} &= \{ \mathbf{v} \mid \mathbf{v} \in \mathbf{V}(\Omega^n), \mathbf{v} = \mathbf{0} \text{ on } \Gamma_-^{n+1/5} \}, \\ \Gamma_-^{n+1/5} &= \{ \mathbf{x} \mid \mathbf{x} \in \gamma^n, \mathbf{u}^{n+1/5}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0 \}, \end{aligned}$$

respectively.

(iii) Find  $\mathbf{u}^{n+3/5} \in \mathbf{V}_0(\Omega^n)$  and  $p^{n+3/5} \in L_0^2(\Omega^n)$  such that

$$\left\{ \begin{array}{l} \rho \int_{\Omega^n} \frac{\mathbf{u}^{n+3/5} - \mathbf{u}^{n+2/5}}{\Delta t} \cdot \mathbf{v} \, d\mathbf{x} \\ + \frac{\mu}{2} \int_{\Omega^n} \nabla \mathbf{u}^{n+3/5} : \nabla \mathbf{v} \, d\mathbf{x} - \int_{\Omega^n} p^{n+3/5} \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ = -\frac{\mu}{2} \int_{\Omega^n} (\nabla \mathbf{u}^{n+2/5})^t : \nabla \mathbf{v} \, d\mathbf{x}, \quad \forall \mathbf{v} \in \mathbf{V}_0(\Omega^n), \\ \int_{\Omega^n} q \nabla \cdot \mathbf{u}^{n+3/5} \, d\mathbf{x} = 0, \quad \forall q \in L^2(\Omega^n). \end{array} \right. \quad (4.44)$$

#### 4.4. Space discretization

At each time step, the sub-problems (4.42)–(4.44) have to be solved in the domain  $\Omega^n$  which is clearly non-polygonal because of the curved part  $\gamma^n$  of the boundary. To approximate velocity and pressure we use here an *iso-parametric* version (discussed in, e.g., [28, Chapter 5]) of the *Bercovier–Pironneau* finite element approximation of the Stokes equation. This approximation was introduced in [5] and is further discussed in [57]; it is a variant of the celebrated Hood–Taylor approximation of the Stokes equation and it relies on a globally continuous approximation of the velocity, piecewise linear on a triangular or tetrahedral finite element mesh, and on a globally continuous approximation of the pressure, piecewise linear on a mesh twice coarser. To describe the *iso-parametric* finite element approximation of the free-boundary problem under consideration, let us consider a generic bounded domain  $\Omega \subset \mathbb{R}^2$ , with a curved boundary  $\partial\Omega$  (like  $\Omega(t)$ , in Fig. 4.1, for example). We introduce a triangulation  $\mathcal{T}_h$  of  $\Omega$  and we decompose it as  $\mathcal{T}_h = \mathcal{T}_{0h} \cup \mathcal{T}_{Rh}$ , where

$$\mathcal{T}_{0h} = \{K \mid K \in \mathcal{T}_h, K \text{ has two vertices on } \partial\Omega \text{ and} \\ \text{the related edge of } K \text{ is curved}\}, \quad (4.45)$$

$$\mathcal{T}_{Rh} = \{K \mid K \in \mathcal{T}_h, \text{ the three edges of } K \text{ are rectilinear}\}. \quad (4.46)$$

Every *rectilinear* triangle  $K \in \mathcal{T}_{Rh}$  is divided into four sub-triangles  $K_{iK}$ ,  $i = 1, 2, 3, 4$ , by joining the mid-points of its edges. On the other hand, every curved triangle  $K \in \mathcal{T}_{0h}$  is approximated by the quadrilateral  $\tilde{K}$  defined by  $a_{1K}$ ,  $a_{2K}$ ,  $a_{23K}$ ,  $a_{3K}$ , with  $a_{23K}$  the mid-point of the arc  $a_{2K}a_{3K}$ , or an approximation of this mid-point, as shown in Fig. 4.3. The quadrilateral  $\tilde{K}$  is then decomposed in four sub-triangles  $K_{iK}$ ,  $i = 1, 2, 3, 4$  and we define the following six-dimensional spaces

$$\tilde{P}_2(\tilde{K}) = \{\varphi \mid \varphi \in \mathcal{C}^0(\tilde{K}), \varphi|_{K_{iK}} \in P_1, \forall i = 1, 2, 3, 4\}, \quad \forall K \in \mathcal{T}_{0h}. \quad (4.47)$$

Similarly, if  $K \in \mathcal{T}_{Rh}$  we associate to it

$$\tilde{P}_2(K) = \{\varphi \mid \varphi \in \mathcal{C}^0(K), \varphi|_{K_{iK}} \in P_1, \forall i = 1, 2, 3, 4\} \quad (4.48)$$

(as usual,  $P_1$  denotes, in (4.47) and (4.48) the space of two variable polynomials of degree  $\leq 1$ ); the spaces defined by (4.47) and (4.48) will be used below to construct the velocity spaces. Similarly, to construct the discrete pressure spaces, we will use the following three-dimensional subspace of  $\tilde{P}_2(\tilde{K})$  if  $K \in \mathcal{T}_{0h}$  (resp., of  $\tilde{P}_2(K)$  if  $K \in \mathcal{T}_{Rh}$ )

$$\tilde{P}_1(\tilde{K}) = \{\varphi \mid \varphi \in \tilde{P}_2(\tilde{K}), \varphi(a_{ijK}) = 1/2[\varphi(a_{iK}) + \varphi(a_{jK})], \\ \forall i, j, 1 \leq i, j \leq 3, i \neq j\}, \quad (4.49)$$

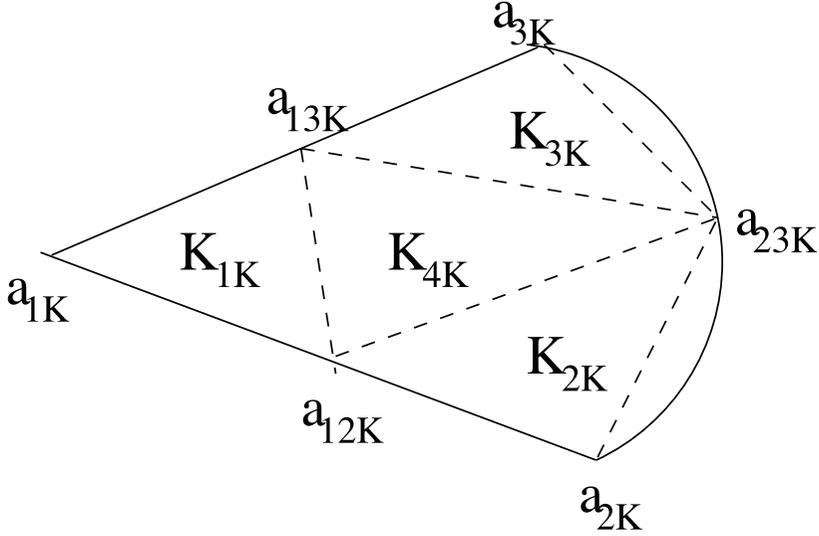


Fig. 4.3. Approximation of a curved triangle  $K$  by  $\tilde{K} = \bigcup_{i=1}^4 K_{iK}$ .

where  $a_{ijK} = a_{jiK}$ ,  $\forall i, j$ ,  $1 \leq i, j \leq 3$ ,  $i \neq j$  (resp.,  $\tilde{P}_1(K) = P_1$ ). Collecting the above pieces of information, we “approximate”  $\mathcal{T}_{0h}$  by  $\tilde{\mathcal{T}}_{0h} = \{\tilde{K}\}_{K \in \mathcal{T}_{0h}}$  and define  $\tilde{\mathcal{T}}_h$  by  $\tilde{\mathcal{T}}_h = \tilde{\mathcal{T}}_{0h} \cup \mathcal{T}_{Rh}$ , and  $\Omega_h$  as the interior of  $\bigcup_{K \in \tilde{\mathcal{T}}_h} K$ . Assuming that  $\Omega = \Omega(t)$  with  $\Omega(t)$  as in Fig. 4.1, we approximate then the pressure and velocity spaces associated to the domain  $\Omega(t)$  by, respectively,

$$\begin{aligned} P_h(\Omega(t)) &= \{q \mid q \in \mathcal{C}^0(\bar{\Omega}_h(t)), q|_K \in P_1, \forall K \in \mathcal{T}_{Rh}, \\ &\quad q|_{\tilde{K}} \in \tilde{P}_1(\tilde{K}), \forall \tilde{K} \in \mathcal{T}_{0h}\}, \end{aligned} \quad (4.50)$$

$$\begin{aligned} \mathbf{V}_h(\Omega(t)) &= \{\mathbf{v} \mid \mathbf{v} \in (\mathcal{C}^0(\bar{\Omega}_h(t)))^2, \mathbf{v}|_K \in (\tilde{P}_2(K))^2, \forall K \in \mathcal{T}_{Rh}, \\ &\quad \mathbf{v}|_{\tilde{K}} \in (\tilde{P}_2(\tilde{K}))^2, \forall \tilde{K} \in \mathcal{T}_{0h}\}. \end{aligned} \quad (4.51)$$

In order to define the space containing the function describing the shape of the free boundary, we need to go into the details of how the triangulations of the flow region are obtained. Let us take a uniform structured triangulation of the reference rectangular domain,  $\hat{\Omega}$ , as shown in Fig. 4.2. The triangulation  $\hat{\mathcal{T}}_h$  of the current domain is obtained through the mapping  $A_t^{-1}$  which corresponds (locally) to a dilation (resp., a compression) in the vertical direction if  $\eta((x_1, t)/H) < 1$  (resp.,  $> 1$ ). The discretization in the  $Ox_1$  direction remains thus the same for  $0 < t < T$ . Moreover, since the triangulation of  $\hat{\Omega}$  is structured and uniform, the number of vertices on the free boundary  $\gamma(t)$  is equal to the number of vertices on the rigid bottom boundary  $\Sigma$ ; the coordinates of the top and bottom vertices differ only by their  $x_2$  component. As a consequence,  $\Sigma$  is divided in intervals of equal

size, and a most natural approximation of the space  $S$  is the finite dimensional space  $S_h$  defined by

$$S_h = \{\eta \mid \eta \in C^0[0, L], \eta|_I \in P_1, \forall I \in I_h, \eta|_{x_1=0} = \eta|_{x_1=L}\},$$

where  $I_h$  is the set of the edges of the velocity triangulation contained in  $\Sigma$ .

#### 4.5. Full discretization

Let us introduce the following sets and spaces:

$$\begin{aligned} \mathbf{V}_{0h}(\Omega_h^n) &= \{\mathbf{v} \mid \mathbf{v} \in \mathbf{V}_h(\Omega_h^n), \mathbf{v} = \mathbf{0} \text{ on } \Sigma, \mathbf{v}|_{x_1=0} = \mathbf{v}|_{x_1=L}\}, \\ P_{0h}(\Omega_h^n) &= \left\{ q \mid q \in P_h(\Omega_h^n), \int_{\Omega_h^n} q \, d\mathbf{x} = 0, q|_{x_1=0} = q|_{x_1=L} \right\}, \\ \mathbf{V}_{h,-}^{n+1/5} &= \left\{ \mathbf{v} \mid \mathbf{v} \in \mathbf{V}_h(\Omega_h^n), \mathbf{v} = \mathbf{0} \text{ on } \Gamma_{h,-}^{n+1/5}, \mathbf{v}|_{x_1=0} = \mathbf{v}|_{x_1=L} \right\}, \\ \Gamma_{h,-}^{n+1/5} &= \left\{ \mathbf{x} \mid \mathbf{x} \in \gamma_h^n, \mathbf{u}_h^{n+1/5}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0 \right\}, \\ \hat{\mathbf{W}}_{h,+}^{n+3/5} &= \left\{ \mathbf{v} \mid \mathbf{v} \in \mathbf{V}_h(\hat{\Omega}), \mathbf{v} = \mathbf{0} \text{ on } \hat{\Gamma}_{h,+}^{n+3/5}, \mathbf{v}|_{\xi_1=0} = \mathbf{v}|_{\xi_1=L} \right\}, \\ \hat{\Gamma}_{h,+}^{n+3/5} &= \left\{ \mathbf{x} \mid \mathbf{x} \in \hat{\gamma}, \mathbf{w}_h^{n+3/5}(\mathbf{x}) \cdot \mathbf{e}_2 > 0 \right\}. \end{aligned}$$

Then, the *fully discrete free-boundary problem* reads as follows (taking advantage of (4.41)–(4.44)):

$$\mathbf{u}_h^0 = \mathbf{u}_{0h}, \quad \Omega_h^0 = \Omega_h(0), \quad \eta_h^0 = \eta_h(0); \quad (4.52)$$

then for  $n \geq 0$ ,  $\mathbf{u}_h^n$ ,  $\Omega_h^n$ ,  $\eta_h^n$  being known, solve the following sub-problems:

(i) Find  $\mathbf{u}_h^{n+1/5} \in \mathbf{V}_{0h}(\Omega_h^n)$  and  $p_h^{n+1/5} \in P_{0h}(\Omega_h^n)$  such that

$$\left\{ \begin{aligned} & \rho \int_{\Omega_h^n} \frac{\mathbf{u}_h^{n+1/5} - \mathbf{u}_h^n}{\Delta t} \cdot \mathbf{v} \, d\mathbf{x} \\ & + \frac{\mu}{2} \int_{\Omega_h^n} \nabla \mathbf{u}_h^{n+1/5} : \nabla \mathbf{v} \, d\mathbf{x} - \int_{\Omega_h^n} p_h^{n+1/5} \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ & = -s \int_0^L \left( 1 + \left| \frac{\partial \eta_h^n}{\partial x_1} \right|^2 \right)^{-1/2} \frac{\partial v_1}{\partial x_1} \, dx_1 \\ & \quad - s \int_0^L \frac{\partial \eta_h^n}{\partial x_1} \left( 1 + \left| \frac{\partial \eta_h^n}{\partial x_1} \right|^2 \right)^{-1/2} \frac{\partial v_2}{\partial x_1} \, dx_1 \\ & + \rho \int_{\Omega_h^n} \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} - \frac{\mu}{2} \int_{\Omega_h^n} (\nabla \mathbf{u}_h^n)^t : \nabla \mathbf{v} \, d\mathbf{x}, \quad \forall \mathbf{v} \in \mathbf{V}_{0h}(\Omega_h^n), \\ & \int_{\Omega_h^n} \nabla \cdot \mathbf{u}_h^{n+1/5} q \, d\mathbf{x} = 0, \quad \forall q \in P_h(\Omega_h^n). \end{aligned} \right. \quad (4.53)$$

- (ii) Find  $\mathbf{u}_h^{n+2/5} \in \mathbf{V}_h(\Omega_h^n)$  via the solution of the following pure advection problem in  $\Omega_h^n \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_{\Omega_h^n} \frac{\partial \mathbf{u}_h}{\partial t} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega_h^n} (\mathbf{u}_h^{n+1/5} \cdot \nabla) \mathbf{u}_h \cdot \mathbf{v} \, d\mathbf{x} = 0, & \forall \mathbf{v} \in \mathbf{V}_{h,-}^{n+1/5}, \\ \mathbf{u}_h(t^n) = \mathbf{u}_h^{n+1/5}, \\ \mathbf{u}_h(t) = \mathbf{u}_h^{n+1/5} & \text{on } \Gamma_{h,-}^{n+1/5} \times (t^n, t^{n+1}), \end{cases} \quad (4.54)$$

and by setting  $\mathbf{u}_h^{n+2/5} = \mathbf{u}_h(t^{n+1})$ .

