Some Facts about Operator-Splitting and Alternating Direction Methods

Roland Glowinski, Tsorng–Whay Pan and Xue–Cheng Tai.

Abstract The main goal of this chapter is to give the reader a (relatively) brief overview of operator-splitting, augmented Lagrangian and ADMM methods and algorithms. Following a general introduction to these methods, we will give several applications in Computational Fluid Dynamics, Computational Physics, and Imaging. These applications will show the flexibility, modularity, robustness and versatility of these methods. Some of these applications will be illustrated by the results of numerical experiments; they will confirm the capabilities of operator-splitting methods concerning the solution of problems still considered complicated by today standards.

1 Introduction

In 2004, the first author of this chapter was awarded the SIAM Von Kármán Prize for his various contributions to Computational Fluid Dynamics, the direct numerical simulation of particulate flow in particular. Consequently, he was asked by some people at SIAM to contribute an article to SIAM Review, related to the Von Kármán lecture he gave at the 2004 SIAM meeting in Portland, Oregon. Since operator-splitting was playing a most crucial role in the results presented during his Portland lecture, he decided to write, jointly with several collaborators (including the second author), a review article on operator-splitting methods, illustrated by several
selected applications. One of the main reasons for that review article was that, to the best of our knowledge at the time, the last comprehensive publication on the subject was [121], a book-size article (266 pages) published in 1990, in the Volume I of the Handbook of Numerical Analysis. Our article was rejected, on the grounds that it was untimely. What is ironical is that the very day (of August 2005) we received the rejection e-mail message, we were having a meeting with computational scientists at Los Alamos National Laboratory (LANL) telling us that one of their main priorities was further investigating the various properties of operator-splitting methods, considering that these methods were (and still are) applied at LANL to solve a large variety of challenging, mostly multi-physics, problems. Another event emphasizing the importance of operator-splitting methods was the December 2005 conference, at Rice University in Houston, commemorating “50 Years of Alternating-Direction Methods” and honoring J. Douglas, D. Peaceman and H. Rachford, the inventors of those particular operator-splitting methods bearing their name. Actually, it was striking to observe during this conference that, at the time, most members of the Partial Differential Equations and Optimization communities were ignoring that most alternating-direction methods for initial value-problems are closely related to primal-dual algorithms such as ADMM (Alternating Direction Methods of Multipliers). In order to create a bridge between these two communities, we updated the failed SIAM Review paper and submitted it elsewhere, leading to [73] (clearly, a publication in a SIAM journal would have had more impact, worldwide). Our goal in this chapter is to present a (kind of) updated variant of [73], less CFD (resp., more ADMM) oriented. It will contain in particular applications to Imaging, a topic barely mentioned in reference [73]. The content of this chapter is as follows:

In Section 2, we will discuss the numerical solution of initial value problems by operator-splitting time-discretization schemes such as Peaceman-Rachford’s, Douglas-Rachford’s, Lie’s, Strang’s, Marchuk-Yanenko’s, and by the fractional $\theta$-scheme, a three-stage variation, introduced in [67] and [68], of Peaceman-Rachford’s scheme. We will conclude this section by some remarks on the parallelization of operator-splitting schemes.

Section 3 will be dedicated to augmented Lagrangian and ADMM algorithms. We will show in particular that some augmented Lagrangian and ADMM algorithms are nothing but disguised operator-splitting methods (justifying thus the ADMM terminology).

Following [73], we will discuss in Section 4 the operator-splitting based direct numerical simulation of particulate flow, in the particular case of mixtures of incompressible viscous fluids and rigid solid particles.

In Section 5, we will discuss the application of operator-splitting methods to the solution of two problems from Physics, namely the Gross-Pitaevskii equation, a nonlinear Schrödinger equation modelling Bose-Einstein condensates, and the Zakharov system, a model for the propagation of Langmuir waves in ionized plasma.

Next, in Section 6, we will discuss applications of augmented Lagrangian and ADMM algorithms to the solution of problems from Imaging, a highly popular topic nowadays (actually, the renewed interest in ADMM type algorithms that we observe...
Operator-Splitting and Alternating Direction Methods

Currently can be largely explained by their application to Image Processing; see [156, 170]).

Finally, in Section 7, we will return to various issues that we left behind in the preceding sections of this chapter: these include augmentation parameter selection, an analysis of the asymptotic behavior of the Peaceman-Rachford and Douglas-Rachford schemes, and various comments concerning high order accurate operator-splitting schemes. Also, owing to the fact that one of the success stories of operator-splitting methods has been the numerical solution of the Navier-Stokes equations modeling viscous flow, we will conclude this section (and the chapter) by providing a (non-exhaustive) list of related references.

In addition to all the other chapters of this volume, material related to operator-splitting, augmented Lagrangian and ADMM algorithms can be found in [72] (see also the references therein). More references will be given in the following sections.

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

2.1 Generalities

Let us consider the following autonomous initial value problem:

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

(1)

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

\[A = \sum_{j=1}^{J} A_j,\]

(2)

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:

\textit{Can we take advantage of decomposition (2) for the solution of (1)?}

It has been known for many years (see for example [36]) that the answer to the above question is definitely yes.

Many schemes have been designed to take advantage of the decomposition (2) when solving (1); several of them will be briefly discussed in the following paragraphs.
2.2 Time-discretization of (1) by Lie’s scheme

Let \( \triangle t (>0) \) be a time-discretization step (for simplicity, we suppose \( \triangle t \) fixed); we denote \( n\triangle t \) by \( t^n \). With \( \phi^n \) denoting an approximation of \( \phi(t^n) \), Lie’s scheme reads as follows (for its derivation see, e.g., [70] (Chapter 6) and Chapter 1, Section 2, of this book):

\[
\phi^0 = \phi_0; \quad (3)
\]

then, for \( n \geq 0 \), \( \phi^n \to \phi^{n+1} \) via

\[
\begin{align*}
\frac{d\phi_j}{dt} + A_j(\phi_j) &= 0 \text{ on } (t^n, t^{n+1}), \\
\phi_j(t^n) &= \phi^{n+(j-1)/J}, \quad \phi_j(t^{n+1}) = \phi_j(t^{n+1}),
\end{align*}
\]

(4)

for \( j = 1, \ldots, J \).

If (1) is taking place in a finite dimensional space and if the operators \( A_j \) are smooth enough, then \( \| \phi(t^n) - \phi^n \| = O(\triangle t) \), function \( \phi \) being the solution of (1).

**Remark 1.** The above scheme applies also for multivalued operators (such as the subdifferentials of proper lower semi-continuous convex functionals), but in such a case first order accuracy is not guaranteed anymore. A related application will be given in Section 2.7.

**Remark 2.** The above scheme is easy to generalize to non-autonomous problems by observing that

\[
\begin{align*}
\frac{d\phi}{dt} + A(\phi, t) &= 0, \\
\phi(0) &= \phi_0
\end{align*}
\]

\[ \Leftrightarrow \]

\[
\begin{align*}
\frac{d\phi}{dt} + A(\phi, \theta) &= 0, \\
\frac{d\theta}{dt} - 1 &= 0, \\
\phi(0) &= \phi_0, \quad \theta(0) = 0.
\end{align*}
\]

**Remark 3.** Scheme (3)-(4) is semi-constructive in the sense that we still have to solve the initial value sub-problems in (4) 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:

\[
\phi^0 = \phi_0; \quad (5)
\]

then, for \( n \geq 0 \), \( \phi^n \to \phi^{n+1} \) via the solution of

\[
\frac{\phi^{n+j/J} - \phi^{n+(j-1)/J}}{\triangle t} + A_j(\phi^{n+j/J}) = 0, \quad (6)
\]

for \( j = 1, \ldots, J \).

Scheme (5)-(6) is known as the Marchuk-Yanenko scheme (see, e.g., refs. [121] and [70] (Chapter 6)) for more details). Several chapters of this volume are making use of the Marchuk-Yanenko scheme.
2.3 Time-discretization of (1) by Strang’s symmetrized scheme

In order to improve the accuracy of Lie’s scheme, G. Strang suggested a symmetrized variant of scheme (3)-4 (ref. [153]). When applied to non-autonomous problems, in the case where \( J = 2 \), we obtain (with \( t^{n+1/2} = (n + 1/2)\Delta t \)):

\[
\phi^0 = \phi_0;
\]

then, for \( n \geq 0 \), \( \phi^n \to \phi^{n+1/2} \to \tilde{\phi}^{n+1/2} \to \phi^{n+1} \) via

\[
\begin{align*}
\frac{d\phi_1}{dt} + A_1(\phi_1, t) &= 0 \text{ on } (t^n, t^{n+1/2}), \\
\phi_1(t^n) &= \phi^n; \tilde{\phi}^{n+1/2} = \phi_1(t^{n+1/2}), \\
\frac{d\phi_2}{dt} + A_2(\phi_2, t^{n+1/2}) &= 0 \text{ on } (0, \Delta t), \\
\phi_2(0) &= \phi^{n+1/2}; \tilde{\phi}^{n+1/2} = \phi_2(\Delta t), \\
\frac{d\phi_1}{dt} + A_1(\phi_1, t) &= 0 \text{ on } (t^{n+1/2}, t^{n+1}), \\
\phi_1(t^{n+1/2}) &= \tilde{\phi}^{n+1/2}; \phi^{n+1} = \phi_1(t^{n+1}).
\end{align*}
\] (7)

If (1) is taking place in a finite dimensional space and if operators \( A_1 \) and \( A_2 \) are smooth enough, then \( \|\phi(t^n) - \phi^n\| = O(\Delta t^2) \), function \( \phi \) being the solution of (1).

Remark 4. In order to preserve the second order accuracy of scheme (7)-(10) (assuming it takes place) we have to solve the initial value problems in (8), (9) and (10) by schemes which are themselves second order accurate (at least); these schemes are highly dependent of the properties of \( A_1 \) and \( A_2 \). The sub-problems (8), (9) and (10) are all particular cases of

\[
\begin{align*}
\frac{d\phi}{dt} + B(\phi, t) &= 0 \text{ on } (t_0, t_f), \\
\phi(t_0) &= \phi_0.
\end{align*}
\] (11)

Suppose now that \( B \) is a (positively) monotone operator; following [70] (Chapter 6), we advocate using for the numerical integration of (11) the second order implicit Runge-Kutta scheme below:

\[
\begin{align*}
\phi^0 &= \phi_0; \\
\text{for } q = 0, \ldots, Q - 1, \phi^q &\to \phi^{q+\theta} \to \phi^{q+1-\theta} \to \phi^{q+1} \text{ via}
\end{align*}
\]

\[
\begin{align*}
\frac{\phi^{q+\theta} - \phi^q}{\theta \tau} + B(\phi^{q+\theta}, t^{q+\theta}) &= 0, \\
\phi^{q+1-\theta} &= \frac{1 - \theta}{\theta} \phi^{q+\theta} + \frac{2\theta - 1}{\theta} \phi^q, \\
\frac{\phi^{q+1} - \phi^{q+1-\theta}}{\theta \tau} + B(\phi^{q+1}, t^{q+1}) &= 0,
\end{align*}
\] (12)
where in (12):

- \( Q(\geq 1) \) is an integer and \( \tau = \frac{t_f - t_0}{Q} \).
- \( \phi^{q+\alpha} \) is an approximation of \( \phi(t^{q+\alpha}) \), with \( t^{q+\alpha} = t_0 + (q + \alpha)\tau \).
- \( \theta = 1 - \frac{1}{\sqrt{2}} \).

It is shown in [70] (Chapter 2) that the implicit Runge-Kutta scheme (12) is stiff \( A \)-stable and "nearly" third-order accurate. It has been used, in particular, in [70] and [162] for the numerical simulation of incompressible viscous flow.

**Remark 5.** The main (if not the unique) drawback of Strang’s symmetrized scheme (7)-(10) concerns its ability at capturing the steady state solutions of (1) (when \( T = +\infty \)), assuming that such solutions do exist. Indeed, the splitting error associated with scheme (7)-(10) prevents using large values of \( \Delta t \) when integrating (1) from \( t = 0 \) to \( t = +\infty \); if the sequence \( \{\phi^n\}_{n \geq 0} \) converges to a limit, this limit is not, in general, a steady state solution of (1), albeit being close to one for small values of \( \Delta t \) (a similar comment applies also to the sequences \( \{\phi^{n+1/2}\}_{n \geq 0} \) and \( \{\hat{\phi}^{n+1/2}\}_{n \geq 0} \)). A simple way to partly-overcome this difficulty is to use variable time discretization steps: for example, in (8), (9) and (10), one can replace \( \Delta t \) by \( \tau_n \) (the sequence \( \{\tau_n\}_{n \geq 0} \) verifying \( \tau_n > 0 \), \( \lim_{n \to \infty} \tau_n = 0 \) and \( \sum_{n=0}^\infty \tau_n = +\infty \)), and then define \( t^{n+1} \) and \( t^{n+1/2} \) by \( t^{n+1} = t^n + \tau_n \forall n \geq 0, t^0 = 0, \) and \( t^{n+1/2} = t^n + \tau^n / 2 \), respectively. A more sophisticated way to fix the asymptotic behavior of scheme (7)-(10) is to proceed as in the chapter by McNamara and Strang in this book (Chapter 3).