- (iii) Find  $\mathbf{u}_h^{n+3/5} \in \mathbf{V}_{0h}(\Omega_h^n)$  and  $p_h^{n+3/5} \in P_{0h}(\Omega_h^n)$  such that

$$\begin{cases} \rho \int_{\Omega_h^n} \frac{\mathbf{u}_h^{n+3/5} - \mathbf{u}_h^{n+2/5}}{\Delta t} \cdot \mathbf{v} \, d\mathbf{x} \\ + \frac{\mu}{2} \int_{\Omega_h^n} \nabla \mathbf{u}_h^{n+3/5} : \nabla \mathbf{v}_h \, d\mathbf{x} - \int_{\Omega_h^n} p_h^{n+3/5} \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ = -\frac{\mu}{2} \int_{\Omega_h^n} (\nabla \mathbf{u}_h^{n+2/5})^t : \nabla \mathbf{v} \, d\mathbf{x}, & \forall \mathbf{v} \in \mathbf{V}_{0h}(\Omega_h^n), \\ \int_{\Omega_h^n} \nabla \cdot \mathbf{u}_h^{n+3/5} q \, d\mathbf{x} = 0, & \forall q \in P_h(\Omega_h^n). \end{cases} \quad (4.55)$$

- (iv) Find  $\hat{\mathbf{u}}_h^{n+4/5} \in \mathbf{V}_h(\hat{\Omega})$  by solving first the following pure advection problem in  $\hat{\Omega} \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_{\hat{\Omega}} \frac{\partial \hat{\mathbf{u}}_h}{\partial t} \cdot \hat{\mathbf{v}} J(\eta_h^n) \, d\xi - \int_{\hat{\Omega}} (\hat{\mathbf{w}}_h^{n+3/5} \cdot \hat{\nabla}) \hat{\mathbf{u}}_h \cdot \hat{\mathbf{v}} J(\eta_h^n) \, d\xi = 0, & \forall \hat{\mathbf{v}} \in \hat{\mathbf{W}}_{h,+}^{n+3/5}, \\ \hat{\mathbf{u}}_h(t^n) = \hat{\mathbf{u}}_h^{n+3/5}, \\ \hat{\mathbf{u}}_h(t) = \hat{\mathbf{u}}_h^{n+3/5} & \text{on } \hat{\Gamma}_{h,+}^{n+3/5} \times (t^n, t^{n+1}). \end{cases} \quad (4.56)$$

Set then  $\hat{\mathbf{u}}_h^{n+4/5} = \hat{\mathbf{u}}_h(t^{n+1})$  and  $p_h^{n+4/5} = p_h^{n+1/5} + p_h^{n+3/5}$ .

- (v) Find the new position of the free surface  $\eta_h^{n+1} \in S_h$  by solving the following problem in  $(0, L) \times (t^n, t^{n+1})$ :

$$\begin{cases} \int_0^L \frac{\partial \eta_h}{\partial t} \phi \, d\xi_1 \\ + \int_0^L \left[ \hat{\mathbf{u}}_{1h}^{n+4/5}(\xi_1, H) \frac{\partial \eta_h}{\partial \xi_1} - \hat{\mathbf{u}}_{2h}^{n+4/5}(\xi_1, H) \right] \phi \, d\xi_1 = 0, & \forall \phi \in S_h, \\ \eta_h(t^n) = \eta_h^n; \end{cases} \quad (4.57)$$

and then by setting

$$\eta^{n+1} = \eta(t^{n+1}). \quad (4.58)$$

- (vi) The new physical domain  $\Omega_h^{n+1}$  is obtained from the reference domain through the mapping  $A_{t^{n+1},h}^{-1}$  defined as

$$A_{t^{n+1},h}^{-1}: \hat{\Omega} \subset \mathbb{R}^2 \rightarrow \Omega_h^{n+1} \subset \mathbb{R}^2, \\ \boldsymbol{\xi} = \{\xi_1, \xi_2\} \rightarrow \mathbf{x} = \{x_1, x_2\} = \begin{cases} x_1 = \xi_1, \\ x_2 = \frac{\eta_h^{n+1}(\xi_1)}{H} \xi_2. \end{cases} \quad (4.59)$$

The velocity  $\mathbf{u}_h^{n+1}$  and the pressure  $p_h^{n+1}$  defined on the new domain  $\Omega_h^{n+1}$  are obtained as follows:

$$\mathbf{u}_h^{n+1}(x_1, x_2) = \hat{\mathbf{u}}_h^{n+4/5}(x_1, Hx_2/\eta_h^{n+1}(x_1)), \\ p_h^{n+1}(x_1, x_2) = p_h^{n+4/5}(x_1, \eta_h^n(x_1)x_2/\eta_h^{n+1}(x_1)), \quad (4.60)$$

with  $\{x_1, x_2\} \in \Omega_h^{n+1}$ .

Several remarks are in order, the main ones being (from our point of view) the following

REMARK 4.1. The boundary integrals in (4.42) are of the following form

$$s \int_{\gamma} \mathcal{H}(\eta) \mathbf{n} \cdot \mathbf{v} \, d\gamma;$$

we evaluated them using the following relations

$$\int_{\gamma} \mathcal{H}(\eta) n_i v_i \, d\gamma = \int_{\gamma} \frac{\partial t_i}{\partial \gamma} v_i \, d\gamma = - \int_{\gamma} t_i \frac{\partial v_i}{\partial \gamma} \, d\gamma, \quad i = 1, 2,$$

where  $\mathbf{t} = \{t_1, t_2\}$  is the unit tangent vector at  $\gamma$ , and where, here,  $\mathbf{v}$  denotes the function  $x_1 \rightarrow \mathbf{v}(x_1, \eta(x_1))$  which is periodic of period  $L$ . Thus,

$$\int_{\gamma} \mathcal{H}(\eta) n_1 v_1 \, d\gamma = - \int_0^L \left(1 + \left|\frac{\partial \eta}{\partial x_1}\right|^2\right)^{-1/2} \frac{\partial v_1}{\partial x_1} \, dx_1, \quad (4.61)$$

and

$$\int_{\gamma} \mathcal{H}(\eta) n_2 v_2 \, d\gamma = - \int_0^L \frac{\partial \eta}{\partial x_1} \left(1 + \left|\frac{\partial \eta}{\partial x_1}\right|^2\right)^{-1/2} \frac{\partial v_2}{\partial x_1} \, dx_1. \quad (4.62)$$

These integrals contain only first order derivatives, then low-order basis functions (1st order in the present case) can be used for their approximation leading to the approximation of surface tension contribution encountered in (4.53). The use of integration by parts to reduce the order of differentiation is quite natural and well in the spirit of the variational approach used here.

REMARK 4.2. To solve problems like (4.57) we have used a Taylor–Galerkin approximation (inspired by the celebrated Lax–Wendroff scheme) on  $(0, L) \times (t^n, t^{n+1})$ , namely

$$\begin{aligned} & \int_0^L (\eta^{n+1} - \eta^n) \phi \, dx_1 \\ &= \Delta t \int_0^L \left( u_2^{n+4/5} - u_1^{n+4/5} \frac{d\eta^n}{dx_1} \right) \phi \, dx_1 - \frac{1}{2} |\tau|^2 \int_0^L u_1^{n+4/5} \frac{du_2^{n+4/5}}{dx_1} \phi \, dx_1 \\ & \quad - \frac{1}{2} |\tau|^2 \int_0^L u_1^{n+4/5} \frac{d\eta^n}{dx_1} \left( u_1^{n+4/5} \frac{d\phi}{dx_1} + \frac{du_1^{n+4/5}}{dx_1} \phi \right) dx_1, \quad \forall \phi \in S_h, \end{aligned} \quad (4.63)$$

The integrals in (4.63) are approximated using piecewise linear polynomial functions associated to the discrete velocity finite element space, i.e., defined from the vertices, located on the (approximate) free surface  $\gamma_h^n$ , of the triangulation used to approximate velocity at time  $t^n$ .

REMARK 4.3. Problems such as (4.53) and (4.55) are *discrete Stokes problems*. Their *iterative solution* by *Uzawa/conjugate gradient algorithms* is discussed, with many details, in, e.g., [28] (see also references therein). The *pure advection* problems (4.43) and (4.54) can be solved by the *method of characteristics* as shown in, e.g., [28] and [57] (see also the many related references therein). An easier to implement alternative to the method of characteristics is provided by a *wave-like equation method* discussed in, e.g., [28, Chapter 6] and [19], [20], [54]. Indeed, the particulate flow related numerical results shown in Section 3.4 have been obtained using the wave-like equation method mentioned just above, a similar comment applying to the numerical results shown in Section 4.6. Owing to the importance of the wave-like equation approach concerning this article, we are going to discuss it right now:

The components of the solution  $\mathbf{u}$  of problem (4.43) are solution of a pure advection problem of the following form

$$\frac{\partial \varphi}{\partial t} + \mathbf{V} \cdot \nabla \varphi = 0 \quad \text{in } \Omega \times (t_0, t_f), \quad (4.64)$$

$$\varphi = g \quad \text{on } \Gamma_- \times (t_0, t_f), \quad \varphi(t) \text{ is } x_1\text{-periodic of period } L, \quad (4.65)$$

$$\varphi(t_0) = \varphi_0, \quad (4.66)$$

where: (i) The functions  $g$  and  $\varphi_0$  are given, with  $\partial g / \partial t = 0$ . (ii)  $\nabla \cdot \mathbf{V} = 0$  (in the sense of distributions) with  $\partial \mathbf{V} / \partial t = \mathbf{0}$ . (iii) The domain  $\Omega$  is like the one shown in Fig. 4.1, namely

$$\Omega = \{ \mathbf{x} \mid \mathbf{x} = \{x_1, x_2\}, x_2 \in (0, \eta(x_1)) \text{ with } x_1 \in (0, L) \}, \quad (4.67)$$

the function  $\eta$  in (4.67) being Lipschitz continuous on  $[0, L]$  and  $x_1$ -periodic of period  $L$ ; we shall denote by  $\gamma$  the subset of the boundary  $\partial \Omega$  of  $\Omega$  defined by

$$\gamma = \{ \mathbf{x} \mid \mathbf{x} = \{x_1, x_2\}, x_2 = \eta(x_1), x_1 \in (0, L) \}.$$

(iv) The subset  $\Gamma_-$  of  $\partial\Omega$  is defined by

$$\Gamma_- = \{\mathbf{x} \mid \mathbf{x} \in \gamma, \mathbf{V}(x) \cdot \mathbf{n}(\mathbf{x}) < 0\},$$

$\mathbf{n}$  being the outward unit normal vector at  $\gamma$ .

We can easily show that (4.64)–(4.66) implies (formally, at least) that for a.e.  $t \in (t_0, t_f)$  we have

$$\int_{\Omega} \frac{\partial^2 \varphi}{\partial t^2} v \, d\mathbf{x} + \int_{\Omega} (\mathbf{V} \cdot \nabla \varphi)(\mathbf{V} \cdot \nabla v) \, d\mathbf{x} + \int_{\gamma \setminus \Gamma_-} \mathbf{V} \cdot \mathbf{n} \frac{\partial \varphi}{\partial t} v \, d\gamma = 0, \quad \forall v \in H_-^1(\Omega), \quad (4.68)$$

to be completed by

$$\varphi = g \quad \text{on } \Gamma_- \times (t_0, t_f), \quad \varphi(t) \text{ is } x_1\text{-periodic of period } L, \quad (4.69)$$

$$\varphi(t_0) = \varphi_0, \quad \frac{\partial \varphi}{\partial t}(t_0) = -\mathbf{V} \cdot \nabla \varphi_0, \quad (4.70)$$

with, in (4.68),  $H_-^1(\Omega) = \{v \mid v \in H^1(\Omega), v = 0 \text{ on } \Gamma_-\}$ . Actually, relation (4.68) implies that

$$\mathbf{V} \cdot \mathbf{n} \left( \frac{\partial \varphi}{\partial t} + \mathbf{V} \cdot \nabla \varphi \right) = 0 \quad \text{on } (\gamma \setminus \Gamma_-) \times (t_0, t_f),$$

which is, clearly, a kind of generalized Neumann boundary condition (perfectly absorbing for the pure advection problem under consideration). As shown in, e.g., [28, Chapter 6], the formulation (4.68)–(4.70) of problem (4.64)–(4.66) is well-suited to solution methods via standard finite element approximations like the globally continuous and piecewise affine ones used to obtain the results presented in Section 4.6 (and 3.4).

REMARK 4.4. To solve the transport problem (4.56) we use again a wave equation approach. We observe that (4.56) does not contain  $\frac{\partial \hat{\mathbf{u}}_h}{\partial \xi_1}$  (since  $\mathbf{w}_h^{n+3/5}$  is parallel to  $\mathbf{e}_2$ ), implying that the above problem reduces to a family (infinite for the continuous problem, finite for the discrete one) of transport problems in one space dimension along the  $\hat{O}\xi_2$  direction. Then for  $\xi_1 \in [0, L)$ , each component of  $\hat{\mathbf{u}}$  is solution of a transport problem of the following form

$$\begin{cases} \frac{\partial \phi}{\partial t} - a\xi_2 \frac{\partial \phi}{\partial \xi_2} = 0 & \text{on } (0, H) \times (t^n, t^{n+1}), \\ \phi(t^n) = \phi_0, \\ \phi(H, t) = b & \text{if } a > 0, \quad t \in (t^n, t^{n+1}), \end{cases} \quad (4.71)$$

with  $a$  and  $b$  constant with respect to  $\xi_2$  and  $t$ . We observe that any smooth solution of (4.71) verifies:

$$\left\{ \begin{array}{l} \frac{\partial^2 \phi}{\partial t^2} - a^2 \xi_2 \frac{\partial}{\partial \xi_2} \left( \xi_2 \frac{\partial \phi}{\partial \xi_2} \right) = 0 \quad \text{on } (0, H) \times (t^n, t^{n+1}), \\ \phi(t^n) = \phi_0, \\ \frac{\partial \phi}{\partial t}(t^n) = a \xi_2 \frac{\partial \phi_0}{\partial \xi_2}, \\ \phi(H, t) = b \quad \text{if } a > 0, \quad t \in (t^n, t^{n+1}), \\ \left[ a \xi_2 \left( \frac{\partial \phi}{\partial t} - a \xi_2 \frac{\partial \phi}{\partial \xi_2} \right) \right]_{\xi_2=H} = 0, \quad \text{if } a \leq 0, \quad t \in (t^n, t^{n+1}). \end{array} \right. \quad (4.72)$$

From a practical point of view we take advantage of the following variational formulation of (4.72), better suited for finite element treatment.

1. Case  $a > 0$ .

$$\left\{ \begin{array}{l} \int_0^H \frac{\partial^2 \phi}{\partial t^2} \frac{v}{\xi_2} d\xi_2 + a^2 \int_0^H \xi_2 \frac{\partial \phi}{\partial \xi_2} \frac{\partial v}{\partial \xi_2} d\xi_2 \\ = 0, \quad \forall v \in \mathcal{D}(0, H), \quad \text{a.e. on } (t^n, t^{n+1}), \\ \phi(t^n) = \phi_0, \\ \frac{\partial \phi}{\partial t}(t^n) = a \xi_2 \frac{\partial \phi_0}{\partial \xi_2}, \\ \phi(H, t) = b, \quad t \in (t^n, t^{n+1}), \end{array} \right. \quad (4.73)$$

with  $\mathcal{D}(0, H) = \{v \mid v \in C^\infty[0, H], v \text{ has a compact support in } (0, H)\}$ .

2. Case  $a \leq 0$ .

$$\left\{ \begin{array}{l} \int_0^H \frac{\partial^2 \phi}{\partial t^2} \frac{v}{\xi_2} d\xi_2 + a^2 \int_0^H \xi_2 \frac{\partial \phi}{\partial \xi_2} \frac{\partial v}{\partial \xi_2} d\xi_2 - a \frac{\partial \phi}{\partial t}(H, t)v(H) = 0, \\ \forall v \in V_0, \quad \text{a.e. on } (t^n, t^{n+1}), \\ \phi(t^n) = \phi_0, \\ \frac{\partial \phi}{\partial t}(t^n) = a \xi_2 \frac{\partial \phi_0}{\partial \xi_2}, \end{array} \right. \quad (4.74)$$

with  $V_0 = \{v \mid v \in C^\infty[0, H], v \text{ has a compact support in } (0, H)\}$ .

REMARK 4.5. The methodology discussed in Sections 4.3 to 4.5 has been applied in [29] to investigate numerically the Hopf bifurcation phenomenon for incompressible viscous flow down inclined planes, discussed by Nishida et al. in [51]. For a general approach concerning the approximation of free surface problems for incompressible viscous fluid with surface tension see, e.g., [72] and [8].

#### 4.6. Numerical experiments

In this section we present the results of numerical experiments where the methodology discussed in Sections 4.3 to 4.5 has been applied to the solution of some test problems.

##### 4.6.1. First test problem: Perturbed flow in an horizontal periodical plane

As a first example, we consider the situation where a liquid of viscosity  $\mu = 1/13$  and density  $\rho = 1$  occupies a periodic region in a half-plane resting above a horizontal line. The liquid is initially at rest but the free-surface on top is not in equilibrium. We assume that the surface tension coefficient is  $s = 1$ , and that  $H = 0.15$  and  $L = 1$ . The initial non-equilibrium position of the free-surface is shown in Fig. 4.5 and it is given by

$$\eta_0(x_1) = \frac{1}{4}e^{-20(x_1-0.5)^2} + H - \frac{1}{4} \int_0^L e^{-20(x-0.5)^2} dx.$$

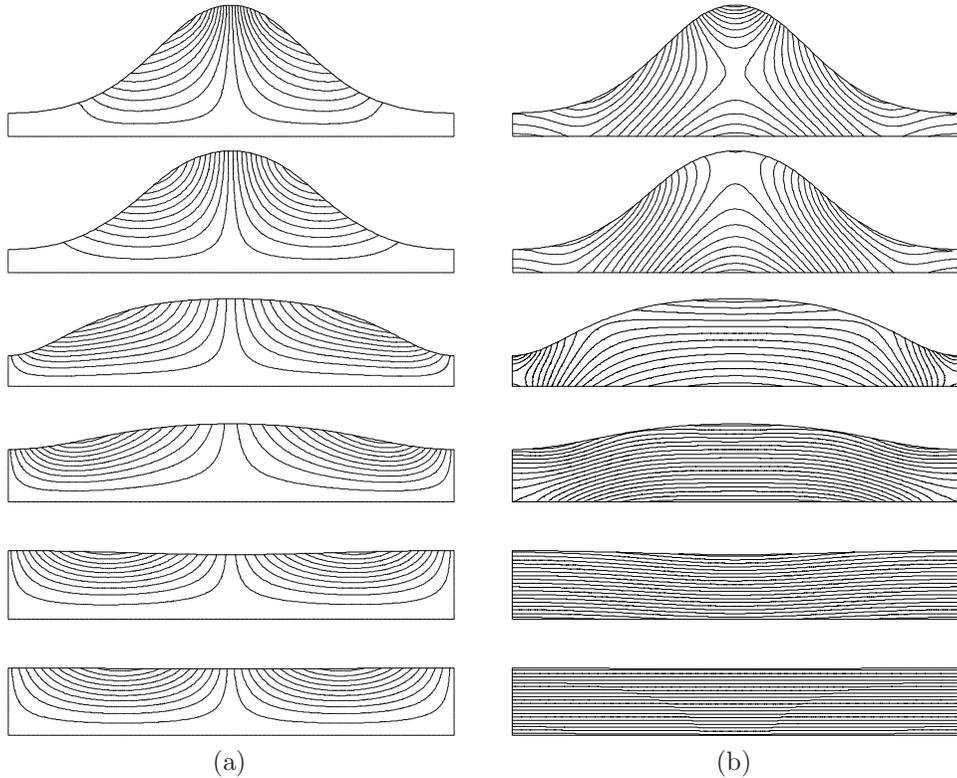


Fig. 4.4. First test problem: the streamlines (left) and isobars (right) at times  $t = 0.05, 0.15, 0.5, 0.75, 1.5,$  and  $2.5$  (from top to bottom).

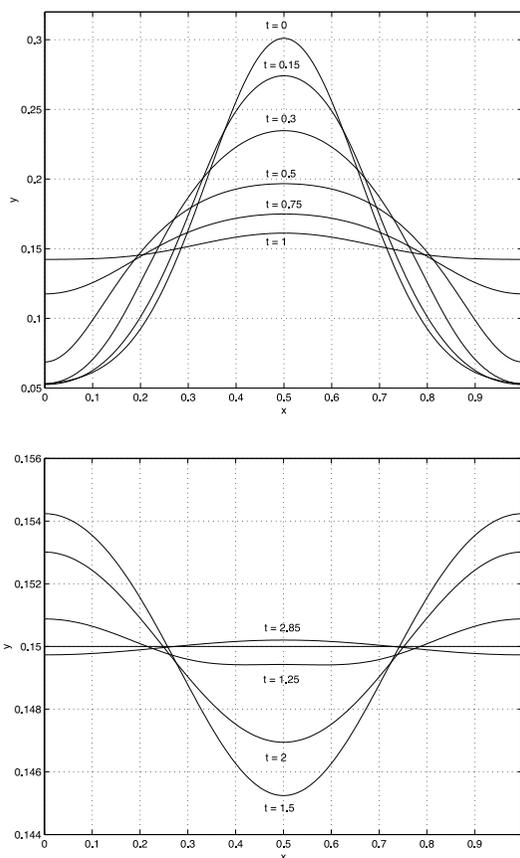


Fig. 4.5. First test problem: evolution of the free surface.

The space-discretization parameters are, in the reference domain  $\hat{\Omega} (= (0, L) \times (0, H))$ ,  $h_p = 1/100$ , and  $h_v = 1/200$  for approximating the pressure and the velocity, respectively. The time-discretization step is  $\Delta t = 10^{-4}$ . For  $t > 0$ , the coupled action of gravity and surface tension puts the fluid into an oscillatory motion leading, as  $t \rightarrow +\infty$ , to an equilibrium state where  $\eta = H = 0.15$  and  $\mathbf{u} = \mathbf{0}$ . Figs. 4.4 (a) and 4.4 (b) show the streamlines and the isobars at different instants of time. Fig. 4.5 shows the evolution of the free-surface. The flow and the shape of the free boundary have a behavior very close, qualitatively, to the one shown in [2] where the simulation of a fluid flow in a three-dimensional annular container is considered.

#### 4.6.2. Second test problem: Perturbed flow in an inclined periodical plane

In the second example we consider the flow of a liquid of viscosity  $\mu = 1/13$ , density  $\rho = 1$ , and surface tension coefficient  $s = 1$ . We suppose that the fluid is above a periodic plane whose inclination with respect to the horizontal is 45 degrees. To simplify the geometry and still employ the methodology used in Section 4.6.1, for the solution of the first test problem, we still keep  $\Omega$  above an horizontal plane but give gravity a 45 degree inclination with respect to the vertical. In this frame, we chose  $H = 0.15$ ,  $L = 1$ , and the same initial shape of the free-surface as the one in the first problem (see Fig. 4.7); we still assume that  $\mathbf{u} = \mathbf{0}$  at time  $t = 0$ . We take  $h_p = 1/100$  and  $h_v = 1/200$  in the reference domain  $\hat{\Omega}$ , and  $\Delta t = 10^{-4}$ . The streamlines and isobars have been visualized in Figs. 4.6 (a) and 4.6 (b), while the evolution of the free surface is shown in Fig. 4.7. This time, the equilibrium position corresponds to  $\eta = H = 0.15$  and for  $\mathbf{u}$  to a velocity distribution of the Poiseuille type (i.e., parabolic) with  $\mathbf{u} = \mathbf{0}$  on the bottom of the flow region and maximum velocity on the free-surface. Depending on the Reynolds number, i.e., the viscosity of the fluid or the inclination of the plane, Hopf bifurcation occurs. This phenomenon has been investigated theoretically in [51], and numerically in [29], using the method discussed in this article.

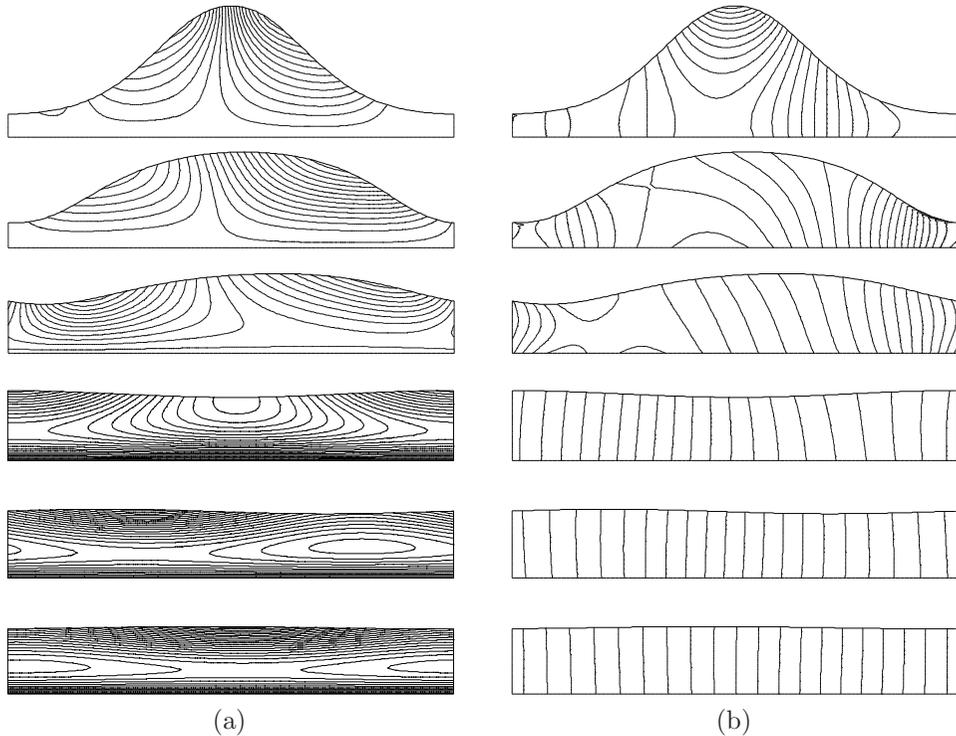


Fig. 4.6. Second test problem: the streamlines (left) and isobars (right) at times  $t = 0.05, 0.4, 0.75, 1.5, 2,$  and  $2.5$  (from top to bottom).

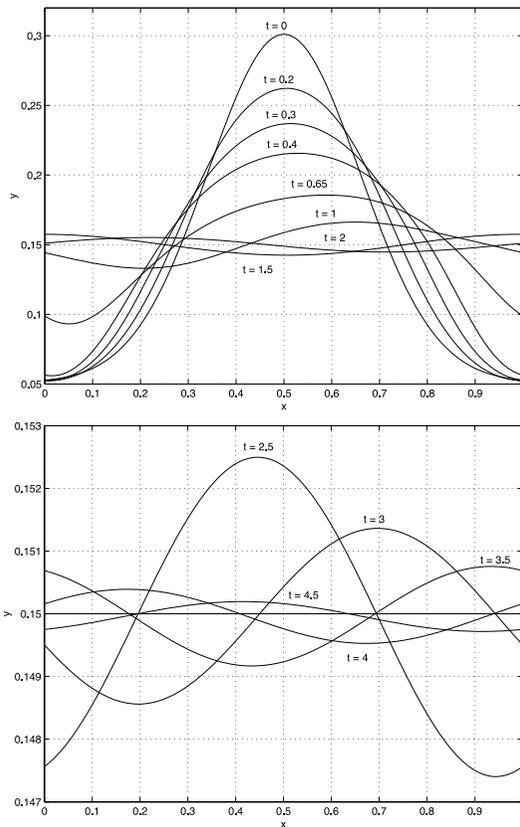


Fig. 4.7. Second test problem: evolution of the free-surface. From top to bottom:  $t = 0$ , 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.15, 0.19, 0.23, 0.5.

## 5. Operator-splitting methods for elliptic Monge–Ampère equations in dimension two

### 5.1. Generalities

These last years have been witnessing a surge of interest in Monge–Ampère equations and their numerical solution. Indeed, beside the important role they play in Differential Geometry, these equations occur in the modeling of various phenomena in Mechanics, Physics, etc. (see [10], [52]) and their “full nonlinearity” (in the sense of, e.g., Caffarelli and Cabré; see ref. [9]) presents an interesting challenge to Numerical Analysts. Our goal, here, is to show that via operator-splitting based methods, and an appropriate reformulation, these equations can be reduced to the solutions of a sequence of classical problem (at least for two-dimensional Dirichlet problems in the elliptic case). In this article, following [16], [18], we will discuss first an augmented Lagrangian approach to the solution of the elliptic Dirichlet–Monge–Ampère problem and then a least-squares based method.