**Remark 6.** More comments on scheme (7)-(10) can be found in, e.g., [70] (Chapter 6), [72] (Chapter 3) and various chapters of this volume, Chapter 3 in particular. Among these comments, the generalization of the above scheme to those situations where \( J \geq 3 \) in (2) has been discussed. Conceptually, the case \( J \geq 3 \) is no more complicated than \( J = 2 \). Focusing on \( J = 3 \), we can return (in a non-unique way) to the case \( J = 2 \) by observing that

\[
A = A_1 + A_2 + A_3 = A_1 + (A_2 + A_3) = (A_1 + A_2) + A_3
\]

\[
= (A_1 + \frac{1}{2}A_2) + (\frac{1}{2}A_2 + A_3).
\]

The first (resp., second and third) arrangement in (13) leads to 5 (resp., 7 and 9) initial value sub-problems per time step. Scheme (7)-(10), combined with the first arrangement in (13), has been applied in [81] to the computation of the periodic solution of a nonlinear integro-differential equation from Electrical Engineering.
Another candidate for the numerical solution of the initial value problem (1), or of its non-autonomous variant
\[
\frac{d\phi}{dt} + A(\phi, t) = 0 \text{ on } (0, T), \\
\phi(0) = \phi_0.
\] (14)
is provided, if \( J = 2 \) in (2), by the Peaceman-Rachford scheme (introduced in [139]). The idea behind the Peaceman-Rachford scheme is quite simple: the notation being like in Sections 2.1, 2.2 and 2.3, one divides the time interval \([t^n, t^{n+1}]\) into two sub-intervals of length \(\frac{\Delta t}{2}\) using the mid-point \(t^{n+1/2}\). Then assuming that the approximate solution \(\phi^n\) is known at \(t^n\) one computes first \(\phi^{n+1/2}\) using over \([t^n, t^{n+1/2}]\) a scheme of the backward Euler type with respect to \(A_1\) and of the forward Euler type with respect to \(A_2\); one proceeds similarly over \([t^{n+1/2}, t^{n+1}]\), switching the roles of \(A_1\) and \(A_2\). The following scheme, due to Peaceman and Rachford (see [139]), realizes precisely this program when applied to the solution of the initial value problem (14):
\[
\begin{align*}
\phi^0 &= \phi_0; \\
\text{for } n \geq 0, \phi^n &\rightarrow \phi^{n+1/2} \rightarrow \phi^{n+1} \text{ via the solution of} \\
\phi^{n+1/2} - \phi^n &= A_1(\phi^{n+1/2}, t^{n+1/2}) + A_2(\phi^n, t^n) = 0, \\
\phi^{n+1} - \phi^{n+1/2} &= A_1(\phi^{n+1/2}, t^{n+1/2}) + A_2(\phi^{n+1}, t^{n+1}) = 0.
\end{align*}
\] (15)
The convergence of the Peaceman-Rachford scheme (15) has been proved in [118] and [84] under quite general monotonicity assumptions concerning the operators \(A_1\) and \(A_2\) (see also [64], [65] and [110]); indeed, \(A_1\) and/or \(A_2\) can be nonlinear, unbounded and even multi-valued. In general, scheme (15) is first order accurate at best; however, if the operators \(A_1\) and \(A_2\) are linear, time independent and commute then scheme (15) is second order accurate (that is \(\|\phi^n - \phi(t^n)\| = O(\|\Delta t^2\|)\), \(\phi\) being the solution of problem (1)). Further properties of scheme (15) can be found in, e.g., [121], [70] (Chapter 2) and [72] (Chapter 3), including its stability, and its asymptotic behavior if \(T = +\infty\); concerning this last issue, a sensible advice is to use another scheme to compute steady state solutions, scheme (15) not being stiff A-stable.

Remark 7. Scheme (15) belongs to the alternating direction method family. The reason of that terminology is well-known: one of the very first applications of scheme (15) was the numerical solution of the heat equation
\[
\frac{\partial \phi}{\partial t} - \frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \phi}{\partial y^2} = f,
\]
completed by initial and boundary conditions. After finite difference discretization, the roles of $A_1$ and $A_2$ were played by the square matrices approximating the operators $-\frac{\partial^2}{\partial x^2}$ and $-\frac{\partial^2}{\partial y^2}$, respectively, explaining the terminology.

**Remark 8.** We observe that operators $A_1$ and $A_2$ play essentially symmetrical roles in scheme (15).

**Remark 9.** For those fairly common situations where operator $A_2$ is uni-valued, but operator $A_1$ is “nasty” (discontinuous and/or multi-valued, etc.), we should use the following equivalent formulation of the Peaceman-Rachford scheme (15):

\[
\begin{align*}
\phi_0 &= \phi_0; \\
\text{for } n \geq 0, \phi^n &\rightarrow \phi^{n+1/2} \rightarrow \phi^{n+1} \text{ via the solution of} \\
\frac{\phi^{n+1/2} - \phi^n}{\Delta t/2} + A_1(\phi^{n+1/2}, t^{n+1/2}) + A_2(\phi^n, t^n) &= 0, \\
\phi^{n+1} - 2\phi^{n+1/2} + \phi^n &= A_2(\phi^{n+1}, t^{n+1}) = A_2(\phi^n, t^n).
\end{align*}
\] (16)

**2.5 Time-discretization of (1) by Douglas-Rachford’s alternating direction method**

We assume that $J = 2$ in (2).

The Douglas-Rachford scheme (introduced in [57]) is a variant of the Peaceman-Rachford scheme (15); when applied to the numerical solution of the initial value problem (14) (the non-autonomous generalization of (1)), it takes the following form:

\[
\begin{align*}
\phi_0 &= \phi_0; \\
\text{for } n \geq 0, \phi^n &\rightarrow \hat{\phi}^{n+1} \rightarrow \phi^{n+1} \text{ via the solution of} \\
\frac{\hat{\phi}^{n+1} - \phi^n}{\Delta t} + A_1(\hat{\phi}^{n+1}, t^{n+1}) + A_2(\phi^n, t^n) &= 0, \\
\phi^{n+1} - 2\phi^{n+1} + \phi^n &= A_2(\phi^{n+1}, t^{n+1}) = A_2(\phi^n, t^n).
\end{align*}
\] (17)

The Douglas-Rachford scheme (17) has clearly a predictor-corrector flavor.

The convergence of the Douglas-Rachford scheme (17) has been proved in [118] and [84] under quite general monotonicity assumptions concerning the operators $A_1$ and $A_2$ (see also [64], [65] and [110]); indeed, $A_1$ and/or $A_2$ can be nonlinear, unbounded and even multi-valued. In general, scheme (17) is first order accurate at best (even if the operators $A_1$ and $A_2$ are linear, time independent and commute, assumptions implying second order accuracy for the Peaceman-Rachford scheme). Further properties of scheme (17) can be found in, e.g., [121], [70] (Chapter 2) and [72] (Chapter 3), including its stability, and its asymptotic behavior if $T = +\infty$. Concerning this last issue, a sensible advice is to use another scheme to compute
steady state solutions, scheme (17) not being stiff A-stable, a property it shares with the Peaceman-Rachford scheme (15).

Remark 10. Unlike the Peaceman-Rachford scheme (15), we observe that the roles played by operators $A_1$ and $A_2$ are non-symmetrical in scheme (17); actually, numerical experiments confirm that fact: for example, for the same $\Delta t$ the speed of convergence to a steady state solution may depend of the choice one makes for $A_1$ and $A_2$. As a rule of thumb, we advocate taking for $A_2$ the operator with the best continuity and monotonicity properties (see, for example, [62] (Chapter 3), [63] (Chapter 3) and [74] (Chapter 3) for more details).

Remark 11. Unlike scheme (15), scheme (17) is easy to generalize to operator decompositions involving more than two operators. Consider thus the numerical integration of (14) when $J \geq 3$ in (2). Following J. Douglas in [54] and [55] we generalize scheme (17) by

$$\phi^0 = \phi_0;$$

then for $n \geq 0$, $\phi^n$ being known, compute $\phi^{n+1/J}$, $\phi^{n+2/J}$, $\phi^{n+3/J}$ via the solution of

$$\phi^{n+1/J} - \phi^n = \frac{1}{\Delta t} \sum_{i=1}^{J-1} A_i (\phi^{n+1/J}, t^{n+1}) + \sum_{i=1}^{J-1} A_i (\phi^n, t^n)$$

(19.1)

and

$$\phi^{n+2/J} - \phi^n = \frac{1}{\Delta t} \sum_{i=1}^{J-1} A_i (\phi^{n+2/J}, t^{n+1}) + \sum_{i=1}^{J-1} A_i (\phi^n, t^n)$$

(19.2)

and

$$\phi^{n+3/J} - \phi^n = \frac{1}{\Delta t} \sum_{i=1}^{J-1} A_i (\phi^{n+3/J}, t^{n+1}) + \sum_{i=1}^{J-1} A_i (\phi^n, t^n)$$

(19.3)

Above, $\phi^{n+i/J}$ and $\phi^{n+j/J}$ denote approximate solutions at steps $i$ and $j$ of the computational process; they do not denote approximations of $\phi(t^{n+i/J})$ and $\phi(t^{n+j/J})$ (unless $i = j = J$).

Remark 12. This is the Douglas-Rachford analog of Remark 9: for those situations where $A_1$ is a “bad” operator (in the sense of Remark 9), we should use (assuming that $A_2$ is uni-valued) the following equivalent formulation of the Douglas-Rachford scheme (17):
\[ \phi^0 = \phi_0; \]
\[ \text{for } n \geq 0, \phi^n \rightarrow \tilde{\phi}^{n+1} \rightarrow \phi^{n+1} \text{ via the solution of} \]
\[ \frac{\phi^{n+1} - \phi^n}{\Delta t} + A_1(\tilde{\phi}^{n+1}, t^{n+1}) + A_2(\phi^n, t^n) = 0, \]
\[ \frac{\phi^{n+1} - \phi^n}{\Delta t} + A_2(\phi^{n+1}, t^{n+1}) = A_2(\phi^n, t^n). \]

(20)

**Remark 13.** To those wondering how to choose between the Peaceman-Rachford and Douglas-Rachford schemes, we will say that, on the basis of many numerical experiments, it seems that the second scheme is more robust and faster for those situations where one of the operators is non-smooth (multi-valued or singular, for example), particularly if one is interested by capturing steady state solutions. Actually, a better advice could be: consider using the fractional $\theta$-scheme to be discussed in Section 2.6, below. Indeed, we have encountered situations where this $\theta$-scheme outperforms both the Peaceman-Rachford and Douglas-Rachford schemes, for steady state computations in particular; such an example is provided by the anisotropic Eikonal equation, a nonlinear hyperbolic problem to be briefly discussed in Section 2.7. We will return to the Peaceman-Rachford vs Douglas-Rachford issue in Section 7.

### 2.6 Time-discretization of (1) by a fractional $\theta$-scheme

This scheme (introduced in [67], [68] for the solution of the Navier-Stokes equations) is a variant of the Peaceman-Rachford scheme (15). Let $\theta$ belong to the open interval $(0, 1/2)$ (in practice, $\theta \in [1/4, 1/3]$); the fractional $\theta$-scheme, applied to the solution of the initial value problem (14) (the non-autonomous generalization of (1)), reads as follows if $A = A_1 + A_2$:

\[ \phi^0 = \phi_0; \]
\[ \text{for } n \geq 0, \phi^n \rightarrow \phi^{n+\theta} \rightarrow \phi^{n+1-\theta} \rightarrow \phi^{n+1} \text{ via the solution of} \]
\[ \frac{\phi^{n+1-\theta} - \phi^{n+\theta}}{\theta \Delta t} + A_1(\phi^{n+\theta}, t^{n+\theta}) + A_2(\phi^n, t^n) = 0, \]
\[ \frac{(1-2\theta) \Delta t}{\theta \Delta t} + A_1(\phi^{n+\theta}, t^{n+\theta}) + A_2(\phi^{n+1-\theta}, t^{n+1-\theta}) = 0, \]
\[ \frac{\phi^{n+1} - \phi^{n+1-\theta}}{\theta \Delta t} + A_1(\phi^{n+1}, t^{n+1}) + A_2(\phi^{n+1-\theta}, t^{n+1-\theta}) = 0. \]

(21)

**Remark 14.** One should avoid confusion between scheme (21) and the following solution method for the initial value problem (14) (with $0 \leq \theta \leq 1$)
Remark 17. The fractional operator leading to the sub-problems which are the easiest to solve (that is, whose
if one takes
Remark 16. and
Runge-Kutta scheme (12), with
A
scheme (21). Since, at each time step, one has to solve two problems (resp., one problem) associated with operator
A
scheme (21) reduces to
forward Euler’s scheme (resp., backward Euler’s scheme, a Crank-Nicolson’s type scheme). Another “interesting” value is \( \theta = 2/3 \) (for reasons detailed in, e.g., [70] (Chapter 2) and [72] (Chapter 3)). By the way, it is to avoid confusion between schemes (21) and (22) that some practitioners (S. Turek, in particular) call the first one a fractional \( \theta \)-scheme. □

The stability and convergence properties of scheme (21) have been discussed in [70] (Chapter 2) and [72] (Chapter 3) for very simple finite dimensional situations where \( A_1 \) and \( A_2 \) are both positive multiples of the same symmetric positive definite matrix. Numerical experiments have shown that the good properties verified by scheme (21) for those simple linear situations, in particular its stiff A-stability for \( \theta \) well-chosen, still hold for more complicated problems, such as the numerical simulation of unsteady incompressible viscous flow modeled by the Navier-Stokes equations (as shown in, e.g., [23], [41], [69] and [70]).