## 5.2. Formulation of the problem

Let  $\Omega$  be a bounded domain of  $\mathbb{R}^2$ ; we denote by  $\Gamma$  the boundary of  $\Omega$ . The two-dimensional Dirichlet problem for the Monge–Ampère equation reads as follows:

$$\det D^2\psi = f \quad \text{in } \Omega, \quad \psi = g \quad \text{on } \Gamma, \quad (5.1)$$

where, in (5.1),  $D^2\psi$  is the Hessian of  $\psi$ , i.e.,  $D^2\psi = \left(\frac{\partial^2\psi}{\partial x_i\partial x_j}\right)_{1 \leq i, j \leq 2}$  and where  $f$  and  $g$  are two given functions, with  $f > 0$ . Unlike the (closely related) Dirichlet problem for the Laplace operator, (5.1) may have multiple solutions (actually, two at most; cf., e.g., [14, Chapter 4]), and the smoothness of the data does not imply the existence of a smooth solution. Concerning the last property, suppose that  $\Omega = (0, 1) \times (0, 1)$  and consider the special case where (5.1) is defined by

$$\frac{\partial^2\psi}{\partial x_1^2} \frac{\partial^2\psi}{\partial x_2^2} - \left| \frac{\partial^2\psi}{\partial x_1\partial x_2} \right|^2 = 1 \quad \text{in } \Omega, \quad \psi = 0 \quad \text{on } \Gamma. \quad (5.2)$$

Problem (5.2) can not have smooth solutions since, for those solutions, the boundary condition  $\psi = 0$  on  $\Gamma$  implies that the product  $\frac{\partial^2\psi}{\partial x_1^2} \frac{\partial^2\psi}{\partial x_2^2}$  and the cross-derivative  $\frac{\partial^2\psi}{\partial x_1\partial x_2}$  vanish at the boundary, implying in turn that  $\det D^2\psi$  is strictly less than one in some neighborhood of  $\Gamma$ . The above (non-existence) result is not a consequence of the non-smoothness of  $\Gamma$ , since a similar non-existence property holds if in (5.2) one replaces the above  $\Omega$  by the ovoid-shaped domain whose  $C^\infty$ -boundary is defined by  $\Gamma = \bigcup_{i=1}^4 \Gamma_i$ , with  $\Gamma_1 = \{x \mid x = \{x_1, x_2\}, x_2 = 0, 0 \leq x_1 \leq 1\}$ ,  $\Gamma_3 = \{x \mid x = \{x_1, x_2\}, x_2 = 1, 0 \leq x_1 \leq 1\}$ ,  $\Gamma_2 = \{x \mid x = \{x_1, x_2\}, x_1 = 1 - \ln 4 / (\ln x_2(1 - x_2)), 0 < x_2 < 1\}$ ,  $\Gamma_4 = \{x \mid x = \{x_1, x_2\}, x_1 = \ln 4 / (\ln x_2(1 - x_2)), 0 < x_2 < 1\}$ . Actually, for the above two  $\Omega$ s the non-existence of solutions for problem (5.2) follows from the non-strict convexity of these domains.

REMARK 5.1. Suppose that  $\Omega$  is simply connected; let us define a vector-valued function  $\mathbf{u}$  by  $\mathbf{u} = \left\{ \frac{\partial\psi}{\partial x_2}, -\frac{\partial\psi}{\partial x_1} \right\} (= \{u_1, u_2\})$ ; problem (5.1) takes then the equivalent formulation

$$\begin{cases} \det \nabla \mathbf{u} = f & \text{in } \Omega, \quad \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega, \\ \mathbf{u} \cdot \mathbf{n} = \frac{dg}{ds} & \text{on } \Gamma, \end{cases} \quad (5.3)$$

where, in (5.3),  $\mathbf{n}$  denotes the outward unit vector normal at  $\Gamma$ , and  $s$  is a counter-clockwise curvilinear abscissa. Once  $\mathbf{u}$  is known, one obtains  $\psi$  via the solution of the following Poisson–Dirichlet problem

$$-\Delta\psi = \frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} \quad \text{in } \Omega, \quad \psi = g \quad \text{on } \Gamma.$$

Problem (5.3) has clearly an incompressible fluid flow flavor,  $\psi$  playing here the role of a stream function. Relations (5.3) can be used to solve problem (5.1) but this approach will not be further investigated here.

### 5.3. An augmented Lagrangian approach for the solution of problem (5.1)

Suppose that in (5.1)  $f \in L^1(\Omega)$  and  $g \in H^{3/2}(\Gamma)$ ; it makes sense then to attempt solving problem (5.1) in  $H^2(\Omega)$  by considering it as a nonlinear biharmonic problem (albeit not a typical one since only one boundary condition is prescribed). A way to do so is to consider the following problem from Calculus of Variations:

$$\min_{\varphi \in E_{fg}} \frac{1}{2} \int_{\Omega} |\Delta \varphi|^2 dx \quad (5.4)$$

with  $E_{fg} = \{\varphi \mid \varphi \in V_g, \det D^2 \varphi = f\}$  and  $V_g = \{\varphi \mid \varphi \in H^2(\Omega), \varphi = g \text{ on } \Gamma\}$ . Problem (5.4) is in turn equivalent to

$$\min_{\{\varphi, \mathbf{q}\} \in \mathcal{E}_{fg}} \frac{1}{2} \int_{\Omega} |\Delta \varphi|^2 dx \quad (5.5)$$

with

$$\mathcal{E}_{fg} = \{\{\varphi, \mathbf{q}\} \mid \varphi \in V_g, \mathbf{q} \in \mathbf{Q}, \mathbf{q} = D^2 \varphi, \det \mathbf{q} = f\},$$

and

$$\mathbf{Q} = \{\mathbf{q} \mid \mathbf{q} = (q_{ij})_{1 \leq i, j \leq 2}, q_{21} = q_{12}, q_{ij} \in L^2(\Omega)\}.$$

Following, e.g., refs. [31], [18], we associate to problem (5.5):

- (i) The augmented Lagrangian functional  $\mathcal{L}_r : (H^2(\Omega) \times \mathbf{Q}) \times \mathbf{Q} \rightarrow \mathbb{R}$  defined, with  $r > 0$ , by

$$\mathcal{L}_r(\varphi, \mathbf{q}; \boldsymbol{\mu}) = \frac{1}{2} \int_{\Omega} |\Delta \varphi|^2 dx + \frac{r}{2} \int_{\Omega} |D^2 \varphi - \mathbf{q}|^2 dx + \int_{\Omega} \boldsymbol{\mu} : (D^2 \varphi - \mathbf{q}) dx, \quad (5.6)$$

with  $\mathbf{S} : \mathbf{T} = \sum_{1 \leq i, j \leq 2} s_{ij} t_{ij}$  if  $\mathbf{S} = (s_{ij})$  and  $\mathbf{T} = (t_{ij})$ .

- (ii) The saddle-point problem

$$\begin{cases} \text{Find } \{\{\psi, \mathbf{p}\}, \boldsymbol{\lambda}\} \in (V_g \times \mathbf{Q}_f) \times \mathbf{Q} \text{ such that} \\ \mathcal{L}_r(\psi, \mathbf{p}; \boldsymbol{\mu}) \leq \mathcal{L}_r(\psi, \mathbf{p}; \boldsymbol{\lambda}) \leq \mathcal{L}_r(\varphi, \mathbf{q}; \boldsymbol{\lambda}), \\ \forall \{\{\varphi, \mathbf{q}\}, \boldsymbol{\mu}\} \in (V_g \times \mathbf{Q}_f) \times \mathbf{Q}, \end{cases} \quad (5.7)$$

with  $\mathbf{Q}_f = \{\mathbf{q} \mid \mathbf{q} \in \mathbf{Q}, \det \mathbf{q} = f\}$ . One can easily show that if  $\{\{\psi, \mathbf{p}\}, \boldsymbol{\lambda}\}$  is a solution of the saddle-point problem (5.7), then  $\psi$  is solution of the Monge–Ampère problem (5.1),  $\mathbf{p} = D^2 \psi$ , and  $\boldsymbol{\lambda}$  is a Lagrange multiplier associated to the relation  $\mathbf{p} - D^2 \psi = \mathbf{0}$ .

Concerning the solution of problem (5.7) we advocate (following, e.g., refs. [31], [16]) the (relatively simple) Douglas–Rachford–Uzawa algorithm below:

$$\{\psi^{-1}, \boldsymbol{\lambda}^0\} \text{ is given in } V_g \times \mathbf{Q}; \quad (5.8)$$

then, for  $n \geq 0$ ,  $\{\psi^{n-1}, \boldsymbol{\lambda}^n\}$  being known in  $V_g \times \mathbf{Q}$ , solve

$$\begin{cases} \mathbf{p}^n \in \mathbf{Q}_f, \\ \mathcal{L}_r(\psi^{n-1}, \mathbf{p}^n; \boldsymbol{\lambda}^n) \leq \mathcal{L}_r(\psi^{n-1}, \mathbf{q}; \boldsymbol{\lambda}^n), \quad \forall \mathbf{q} \in \mathbf{Q}_f, \end{cases} \quad (5.9)$$

$$\begin{cases} \psi^n \in V_g, \\ \mathcal{L}_r(\psi^n, \mathbf{p}^n; \boldsymbol{\lambda}^n) \leq \mathcal{L}_r(\varphi, \mathbf{p}^n; \boldsymbol{\lambda}^n), \quad \forall \varphi \in V_g, \end{cases} \quad (5.10)$$

and update  $\boldsymbol{\lambda}^n$  by

$$\boldsymbol{\lambda}^{n+1} = \boldsymbol{\lambda}^n + r(D^2\psi^n - \mathbf{p}^n). \quad (5.11)$$

Algorithm (5.8)–(5.11) deserves many comments, among them:

- Concerning the initialization of algorithm (5.8)–(5.11), we advocate  $\boldsymbol{\lambda}^0 = \mathbf{0}$  and  $\psi^{-1}$  is the solution of the Poisson–Dirichlet problem

$$-\Delta\psi^{-1} = f^{1/2} \quad \text{in } \Omega, \quad \psi^{-1} = g \quad \text{on } \Gamma, \quad (5.12)$$

the rationale for such a choice being given in ref. [18].

- Problem (5.9) can be solved point-wise (in practice at the grid points of a finite element or finite difference mesh). Indeed, (5.9) reduces, a.e. on  $\Omega$ , to the solution of a finite dimensional problem of the following type:

$$\min_{\mathbf{z}} \left\{ \frac{r}{2} (z_1^2 + z_2^2 + 2z_3^2) - \mathbf{b}_n(x) \cdot \mathbf{z} \right\} \quad (5.13)$$

with  $\mathbf{z} = \{z_i\}_{i=1}^3 \in \{\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^3, z_1 z_2 - z_3^2 = f(x)\}$ . The solution of problem (5.13) (a generalized eigenvalue problem) is discussed in [18].

- Problem (5.10) reduces to a linear variational problem of the following type:

$$\begin{cases} \psi^n \in V_g, \\ \int_{\Omega} \Delta\psi^n \Delta\varphi \, dx + r \int_{\Omega} D^2\psi^n : D^2\varphi \, dx = L_n(\varphi), \\ \forall \varphi \in V_0 \quad (= H^2(\Omega) \cap H_0^1(\Omega)), \end{cases} \quad (5.14)$$

functional  $L_n(\cdot)$  being linear and continuous over  $V_0$ . The unique solution of problem (5.14) can be computed by a conjugate gradient algorithm operating in  $V_g$  and  $V_0$  equipped with the scalar product  $\{v, w\} \rightarrow \int_{\Omega} \Delta v \Delta w \, dx$  and the corresponding norm. Such an algorithm is described in [18]; its most important feature is that for well chosen finite element approximations, its discrete variants require no more than the solution of two discrete Poisson–Dirichlet problems at each iteration (see [18] for details).

- Suppose that problem (5.1) has no solution in  $H^2(\Omega)$  while  $V_g$  and  $\mathbf{Q}_f$  are both nonempty (as it is the case for problem (5.2)). In that case, we expect the arithmetic divergence of sequence  $\{\lambda^n\}_n$  and the geometric convergence of sequence  $\{\{\psi^n, \mathbf{p}^n\}\}_n$  to a pair  $\{\psi, \mathbf{p}\} \in V_g \times \mathbf{Q}_f$  such that

$$\begin{cases} \{\psi, \mathbf{p}\} \text{ minimizes (locally or globally) the functional} \\ \{\varphi, \mathbf{q}\} \rightarrow \|D^2\varphi - \mathbf{q}\|_{(L^2(\Omega))^4} \text{ over the set } V_g \times \mathbf{Q}_f. \end{cases} \quad (5.15)$$

The rationale for this prediction is discussed at length in [18], and, indeed, when applying algorithm (5.8)–(5.11) to the solution of problem (5.2), the numerical results show that the above algorithm behaves as expected; these numerical results are reported in Section 5.4 (see also [16], [18]). On the basis of these results it seems that for parameter  $r$  well chosen, algorithm (5.8)–(5.11) produces either a solution of problem (5.1) if such a solution exists in  $H^2(\Omega)$ , or a generalized solution in the sense of (5.15), i.e., a least squares solution, if (5.1) is without solution in  $H^2(\Omega)$ , while  $V_g$  and  $\mathbf{Q}_f$  are nonempty. This result justifies the least squares methodology discussed in Section 5.5 (see also [17]). From a geometrical point of view problem (5.1) has solutions in  $H^2(\Omega)$  if  $D^2V_g$  and  $\mathbf{Q}_f$  (both subsets of space  $\mathbf{Q}$ ) intersect; such a situation has been visualized on Fig. 5.1. Fig. 5.2 corresponds to a situation where problem (5.1) has no solution in  $H^2(\Omega)$ ; a generalized solution, in the sense of (5.15) has been visualized on this figure.

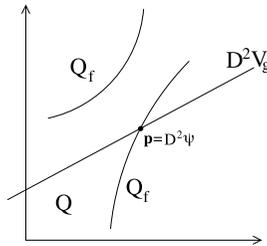


Fig. 5.1. Problem (5.1) has a solution in  $H^2(\Omega)$ .

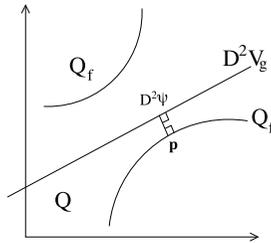


Fig. 5.2. Problem (5.1) has no solution in  $H^2(\Omega)$ .

- The saddle-point formulation (5.7) of problem (5.1) is a mixed variational formulation where the “nonlinearity burden” has been transferred from  $\psi$  to  $\mathbf{p}$ , making it purely algebraic. Actually, this approach (this is even more true for the discrete analogues of problems (5.1) and (5.7); see [18] for details) provides a solution method where instead of solving (5.1) directly (i.e., without introducing additional functions, like  $\mathbf{p}$ ), we solve it via its associated Pfaff system (see, e.g., [22, Chapter A, V]), namely:

$$d\psi - u_1 dx_1 - u_2 dx_2 = 0 \quad \text{in } \Omega, \quad (5.16)$$

$$du_1 - p_{11} dx_1 - p_{12} dx_2 = 0 \quad \text{in } \Omega, \quad (5.17)$$

$$du_2 - p_{12} dx_1 - p_{22} dx_2 = 0 \quad \text{in } \Omega, \quad (5.18)$$

$$p_{11} p_{22} - p_{12}^2 = f, \quad (5.19)$$

completed by the boundary condition

$$\psi = g \quad \text{on } \Gamma. \quad (5.20)$$

System (5.16)–(5.20) provides clearly a mixed formulation of (5.1).

- Let us return to problem (IVP) of Section 2.1 and suppose that  $A = A_1 + A_2$ . An alternative to the Lie’s and Strang’s schemes is provided by the following (first order) Douglas–Rachford scheme:

$$\varphi^0 = \varphi_0; \quad (5.21)$$

then, for  $n \geq 0$ ,  $\varphi^n$  being known, solve

$$(\varphi^{n+1/2} - \varphi^n)/\tau + A_1(\varphi^{n+1/2}) + A_2(\varphi^n) = 0, \quad (5.22)$$

$$(\varphi^{n+1} - \varphi^{n+1/2})/\tau + A_1(\varphi^{n+1/2}) + A_2(\varphi^{n+1}) = 0. \quad (5.23)$$

The basic properties of scheme (5.21)–(5.23), can be found in, e.g., [28, Chapter 2] (see also the references therein). It is shown in, e.g., ref. [31] (see also [24]) that algorithms such as (5.8)–(5.11) are in fact “disguised” Douglas–Rachford algorithms, with parameter  $r$  the reciprocal of a time step (see the two above references for details). We will return on this issue in Remark 5.2, below.