Remark 15. We observe that operators \( A_1 \) and \( A_2 \) play non-symmetrical roles in scheme (21). Since, at each time step, one has to solve two problems (resp., one problem) associated with operator \( A_1 \) (resp., \( A_2 \)) a natural choice is to take for \( A_1 \) the operator leading to the sub-problems which are the easiest to solve (that is, whose solution is the least time consuming). Less naive criteria may be used to choose \( A_1 \) and \( A_2 \), such as the regularity (or lack of regularity) of these operators.

Remark 16. If one takes \( A_1 = A \) and \( A_2 = 0 \) in (21), the above scheme reduces to the Runge-Kutta scheme (12), with \( A \) replacing \( B \).

Remark 17. The fractional \( \theta \)-scheme (21) is a symmetrized scheme. From that point of view, it has some analogies with Strang’s symmetrized scheme (7)-(10), discussed in Section 2.3.

Remark 18. This is the fractional \( \theta \)-scheme analog of Remarks 9 and 12. For those situations where \( A_1 \) is a “bad” operator (in the sense of Remark 9), we advocate using the following equivalent formulation of the \( \theta \)-scheme (21):

\[
\begin{align*}
\phi^0 &= \phi_0; \\
\text{for } n \geq 0, &\quad \phi^n \rightarrow \phi^{n+\theta} \rightarrow \phi^{n+1-\theta} \rightarrow \phi^{n+1} \text{ via the solution of } \\
&\phi^{n+\theta} - \phi^n + \frac{\theta}{1 + \theta} A_1(\phi^{n+\theta}, t^{n+\theta}) + A_2(\phi^n, t^n) = 0, \\
&\theta(1 - \theta) \phi^{n+1-\theta} - (1 - \theta) \phi^{n+\theta} + (1 - 2\theta) \phi^n + \frac{\theta}{1 + \theta} A_2(\phi^{n+1-\theta}, t^{n+1-\theta}) = A_2(\phi^n, t^n), \\
&\phi^{n+1} - \phi^{n+1-\theta} + \frac{\theta}{1 + \theta} A_1(\phi^{n+1-\theta}, t^{n+1}) + A_2(\phi^{n+1-\theta}, t^{n+1}) = 0. 
\end{align*}
\]
2.7 Two applications: smallest eigenvalue computation and solution of an anisotropic Eikonal equation

2.7.1 Synopsis

It is not an exaggeration to say that applications of operator-splitting methods are everywhere, new ones occurring “almost” every day; indeed, some well-known methods and algorithms are disguised operator-splitting schemes as we will show in Section 2.7.2, concerning the computation of the smallest eigenvalue of a real symmetric matrix. In Section 2.7.3, we will apply the fractional $\theta$-scheme (21) to the solution of an Eikonal equation modeling wave propagation in anisotropic media. More applications will be discussed in Sections 4 and 5.

2.7.2 Application to some eigenvalue computation

Suppose that $A$ is a real $d \times d$ symmetric matrix. Ordering the eigenvalues of $A$ as follows: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$, our goal is to compute $\lambda_1$. We have (with obvious notation)

$$\lambda_1 = \min_{v \in S} v^tAv,$$

with $S = \{v | v \in \mathbb{R}^d, ||v|| = 1\}$, the norm in (24) being the canonical Euclidean one. The constrained minimization problem in (24) is equivalent to

$$\min_{v \in \mathbb{R}^d} \left[ \frac{1}{2} v^tAv + I_S(v) \right],$$

where, in (25), the functional $I_S : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined as follows

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

implying that $I_S$ is the indicator functional of the sphere $S$. Suppose that $u$ is a solution of (25) (that is a minimizer of the functional in (25)); we have then

$$Au + \partial I_S(u) \ni 0,$$

(26)

$\partial I_S(u)$ in (26) being a (kind of) generalized gradient of functional $I_S$ at $u$ ($\partial I_S$ is a multivalued operator). Next, we associate with the (necessary) optimality system (26) the following initial value problem (flow in the Dynamical System terminology):

$$\left\{ \begin{array}{l} \frac{du}{dt} + Au + \partial I_S(u) \ni 0 \text{ in } (0, +\infty), \\ u(0) = u_0. \end{array} \right.$$
If one applies the Marchuk-Yanenko scheme (5)-(6) to the solution of problem (27), one obtains (with $\tau = \triangle t$):

$$
\begin{cases}
    u^0 = u_0, \\
    \text{for } n \geq 0, \quad u^n \rightarrow u^{n+1/2} \rightarrow u^{n+1} \text{ via the solution of} \\
    \frac{u^{n+1/2} - u^n}{\tau} + Au^{n+1/2} = 0, \\
    \frac{u^{n+1} - u^{n+1/2}}{\tau} + \partial I_S(u^{n+1}) \ni 0.
\end{cases}
$$

The first finite difference equation in (28) implies

$$u^{n+1/2} = (I + \tau A)^{-1} u^n. \quad (29)$$

On the other hand, the second finite difference equation in (28) can be interpreted as a necessary optimality condition for the following minimization problem

$$\min_{v \in S} \left[ \frac{1}{2} \|v\|_2^2 - v^T u^{n+1/2} \right]. \quad (30)$$

Since $\|v\| = 1$ over $S$, the solution of problem (30) is given by

$$u^{n+1} = \frac{u^{n+1/2}}{\|u^{n+1/2}\|}. \quad (31)$$

It follows from (29) and (31) that algorithm (28) is nothing but the inverse power method with shift, a well-known algorithm from Numerical Linear Algebra. Indeed, if

$$0 < \tau < \frac{1}{\max(0, -\lambda_1)},$$

and if the projection of $u_0$ on the vector space spanned by the eigenvectors of $A$ associated with $\lambda_1$ is different from $0$, we can easily prove that the sequence $\{u^n\}_{n \geq 0}$ converges to an eigenvector of $A$ associated with $\lambda_1$ and also that

$$\lim_{n \to +\infty} (u^n)^T A u^n = \lambda_1.$$

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, Finance, etc., including some nonlinear eigenvalue problems, as shown in, e.g., [72] (Chapter 7).
2.7.3 Application to the solution of an anisotropic Eikonal equation from acoustics

The next application of operator-splitting, that we are going to (briefly) consider in this chapter, was brought to our attention recently (December 2014) by our colleagues S. Leung and J. Qian. It concerns the numerical solution of the following nonlinear hyperbolic partial differential equation

\[ |\nabla \tau| - \frac{|1 - V \cdot \nabla \tau|}{c} = 0 \text{ in } \Omega, \quad (32) \]

encountered in Acoustics and known as the anisotropic Eikonal equation. In (32), we have (see [40] for more details):

- \( \Omega \subset \mathbb{R}^d \), with \( d \geq 2 \).
- \( \tau(x) \) is the time of \( 1^{st} \) arrival of the wave front at \( x \in \Omega \).
- \( c > 0 \) is the wave propagation speed in the medium filling \( \Omega \), assuming that this medium is at rest (the so-called background medium).
- Assuming that the ambient medium is moving, \( V \) is its moving velocity; we assume that \( V \in (L^\infty(\Omega))^d \).

Fast-sweeping methods have been developed for the efficient numerical solution of the classical Eikonal equation

\[ |\nabla \tau| = \frac{1}{c} \text{ in } \Omega, \quad (33) \]

(see, e.g., [104] and [181]); these methods provide automatically viscosity solutions in the sense of Crandall and Lions (see [38] for this notion). Unfortunately, as shown in [40], the fast sweeping methods developed for the solution of (33) cannot handle (32), unless one modifies them significantly, as done in [40]. Actually, there exists an alternative, simpler to implement, to the method developed in [40]: it relies on the operator-splitting methods discussed in Sections 2.3, 2.4, 2.5 and 2.6, and takes advantage of the fact that the fast-sweeping methods developed for the solution of (33) can be easily modified in order to handle equations such as

\[ \alpha \tau - \beta \nabla^2 \tau + |\nabla \tau| = f \quad (34) \]

and

\[ \alpha \tau - \beta \nabla^2 \tau - \frac{|1 - V \cdot \nabla \tau|}{c} = f, \quad (35) \]

with \( \alpha > 0 \) and \( \beta \geq 0 \). Therefore, in order to solve problem (32), we associate with it the following initial value problem:

\[
\begin{cases}
(I - \varepsilon \nabla^2) \frac{\partial \tau}{\partial t} + |\nabla \tau| - \frac{|1 - V \cdot \nabla \tau|}{c} = 0 \text{ in } \Omega \times (0, +\infty), \\
\tau(0) = \tau_0,
\end{cases}
\quad (36)
\]
whose steady state solutions are also solutions of (32). In (36), $\varepsilon$ is a non-negative parameter (a regularizing one if $\varepsilon > 0$) and $\tau(t)$ denotes the function $t \rightarrow \tau(x,t)$. Actually, additional conditions are required to have solution uniqueness, typical ones being $\tau$ specified on a subset of $\Omega$ ($= \Omega \cup \partial \Omega$), possibly reduced to just one point (a point source for the wave). A typical choice for $\tau_0$ is the corresponding solution of problem (33).

The results reported in [75] show that, with $\theta = 1/3$, the fractional $\theta$-scheme discussed in Section 2.6 outperforms the Strang’s, Peaceman-Rachford’s and Douglas-Rachford’s schemes when applied to the computation of the steady state solutions of (36). The resulting algorithm reads as follows:

$$
\begin{cases}
\tau^0 = \tau_0; \\
\text{for } n \geq 0, \tau^n \rightarrow \tau^{n+\theta} \rightarrow \tau^{n+1-\theta} \rightarrow \tau^{n+1} \text{ via the solution of} \\
(I - \varepsilon \nabla^2) \frac{\tau^{n+\theta} - \tau^n}{\theta \Delta t} + |\nabla \tau^{n+\theta}| - \frac{c}{|1 - \mathbf{V} \cdot \nabla \tau^n|} = 0, \\
(I - \varepsilon \nabla^2) \frac{\tau^{n+1-\theta} - \tau^{n+\theta}}{(1 - 2\theta) \Delta t} + |\nabla \tau^{n+\theta}| - \frac{c}{|1 - \mathbf{V} \cdot \nabla \tau^{n+1-\theta}|} = 0, \\
(I - \varepsilon \nabla^2) \frac{\tau^{n+1} - \tau^{n+1-\theta}}{\theta \Delta t} + |\nabla \tau^{n+1}| - \frac{c}{|1 - \mathbf{V} \cdot \nabla \tau^{n+1-\theta}|} = 0.
\end{cases}
$$

The three problems in (37) being particular cases of (34) and (35), their finite difference analogues can be solved by fast-sweeping algorithms. Physical considerations suggest that $\Delta t$ has to be of the order of the space discretization $h$. Actually, the numerical results reported in [75] show that, unlike the other schemes discussed in Sections 2.2 to 2.5, scheme (37), with $\theta = 1/3$, has very good convergence properties, even for large values of the ratio $\Delta t / h$ (100, typically). If $\varepsilon = 0$ (resp., $h^2$), these numerical experiments suggest that the number of iterations (time steps) necessary to achieve convergence to a steady state solution, varies (roughly) like $h^{-1/2}$ (resp., $h^{-1/3}$), for two and three-dimensional test problems (see [75] for further results and more details). Clearly, preconditioning does pay here (a well-known fact, in general).

Remark 19. Some readers may wonder why the authors of [75] gave the role of $A_1$ (resp., $A_2$) to the operator $\tau \rightarrow |\nabla \tau|$ (resp., $\tau \rightarrow -\frac{1}{c} |1 - \mathbf{V} \cdot \nabla \tau|$), and not the other way around. Let us say to these readers that the main reason behind that choice was preliminary numerical experiments showing that, for the same values of $\alpha$ and $\beta$, problem (34) is cheaper to solve than problem (35).

### 2.8 Time-discretization of (1) by a parallel splitting scheme

The splitting schemes presented so far have a sequential nature, i.e. the subproblems associated with the decomposed operators are solved in a sequential man-
ner. Actually, it is also possible to solve the sub-problems in parallel, as shown just below, using the following variant of Marchuk-Yanenko’s scheme:

\[
\begin{align*}
\phi^0 &= \phi_0; \\
\text{for } n \geq 0, \text{ we obtain } \phi^{n+1} \text{ from } \phi^n \text{ by solving first} \\
& \quad \frac{\phi^{n+j/2J} - \phi^n}{J \Delta t} + A_j(\phi^{n+j/2J}, t^{n+1}) = 0, \text{ for } j = 1, \ldots, J, \\
\phi^{n+1} \text{ being then obtained by averaging as follows} \\
& \quad \phi^{n+1} = \frac{1}{J} \sum_{j=1}^{J} \phi^{n+j/2J}.
\end{align*}
\]

Scheme (38) is nothing but Algorithm 5.1 in [119]. Under suitable conditions, it has been proved in the above reference that scheme (38) is first order accurate, that is \( \| \phi^n - \phi(t^n) \| = O(\Delta t) \). A parallelizable algorithm with second order accuracy is presented also in [119]. The main advantage of the above schemes is that the sub-problems can be solved in parallel. Clearly, this parallel splitting idea can be used for computing the steady state solutions of (1). As observed in [155], the sub-problems (or at least some of them) can also be solved in parallel if the corresponding operator \( A_j \) has the right decomposition properties.