- The finite element implementation of algorithm (5.8)–(5.11) is discussed in ref. [18]; it relies on a mixed finite element approximation of problems (5.1) and (5.7) where  $\psi$  and the four components of the tensor-valued functions  $\mathbf{p}$  and  $\boldsymbol{\lambda}$  are approximated by functions belonging to the finite dimensional space  $V_h$  defined by

$$V_h = \{v \mid v \in C^0(\bar{\Omega}), v|_T \in P_1, \forall T \in \mathcal{T}_h\}, \quad (5.24)$$

where in (5.24):  $P_1$  is the space of the two-variable polynomials of degree  $\leq 1$ ,  $\mathcal{T}_h$  is a triangulation of  $\Omega$ , and  $h$  is a space-discretization step. The numerical results shown in Section 5.4 have been obtained using the above approximation.

REMARK 5.2. Since the link between *augmented Lagrangian* methods and *operator-splitting* ones has been mentioned a few times already, we are feeling obliged to show it. To reach that goal we consider the following problem:

$$A(\varphi) = 0, \quad (5.25)$$

where  $A$  is an operator mapping a real Hilbert space  $H$  into itself; we denote by  $(\cdot, \cdot)$  the scalar product of  $H$  and by  $\|\cdot\|$  the corresponding norm. We suppose that  $A$  is the differential of a *convex functional*  $J: H \rightarrow \mathbb{R}$  (i.e.,  $J' = A$ ) implying that if  $\varphi$  is a solution of problem (5.25) it is also a solution of the *equivalent minimization problem*

$$\begin{cases} \varphi \in H, \\ J(\varphi) \leq J(\theta), \quad \forall \theta \in H. \end{cases} \quad (5.26)$$

Suppose now that  $J = J_1 + J_2$ , the functionals  $J_1$  and  $J_2$  being both *convex* and *differentiable*; if we denote  $J'_1$  and  $J'_2$  by  $A_1$  and  $A_2$ , respectively, problem (5.25) can also be written as

$$A_1(\varphi) + A_2(\varphi) = 0, \quad (5.27)$$

Observe now that problem (5.26) is *equivalent* to

$$\begin{cases} \{\varphi, p\} \in \mathcal{H}, \\ j(\varphi, p) \leq j(\theta, q), \quad \forall \{\theta, q\} \in \mathcal{H} \end{cases} \quad (5.28)$$

where  $j(\theta, q) = J_1(\theta) + J_2(q)$  and  $\mathcal{H} = \{\{\theta, q\} \mid \{\theta, q\} \in H \times H, \theta - q = 0\}$ . Let  $r$  be a positive parameter; we associate to  $j(\cdot, \cdot)$  and  $\mathcal{H}$  the following *augmented Lagrangian* and *saddle-point* problem:

$$L_r(\theta, q; \mu) = j(\theta, q) + \frac{r}{2} \|\theta - q\|^2 + (\mu, \theta - q), \quad (5.29)$$

$$\begin{cases} \{\{\theta, q\}, \lambda\} \in (H \times H) \times H, \\ L_r(\varphi, p; \mu) \leq L_r(\varphi, p; \lambda) \leq L_r(\theta, q; \lambda), \quad \forall \{\{\theta, q\}, \mu\} \in (H \times H) \times H. \end{cases} \quad (5.30)$$

It is a simple exercise to show that if  $\{\{\varphi, p\}, \lambda\}$  is a saddle-point of  $L_r$  over  $(H \times H) \times H$ , it is also a solution of the following system:

$$r(\varphi - p) + A_1(\varphi) + \lambda = 0, \quad (5.31)$$

$$r(p - \varphi) + A_2(p) - \lambda = 0, \quad (5.32)$$

$$\varphi - p = 0. \quad (5.33)$$

The relations (5.31)–(5.33) imply in turn that  $\varphi$  is solution of (5.27), i.e., (since  $A = A_1 + A_2$ ), of (5.25). Following, e.g., refs. [31] and [24], we apply the following variant of algorithm (5.8)–(5.11) to the solution of the saddle-point problem (5.30):

$$\{\varphi^{-1}, \lambda^0\} \text{ is given in } H \times H; \quad (5.34)$$

for  $n \geq 0$ , assuming that  $\{\varphi^{n-1}, \lambda^n\}$  is known, solve

$$p^n \in H; \quad L_r(\varphi^{n-1}, p^n; \lambda^n) \leq L_r(\varphi^{n-1}, q; \lambda^n), \quad \forall q \in H, \quad (5.35)$$

$$\varphi^n \in H; \quad L_r(\varphi^n, p^n; \lambda^n) \leq L_r(\theta, p^n; \lambda^n), \quad \forall \theta \in H, \quad (5.36)$$

$$\lambda^{n+1} = \lambda^n + r(\varphi^n - p^n). \quad (5.37)$$

Taking advantage of the differentiability of the functional  $L_r$  with respect to  $\theta$  and  $q$ , we can replace (5.35) and (5.36) by

$$r(p^n - \varphi^{n-1}) + A_2(p^n) - \lambda^n = 0 \quad (5.38)$$

and

$$r(\varphi^n - p^n) + A_1(\varphi^n) + \lambda^n = 0, \quad (5.39)$$

respectively. Comparing relations (5.37) and (5.39) shows that  $\lambda^{n+1} = -A_1(\varphi^n)$ , which implies in turn that

$$\lambda^n = -A_1(\varphi^{n-1}). \quad (5.40)$$

It follows then from relations (5.38), (5.39) and (5.40) that

$$r(p^n - \varphi^{n-1}) + A_1(\varphi^{n-1}) + A_2(p^n) = 0, \quad (5.41)$$

$$r(\varphi^n - \varphi^{n-1}) + A_1(\varphi^n) + A_2(p^n) = 0. \quad (5.42)$$

Define  $\tau$  by  $\tau = 1/r$  and denote  $p^n$  by  $\varphi^{n-1/2}$ ; replacing  $n$  by  $n+1$  in the two above relations we obtain then

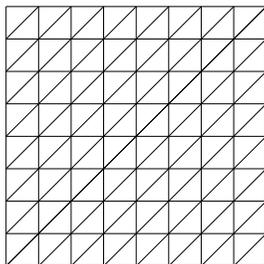
$$(\varphi^{n+1/2} - \varphi^n)/\tau + A_1(\varphi^n) + A_2(\varphi^{n+1/2}) = 0, \quad (5.43)$$

$$(\varphi^{n+1} - \varphi^n)/\tau + A_1(\varphi^{n+1}) + A_2(\varphi^{n+1/2}) = 0 \quad (5.44)$$

i.e., an obvious variant of the Douglas–Rachford scheme (5.22), (5.23), justifying thus the terminology Douglas–Rachford–Uzawa that we used for algorithm (5.8)–(5.11) (to the best of our knowledge, those links between augmented Lagrangian algorithms and alternating direction schemes were reported for the first time in [11]).

#### 5.4. Numerical results obtained via the augmented Lagrangian methodology

We are going to apply the methodology (briefly) discussed in Section 5.3 to the solution of three test problems. For all these test problems, we shall assume that  $\Omega = (0, 1) \times (0, 1)$  and that  $\mathcal{T}_h$  is a uniform triangulation like the one in Fig. 5.3, with  $h$  the length of the edges of  $\mathcal{T}_h$  adjacent to the right angles.

Fig. 5.3. A uniform triangulation of  $\Omega$  ( $h = 1/8$ ).

The **first test problem** is defined as follows:

- (i)  $f(x) = \frac{R^2}{(R^2 - |x|^2)^2}$ ,  $\forall x \in \Omega$ , with  $R \geq \sqrt{2}$  and  $|x| = (x_1^2 + x_2^2)^{1/2}$ .
- (ii)  $g(x) = \sqrt{R^2 - |x|^2}$ ,  $\forall x \in \Gamma$  ( $= \partial\Omega$ ).

If the above data hold, function  $\psi$  given by

$$\psi(x) = \sqrt{R^2 - |x|^2}, \quad \forall x \in \Omega, \quad (5.45)$$

is solution to the corresponding Monge–Ampère problem (5.1). The graph of function  $\psi$  is clearly a piece of the sphere of radius  $R$  centered at  $\{0, 0, 0\}$ . If  $R > \sqrt{2}$  we have  $\psi = C^\infty(\bar{\Omega})$ ; on the other hand, if  $R = \sqrt{2}$  we have no better than  $\psi \in W^{1,p}(\Omega)$  with  $p \in [1, 4)$ , implying that in that particular case  $\psi$  does not have the  $H^2$ -regularity. When applying the computational methods discussed in Section 5.3 to the solution of the above problem (with  $r = 1$  in algorithm (5.8)–(5.11)), we obtain if  $R = 2$ , and after 78 iterations of the discrete variant of the above algorithm, the results displayed in Table 5.1:

Table 5.1. Results for the 1st test problem ( $R = 2$ ).

$h$	$\ \psi_h^c - \psi\ _{0,\Omega}$	$\ \mathbf{D}_h^2 \psi_h^c - \mathbf{p}_h^c\ _{0,\Omega}$
1/32	$4.45 \times 10^{-6}$	$9.48 \times 10^{-7}$
1/64	$1.14 \times 10^{-6}$	$1.35 \times 10^{-6}$
1/128	$2.97 \times 10^{-7}$	$1.58 \times 10^{-6}$

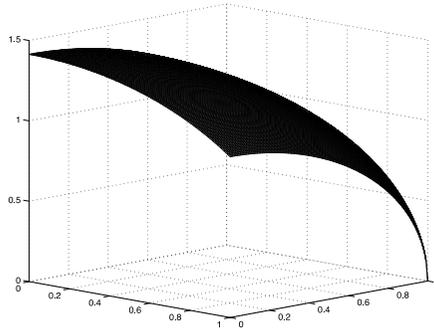
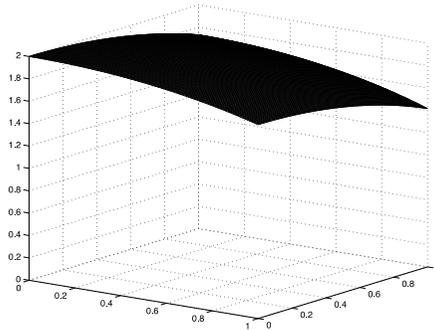
In the Table 5.1,  $\psi_h^c$  is the computed approximate solution,  $\mathbf{D}_h^2 \psi_h^c$  is the corresponding discrete Hessian, and  $\mathbf{p}_h^c$  is the computed approximation of tensor  $\mathbf{p}$ . The above results strongly suggest second order accuracy (a textbook one, indeed), which is in some sense optimal considering the type of finite element approximations we are using. If we take  $R = \sqrt{2}$ , our methodology which has been designed to solve the Monge–Ampère equation (5.1) in  $H^2(\Omega)$  is unable to capture any solution of the above problem, the corresponding algorithm (5.8)–(5.11) being divergent for any value of  $r$ . The same troubles persist if one takes  $R = \sqrt{2} + 10^{-2}$ ; on the other hand, if one takes  $R = \sqrt{2} + 10^{-1}$ , things are back to normal since, using again  $r = 1$ , we obtain after 117 iterations of algorithm (5.8)–(5.11) the results summarized in Table 5.2.

Table 5.2. Results for the 1st test problem ( $R = \sqrt{2} + 10^{-1}$ ).

$h$	$\ \psi_h^c - \psi\ _{0,\Omega}$	$\ \mathbf{D}_h^2 \psi_h^c - \mathbf{P}_h^c\ _{0,\Omega}$
1/32	$2.20 \times 10^{-5}$	$9.68 \times 10^{-7}$
1/64	$5.51 \times 10^{-6}$	$1.54 \times 10^{-6}$
1/128	$1.37 \times 10^{-6}$	$2.04 \times 10^{-6}$

The above results show that second order accuracy holds again. However, the second order derivatives of  $\psi$  being larger for  $R = \sqrt{2} + 10^{-1}$  than for  $R = 2$ , the corresponding approximations errors are also larger. On Figs. 5.4 to 5.7 we have visualized, respectively:

- (i) The graph of  $\psi$  when  $R = \sqrt{2}$ ; the singularity of  $\nabla\psi$  at  $\{1, 1\}$  appears clearly on Fig. 5.4.
- (ii) The graph of  $\psi_h^c$  corresponding to  $h = 1/128$  and  $R = 2$ .
- (iii) The graph of  $\psi_h^c$  corresponding to  $h = 1/128$  and  $R = \sqrt{2} + 10^{-1}$ .
- (iv) The graph of  $f$  when  $R = \sqrt{2} + 10^{-1}$ .

Fig. 5.4. First test problem: graph of  $\psi$  when  $R = \sqrt{2}$ .Fig. 5.5. First test problem: graph of  $\psi_h^c$  when  $R = 2$  and  $h = 1/128$ .

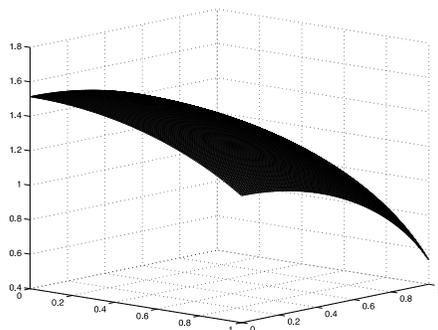


Fig. 5.6. First test problem: graph of  $\psi_h^c$  when  $R = \sqrt{2} + 10^{-1}$  and  $h = 1/128$ .

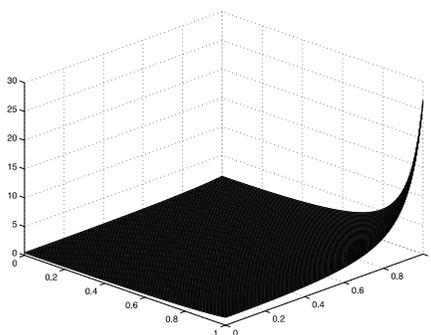


Fig. 5.7. First test problem: graph of  $f$  when  $R = \sqrt{2} + 10^{-1}$ .

REMARK 5.3. When computing the approximate solutions for  $h = 1/32$ , we stopped the iterations of algorithm (5.8)–(5.11) as soon as

$$\|\mathbf{D}_h^2 \psi_h^n - \mathbf{p}_h^n\|_{0,\Omega} \leq 10^{-6}. \quad (5.46)$$

The corresponding number of iterations is 78 for  $R = 2$ , and 117 for  $R = \sqrt{2} + 10^{-1}$  (we did not try to find the optimal value of  $r$ , or to use a variable  $r$  strategy). Next, when computing the approximate solutions for  $h = 1/64$  and  $1/128$ , we stopped iterating once the iteration numbers associated to  $h = 1/32$  were reached (actually, using (5.46) as stopping criteria for  $h = 1/64$  and  $1/128$  did not change much the approximation errors shown in Tables 5.1 and 5.2).