3 Augmented Lagrangian algorithms and Alternating Direction Methods of Multipliers

3.1 Introduction

It is our opinion that a review chapter like this one has to include some material about augmented Lagrangian algorithms, including of course their relationships with alternating direction methods. On the other hand, since augmented Lagrangian algorithms and alternating direction methods of multipliers, and their last known developments, are discussed, with many details, in other chapters of this book, we will not say much about these methods in this section. However, we will give enough information so that the reader may follow Section 6 (dedicated to Image Processing) without spending too much time consulting the other chapters (or other references).

In Section 3.2 we will introduce several augmented Lagrangian algorithms, and show in Section 3.3 how these algorithms relate to the alternating direction methods discussed in Sections 2.4 (Peaceman-Rachford’s) and 2.5 (Douglas-Rachford’s).

This section is largely inspired by Chapter 4 of [72].
3.2 Decomposition-coordination methods by augmented Lagrangians

3.2.1 Abstract problem formulation. Some examples

A large number of problems in Mathematics, Physics, Engineering, Economics, Data Processing, Imaging, etc. can be formulated as

$$u = \arg \min_{v \in V} [F(Bv) + G(v)], \quad (39)$$

where: (i) $V$ and $H$ are Banach spaces. (ii) $B \in L(V, H)$. (iii) $F : H \rightarrow \mathbb{R} \cup \{+\infty\}$ and $G : V \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, lower semi-continuous and convex functionals verifying $\text{dom}(F \circ B) \cap \text{dom}(G) \neq \emptyset$, implying that problem (39) may have solutions.

**Example 1.** This first example concerns the following variational problem:

$$u = \arg \min_{v \in H^1_0(\Omega)} \left[ \frac{\mu}{2} \int_\Omega |\nabla v|^2 \, dx + \tau \int_\Omega |\nabla v| \, dx - \varpi \int_\Omega v \, dx \right], \quad (40)$$

where: (i) $\Omega$ is a bounded domain (that is a bounded open connected subset) of $\mathbb{R}^2$; we denote by $\Gamma$ the boundary of $\Omega$. (ii) $dx = dx_1 \, dx_2$. (iii) $\mu$ and $\tau$, are two positive constants. (iv) $|\nabla v|^2 = \left| \frac{\partial v}{\partial x_1} \right|^2 + \left| \frac{\partial v}{\partial x_2} \right|^2$ (v) The space $H^1_0(\Omega)$ (a Sobolev space) is defined by

$$H^1_0(\Omega) = \{ v \in L^2(\Omega), \partial v / \partial x_i \in L^2(\Omega), \forall i = 1, 2, v|_\Gamma = 0 \}, \quad (41)$$

the two derivatives in (41) being in the sense of distributions (see, e.g., [148], [157] for this notion). Since $\Omega$ is bounded, $H^1_0(\Omega)$ is a Hilbert space for the inner product $\{ v, w \} \rightarrow \int_\Omega \nabla v \cdot \nabla w \, dx$, and the associated norm. Problem (40) is a well-known problem from non-Newtonian fluid mechanics; it models the flow of an incompressible visco-plastic fluid (of the Bingham type) in an infinitely long cylinder of cross-section $\Omega$, $\varpi$ being the pressure drop per unit length and $u$ the flow axial velocity. In (40), $\mu$ denotes the fluid viscosity and $\tau$, its plasticity yield (see, e.g., [59] and [83] for further information on visco-plastic fluid flows; see also the references therein). It follows from, e.g., [66] and [72], that the variational problem (40) has a unique solution.

Problem (40) is a particular case of (39) with $V = H^1_0(\Omega)$, $H = (L^2(\Omega))^2$, $B = \nabla$, $F(q) = \int_\Omega q \, dx$, and $G(v) = \frac{\mu}{2} \int_\Omega |\nabla v|^2 \, dx - \varpi \int_\Omega v \, dx$; other decompositions are possible.

Close variants of problem (40) are encountered in imaging, as shown in Section 6 (and other chapters of this volume).

**Example 2.** It concerns the following variant of problem (40):
\[ u = \arg \min_{v \in K} \left[ \frac{\mu}{2} \int_{\Omega} |\nabla v|^2 \, dx - C \int_{\Omega} v \, dx \right], \quad (42) \]

where \( \Omega \) is a bounded domain of \( \mathbb{R}^2 \), \( \mu \) is a positive constant and
\[
K = \{ v \in H^1_0(\Omega), |\nabla v| \leq 1 \text{ a.e. in } \Omega \}.
\]

It is a classical result (see, e.g., [59]) that (42) models, in an appropriate system of mechanical units, the torsion of an infinitely long cylinder of cross-section \( \Omega \), made of an elastic-plastic material, \( C \) being the torsion angle per unit length and \( u \) a stress potential. It follows from, e.g., [66] and [72], that the variational problem (42) has a unique solution.

Problem (42) is a particular case of problem (39) with
\[
W = \{(v,q) \in V \times H, Bv - q = 0\},
\]
\[
V = H^1_0(\Omega), \quad H = (L^2(\Omega))^2, \quad B = \nabla, \quad G(v) = \frac{\mu}{2} \int_{\Omega} |\nabla v|^2 \, dx - C \int_{\Omega} v \, dx, \quad F(q) = I_K(q), \quad I_K(\cdot) \text{ being the indicator functional of the closed convex non-empty subset } K \text{ of } H \text{ defined as}
\]
\[
I_K(q) = \begin{cases} 
0 & \text{if } q \in K, \\
+\infty & \text{otherwise},
\end{cases}
\]

implying, from the properties of \( K \), that \( I_K : H \to \mathbb{R} \cup \{+\infty\} \) is convex, proper and lower semi-continuous. \( \square \)

Other decompositions are possible.

**Remark 20.** We recall that, we have, (from the definition of indicator functionals)
\[
I_K(q) = \begin{cases} 
0 & \text{if } q \in K, \\
+\infty & \text{otherwise},
\end{cases}
\]

implying, from the properties of \( \mathcal{K} \), that \( I_K : H \to \mathbb{R} \cup \{+\infty\} \) is convex, proper and lower semi-continuous. \( \square \)

Numerical methods for the solution of problem (42) can be found in, e.g., [66] and [76].

### 3.2.2 Primal-dual methods for the solution of problem (39): ADMM algorithms

In order to solve problem (39), we are going to use a strategy introduced in [77] and [78] (to the best of our knowledge). The starting point is the obvious equivalence between (39) and the following linearly constrained optimization problem:
\[
\{u, Bu\} = \arg \min_{(v,q) \in W} j(v,q), \quad (43)
\]

where
\[
j(v,q) = F(q) + G(v),
\]
and
\[
W = \{ (v,q) \in V \times H, Bv - q = 0\}.
\]
From now on, we will assume that $V$ and $H$ are (real) Hilbert spaces, the $H$-norm being denoted by $|\cdot|$ and the associated inner-product by $(\cdot, \cdot)$. The next step is quite natural: we associate with the minimization problem (43) a Lagrangian functional $\mathcal{L}$ defined by

$$\mathcal{L}(v, q; \mu) = j(v, q) + (\mu, Bv - q),$$

and an augmented Lagrangian functional $\mathcal{L}_r$ defined (with $r > 0$) by

$$\mathcal{L}_r(v, q; \mu) = \mathcal{L}(v, q; \mu) + \frac{r}{2} |Bv - q|^2. \quad (44)$$

One can easily prove that the functionals $\mathcal{L}$ and $\mathcal{L}_r$ share the same saddle-points over $(V \times H) \times H$, and also that, if $\{\{u, p\}, \lambda\}$ is such a saddle-point, then $u$ is a solution of problem (39) and $p = Bu$. A classical algorithm to compute saddle-points is the so-called Uzawa algorithm, popularized by [3] (a book dedicated to the study of Economics equilibria), and further discussed in, e.g., [76]. Applying a close variant of the Uzawa algorithm to the computation of the saddle-points of $\mathcal{L}_r$ over $(V \times H) \times H$, we obtain

$$\begin{align*}
\lambda^0 \text{ is given in } H; \\
\text{for } n \geq 0, \lambda^n \to \{u^n, p^n\} \to \lambda^{n+1} \text{ via} \\
\{u^n, p^n\} = \arg \min_{\{v, q\} \in V \times H} \mathcal{L}_r(v, q; \lambda^n), \\
\lambda^{n+1} = \lambda^n + \rho (Bu^n - p^n),
\end{align*} \quad (45)$$

an algorithm called ALG1 by some practitioners, following a terminology introduced in [78] (an alternative name could have been augmented Lagrangian Uzawa algorithm which summarizes quite well what algorithm (45) is all about).

Concerning the convergence of ALG1 it has been proved in, e.g., [62], [63], [66] and [74] (see also [78]), that if:

(i) $\mathcal{L}$ has a saddle-point $\{\{u, p\}, \lambda\}$ over $(V \times H) \times H$.

(ii) $B$ is an injection and $R(B)$ is closed in $H$.

(iii) $\lim_{|q| \to +\infty} \frac{F(q)}{|q|} = +\infty$.

(iv) $F = F_0 + F_1$ with $F_0$ and $F_1$ proper, lower semi-continuous and convex, with $F_0$ Gateaux-differentiable, and uniformly convex on the bounded sets of $H$ (the above properties imply that problem (39) has a unique solution), then we have, $\forall \ r > 0$ and if

$$0 < \rho < 2r,$$

the following convergence result

$$\lim_{n \to +\infty} \{u^n, p^n\} = \{u, Bu\} \text{ in } V \times H, \quad (46)$$

where $u$ is the solution of problem (39); moreover, the convergence result (46) holds $\forall \ \lambda^0 \in H$. The convergence of the multiplier sequence $\{\lambda^n\}_{n \geq 0}$ is no better than
weak in general, implying that the criterion used to stop ALG1 has to been chosen carefully. Of course, in finite dimension, the properties of $B$, $F$ and $G$ implying convergence are less demanding than in infinite dimension; for example, the existence of a solution to problem (39) is sufficient to imply the existence of a saddle-point.

The main difficulty with the Uzawa algorithm (45) is clearly the solution of the minimization problem it contains. An obvious choice to solve this problem is to use a relaxation method (as advocated in [77], [78]). Suppose that, as advocated in the two above references (which show that, indeed, for the nonlinear elliptic problem discussed there the number of relaxation iterations reduces quickly to two), we limit the number of relaxation iterations to one when solving the minimization problem in (45): we obtain then the following primal-dual algorithm (called ALG2 by some practitioners):

\[
\{u^{-1}, \lambda^0\} \text{ is given in } V \times H; \tag{47}
\]

for \( n \geq 0 \), \( \{u^{n-1}, \lambda^n\} \to p^n \to u^n \to \lambda^{n+1} \) via

\[
p^n = \arg\min_{q \in H} \mathcal{L}_r(u^{n-1}, q; \lambda^n), \tag{48}
\]

\[
u^n = \arg\min_{v \in V} \mathcal{L}_r(v, p^n; \lambda^n), \tag{49}
\]

\[
\lambda^{n+1} = \lambda^n + \rho (Bu^n - p^n). \tag{50}
\]

Assuming that

\[ 0 < \rho < \frac{1 + \sqrt{5}}{2}r, \]

with the other assumptions implying the convergence of ALG1 still holding, we have

\[
\lim_{n \to +\infty} \{u^n, p^n\} = \{u, Bu\} \text{ in } V \times H,
\]

where \( u \) is the solution of problem (39). Convergence proofs can be found in [62], [63], [66] and [74].

A simple variant (called ALG3) of algorithm (47)-(50) is obtained by updating the multiplier a first time immediately after (48); we obtain then

\[
\{u^{-1}, \lambda^0\} \text{ is given in } V \times H; \tag{51}
\]

for \( n \geq 0 \), \( \{u^{n-1}, \lambda^n\} \to p^n \to \lambda^{n+1/2} \to u^n \to \lambda^{n+1} \) via

\[
p^n = \arg\min_{q \in H} \mathcal{L}_r(u^{n-1}, q; \lambda^n), \tag{52}
\]

\[ \lambda^{n+1/2} = \lambda^n + \rho (Bu^{n-1} - p^n). \tag{53}\]

\[
u^n = \arg\min_{v \in V} \mathcal{L}_r(v, p^n; \lambda^{n+1/2}), \tag{54}
\]

\[ \lambda^{n+1} = \lambda^{n+1/2} + \rho (Bu^n - p^n). \tag{55}\]
Most practitioners prefer ALG2 to ALG3, the main reason being that ALG2 is more robust than ALG3, in general.

**Remark 21.** If one takes \( \rho = r \) in (47)-(50) and (51)-(55), the algorithms we obtain belong to the Alternating Direction Methods of Multipliers (ADMM) family (a terminology we will justify in Section 3.3). The convergence of ADMM related algorithms is rather well established in the convex case (see, for example, [18], [61], [95]; see also the references therein and other chapters of this book, the one by M. Burger, A. Sawatzky & G. Steidl in particular). On the other hand, one is still lacking a general theory for the convergence of algorithms such as ALG1, ALG2 and ALG3 when applied to the solution of non-convex variational problems. Nevertheless, the above algorithms have been successfully applied to the solution of non-convex problems as shown, for example, in [42], [72] (Chapter 4), [74], and other chapters of this book, Chapters 7 and 8, in particular.

**Remark 22.** An important issue with the above primal-dual algorithms is how to vary \( r \) and \( \rho \) **dynamically** in order to improve the speed of convergence of these algorithms. This issue has been addressed in, e.g., [18], [34], [45], [46] (see also the references therein).

**Remark 23.** An overlooked ([34] being a notable exception) property of primal-dual algorithms such as ALG1, ALG2 and ALG3 is that they may be constructive still, in those not so uncommon situations where in (39) one has \( \text{dom}(F \circ B) \cap \text{dom}(G) = \emptyset \), implying that problem (39) has no solutions, strictly speaking. On the basis of the numerical results reported in [42] (see also [72] (Chapter 4) and Chapter 8 of this volume), we conjecture that if the parameters \( \rho \) and \( r \) are properly chosen, the sequence \( \{u^n, p^n\}_{n \geq 0} \) converges to a pair \( \{u, p\} \) minimizing the functional

\[
\{v, q\} \mapsto G(v) + F(q)
\]

over the set

\[
\{\{v, q\} \mid v, q \in \text{dom}(G) \times \text{dom}(F), |Bv - q| = \min_{\{w, \omega\} \in \text{dom}(G) \times \text{dom}(F)} |Bw - \omega|\},
\]

while the sequence \( \{\lambda^n\}_{n \geq 0} \) diverges **arithmetically** (that is, \( |\lambda^n| \to +\infty \) like \( n \) multiplied by a positive constant, that is slowly). If the above convergence/divergence result holds true (which seems to be the case for the non-convex problem discussed in [42]), it implies that the above primal-dual algorithms solve problem (39) in a **least-squares sense**, a most remarkable property indeed, testifying of the robustness of these algorithms. The above results look natural, but the optimization experts we consulted had trouble to give us a precise reference (or a proof).

**Remark 24.** We encountered situations (in incompressible finite elasticity in particular; see, e.g., [74] for details) where a safe way to proceed with the above primal-dual algorithms is as follows: Employ ALG1 with a well-balanced (that is neither to small nor too large) stopping criterion for the relaxation algorithm used to solve the minimization problem in (45); it has been observed quite often that the number
of relaxation iterations necessary to compute \( \{u^n, p^n\} \) from \( \lambda^n \) goes down quickly to one or two (an observation at the origin of ALG2), implying that starting with ALG1, the algorithm switches automatically to ALG2. It is not uncommon that this implementation of ALG1 produces an algorithm faster (CPU-wise) than ALG2 and ALG3, when solving “hard” problems. □

Further information on the convergence of Lagrange multiplier based iterative methods can be found in other chapters of this volume, and in, e.g., [45], [60], [62], [63], [66], [74] and [100] (see also the many references therein).

3.3 On the relationship between Alternating Direction Methods and ALG2, ALG3

As reported in [71] and [72] (Chapter 4) some previously unknown relationships between alternating direction methods and augmented Lagrangian algorithms were identified in 1975 by T.F. Chan and the first author of this chapter, while investigating the numerical solution of some simple nonlinear elliptic problems by various iterative methods (see [30] for details). Indeed, let us consider the particular case of problem (39) where \( V = H, B = I \) and \( F \) and \( G \) are both differentiable over \( V \); then, assuming that \( \rho = r \), ALG2 (that is algorithm (47)-(50)) takes the following form:

\[
\{u^{-1}, \lambda^0\} \text{ is given in } V \times H; \tag{56}
\]

for \( n \geq 0, \{u^{n-1}, \lambda^n\} \rightarrow p^n \rightarrow u^n \rightarrow \lambda^{n+1} \) via

\[
r(p^n - u^{n-1}) + DF(p^n) - \lambda^n = 0, \tag{57}
\]
\[
r(u^n - p^n) + DG(u^n) + \lambda^n = 0, \tag{58}
\]
\[
\lambda^{n+1} = \lambda^n + r(u^n - p^n), \tag{59}
\]

where \( DF \) (resp., \( DG \)) denotes the differential of \( F \) (resp., \( G \)). By elimination of \( \lambda^n \) and \( \lambda^{n+1} \) in (57)-(59), we obtain

\[
r(p^n - u^{n-1}) + DF(p^n) + DG(u^{n-1}) = 0,
\]
\[
r(u^n - u^{n-1}) + DF(p^n) + DG(u^n) = 0,
\]

which imply in turn (after changing \( n - 1 \) in \( n \)):

\[
r(p^{n+1} - u^n) + DF(p^{n+1}) + DG(u^n) = 0, \tag{60}
\]
\[
r(u^{n+1} - u^n) + DF(p^{n+1}) + DG(u^{n+1}) = 0. \tag{61}
\]

Comparing to (17) shows that in this particular case, ALG2 is a disguised form of the Douglas-Rachford scheme discussed in Section 2.5, with \( r = 1/\Delta t \) and \( DF \) (resp., \( DG \)) playing the role of \( A_1 \) (resp., \( A_2 \)). A similar interpretation holds for ALG3: in-
deed, if we assume again that $V = H$, $B = I$ and $F$ and $G$ are differentiable, then, if $\rho = r$, algorithm (51)-(55) reduces to the Peaceman-Rachford scheme (15) discussed in Section 2.4. The above equivalence result can be generalized to situations where $F$ and/or $G$ are not differentiable.

The reasons for which ALG2 and ALG3 are called Alternating Direction Methods of Multipliers (ADMM) by many practitioners should be clear now. For further information and details on these primal-dual equivalences, see the discussion by M. Yan and W. Yin in Chapter 5 of this book.

4 Operator-splitting methods for the direct numerical simulation of particulate flow

4.1 Generalities. Problem formulation

It is the (necessarily biased) opinion of the authors of this chapter that the direct numerical simulation of particulate flow has been one of the success stories of operator-splitting methods, justifying thus a dedicated section in this chapter, despite the fact that this story has been told in several publications (see, e.g., [70] (Chapters 8 & 9), [73] and [79], and the references therein). For simplicity, we will discuss only the one-particle case (however, the results of numerical experiments involving more than one particle will be presented).

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 $I$ at the center of mass (see Figure 1, for additional details).

The fluid occupies the region $\Omega \setminus B$ and we suppose that distance $(\partial B(0), \Gamma) > 0$.

From now on, $x = \{x_i\}_{i=1}^d$ will denote the generic point of $\mathbb{R}^d$, $dx = dx_1 \ldots dx_d$, while $\phi(t)$ will denote the function $x \rightarrow \phi(x,t)$. Assuming that the only external force is gravity, the fluid flow-rigid body motion coupling is modeled by

\[
\begin{align*}
\rho_f \left( \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right) - \mu_f \nabla^2 \mathbf{u} + \nabla p &= \rho_f g \text{ in } \Omega \setminus B(t), \text{ a.e. } t \in (0, T), \\
\nabla \cdot \mathbf{u}(t) &= 0 \text{ in } \Omega \setminus B(t), \text{ a.e. } t \in (0, T), \\
\mathbf{u}(t) &= \mathbf{u}_f(t) \text{ on } \Gamma, \text{ a.e. } t \in (0, T), \text{ with } \int_\Gamma \mathbf{u}_f(t) \cdot \mathbf{n} d\Gamma = 0, \\
\mathbf{u}(0) &= \mathbf{u}_0 \text{ in } \Omega \setminus B(0) \text{ with } \nabla \cdot \mathbf{u}_0 = 0,
\end{align*}
\]
\begin{align}
\frac{dG}{dt} &= V, \quad (66) \\
M \frac{dV}{dt} &= Mg + R_H, \quad (67) \\
\frac{d(10)}{dt} &= T_H, \quad (68) \\
G(0) &= G_0, V(0) = V_0, \omega(0) = \omega_0, B(0) = B_0. \quad (69)
\end{align}

In relations (62)-(69):

- Vector \( u = \{u_i\}_{i=1}^d \) is the fluid flow velocity and \( p \) is the pressure.
- \( u_0 \) and \( u_{\Gamma} \) are given vector-valued functions.
- \( V \) is the velocity of the center of mass of body \( B \), while \( \omega \) is the angular velocity.
- \( R_H \) and \( T_H \) denote, respectively, the resultant and the torque of the hydrodynamical forces, namely the forces that the fluid exerts on \( B \); we have, actually,

\[ R_H = \int_{\partial B} \sigma n d\gamma \quad \text{and} \quad T_H = \int_{\partial B} \overrightarrow{Gx} \times \sigma n d\gamma. \quad (70) \]

In (70) the stress-tensor \( \sigma \) is defined by \( \sigma = 2\mu D(u) - p\mathbf{1}_d \), with \( D(v) = \frac{1}{2}(\nabla v + (\nabla v)^t) \), while \( n \) is a unit normal vector at \( \partial B \) and \( \mathbf{1}_d \) 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) the no-slip boundary condition holds, namely
\[ \mathbf{u}(x,t) = \mathbf{V}(t) + \mathbf{\omega}(t) \times \mathbf{G}(t)x, \quad \forall x \in \partial B(t). \]  

Remark 25. System (62)-(65) (resp., (66)-(69)) 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 (62)-(69) has been proved in [52], assuming that, at \( t = 0 \), the particles do not touch \( \Gamma \) and each other (see also [87] and [145]). Concerning the numerical solution of (62)-(69), (71) several approaches are encountered in the literature, among them: (i) The Arbitrary Lagrange-Euler (ALE) methods; these methods, which rely on moving meshes, are discussed in, e.g., [98], [103] and [127]. (ii) The fictitious boundary method discussed in, e.g., [165], and (iii) the non-boundary fitted fictitious domain methods discussed in, e.g., [70], [79] and [140], [141] (and in Section 4.2, hereafter). Among other things, the methods in (ii) and (iii) have in common that the meshes used for the flow computations do not have to match the boundary of the particles.

Remark 26. Even if theory suggests that collisions may never take place in finite time (if we assume that the particles have smooth shapes and that the flow is still modeled by the Navier-Stokes equations as long as the particles do not touch each other, or \( \Gamma \)), near collisions take place, and after discretization particles may collide. These phenomena can be handled by introducing (as done in, e.g., [70] (Chapter 8) and [79]) well-chosen short range repulsion potentials reminiscent of those encountered in Molecular Dynamics, or by using Kuhn-Tucker multipliers to authorize particle motions with contact but no overlapping (as done in, e.g., [128] and [129]).

More information on the numerical treatment of particles in flow, can be found in, e.g., [152] (and the references therein), and of course in Google.

### 4.2 A fictitious domain formulation

Considering the fluid-rigid body mixture as a unique (heterogeneous) medium we are going to derive a fictitious domain based variational formulation to model its motion. The principle of this derivation is pretty simple: it relies on the following steps (see, e.g., [70] and [79] for more details), where in Step a we denote by \( \mathbf{S} : \mathbf{T} \) the Fröbenius inner product of the tensors \( \mathbf{S} \) and \( \mathbf{T} \), that is (with obvious notation)

\[ \mathbf{S} : \mathbf{T} = \sum_{1 \leq i,j \leq d} s_{ij}t_{ij}. \]

**Step a.** Start from the following global weak formulation (of the virtual power type):
\[
\begin{align*}
\rho_f & \int_{\Omega_{\beta(t)}} \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] \cdot \mathbf{v} \, d\mathbf{x} + 2\mu_f \int_{\Omega_{\beta(t)}} \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) \, d\mathbf{x} \\
& - \int_{\Omega_{\beta(t)}} p \nabla \cdot \mathbf{v} \, d\mathbf{x} + M \frac{d\mathbf{Y}}{dt} \cdot \mathbf{Y} + \frac{d(\mathbf{1}\omega \cdot \theta)}{dt} \cdot \theta \\
& = \rho_f \int_{\Omega_{\beta(t)}} \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} + Mg \cdot \mathbf{Y},
\end{align*}
\]

\[
\forall \{\mathbf{v}, \mathbf{Y}, \theta\} \in (H^1(\Omega_{\beta(t)}))^d \times \mathbb{R}^d \times \Theta \quad \text{and verifying}
\]

\[
v = 0 \text{ on } \Gamma, \quad \mathbf{v}(\mathbf{x}) = \mathbf{Y} + \theta \times \mathbf{G}(t)\mathbf{x}, \forall \mathbf{x} \in \partial B(t), \quad t \in (0,T),
\]

\[
\text{with } \Theta = \mathbb{R}^3 \text{ if } d = 3, \quad \Theta = \{(0,0,\theta) \mid \theta \in \mathbb{R}\} \text{ if } d = 2.
\]

\[
\int_{\Omega_{\beta(t)}} q \nabla \cdot \mathbf{u}(t) \, d\mathbf{x} = 0, \forall q \in L^2(\Omega_{\beta(t)}), \quad t \in (0,T),
\]

\[
\mathbf{u}(t) = \mathbf{u}_f(t) \text{ on } \Gamma, \quad t \in (0,T),
\]

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

\[
\frac{d\mathbf{G}}{dt} = \mathbf{V},
\]

\[
\mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x}), \forall \mathbf{x} \in \Omega_{\beta(0)},
\]

\[
\mathbf{G}(0) = \mathbf{G}_0, \quad \mathbf{V}(0) = \mathbf{V}_0, \quad \overrightarrow{\mathbf{\omega}}(0) = \overrightarrow{\mathbf{\omega}}_0, \quad B(0) = B_0.
\]

**Step b.** Fill \(B\) with the surrounding fluid and impose a rigid body motion to the fluid inside \(B\). 

**Step c.** Modify the global weak formulation (72)-(78) accordingly, taking advantage of the fact that if \(\mathbf{v}\) is a rigid body motion velocity field, then \(\nabla \cdot \mathbf{v} = 0\) and \(\mathbf{D}(\mathbf{v}) = 0\). 