The **second test problem** is defined as follows:

- (i)  $f(x) = 1/|x|$ ,  $\forall x \in \Omega$ .
- (ii)  $g(x) = \frac{(2|x|)^{\frac{3}{2}}}{3}$ ,  $\forall x \in \Gamma$ .

With these data, a solution to the Monge–Ampère problem (5.1) is the function  $\psi$  defined by

$$\psi(x) = \frac{(2|x|)^{\frac{3}{2}}}{3}, \quad \forall x \in \Omega. \quad (5.47)$$

One can easily check that  $\psi \notin C^2(\bar{\Omega})$ ; however, since  $\psi \in W^{2,p}(\Omega)$ ,  $\forall p \in [1, 4)$ , it has, in principle, enough regularity so that we can apply algorithm (5.8)–(5.11) to the solution of the corresponding problem (5.1). Indeed, despite the singularity of function  $f$  at  $\{0, 0\}$  (see Fig. 5.8), algorithm (5.8)–(5.11), with  $r = 1$ , provides after 160 iterations the results summarized in the following Table 5.3:

Table 5.3. Results for the 2nd test problem.

$h$	$\ \psi_h^c - \psi\ _{0,\Omega}$	$\ \mathbf{D}_h^2 \psi_h^c - \mathbf{P}_h^c\ _{0,\Omega}$
1/32	$5.56 \times 10^{-5}$	$9.91 \times 10^{-7}$
1/64	$1.50 \times 10^{-5}$	$1.60 \times 10^{-6}$
1/128	$3.94 \times 10^{-6}$	$2.02 \times 10^{-6}$

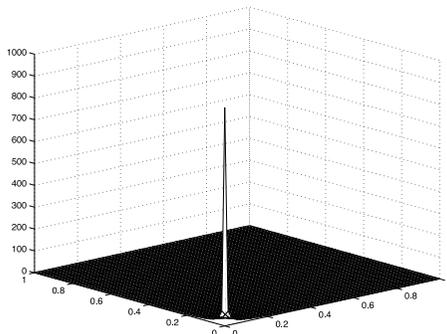


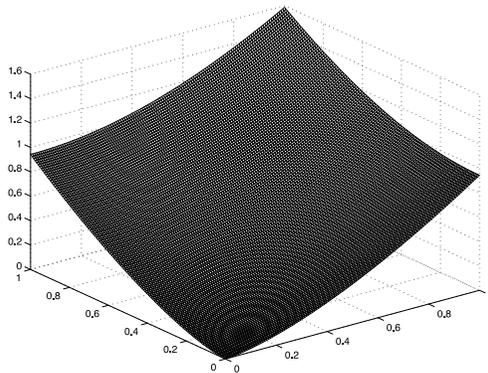
Fig. 5.8. Second test problem: graph of  $f$ .

From the above results we can infer that second order accuracy still holds. On Fig. 5.8 (resp., 5.9) we have visualized function  $f$  (resp., the computed approximate solution obtained with  $h = 1/128$ ).

The **third test problem** is—by far—the more interesting since we consider this time the solution of problem (5.2), namely:

$$\frac{\partial^2 \psi}{\partial x_1^2} \frac{\partial^2 \psi}{\partial x_2^2} - \left| \frac{\partial^2 \psi}{\partial x_1 \partial x_2} \right|^2 = 1 \quad \text{in } \Omega, \quad \psi = 0 \quad \text{on } \Gamma.$$

Despite the smoothness of its data, the above problem has no smooth solution, the troubles coming from the non-strict convexity of  $\Omega = (0, 1) \times (0, 1)$ . When applying algorithm (5.8)–(5.11) (in fact a discrete variant of it) to the solution of problem (5.2) we observe the following phenomena:

Fig. 5.9. Second test problem: graph of  $\psi_h^c$  ( $h = 1/128$ ).

- (i) For  $r$  sufficiently small (here,  $r \leq 2$  is fine) the sequence  $\{\{\psi_h^n, \mathbf{p}_h^n\}\}_{n \geq 0}$  converges geometrically (albeit slowly (approximately 1,000 iterations if  $r = 2$ )) to a limit  $\{\psi_h^c, \mathbf{p}_h^c\}$  while sequence  $\{\lambda_h^n\}_{n \geq 0}$  diverges arithmetically.
- (ii) A close inspection of the numerical results shows that the curvature of the graph of  $\psi_h^c$  becomes negative close to the corners, in violation of the Monge–Ampère equation; actually, as expected, it is violated also along the boundary, since  $\|\mathbf{D}_h^2 \psi_h^c - \mathbf{p}_h^c\|_{0, \Omega} = 1.8 \times 10^{-2}$  if  $h = 1/32$ ,  $3.3 \times 10^{-2}$  if  $h = 1/64$ ,  $4.2 \times 10^{-2}$  if  $h = 1/128$ , while  $\|\mathbf{D}_h^2 \psi_h^c - \mathbf{p}_h^c\|_{0, \Omega_1} = 2.7 \times 10^{-4}$  if  $h = 1/32$ ,  $4.1 \times 10^{-4}$  if  $h = 1/64$ ,  $4.9 \times 10^{-4}$  if  $h = 1/128$ , and  $\|\mathbf{D}_h^2 \psi_h^c - \mathbf{p}_h^c\|_{0, \Omega_2} = 4.4 \times 10^{-5}$  if  $h = 1/32$ ,  $4.9 \times 10^{-5}$  if  $h = 1/64$ ,  $5.1 \times 10^{-5}$  if  $h = 1/128$ , where  $\Omega_1 = (1/8, 7/8)^2$  and  $\Omega_2 = (1/4, 3/4)^2$ . These results suggest that  $\det D^2 \psi = 1$  is “almost” verified in  $\Omega_2$ .

The graph of  $\psi_h^c$  obtained with  $h = 1/64$  has been shown on Fig. 5.10, while the intersections of this graph with the planes  $x_1 = 1/2$  and  $x_1 = x_2$  have been shown on Figs. 5.11 and 5.12, respectively, for  $h = 1/32$ ,  $1/64$ , and  $1/128$ . Since  $\psi_h^c$  does not vary very much with  $h$ , we suspect that, according to Section 5.3, what we have here is a (good) approximation of one of those functions of  $H^2(\Omega) \cap H_0^1(\Omega)$  whose Hessian is at a minimal  $L^2$ -distance (global or local) from the set  $\mathbf{Q}_f$  defined in Section 5.3. Assuming that the above is true we can claim that the solution-less problem (5.2) has been solved in a *least squares sense* in the functional space  $H^2(\Omega)$ , leading to a (not so novel in general, but possibly new in the Monge–Ampère “environment”) concept of *generalized solution*. Actually, our intuition concerning the least squares interpretation of the solutions obtained, via algorithm (5.8)–(5.11), when  $V_g$  and  $\mathbf{Q}_f$  are both nonempty, seems to be right as we shall see in Section 5.5, below, where the least-squares solution of problem (5.1) will be discussed.

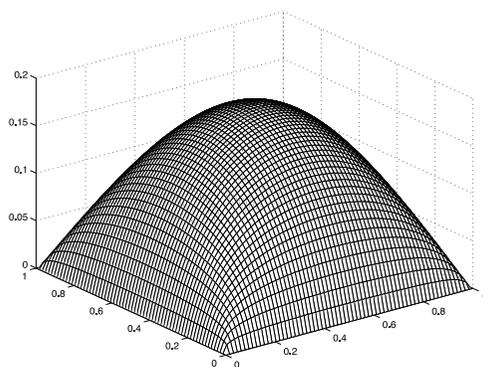


Fig. 5.10. Third test problem: graph of the computed solution  $\psi_h^c$  ( $h = 1/64$ ).

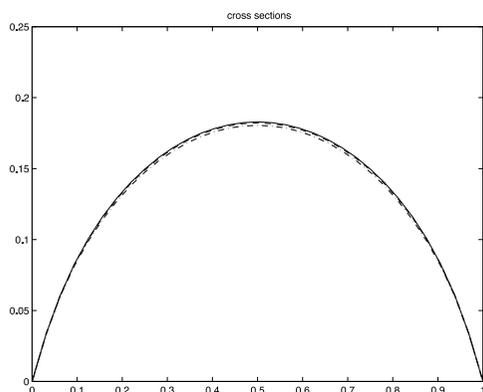


Fig. 5.11. Third test problem: graphs of the computed solutions  $\psi_h^c$  restricted to the plane  $x_1 = 1/2$ .

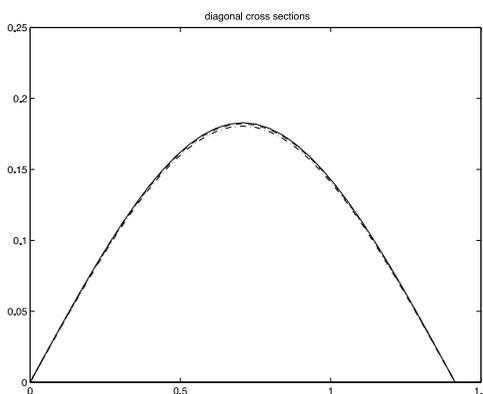


Fig. 5.12. Third test problem: graphs of the computed solutions  $\psi_h^c$  restricted to the plane  $x_1 = x_2$ .

REMARK 5.4. When applying algorithm (5.8)–(5.11) to the solution of problem (5.1) we have to solve at each iteration a discrete analogue of the linear variational problem (5.14); to solve this finite dimensional problem, we have used the *preconditioned conjugate gradient algorithm*, converging typically 5 to 7 iterations, the preconditioning requiring the solution of two discrete Poisson–Dirichlet problems per conjugate gradient iteration. The mesh being uniform we have used *Fast Poisson Solvers* to achieve preconditioning.

## 5.5. A least-squares approach for the solution of problem (5.1)

### 5.5.1. Two least-squares formulation of problem (5.1)

The above mentioned behavior of algorithm (5.8)–(5.11) strongly suggests to look at least-squares methods for the solution of (5.1). Such a method has been investigated in [18]; it relies on the following (brute force) least-squares formulation of (5.1):

$$\min_{\varphi \in V_g} j_1(\varphi), \quad (\text{LSQ1})$$

with

$$j_1(\varphi) = \begin{cases} \frac{1}{2} \int_{\Omega} |\det D^2 \varphi - f|^2 dx, & \text{if } (\det D^2 \varphi - f) \in L^2(\Omega), \\ +\infty, & \text{otherwise.} \end{cases} \quad (5.48)$$

The solution of (5.1) via the nonlinear least-squares formulation (LSQ1) is discussed in [18]; it relies on iterative methods whose convergence, however, is not as clear cut as the convergence of the discrete variants of algorithm (5.8)–(5.11). Actually, Section 5.3 suggests an alternative (and more natural) least squares formulation, namely

$$\min_{\{\varphi, \mathbf{q}\} \in V_g \times \mathbf{Q}_f} j_2(\varphi, \mathbf{q}), \quad (\text{LSQ2})$$

with

$$j_2(\varphi, \mathbf{q}) = \frac{1}{2} \int_{\Omega} |D^2 \varphi - \mathbf{q}|^2 dx. \quad (5.49)$$

### 5.5.2. On the iterative solution of problem (LSQ2) and related issues

Let us define the (non-convex) functional  $I_{\mathbf{Q}_f} : \mathbf{Q} \rightarrow \mathbb{R} \cup \{+\infty\}$  by

$$I_{\mathbf{Q}_f} = \begin{cases} 0, & \text{if } \mathbf{q} \in \mathbf{Q}_f; \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.50)$$

in other words,  $I_{\mathbf{Q}_f}(\cdot)$  is the *indicator functional* of the set  $\mathbf{Q}_f$  in  $\mathbf{Q}$ . Problem (LSQ2) is thus clearly *equivalent* to the following minimization problem in  $V_g \times \mathbf{Q}$ :

$$\min_{\{\varphi, \mathbf{q}\} \in V_g \times \mathbf{Q}} [j_2(\varphi, \mathbf{q}) + I_{\mathbf{Q}_f}(\mathbf{q})], \quad (5.51)$$

whose (formal) *Euler–Lagrange* equation reads as follows at a solution  $\{\psi, \mathbf{p}\}$  of problem (LSQ2):

$$\begin{cases} \{\psi, \mathbf{p}\} \in V_g \times \mathbf{Q}, \\ \int_{\Omega} (D^2\psi - \mathbf{p}) : (D^2\varphi - \mathbf{q}) \, dx + \langle \partial I_{\mathbf{Q}_f}(\mathbf{p}), \mathbf{q} \rangle = 0, \\ \forall \{\varphi, \mathbf{q}\} \in V_0 \times \mathbf{Q}, \end{cases} \quad (5.52)$$

with  $V_0 = H^2(\Omega) \cap H_0^1(\Omega)$  and  $\partial I_{\mathbf{Q}_f}(\mathbf{p})$  a (kind of) generalized differential of functional  $I_{\mathbf{Q}_f}(\cdot)$  at  $\mathbf{p}$ . Classically, we associate to (5.52) the following initial value problem (flow in the terminology of Dynamical Systems) since its steady state solutions solve problem (5.52):

$$\{\psi(0), \mathbf{p}(0)\} = \{\psi_0, \mathbf{p}_0\} \in V_g \times \mathbf{Q}, \quad (5.53)$$

$$\begin{cases} \{\psi(t), \mathbf{p}(t)\} \in V_g \times \mathbf{Q}, \quad \forall t \in (0, +\infty), \\ \int_{\Omega} \Delta(\partial\psi/\partial t) \Delta\varphi \, dx + \int_{\Omega} (D^2\psi - \mathbf{p}) : D^2\varphi \, dx = 0, \quad \forall \varphi \in V_0, \\ \int_{\Omega} (\partial\mathbf{p}/\partial t) : \mathbf{q} \, dx + \int_{\Omega} (\mathbf{p} - D^2\psi) : \mathbf{q} \, dx + \langle \partial I_{\mathbf{Q}_f}(\mathbf{p}), \mathbf{q} \rangle = 0, \quad \forall \mathbf{q} \in \mathbf{Q}, \end{cases} \quad (5.54)$$

the idea being to capture the steady state solutions of (5.54) via the integration of (5.53), (5.54) from  $t = 0$  to  $t = +\infty$ . Concerning the initialization of (5.53), (5.54), following Section 5.3 we advocate for  $\psi_0$  the unique solution in  $V_g$  of the Dirichlet problem  $-\Delta\psi_0 = \sqrt{f}$  in  $\Omega$ ,  $\psi_0 = g$  on  $\partial\Omega$  and  $\mathbf{p}_0 = D^2\psi_0$ . Let  $\tau (> 0)$  be a *time-discretization step*; applying to (5.53), (5.54) an operator-splitting scheme à la Marchuk–Yanenko (see, e.g., Section 2 and the references therein) we obtain the following iterative method:

$$\{\psi^0, \mathbf{p}^0\} = \{\psi_0, \mathbf{p}_0\}; \quad (5.55)$$

for  $n \geq 0$ ,  $\{\psi^n, \mathbf{p}^n\}$  being known, compute  $\{\psi^{n+1}, \mathbf{p}^{n+1}\}$  as follows:

$$\frac{\mathbf{p}^{n+1} - \mathbf{p}^n}{\tau} + \mathbf{p}^{n+1} + \partial I_{\mathbf{Q}_f}(\mathbf{p}^{n+1}) = D^2\psi^n, \quad (5.56)$$

$$\begin{cases} \psi^{n+1} \in V_g, \\ \int_{\Omega} \Delta[(\psi^{n+1} - \psi^n)/\tau] \Delta\varphi \, dx + \int_{\Omega} D^2\psi^{n+1} : D^2\varphi \, dx \\ = \int_{\Omega} \mathbf{p}^{n+1} : D^2\varphi \, dx, \quad \forall \varphi \in V_0. \end{cases} \quad (5.57)$$

Relation (5.56) is a *necessary optimality condition* for the following minimization problem:

$$\min_{\mathbf{q} \in \mathbf{Q}_f} \left[ \frac{1}{2}(1 + \tau) \int_{\Omega} |\mathbf{q}|^2 dx - \int_{\Omega} (\mathbf{p}^n + \tau D^2 \psi^n) : \mathbf{q} dx \right], \quad (\text{NLP})$$

while (5.57) characterizes  $\psi^{n+1}$  as the solution of

$$\min_{\varphi \in V_g} \left[ \frac{1}{2} \int_{\Omega} (|\Delta \varphi|^2 + \tau |D^2 \varphi|^2) dx - \int_{\Omega} (\Delta \psi^n \Delta \varphi + \tau \mathbf{p}^{n+1} : D^2 \varphi) dx \right]. \quad (\text{LQP})$$

Each problem (NLP) being a variant of problem (5.9) (see Section 5.3) can be solved point-wise (in practice at the vertices of a finite element or finite difference mesh); to obtain  $\mathbf{p}^{n+1}$  from  $\mathbf{p}^n$  and  $\psi^n$  we have to solve point-wise a minimization problem similar to (5.13), an issue already addressed in Section 5.3. Each problem (LQP) is equivalent to (5.57), a well-posed linear variational problem. Problem (5.57) can be solved by a conjugate gradient algorithm operating in  $V_g$  and  $V_0$  equipped with the scalar product  $\{v, w\} \rightarrow \int_{\Omega} \Delta v \Delta w dx$ . It follows from, e.g., [28, Chapter 3] that such an algorithm reads as follows (the subscript  $k$  numbers the conjugate gradient iterations):

Take

$$\psi_0^{n+1} = \psi^n; \quad (5.58)$$

solve

$$\left\{ \begin{array}{l} g_0^{n+1} \in V_0, \\ \int_{\Omega} \Delta g_0^{n+1} \Delta \varphi dx = \int_{\Omega} (\Delta \psi^n \Delta \varphi + \tau D^2 \psi^n : D^2 \varphi) dx \\ \quad - \int_{\Omega} (\Delta \psi^n \Delta \varphi + \tau \mathbf{p}^{n+1} : D^2 \varphi) dx, \quad \forall \varphi \in V_0. \end{array} \right. \quad (5.59)$$

If  $g_0^{n+1} = 0$  (in practice if  $\int_{\Omega} |\Delta g_0^{n+1}|^2 dx / \int_{\Omega} |\Delta \psi^n|^2 dx \leq \epsilon$ ) take  $\psi^{n+1} = \psi^n$ ; otherwise, set

$$w_0^{n+1} = g_0^{n+1}. \quad (5.60)$$

For  $k \geq 0$  assuming that  $\psi_k^{n+1}$ ,  $g_k^{n+1}$ , and  $w_k^{n+1}$  are known with the last two different from 0, solve

$$\left\{ \begin{array}{l} \bar{g}_k^{n+1} \in V_0, \\ \int_{\Omega} \Delta \bar{g}_k^{n+1} \Delta \varphi dx = \int_{\Omega} (\Delta w_k^{n+1} \Delta \varphi + \tau D^2 w_k^{n+1} : D^2 \varphi) dx, \\ \forall \varphi \in V_0, \end{array} \right. \quad (5.61)$$

and compute

$$\rho_k^{n+1} = \int_{\Omega} |\Delta g_k^{n+1}|^2 dx \Big/ \int_{\Omega} \Delta \bar{g}_k^{n+1} \Delta w_k^{n+1} dx, \quad (5.62)$$

and set

$$\psi_{k+1}^{n+1} = \psi_k^{n+1} - \rho_k^{n+1} w_k^{n+1}, \quad (5.63)$$

$$g_{k+1}^{n+1} = g_k^{n+1} - \rho_k^{n+1} \bar{g}_k^{n+1}. \quad (5.64)$$

If  $\int_{\Omega} |\Delta g_{k+1}^{n+1}|^2 dx / \int_{\Omega} |\Delta g_0^{n+1}|^2 dx \leq \epsilon$  take  $\psi^{n+1} = \psi_{k+1}^{n+1}$ ; else, compute

$$\gamma_k^{n+1} = \int_{\Omega} |\Delta g_{k+1}^{n+1}|^2 dx \Big/ \int_{\Omega} |\Delta g_k^{n+1}|^2 dx, \quad (5.65)$$

and then

$$w_{k+1}^{n+1} = g_{k+1}^{n+1} + \gamma_k^{n+1} w_k^{n+1}. \quad (5.66)$$

Do  $k = k + 1$  and go to (5.61).  $\square$

From a practical point of view algorithm (5.58)–(5.66) is not particularly difficult to implement; indeed, after an appropriate space discretization, each iteration will require the solution of two discrete Poisson–Dirichlet problems in order to solve the discrete analogues of the biharmonic problems (5.59) and (5.61) (see ref. [18] for details concerning the solution of a closely related problem, namely problem (5.10) of Section 5.3). Concerning precisely the space approximation of problem (LSQ2), we have used a mixed finite element discretization closely related to the one briefly discussed in Section 5.3, which is itself directly inspired from those approximations employed in [31], [24], [32] for the numerical simulation of two-dimensional *Bingham visco-plastic flow* using the *stream function formulation*. With this approach  $\varphi$ ,  $\mathbf{q}$ ,  $\psi$ ,  $\mathbf{p}$  are approximated by continuous piecewise linear approximations associated to a finite element triangulation of  $\Omega$ . The condition  $\det \mathbf{q} = f$  is imposed at the vertices of this triangulation.

REMARK 5.5. Algorithm (5.58)–(5.66) is clearly of the *relaxation type*. Actually, when  $\tau \rightarrow +\infty$ , we recover at the limit an algorithm very close to the *nonlinear block Gauss–Seidel* one discussed in, e.g., [32], [27].

### 5.5.3. Numerical experiments

The least-squares method discussed in Sections 5.5.1 and 5.5.2 has been applied to the solution of three type (5.1) problems, with  $\Omega = (0, 1)^2$ ; these test problems have been already encountered in Section 5.4. The first test problem can be formulated as follows (with  $|x| = (x_1^2 + x_2^2)^{\frac{1}{2}}$  and  $R \geq \sqrt{2}$ ):

$$\det D^2 \psi = \frac{R^2}{(R^2 - |x|^2)^2} \quad \text{in } \Omega, \quad \psi = (R^2 - |x|^2)^{\frac{1}{2}} \quad \text{on } \partial\Omega. \quad (5.67)$$

The function  $\psi$  defined by  $\psi(x) = (R^2 - |x|^2)^{\frac{1}{2}}$  is a solution of problem (5.67); as already mentioned (in Section 5.4) its graph is a piece of the *sphere* of center  $\mathbf{0}$  and radius  $R$ . We have discretized problem (5.67) relying, as in Section 5.4, on a mixed variational formulation associated to uniform triangulations of  $\Omega$  (like the one shown on Fig. 5.3, but finer); the uniformity of the mesh allows us to solve the various elliptic problems encountered at each iteration of (5.58)–(5.66) by fast Poisson and Helmholtz solvers taking advantage of the decomposition properties of the discrete analogues of the biharmonic problems (5.59) and (5.61) (see [18] for details). The finite element analogue of algorithm (5.55)–(5.57) diverges if  $R = \sqrt{2}$  (which is not surprising since the corresponding  $\psi \notin H^2(\Omega)$ ); on the other hand, for  $R = 2$  we have a quite fast convergence as long as  $\tau$  is large enough, the corresponding results being reported on Table 5.4, (we stopped iterating as soon as  $\|D_h^2 \psi_h^n - \mathbf{p}_h^n\| \leq 10^{-6} \psi_h^n$  and  $\mathbf{p}_h^n$  being the computed approximations of  $\psi^n$  and  $\mathbf{p}^n$ , respectively).

Table 5.4. First test problem.

$h$	$\tau$	$n_{it}$	$\ D_h^2 \psi_h^c - \mathbf{p}_h^c\ _{\mathbf{Q}}$	$\ \psi_h^c - \psi\ _{L^2(\Omega)}$
1/32	0.1	517	$0.9813 \times 10^{-6}$	$0.450 \times 10^{-5}$
1/32	1	73	$0.9618 \times 10^{-6}$	$0.449 \times 10^{-5}$
1/32	10	28	$0.7045 \times 10^{-6}$	$0.450 \times 10^{-5}$
1/32	100	21	$0.6773 \times 10^{-6}$	$0.449 \times 10^{-5}$
1/32	1,000	22	$0.8508 \times 10^{-6}$	$0.449 \times 10^{-5}$
1/32	10,000	22	$0.8301 \times 10^{-6}$	$0.449 \times 10^{-5}$
1/64	1	76	$0.9624 \times 10^{-6}$	$0.113 \times 10^{-5}$
1/64	10	29	$0.8547 \times 10^{-6}$	$0.113 \times 10^{-5}$
1/64	100	24	$0.8094 \times 10^{-6}$	$0.113 \times 10^{-5}$

Above,  $\{\psi_h^c, \mathbf{p}_h^c\}$  is the computed approximate solution,  $h$  the space discretization step and  $n_{it}$  the number of iterations necessary to achieve convergence. Table 5.4 clearly suggests that: (i) For  $\tau$  large enough the speed of convergence is essentially independent of  $\tau$ . (ii) The speed of convergence is essentially independent of  $h$ . (iii) The  $L^2(\Omega)$ -approximation error is  $O(h^2)$ . By comparing the above results to those reported in [18] (and Section 5.4), concerning the solution of problem (5.67) by the augmented Lagrangian algorithm (5.8)–(5.11), we can add to (i)–(iii), above, that the new approach is easier to implement, is more robust, and provides the same approximate solutions, but faster (for  $\tau$  large enough); it avoids also the adjustment of parameter  $r$ , a delicate issue, particularly if one looks for an optimal value. Similarly, the new methodology is easier to implement and leads to faster algorithms than those derived from (LSQ1), another least-squares approach.

The second test problem is defined by

$$\det D^2 \psi = \frac{1}{|x|} \quad \text{in } \Omega, \quad \psi = \frac{(2|x|)^{\frac{3}{2}}}{3} \quad \text{on } \partial\Omega. \quad (5.68)$$

With these data, the function  $\psi$  defined by  $\psi(x) = \frac{(2|x|)^{\frac{3}{2}}}{3}$  is solution of (5.68). As previously mentioned,  $\psi \in W^{2,p}(\Omega)$ ,  $\forall p \in [1, 4)$ , but does not have the  $C^2(\bar{\Omega})$ -regularity. Using the same approximation and algorithm as the first test problem, we obtain then the results reported in Table 5.5.

Table 5.5. Second test problem.

$h$	$\tau$	$n_{it}$	$\ D_h^2 \psi_h^c - \mathbf{p}_h^c\ _{\mathbf{Q}}$	$\ \psi_h^c - \psi\ _{L^2(\Omega)}$
1/32	1	145	$0.9381 \times 10^{-6}$	$0.556 \times 10^{-4}$
1/32	10	56	$0.9290 \times 10^{-6}$	$0.556 \times 10^{-4}$
1/32	100	46	$0.9285 \times 10^{-6}$	$0.556 \times 10^{-4}$
1/32	1,000	45	$0.9405 \times 10^{-6}$	$0.556 \times 10^{-4}$
1/64	1	151	$0.9500 \times 10^{-6}$	$0.145 \times 10^{-4}$
1/64	10	58	$0.9974 \times 10^{-6}$	$0.145 \times 10^{-4}$
1/64	100	49	$0.9531 \times 10^{-6}$	$0.145 \times 10^{-4}$
1/64	1,000	48	$0.9884 \times 10^{-6}$	$0.145 \times 10^{-4}$

The various comments we have done concerning the solution of the first test problem still apply here.

The third test problem, namely

$$\det D^2 \psi = 1 \quad \text{in } \Omega, \quad \psi = 0 \quad \text{on } \partial\Omega, \quad (5.69)$$

has no solution in  $H^2(\Omega)$ , despite the smoothness of the data. Its *augmented Lagrangian* solution was addressed in Section 5.4, and, indeed, one of our goals here is to use algorithm (5.55)–(5.57) to check the convergence properties of the augmented Lagrangian algorithm (5.8)–(5.11), discussed in Section 5.3, when problem (5.1) has no solution in  $H^2(\Omega)$ , despite the fact that neither  $V_g$  nor  $\mathbf{Q}_f$  are empty (which is clearly the case for problem (5.69)). In Sections 5.3 and 5.4 we conjectured that if the above situation prevails, then the sequence  $\{\{\psi^n, \mathbf{p}^n\}\}_n$  generated by algorithm (5.8)–(5.11) converges to a pair  $\{\psi, \mathbf{p}\}$  minimizing precisely the functional  $\{\varphi, \mathbf{q}\} \rightarrow \int_{\Omega} |D^2 \varphi - \mathbf{q}|^2 dx$  over the set  $V_g \times \mathbf{Q}_f$ . We have reported in Table 5.6 the results produced by algorithm (5.55)–(5.57) (indeed a discrete analogue of it) using  $\|\psi_h^{n+1} - \psi_h^n\|_{L^2(\Omega)} \leq 10^{-7}$  as the stopping criterion.

It is clear from Table 5.6 that the convergence is clearly slower than for the two first test problems, however some important features remain such as: the number of iterations necessary to achieve convergence is essentially independent of  $\tau$ , as long as this last parameter is large enough, and increases slowly with  $h$  (actually like  $\sqrt{h}$ ). Most importantly (from a conceptual point of view), the solutions computed via formulation (LSQ2) and algorithm (5.55)–(5.57) coincide, essentially, with those obtained via the augmented Lagrangian algorithm (5.8)–(5.11); this is a result we were looking for, in order to clarify the convergence properties of algorithm (5.8)–(5.11) when problem (5.1) has no solution in  $H^2(\Omega)$  while  $V_g$  and  $\mathbf{Q}_f$  are both nonempty.

Table 5.6. Third test problem.