**Step d.** 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:

\[
\begin{align*}
\rho_f & \int_\Omega \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] \cdot \mathbf{v} \, d\mathbf{x} + 2\mu_f \int_\Omega \mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{v}) \, d\mathbf{x} - \int_\Omega p \nabla \cdot \mathbf{v} \, d\mathbf{x} \\
& + (1 - \rho_f/\rho_s) \left[ M \frac{d\mathbf{Y}}{dt} \cdot \mathbf{Y} + \frac{d(\mathbf{1}\omega \cdot \theta)}{dt} \cdot \theta \right] + \langle \lambda, \mathbf{v} - \theta \times \mathbf{G}(t)\mathbf{x} \rangle >_{B(t)} \\
& = \rho_f \int_\Omega \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} + (1 - \rho_f/\rho_s)Mg \cdot \mathbf{Y}, \quad \forall \{\mathbf{v}, \mathbf{Y}, \theta\} \in (H^1(\Omega))^d \times \mathbb{R}^d \times \Theta,
\end{align*}
\]

\[
\text{with } \Theta = \mathbb{R}^3 \text{ if } d = 3, \quad \Theta = \{(0,0,\theta) \mid \theta \in \mathbb{R}\} \text{ if } d = 2.
\]
\[ \int_{\Omega} q \nabla \cdot u(t) \, dx = 0, \forall q \in L^2(\Omega), \, t \in (0, T), \quad (80) \]
\[ u(t) = u(0) \text{ on } \Gamma, \, t \in (0, T), \quad (81) \]
\[
\begin{cases} 
\langle \mu, u(x,t) - \nabla(t) - \omega(t) \times G(t)x \rangle_{B(t)} = 0, \\
\forall \mu \in \Lambda(t) (= (H^1(B(t)))^d), \, t \in (0, T), \\
\end{cases} \quad (82) \]
\[ \frac{dG}{dt} = \nu, \quad (83) \]
\[
\begin{cases} 
G(0) = G_0, \, V(0) = V_0, \, \omega(0) = \omega_0, \, B(0) = B_0, \\
u(x, 0) = u_0(x), \forall x \in \Omega \setminus \bar{B}_0, \, u(x, 0) = V_0 + \omega_0 \times \rightarrow G_0, \forall x \in \bar{B}_0. \\
\end{cases} \quad (84) \]

From a theoretical point of view, a natural choice for \( \langle \cdot, \cdot \rangle_{B(t)} \) is provided by, e.g.,
\[ \langle \mu, v \rangle_{B(t)} = \int_{B(t)} [\mu \cdot v + l^2 D(\mu) : D(v)] \, dx; \quad (85) \]
in (85), \( l \) is a characteristic length, the diameter of \( B \), for example. In practice, following [70] (Chapter 8) and [79], one makes things much simpler by approximating \( \Lambda(t) \) by
\[ \Lambda_h(t) = \{ \mu \mid \mu = \sum_{j=1}^{N(t)} \mu_j \delta(x - x_j), \text{ with } \mu_j \in \mathbb{R}^d, \forall j = 1, \ldots, N \}, \quad (86) \]
and the pairing in (85) by
\[ \langle \mu, v \rangle_{(B(t), h)} = \sum_{j=1}^{N(t)} \mu_j \cdot v(x_j). \quad (87) \]

In (86), (87), \( x \rightarrow \delta(x - x_j) \) is the Dirac measure at \( x_j \), and the set \( \{ x_j \}_{j=1}^N \) 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.

A variant of the above fictitious domain approach is discussed in [140] and [141]; after an appropriate elimination, it does not make use of Lagrange multipliers to force the rigid body motion of the particles, but uses instead projections on velocity subspaces where the rigid body motion velocity property is verified over the particles (see [140] and [141] for details).
4.3 Solving problem (79)-(84) by operator-splitting

We do not consider collisions; after (formal) elimination of \( p \) and \( \lambda \), problem (79)-(84) reduces to an initial value problem of the following form

\[
\frac{dX}{dt} + \sum_{j=1}^{J} A_j(X,t) = 0 \text{ on } (0,T), \quad X(0) = X_0,
\]

where \( X = \{ u, V, \omega, G \} \) (or \( \{ u, V, I, \omega, G \} \)). A typical situation will be the one where, with \( J = 4 \), operator \( A_1 \) will be associated with incompressibility, \( A_2 \) with advection, \( A_3 \) with diffusion, \( A_4 \) with the fictitious domain treatment of the rigid body motion; other decompositions are possible as shown in, e.g., [70] (Chapter 8) and [79] (both references include a collision operator). The Lie’s scheme (3), (4) applies “beautifully” to the solution of problem (79)-(84). The resulting method is quite modular implying that different space and time approximations can be used to treat the various sub-problems encountered at each time step; the only constraint is that two successive steps have to communicate (by projection in general to provide the initial condition required by each initial value sub-problem).

4.4 Numerical experiments

4.4.1 Generalities

The methodology we described (briefly) in the above paragraphs has been validated by numerous experiments (see, in particular, [70] (Chapters 8 & 9), [73], [79], [97] [137] and the related publications reported in http://www.math.uh.edu/~pan/). In this chapter, we will consider two test problems (borrowed from [73] (Section 3.4)): The first test problem involves three particles, while the second one concerns a channel flow with 300 particles. The fictitious domain/operator-splitting approach has made the solution of these problems (almost) routine nowadays. All the flow computations have been done using the Bercovier-Pironneau finite element approximation; namely (see [70] (Chapters 5, 8 and 9) for details), we used a globally continuous piecewise affine approximation of the velocity (resp., the pressure) associated with a triangulation (in 2-D) or tetrahedral partition (in 3-D) \( T_h \) (resp., \( T_{2h} \)) of \( \Omega \), \( h \) being a space discretization step. The pressure mesh is thus twice coarser than the velocity one. The calculations have been done using uniform partitions \( T_h \) and \( T_{2h} \).

4.4.2 First test problem: Settling of three balls in a vertical narrow tube

Our goal in this sub-section is to discuss the interaction of three identical balls settling in a narrow tube of rectangular cross-section, 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 cylinder of length 6 whose cross-section is the rectangle $\Omega = (0, 1.5) \times (0, 0.25)$; 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, but less satisfying, alternative would be to specify periodicity in the vertical direction). At time $t = 0$, we suppose that the truncated cylinder coincides with the “box” $(0, 1.5) \times (0, 0.25) \times (0, 6)$, and the centers of the balls are on the vertical axis of the cylinder at the points $x_1 = 0.75, x_2 = 0.125, x_3 = 1, 1.3$ and 1.6. The parameters for this test case are $\rho_s = 1.1, \rho_f = 1, \mu_f = 1$, the diameter of the balls being $d = 0.2$. 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 0, while the three balls are released from rest. The velocity on the cylinder wall is 0. On the time interval $[0, 15]$ the drafting, kissing and tumbling phenomenon (a terminology introduced by D.D. Joseph) has been observed several times before a stable quasi-horizontal configuration takes place, as shown in Figures 2, 3 and 4. The averaged vertical velocity of the balls is 2.4653 on the time interval $[13, 15]$, while the \textit{averaged particle Reynolds number} is 49.304 on the same time interval, a clear evidence that inertia has to be taken into account.
Fig. 3 Relative positions 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)
4.4.3 Motion of 300 neutrally buoyant disks in a two-dimensional horizontal channel

This second test problem involving 300 particles and a solid volume/fluid volume of the order of 0.38, collisions (or near-collisions) have to be accounted for in the simulations; to do so, we have used the methods discussed in [70] (Chapter 8) and [79]. Another peculiarity of this test problem is that \( \rho_s = \rho_f \) for all the 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 (79); however this difficulty can be overcome as shown in [136]. 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 \( \rho_f = 1 \) and viscosity \( \mu_f = 1 \), with 300 identical rigid solid disks of density \( \rho_f = 1 \) and diameter 0.9. (c) At \( t = 0 \), 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 10/9 per unit length (without the particles the corresponding steady flow would have been of the Poiseuille type with 20 as maximal flow speed). (e) The boundary conditions are given by \( \mathbf{u}(x_1,x_2,t) = 0 \) 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 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 in the \( \partial x_1 \) 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 Figures 5. These figures show that, initially, we have the sliding motion of horizontal particle layers, then after some critical time a chaotic flow-motion takes place in very few time units, the highest particle concentration being along the channel axis (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. On Figure 6 we show the averaged solid fraction as a function of $x_2$, the averaging space-time set being $\{x_1, t\} | 0 \leq x_1 \leq 42, 380 \leq t \leq 400 \}$; the particle aggregation along the channel horizontal symmetry axis appears very clearly from this figure since the solid fraction is close to 0.58 at $x_2 = 6$ while the global solid fraction is 0.38 (vertical line in the figure). Finally, we have visualized on Figure 7 the $x_1$-averaged horizontal component of the mixture velocity at $t = 400$, as a function of $x_2$. The dashed line corresponds to a horizontal velocity distribution of the steady flow of the same fluid, with no particle in the channel, for the same pressure drop; the corresponding velocity profile is (of course) of the Poiseuille 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 [136] 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 perfectly plastic material). Figures 5, 6 and 7 show also that, as well-known, some order may be found in chaos.

For more details and further results and comments on pressure driven neutrally buoyant particulate flows in two-dimensional channels (including simulations with much larger numbers of particles, the largest one being 1,200) see [70] (Chapter 9) and [97].
5 Operator-splitting methods for the numerical solution of nonlinear problems from condensate and plasma physics

5.1 Introduction

Operator-splitting methods have been quite successful at solving problems in Computational Physics, beside those from Computational Mechanics (CFD in particular). Among these successful applications let us mention those involving nonlinear Schrödinger equations, as shown, for example, by [9], [10], [44] and [102]. On the basis of some very inspiring articles (see, e.g., [9], [10] and [102]) he wrote on the above topic, the editors asked their colleague Peter Markowich to contribute a related chapter for this book; unfortunately, Professor Markowich being busy elsewhere had to say no. Considering the importance of nonlinear Schrödinger related problems, it was decided to (briefly) discuss in this chapter the solution of some of them by operator-splitting methods (see also Chapter 18 on the propagation of laser pulses along optical fibers). In Section 5.2, we will discuss the operator-splitting solution of the celebrated Gross-Pitaevskii equation for Bose-Einstein condensates, then, in Section 5.3, we will discuss the solution of the Zakharov system modeling the propagation of Langmuir waves in ionized plasma.

5.2 On the solution of the Gross-Pitaevskii equation

A Bose-Einstein condensate (BEC) is a state of matter of a dilute gas of bosons cooled to temperatures very close to absolute zero. Under such conditions, a large fraction of the bosons occupies the lowest quantum state, at which point macroscopic quantum phenomena become apparent. The existence of Bose-Einstein condensates was predicted in the mid-1920s by S. N. Bose and A. Einstein. If dilute enough, the time evolution of a BEC is described by the following Gross-Pitaevskii equation (definitely of the nonlinear Schrödinger type and given here in a-dimensional form (following [9])):

\[ i \varepsilon \frac{\partial \psi}{\partial t} = -\frac{\varepsilon^2}{2} \nabla^2 \psi + V_d(x) \psi + K_d |\psi|^2 \psi \text{ in } \Omega \times (0, T), \tag{89} \]

where, in (89), \( \psi \) is a complex-valued function of \( x \) and \( t \), \( i = \sqrt{-1} \), \( \Omega \) is an open connected subset of \( \mathbb{R}^d \) (with \( d = 1, 2 \) or \( 3 \)), the real-valued function \( V_d \) denotes an external potential, and the real-valued parameter \( K_d \) is representative of the particles interactions. Equation (89) has to be completed by boundary and initial conditions. Equation (89) has motivated a very large literature from both physical and mathematical points of view. Let us mention among many others [1], [9], [125] and [126] (see also the many references therein). To solve equation (89) numerically we need to complete it by boundary and initial conditions: from now on, we will assume that
Let us denote by \( \psi \) \( x \in \Omega \),
\[
\psi(x,0) = \psi_0(x), \quad x \in \Omega, \tag{90}
\]
and (denoting by \( \Gamma \) the boundary of \( \Omega \))
\[
\psi = 0 \text{ on } \Gamma \times (0,T). \tag{91}
\]

The boundary conditions in (91) have been chosen for their simplicity, and also to provide an alternative to the periodic boundary conditions considered in [9]. An important (and very easy to prove) property of the solution of the initial boundary value problem (89)-(91) reads as:
\[
\frac{d}{dt} \int \Omega |\psi(x,t)|^2 \, dx = 0 \text{ on } [0,T],
\]

implying that
\[
\int \Omega |\psi(x,t)|^2 \, dx = \int \Omega |\psi_0(x)|^2 \, dx \text{ on } [0,T]. \tag{92}
\]

As done before, we denote by \( \psi(t) \) the function \( x \rightarrow \psi(x,t) \). Let \( \Delta t > 0 \) be a time discretization step and denote \((n + \alpha)\Delta t\) by \( t^{n+\alpha} \); applying to problem (89)-(91) the Strang’s symmetrized scheme (7)-(10) of Section 2.3, we obtain:
\[
\psi^n = \psi_0; \tag{93}
\]
for \( n \geq 0 \), \( \psi^n \rightarrow \psi^{n+1/2} \rightarrow \psi^{n+1/2} \rightarrow \psi^{n+1} \) as follows
\[
\begin{aligned}
&i \cdot \frac{\partial \psi^n}{\partial t} + \frac{\epsilon}{2} \nabla^2 \psi = 0 \text{ in } \Omega \times (t^n, t^{n+1/2}), \\
&\psi = 0 \text{ on } \Gamma \times (t^n, t^{n+1/2}), \\
&\psi(t^n) = \psi^n; \quad \psi^{n+1/2} = \psi(t^{n+1/2}), \\
&i \cdot \frac{\partial \psi^n}{\partial t} = V_d(x) \psi^n + K_d |\psi^n|^2 \psi^n \text{ in } \Omega \times (0, \Delta t), \\
&\psi = 0 \text{ on } \Gamma \times (0, \Delta t), \\
&\psi(0) = \psi^{n+1/2}; \quad \psi^{n+1/2} = \psi(\Delta t), \tag{95}
\end{aligned}
\]
\[
\begin{aligned}
&i \cdot \frac{\partial \psi^n}{\partial t} + \frac{\epsilon}{2} \nabla^2 \psi = 0 \text{ in } \Omega \times (t^{n+1/2}, t^{n+1}), \\
&\psi = 0 \text{ on } \Gamma \times (t^{n+1/2}, t^{n+1}), \\
&\psi(t^{n+1/2}) = \psi^{n+1/2}; \quad \psi^{n+1} = \psi(t^{n+1}). \tag{96}
\end{aligned}
\]

On the solution of (95): Let us denote by \( \psi_1 \) (resp., \( \psi_2 \)) the real (resp., imaginary) part of \( \psi \); from (95), we have
\[
\begin{aligned}
&\frac{\partial \psi_1}{\partial t} = V_d(x) \psi_2 + K_d |\psi_1|^2 \psi_2 \text{ in } \Omega \times (0, \Delta t), \\
&\frac{\partial \psi_2}{\partial t} = -V_d(x) \psi_1 - K_d |\psi_1|^2 \psi_1 \text{ in } \Omega \times (0, \Delta t). \tag{97}
\end{aligned}
\]
Multiplying by $\psi_1$ (resp., $\psi_2$) the $1$st (resp., the $2$nd) equation in (97), we obtain by addition

$$\frac{\partial}{\partial t} |\psi(x,t)|^2 = 0 \text{ on } (0, \triangle t), \text{ a.e. } x \in \Omega,$$

which implies in turn that

$$|\psi(x,t)| = |\psi(x,0)| = |\psi^{n+1/2}| \text{ on } (0, \triangle t), \text{ a.e. } x \in \Omega. \quad (98)$$

It follows then from (95) and (98) that

$$\begin{align*}
  \psi^{n+1/2} \text{ is a very classical one. Its solution is obviously given by } \\
  \phi(t) = e^{i\frac{\epsilon}{2}t} \phi_0, \forall t \in [t_0,t_f].
\end{align*} \quad (101)$$

Suppose that $\Omega = (0,a) \times (0,b) \times (0,c)$ with $0 < a < +\infty$, $0 < b < +\infty$ and $0 < c < +\infty$; since the eigenvalues, and related eigenfunctions, of the negative Laplace operator $-\nabla^2$, associated with the homogeneous Dirichlet boundary conditions are known explicitly, and given, for $p$, $q$ and $r$ positive integers, by

$$\begin{align*}
  \lambda_{pqr} &= \pi^2 \left( \frac{p^2}{a^2} + \frac{q^2}{b^2} + \frac{r^2}{c^2} \right), \\
  w_{pqr}(x_1,x_2,x_3) &= 2 \sqrt{\frac{2}{abc}} \sin \left( \frac{p\pi x_1}{a} \right) \sin \left( \frac{q\pi x_2}{b} \right) \sin \left( \frac{r\pi x_3}{c} \right)
\end{align*} \quad (102)$$

we have then $\int_{\Omega} |w_{pqr}(x)|^2 dx = 1$ it follows from (101) that

$$\phi(x,t) = \sum_{1 \leq p,q,r < +\infty} \phi_{pqr}^0 e^{-i\frac{\epsilon}{2}\lambda_{pqr}(t-t_0)} w_{pqr}(x), \text{ with } \phi_{pqr}^0 = \int_{\Omega} w_{pqr}(y) \phi_0(y) dy. \quad (103)$$
In practice, one takes $1 \leq p \leq P$, $1 \leq q \leq Q$, $1 \leq r \leq R$, and uses the Fast Fourier Transform (FFT) to compute the coefficients $\phi_{pq}^0$ and then $\phi(x,t)$.

For those more general situations where the solutions of the following linear eigenvalue problem

$$
\begin{cases}
\{w, \lambda\} \in H^1_0(\Omega) \times \mathbb{R}, & \int_{\Omega} |w(x)|^2 \, dx = 1, \; \lambda > 0, \\
\int_{\Omega} \nabla w \cdot \nabla \varphi \, dx = \lambda \int_{\Omega} w \varphi \, dx, & \forall \varphi \in H^1_0(\Omega),
\end{cases}
$$

are not known explicitly, one still has several options to solve (100), an obvious one being:

Approximate (104) by

$$
\begin{cases}
\{w, \lambda\} \in V_h \times \mathbb{R}, & \int_{\Omega} |w(x)|^2 \, dx = 1, \; \lambda > 0, \\
\int_{\Omega} \nabla w \cdot \nabla \varphi \, dx = \lambda \int_{\Omega} w \varphi \, dx, & \forall \varphi \in V_h,
\end{cases}
$$

where $V_h$ is a finite dimensional sub-space of $H^1_0(\Omega)$. Then, as in, e.g., [17], [82] use an eigensolver (like the one discussed in [113]) to compute the first $Q \leq N = \dim V_h$ eigen-pairs solutions of (105), such that (with obvious notation)

$$
\int_{\Omega} w_p w_q \, dx = 0 \quad \forall p, q, 1 \leq p, q \leq Q, \; p \neq q,
$$

and denote by $V_Q$ the finite dimensional space span by the basis $\{w_q\}_{q=1}^Q$. Next, proceeding as in the continuous case we approximate the solution of problem (100) by $\phi_Q$ defined by

$$
\phi_Q(x,t) = \sum_{q=1}^Q \phi_q^0 e^{-i \varepsilon \lambda_q^2 (t-t_0)} w_q(x), \quad \text{with} \quad \phi_q^0 = \int_{\Omega} w_q(y) \phi_0(y) \, dy.
$$

For the space $V_h$ in (105), we can use these finite element approximations of $H^1_0(\Omega)$ discussed for example in [37], [66] (Appendix 1) and [72] (Chapter 1) (see also the references therein).

Another approach, less obvious but still natural, is to observe that if $\phi$, the unique solution of (100) is smooth enough, it is also the unique solution of

$$
\begin{cases}
\frac{\partial^2 \phi}{\partial t^2} + \frac{\varepsilon^2}{4} \nabla^4 \phi = 0 \quad \text{in} \; \Omega \times (t_0,t_f), \\
\phi = 0 \quad \text{and} \; \nabla^2 \phi = 0 \quad \text{on} \; \Gamma \times (t_0,t_f), \\
\phi(t_0) = \phi_0, \quad \frac{\partial \phi}{\partial t}(t_0) = i \varepsilon \nabla^2 \phi_0(= \phi_1),
\end{cases}
$$

a well-known model in elasto-dynamics (vibrations of simply supported plates).

From $Q$, a positive integer, we define a time discretization step $\tau$ by $\tau = \frac{t_f - t_0}{Q}$.

The initial-boundary value problem (107) is clearly equivalent to
\[
\begin{align*}
\frac{\partial \phi}{\partial t} &= v \quad \text{in } \Omega \times (t_0, t_f), \\
\frac{\partial v}{\partial t} + \frac{e^2}{4} \nabla^4 \phi &= 0 \quad \text{in } \Omega \times (t_0, t_f), \\
\phi &= 0 \text{ and } \nabla^2 \phi = 0 \quad \text{on } \Gamma \times (t_0, t_f), \\
\phi(t_0) &= \phi_0, \quad v(t_0) = \frac{i e}{2} \nabla^2 \phi_0 (= v_0).
\end{align*}
\]

(108)

A time-discretization scheme for (107) (via (108)), combining good accuracy, stability and energy conservation properties (see, e.g., [14]) reads as follows (with \(\{\phi^q, v^q\}\) an approximation of \(\{\phi, v\}\) at \(t^q = t_0 + q\tau\)):

\[
\begin{align*}
\phi^0 &= \phi_0, \quad v^0 = v_0; \\
&\text{for } q = 0, \cdots, Q - 1, \text{ compute } \{\phi^{q+1}, v^{q+1}\} \text{ from } \{\phi^q, v^q\} \text{ via the solution of} \\
&\begin{cases}
\phi^{q+1} - \phi^q = \frac{1}{2}(v^{q+1} + v^q), \\
v^{q+1} - v^q = \frac{\tau}{8} \nabla^4 (\phi^{q+1} + \phi^q) = 0 \quad \text{in } \Omega, \\
\phi^{q+1} = 0 \text{ and } \nabla^2 \phi^{q+1} = 0 \quad \text{on } \Gamma.
\end{cases}
\end{align*}
\]

(109)

By elimination of \(v^{q+1}\) it follows from (109) that \(\phi^{q+1}\) is solution of

\[
\begin{align*}
\phi^{q+1} + \frac{(\tau e)^2}{8} \nabla^4 \phi^{q+1} &= \phi^q + \tau v^q - \frac{(\tau e)^2}{8} \nabla^4 \phi^q \quad \text{in } \Omega, \\
\phi^{q+1} &= 0 \text{ and } \nabla^2 \phi^{q+1} = 0 \quad \text{on } \Gamma,
\end{align*}
\]

(110)

a bi-harmonic problem which is well-posed in \(H^1_0(\Omega) \cap H^2(\Omega)\). Next, one obtains easily \(v^{q+1}\) from

\[v^{q+1} = \frac{\tau}{2} (\phi^{q+1} - \phi^q) - v^q.\]

For the solution of the bi-harmonic problem (110) we advocate those mixed finite element approximations and conjugate gradient algorithms used in various chapters of [72] (see also the references therein).

5.3 On the solution of Zakharov systems

In 1972, V.E. Zakharov introduced a mathematical model describing the propagation of Langmuir waves in ionized plasma (ref. [179]). This model reads as follows (after rescaling):

\[
\begin{align*}
\frac{i}{\tau} \frac{\partial u}{\partial t} + \nabla^2 u &= u n, \\
\frac{\partial^2 n}{\partial t^2} - \nabla^2 n + \nabla^2 (|u|^2) &= 0,
\end{align*}
\]

(111)
where the complex-valued function $u$ is associated with a highly oscillating electric field, while the real-valued function $n$ denotes the fluctuation of the plasma-ion density from its equilibrium state. In this section, following [102], we will apply the symmetrized Strang operator-splitting scheme (previously discussed in Section 2.3 of this chapter) to the following generalization of the above equations:

$$
\begin{align*}
  \frac{i}{c} \frac{\partial u}{\partial t} + \nabla^2 u + 2\lambda |u|^2 u + 2un &= 0, \\
  \frac{1}{c^2} \frac{\partial^2 n}{\partial t^2} - \nabla^2 n + \mu \nabla^2 (|u|^2) &= 0,
\end{align*}
$$

(112)

where $\lambda$ and $\mu$ are real numbers and $c(>0)$ is the wave propagation speed. Following again [102], we will assume, for simplicity, that the physical phenomenon modelled by (112) takes place on the bounded interval $(0, L)$, with $u, n, \frac{\partial u}{\partial x}$ and $\frac{\partial n}{\partial x}$ space-periodic, during the time interval $[0, T]$. Thus, (112), completed by initial conditions, reduces to:

$$
\begin{align*}
  \frac{i}{c} \frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} + 2\lambda |u|^2 u + 2un &= 0 \text{ in } (0, L) \times (0, T), \\
  \frac{1}{c^2} \frac{\partial^2 n}{\partial t^2} - \frac{\partial^2 n}{\partial x^2} + \mu \frac{\partial^2}{\partial x^2} (|u|^2) &= 0 \text{ in } (0, L) \times (0, T), \\
  u(0, t) &= u(L, t), \quad \frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) \text{ on } (0, T), \\
  n(0, t) &= n(L, t), \quad \frac{\partial n}{\partial x}(0, t) = \frac{\partial n}{\partial x}(L, t) \text{ on } (0, T), \\
  u(0) &= u_0, \quad n(0) = n_0, \quad \frac{\partial n}{\partial t}(0) = n_1.
\end{align*}
$$

(113)

As done previously in this chapter, we denote by $\phi(t)$ the function $x \rightarrow \phi(x, t)$.