$h$	$\tau$	$n_{it}$	$\ D_h^2 \psi_h^c - \mathbf{P}_h^c\ _{\mathbf{Q}}$
1/32	1	4,977	$0.1054 \times 10^{-1}$
1/32	100	3,297	$0.4980 \times 10^{-2}$
1/32	1,000	3,275	$0.4904 \times 10^{-2}$
1/32	10,000	3,273	$0.4896 \times 10^{-2}$
1/64	1	6,575	$0.1993 \times 10^{-1}$
1/64	100	4,555	$0.1321 \times 10^{-1}$
1/64	1,000	4,527	$0.1312 \times 10^{-1}$
1/128	100	5,402	$0.1841 \times 10^{-1}$
1/128	1,000	5,372	$0.1830 \times 10^{-1}$

REMARK 5.6. An evidence that the augmented Lagrangian and least squares approaches produce, essentially, the same results for the third test problem is the fact that  $\|\psi_h^{LS} - \psi_h^{AL}\|_{L^2(\Omega)}$  is of the order of  $10^{-5}$  (the superscript *LS* (resp., *AL*) being associated to the least-squares (resp., augmented Lagrangian) solution).

*Acknowledgments.* The authors would like to thank J.D. Benamou, Y. Brenier, A. Caboussat, L.A. Caffarelli, B. Dacorogna, J. Cagnol, G.P. Galdi, D.D. Joseph, D.B. Kothe, P. Le Tallec, P.L. Lions, P. Muscarello, M. Padula, J. Rappaz, T.E. Tezduyar, and J.P. Zolezio for assistance and helpful comments and suggestions. This work was partially supported by the NSF (grants ECS-9527123, CTS-9873236, DMS-9902035, DMS-0209066, DMS-0412267, DMS-0443549), and DOE/LACSI (grant R71700K-292-000-99).

## References

- [1] G. Allain, Small-time existence for the Navier–Stokes equations with a free surface. *Appl. Math. Optim.*, **16** (1987), 37–50.
- [2] E. Bänsch, Finite element discretization of the Navier–Stokes equations with a free capillary surface. *Numer. Math.*, **88** (2001), 203–235.
- [3] E. Bänsch and B. Höhn, Numerical simulation of a silicon floating zone with a free capillary surface. *Scientific Computing in Chemical Engineering II*, Vol. 1, F. Keli, W. Mackens, H. Voss and J. Werther eds., 1999, 328–335.
- [4] J.T. Beale, The initial value problem for the Navier–Stokes equations with a free-surface. *Comm. Pure Appl. Math.*, **34** (1981), 359–392.
- [5] M. Bercovier and O. Pironneau, Error estimates for finite element method solution of the Stokes problem in the primitive variables. *Numer. Math.*, **33** (1979), 211–224.
- [6] J.U. Brackbill, D.B. Kothe and C. Zemach, A continuum method for modeling surface tension. *J. Comput. Phys.*, **100** (1992), 335–354.
- [7] A. Caboussat, Analysis and Numerical Simulation of Free Surface Flows. Ph.D. Dissertation, Ecole Polytechnique Fédérale de Lausanne, Department of Mathematics, Lausanne, Switzerland, 2003.
- [8] A. Caboussat, A numerical method for the simulation of free surface flows with surface tension. *Comp. and Fluids*, **35** (2006), 1205–1216.
- [9] L.A. Caffarelli and X. Cabré, Fully Nonlinear Elliptic Equations. American Mathematical Society, Providence, RI, 1995.
- [10] L.A. Caffarelli and M. Milman, eds., Monge–Ampère Equation: Application to Geometry and Optimization. American Math. Society, Providence, RI, 1999.

- [11] T.F. Chan and R. Glowinski, Numerical methods for a class of mildly nonlinear elliptic equations. *Atas do Decimo Primeiro Coloquio Brasileiro do Matematicas*, Vol. I, C.N.D.T./IMPA, Rio do Janeiro, 1978, 279–318.
- [12] A.J. Chorin, Numerical study of a slightly viscous flow. *J. Fluid Mech.*, **57** (1973), 785–796.
- [13] A.J. Chorin, T.J.R. Hughes, M.F. McCracken and J.E. Marsden, Product formulas and numerical algorithms. *Comm. Pure Appl. Maths.*, **31** (1978), 205–256.
- [14] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II. Wiley Interscience, New York, NY, 1989.
- [15] C. Cuvelier and R.M. Schulkes, Some numerical methods for the computation of capillary free boundaries governed by the Navier–Stokes equations. *SIAM Review*, **32** (1990), 355–423.
- [16] E.J. Dean and R. Glowinski, Numerical solution of the two-dimensional elliptic Monge–Ampère equation with Dirichlet boundary conditions: an augmented Lagrangian approach. *C.R. Acad. Sci. Paris, Sér. I*, **336** (2003), 779–784.
- [17] E.J. Dean and R. Glowinski, Numerical solution of the two-dimensional elliptic Monge–Ampère equation with Dirichlet boundary conditions: a least squares approach. *C.R. Acad. Sci. Paris, Sér. I*, **339** (2004), 887–892.
- [18] E.J. Dean and R. Glowinski, Numerical methods for fully nonlinear elliptic equations of the Monge–Ampère type. *Comp. Meth. Appl. Mech. Engin.*, **195** (2006), 1344–1386.
- [19] E.J. Dean and R. Glowinski, A wave equation approach to the numerical solution of the Navier–Stokes equations for incompressible viscous flow. *C.R. Acad. Sci. Paris, Ser. I*, **325** (1997), 789–797.
- [20] E.J. Dean, R. Glowinski and T.W. Pan, A wave equation approach to the numerical simulation of incompressible viscous flow modeled by the Navier–Stokes equations. *Mathematical and Numerical Aspects of Wave Propagation*, J.A. De Santo ed., SIAM, Philadelphia, PA, 1998, 65–74.
- [21] B. Desjardin and M.J. Esteban, On weak solution for fluid-rigid structure interaction: compressible and incompressible models. *Arch. Rational Mech. Anal.*, **146** (1999), 59–71.
- [22] J. Dieudonné, *Panorama des Mathématiques Pures: Le Choix Bourbachique*. Editions Jacques Gabay, Paris, 2003.
- [23] J. Douglas and H.H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables. *Trans. Am. Math. Soc.*, **82** (1956), 421–439.
- [24] M. Fortin and R. Glowinski, *Augmented Lagrangians Methods: Application to the Numerical Solution of Boundary-Value Problems*. North-Holland, Amsterdam, 1983.
- [25] F. Foss, On the exact point-wise interior controllability of the scalar wave equation and solution of nonlinear elliptic eigenproblems. PhD dissertation, Department of Mathematics, University of Houston, Houston, Texas, 2006.
- [26] V. Girault and P.A. Raviart, *Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms*. Springer-Verlag, Berlin, 1986.
- [27] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*. Springer-Verlag, New York, NY, 1984.
- [28] R. Glowinski, *Finite Element Methods for Incompressible Viscous Flow*. *Handbook of Numerical Analysis*, Vol. IX, P.G. Ciarlet and J.L. Lions eds., North-Holland, Amsterdam, 2003, 3–1176.
- [29] R. Glowinski and G. Guidoboni, Hopf bifurcation in viscous incompressible flow down an inclined plane: a numerical approach. *J. Math. Fluid Mech.*, **9** (2007), 1–21.
- [30] R. Glowinski and L.H. Juárez, Finite element method and operator-splitting for a time-dependent viscous incompressible free-surface flow. *Comp. Fluid Dyn. J.*, **12** (2003), 459–468.
- [31] R. Glowinski and P. Le Tallec, *Augmented Lagrangians and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia, PA, 1989.
- [32] R. Glowinski, J.L. Lions and R. Tremolières, *Numerical Analysis of Variational Inequalities*. North-Holland, Amsterdam, 1981.
- [33] R. Glowinski, T.W. Pan, T.I. Hesla, D.D. Joseph and J. Périaux, A fictitious domain approach to the direct numerical simulation of incompressible viscous fluid flow past moving rigid bodies: application to particulate flow. *J. Comp. Phys.*, **169** (2001), 363–426.
- [34] C. Grandmont and Y. Maday, Existence for an unsteady fluid-structure interaction problem. *Math. Model. Num. Anal.*, **34** (2000), 609–636.

- [35] H.H. Hu., N.A. Patankar and M.Y. Zhu, Direct numerical simulation of fluid-solid systems using arbitrary Lagrangian–Eulerian techniques. *J. Comp. Phys.*, **169** (2001), 427–462.
- [36] A.A. Johnson and T.E. Tezduyar, Mesh update strategies in parallel finite element computations of flow problems with moving boundaries and interfaces. *Comp. Meth. Appl. Mech. Eng.*, **119** (1994), 73–94.
- [37] A.A. Johnson and T. Tezduyar, 3-D simulations of fluid-particle interactions with the number of particles reaching 100. *Comp. Methods Appl. Mech. Engrg.*, **145** (1997), 301–321.
- [38] D.D. Joseph and L. Preziosi, Stability of rigid motions and coating films in bi-component flows of immiscible liquids. *J. Fluid Mech.*, **185** (1987), 323–351.
- [39] L.H. Juárez, P. Saavedra and M. Salazar, Computational study of a free-boundary model. *Advances in Optimization and Numerical Analysis*, S. Gomez and J.P. Hennart eds., Kluwer, Dordrecht, 1994, 245–260.
- [40] J.B. Keller and M.J. Miksis, Surface tension driven flows. *SIAM J. Appl. Math.*, **43** (1983), 268–277.
- [41] S.F. Kistler and L.E. Scriven, Coating flow theory by finite element methods and asymptotic analysis of the Navier–Stokes system. *Inter. J. Numer. Meth. Fluids*, **4** (1984), 207–229.
- [42] P. Le Tallec, Numerical methods for nonlinear three-dimensional elasticity. *Handbook of Numerical Analysis*, Vol. III, P.G. Ciarlet and J.L. Lions eds., North-Holland, Amsterdam, 1994, 465–622.
- [43] G.I. Marchuk, Splitting and alternating direction methods. *Handbook of Numerical Analysis*, Vol. I, P.G. Ciarlet and J.L. Lions eds., North-Holland, Amsterdam, 1990, 197–462.
- [44] M. Marion and R. Temam, Navier–Stokes equations. *Handbook of Numerical Analysis*, Vol. VI, P.G. Ciarlet and J.L. Lions eds., North-Holland, Amsterdam, 1998, 503–689.
- [45] V. Maronnier, M. Picasso and J. Rappaz, Numerical simulation of free-surface flows. *J. Comp. Phys.*, **155** (1999), 439–455.
- [46] V. Maronnier, M. Picasso and J. Rappaz, Numerical simulation of three dimensional free-surface flows. *Int. J. Numer. Meth. Fluids*, **42** (2003), 697–716.
- [47] B. Maury, A characteristics-ALE method for the unsteady 3-D Navier–Stokes equations with a free-surface. *Int. J. Comp. Fluids Dynamics*, **6** (1996), 175–188.
- [48] B. Maury, Direct simulation of 2-D fluid-particle flows in bi-periodic domains, *J. Comp. Phys.*, **156** (1999), 325–351.
- [49] G. Müller and A. Ostrogorsky, Convection in melt growth. *Handbook of Crystal Growth 2B*, D.T. Hurle ed., North-Holland, 1994, 709–819.
- [50] J.A. Nietsche, Free-boundary problems for Stokes flows and finite element methods. *Ecuadiff 6, Lecture Notes in Math.*, **1192**, Springer-Verlag, Berlin, 1986, 327–332.
- [51] T. Nishida, Y. Teramoto and H. Yoshihara, Hopf bifurcation in viscous incompressible flow down an inclined plane. *J. Math. Fluid Mech.*, **7** (2005), 29–71.
- [52] J.R. Ockendon, S. Howison, A. Lacey and A. Movchan, *Applied Partial Differential Equations*. Oxford University Press, Oxford, UK, 1999.
- [53] M. Padula and V.A. Solonikov, On Rayleigh–Taylor stability. *Ann. Univ. Ferrara, Sez. VII, Sc. Mat.*, XLXI, 2000, 307–336.
- [54] T.W. Pan and R. Glowinski, A projection/wave-like equation method for the numerical simulation of incompressible viscous fluid flow modeled by the Navier–Stokes equations. *Comput. Fluid Dyn. J.*, **9** (2000), 28–42.
- [55] T.W. Pan and R. Glowinski, Direct simulation of the motion of neutrally buoyant circular cylinders in plane Poiseuille flow. *J. Comp. Phys.*, **181** (2002), 260–279.
- [56] D.H. Peaceman and H.H. Rachford, The numerical solution of parabolic and elliptic differential equations. *J. Soc. Ind. Appl. Math.*, **3** (1955), 28–41.
- [57] O. Pironneau, *Finite Element Methods for Fluids*. J. Wiley, Chichester, U.K., 1989.
- [58] C. Pozrikidis, The flow of a liquid film along a periodic wall. *J. Fluid Mech.*, **188** (1988), 275–300.
- [59] W.G. Pritchard, Instability and chaotic behavior in a free-surface flow. *J. Fluid Mech.*, **165** (1986), 1–60.
- [60] W.G. Pritchard, L.R. Scott and S.J. Tavener, Numerical and asymptotic methods for certain viscous free-surface flows. *Phil. Trans. Royal Soc. London, A*, **340** (1992), 1–45.
- [61] V.V. Pukhnachëv, Hydrodynamic free-boundary problems. *Nonlinear Partial Differential and their Applications*, Collège de France Seminar, Paris, Vol. III, Pitman, Boston, 1982, 301–308.

- [62] P. Saavedra and L.R. Scott, Variational formulation of a model free-boundary problem. *Math. Comp.*, **57** (1991), 451–475.
- [63] H. Saito and L.E. Scriven, Study of coating flow by the finite element method. *J. Comp. Phys.*, **42** (1981), 53–76.
- [64] J.A. San Martin, V. Starovoitov and M. Tucsnak, Global weak solutions for the two-dimensional motion of several rigid bodies in an incompressible viscous fluid. *Arch. Ration. Mech. Anal.*, **161** (2002), 113–147.
- [65] D. Schwabe, Surface-tension driven flow in crystal growth melts. *Crystal Growth Properties and Applications 11*, Springer-Verlag, Berlin, 1988.
- [66] H. Sigurgeirson, A.M. Stuart and J. Wan, Collision detection for particles in flow. *J. Comp. Phys.*, **172** (2001), 766–807.
- [67] V.A. Solonikov, On the Stokes equations in domains with non-smooth boundaries and on viscous incompressible flow with a free surface. *Nonlinear Partial Differential and their Applications*, Collège de France Seminar, Paris, Vol. III, Pitman, Boston, 1982, 340–423.
- [68] G. Strang, On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, **5** (1968), 506–517.
- [69] T.E. Tezduyar, Stabilized finite element formulations for incompressible flow computations. *Adv. Appl. Mech.*, **28** (1992), 1–44.
- [70] T.E. Tezduyar, M. Behr, S. Mittal and J. Liou, A new strategy for finite element computations involving moving boundaries and interfaces—The deforming-spatial-domain/space-time procedure: computation of free-surface flows, two-liquid flows, and flows with drifting cylinders. *Comp. Meth. Appl. Mech. Eng.*, **94** (1992), 353–371.
- [71] S. Turek, A comparative study of time-stepping techniques for the incompressible Navier–Stokes equations: from fully implicit nonlinear schemes to semi-implicit projection methods. *Int. J. Numer. Meth. Fluids*, **22** (1996), 987–1011.
- [72] M.W. Williams, D.B. Kothe and E.G. Puckett, Accuracy and convergence of continuum surface tension models. *Fluid Dynamics at Interface*, W. Shyy and R. Narayanan, eds., Cambridge University Press, 1999, 294–305.