Remark 27. Albeit considered by some as too simple from a physical point of view, space-periodic boundary conditions are common in plasma physics. They have been used for example in [131], a most celebrated article dedicated to the mathematical analysis of the behavior of plasma entropy (see also [163] which relates a discussion that C. Villani had with E. Lieb concerning precisely the use of space-periodic boundary conditions in plasma physics). $\square$

From the rich structure of the Zakharov’s system (113) it is not surprising that a variety of operator-splitting schemes can be applied to its numerical solution, several of these schemes being described in [102] (see also the references therein concerning splitting schemes not described in [102]). A first step to the application of operator-splitting scheme to the time-discretization of problem (113) is to introduce the function $p = \frac{\partial n}{\partial t}$ and to rewrite (113) as:
\begin{align}
\begin{cases}
\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} + 2\lambda |u|^2 u + 2un = 0 \text{ in } (0, L) \times (0, T), \\
\frac{\partial q}{\partial t} - p = 0 \text{ in } (0, L) \times (0, T), \\
1 \frac{\partial p}{\partial t} - \frac{\partial^2 u}{\partial x^2} + \mu \frac{\partial^2}{\partial x^2}(|u|^2) = 0 \text{ in } (0, L) \times (0, T), \\
u(0, t) = u(L, t), \quad \frac{\partial u}{\partial n}(0, t) = \frac{\partial u}{\partial n}(L, t) \text{ on } (0, T), \\
n(0, t) = n(L, t), \quad \frac{\partial n}{\partial x}(0, t) = \frac{\partial n}{\partial x}(L, t) \text{ on } (0, T), \\
u(0) = u_0, \quad n(0) = n_0, \quad p(0) = n_1.
\end{cases}
\end{align}

(114)

Applying the Strang’s symmetrized scheme to the time-discretization of problem (114), one obtains (among other possible schemes, and with \( t^{q+} = (q + \alpha) \Delta t \)):

\begin{align}
\{u^0, n^0, p^0\} = \{u_0, n_0, n_1\}.
\end{align}

(115)

For \( q \geq 0 \), \( \{u^q, n^q, p^q\} \rightarrow \{u^{q+1/2}, n^{q+1/2}, p^{q+1/2}\} \rightarrow \{\hat{u}^{q+1/2}, \hat{n}^{q+1/2}, \hat{p}^{q+1/2}\} \rightarrow \{u^{q+1}, n^{q+1}, p^{q+1}\} \) via

\begin{align}
\begin{cases}
\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} = 0 \text{ in } (0, L) \times (t^q, t^{q+1/2}), \\
\frac{\partial n}{\partial t} - p = 0 \text{ in } (0, L) \times (t^q, t^{q+1/2}), \\
\frac{\partial p}{\partial t} - \frac{\partial^2 n}{\partial x^2} = 0 \text{ in } (0, L) \times (t^q, t^{q+1/2}), \\
u(t^q) = u^q, \quad n(t^q) = n^q, \quad p(t^q) = p^q; \\
u^{q+1/2} = u(t^{q+1/2}), n^{q+1/2} = n(t^{q+1/2}), p^{q+1/2} = p(t^{q+1/2}).
\end{cases}
\end{align}

(116)
\[
\begin{aligned}
& \left\{ \begin{array}{l}
\frac{i}{\partial t} \frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} = 0 \text{ in } (0, L) \times (t^{q+1/2}, t^{q+1}), \\
\frac{\partial n}{\partial t} - \frac{P}{2} = 0 \text{ in } (0, L) \times (t^{q+1/2}, t^{q+1}), \\
\frac{1}{c^2} \frac{\partial^2 n}{\partial t^2} - \frac{\partial^2 n}{\partial x^2} = 0 \text{ in } (0, L) \times (t^{q+1/2}, t^{q+1}), \\
\end{array} \right. \\
& u(0, t) = u(L, t), \quad \frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) \text{ on } (t^{q+1/2}, t^{q+1}), \\
& n(0, t) = n(L, t), \quad \frac{\partial n}{\partial x}(0, t) = \frac{\partial n}{\partial x}(L, t), \quad p(0, t) = p(L, t) \text{ on } (t^{q+1/2}, t^{q+1}), \\
& u(t^{q+1/2}) = \hat{u}^{q+1/2}, \quad n(t^{q+1/2}) = \hat{n}^{q+1/2}, \quad p(t^{q+1/2}) = \hat{p}^{q+1/2}, \\
& u^{q+1} = u(t^{q+1}), n^{q+1} = n(t^{q+1}), p^{q+1} = p(t^{q+1}).
\end{aligned}
\]

Scheme (115)-(118) is clearly equivalent to

\[
\{u^0, n^0, p^0\} = \{u_0, n_0, n_1\}.
\]

For \( q \geq 0 \), \( \{u^q, n^q, p^q\} \rightarrow \{u^{q+1/2}, n^{q+1/2}, p^{q+1/2}\} \rightarrow \{\hat{u}^{q+1/2}, \hat{n}^{q+1/2}, \hat{p}^{q+1/2}\} \rightarrow \{u^{q+1}, n^{q+1}, p^{q+1}\} \) via

\[
\begin{aligned}
& \left\{ \begin{array}{l}
\frac{i}{\partial t} \frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} = 0 \text{ in } (0, L) \times (t^q, t^{q+1/2}), \\
2 \frac{\partial^2 n}{\partial t^2} - \frac{\partial^2 n}{\partial x^2} = 0 \text{ in } (0, L) \times (t^q, t^{q+1/2}), \\
u(0, t) = u(L, t), \quad \frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) \text{ on } (t^q, t^{q+1/2}), \\
n(0, t) = n(L, t), \quad \frac{\partial n}{\partial x}(0, t) = \frac{\partial n}{\partial x}(L, t) \text{ on } (t^q, t^{q+1/2}), \\
u(t^q) = u^q, \quad n(t^q) = n_q, \quad \frac{\partial n}{\partial t}(t^q) = p^q/2; \\
u^{q+1/2} = u(t^{q+1/2}), \quad n^{q+1/2} = n(t^{q+1/2}), \quad p^{q+1/2} = 2 \frac{\partial n}{\partial t}(t^{q+1/2}),
\end{array} \right. \\
& \left\{ \begin{array}{l}
\frac{i}{\partial t} \frac{\partial u}{\partial t} + 2\lambda |u|^2 u + 2un = 0 \text{ in } (0, L) \times (0, \Delta t), \\
2 \frac{\partial^2 n}{\partial t^2} + \mu \frac{\partial^2 (|u|^2)}{\partial x^2} = 0 \text{ in } (0, L) \times (0, \Delta t), \\
u(0, t) = u(L, t), \quad \frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) \text{ on } (0, \Delta t), \\
n(0, t) = n(L, t), \quad \frac{\partial n}{\partial x}(0, t) = \frac{\partial n}{\partial x}(L, t) \text{ on } (0, \Delta t),
\end{array} \right. \\
u(0) = u^{q+1/2}, \quad n(0) = n^{q+1/2}, \quad \frac{\partial n}{\partial t}(0) = p^{q+1/2}/2; \\
\hat{u}^{q+1/2} = u(\Delta t), \quad \hat{n}^{q+1/2} = n(\Delta t), \quad \hat{p}^{q+1/2} = 2 \frac{\partial n}{\partial t}(\Delta t),
\end{aligned}
\]
\[
\begin{aligned}
\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} &= 0 \quad \text{in } (0, L) \times (t_{q+1/2}, t_{q+1}), \\
\frac{\partial^2 n}{\partial t^2} - \frac{\partial^2 n}{\partial x^2} &= 0 \quad \text{in } (0, L) \times (t_{q+1/2}, t_{q+1}), \\
u(0,t) = u(L,t), \quad \frac{\partial u}{\partial x}(0,t) = \frac{\partial u}{\partial x}(L,t) \quad \text{on } (t_{q+1/2}, t_{q+1}), \\
n(0,t) = n(L,t), \quad \frac{\partial n}{\partial x}(0,t) = \frac{\partial n}{\partial x}(L,t) \quad \text{on } (t_{q+1/2}, t_{q+1}), \\
u(t_{q+1/2}) = \hat{v}^{q+1/2}, \quad n(t_{q+1/2}) = \hat{n}^{q+1/2}, \quad \frac{\partial n}{\partial t}(t_{q+1/2}) = \hat{p}^{q+1/2}/2; \\
u^{q+1} = u(t_{q+1}), n^{q+1} = n(t_{q+1}), p^{q+1} = 2\frac{\partial n}{\partial t}(t_{q+1}).
\end{aligned}
\]

The linear Schrödinger and wave equations in (120) and (122) are uncoupled, implying that they can be solved by a variety of classical spectral or finite difference methods taking advantage of the space-periodic boundary conditions. On the other hand, the nonlinear system (121) can be solved pointwise: Indeed, since \(u\) and \(n\) are real-valued functions, it follows from the first and fifth equations in (121) that

\[
|u(x,t)| = |u^{q+1/2}(x)|, \quad \forall t \in [0, \triangle t], \quad x \in [0, L].
\]

(123)

It follows then from (121) and (123) that the solution \(n\) in (121) is also a solution of the following linear problem

\[
\begin{aligned}
\frac{\partial^2 n}{\partial t^2} &= -\frac{\mu}{2} \frac{\partial^2}{\partial x^2}(|u^{q+1/2}|^2) \quad \text{in } (0, L) \times (0, \triangle t), \\
n(0,t) = n(L,t) \quad \text{on } (0, \triangle t), \\
n(0) = n^{q+1/2}, \quad \frac{\partial n}{\partial t}(0) = p^{q+1/2}/2.
\end{aligned}
\]

(124)

The closed form solution of (124) is given by

\[
n(x,t) = n^{q+1/2}(x) + \frac{1}{2}p^{q+1/2}(x)t - \frac{\mu}{4}c^2 \frac{\partial^2}{\partial x^2}(|u^{q+1/2}|^2)t^2 \quad \text{on } (0, L) \times (0, \triangle t),
\]

(125)

implying, in particular, that

\[
\hat{n}^{q+1/2} = n^{q+1/2} + \frac{\lambda t}{2}p^{q+1/2} - \frac{\mu}{4}(c\triangle t)^2 \frac{\partial^2}{\partial x^2}(|u^{q+1/2}|^2).
\]

Finally, to obtain the \(u\) solution of system (121), we observe that (\(n\) being known from (125)) it is the unique solution of the following non-autonomous linear initial value problem

\[
\begin{aligned}
\frac{\partial u}{\partial t} + 2(\lambda |u^{q+1/2}|^2 + n) u &= 0 \quad \text{in } (0, L) \times (0, \triangle t), \\
u(0,t) = u(L,t) \quad \text{on } (0, \triangle t), \\
u(0) = u^{q+1/2},
\end{aligned}
\]

(126)
a particular case of
\[
\begin{cases}
\frac{\partial \phi}{\partial t} + 2(\lambda |\psi|^2 + v)\phi = 0 \text{ in } (0,L) \times (t_0, t_f), \\
\phi(0,t) = \phi(L,t) \text{ on } (t_0, t_f), \\
\phi(t_0) = \phi_0,
\end{cases}
\] (127)

\(\psi\) (resp., \(v\)) being a given complex (resp., real)-valued function of \(x\) (resp., of \(\{x,t\}\)).

With \(M \geq 1\) an integer, let us define \(\tau\), a time-discretization step, by \(\tau = \frac{t_f - t_0}{M}\), and \(t^m = t_0 + m\tau\). To solve (127) we advocate the following time-discretization scheme of the Crank-Nicolson type:
\[
\phi^0 = \phi_0.
\] (128)

For \(m = 0, \ldots, M - 1\), \(\phi^m \rightarrow \phi^{m+1}\) via the solution of
\[
\begin{cases}
\frac{\phi^{m+1} - \phi^m}{\tau} + \frac{\lambda |\psi|^2 + v(t^{m+1}) + v(t^m)}{2} (\phi^{m+1} + \phi^m) = 0 \text{ in } (0,L), \\
\phi^{m+1}(0) = \phi^{m+1}(L).
\end{cases}
\] (129)

Problem (129), can be solved point-wise (in practice at the grid-points of a finite difference one-dimensional “grid”). Scheme (128)-(129) is \textit{second-order accurate} and \textit{modulus preserving} (that is, verifies \(|\phi^{m+1}| = |\phi^m|, \forall m = 0, \ldots, M - 1\)). On \([0,L]\), \(\phi^{m+1}(x)\) is obtained via the solution of a \(2 \times 2\) linear system (for those who prefer to use real arithmetic).

\textbf{Remark 28.} In [102], one advocates using instead of \(n\) the function \(n - \mu |u|^2\). The numerical results reported in the above publication clearly show that operator-splitting provides efficient methods for the numerical solution of the Zakharov’s system (112).

6 Applications of augmented Lagrangian and ADMM algorithms to the solution of problems from Imaging

6.1 Variational models for image processing

6.1.1 Generalities

Usually, image processing refers to the processing and analysis of digital images. Variational models have become an essential part of image processing, such models relying on the minimization of a well-chosen energy functional, the minimization problem reading typically as
\[
u = \arg \min_{\nu \in V} \left[ E_{\text{fitting}}(\nu) + E_{\text{regularizing}}(\nu) \right].
\] (130)
As shown above, the energy functional has two parts, namely a fitting part and a regularizing one. In the following we will present various variational image processing models and show that the operator-splitting and ADMM methodology provides efficient methods for the numerical solution of the related minimization problems. We will start our discussion with the well-known Rudin-Osher-Fatemi (ROF) model, and then follow with the presentation of some higher order models. Before going into more details, some remarks are in order, namely:

**Remark 29.** Most of the models we are going to consider below are not fully understood yet from a mathematical point of view, two of the main issues being, in (130), the choice of the space $V$ and the weak-continuity properties of the energy functional. This will not prevent us to use these continuous models, for the simplicity of their formalism which facilitates the derivation of algorithms whose discrete analogues have provable convergence properties.

**Remark 30.** For image processing problems, the computational domain is always a rectangle, the image pixels providing a natural mesh for space discretization. This particularity makes easy, in general, the finite difference discretization of problem (130) and the implementation of iterative solution algorithms. The methodology we are going to discuss is not restricted to rectangular domains, however for domains with curved boundaries using finite-difference discretization may become complicated near the boundary; an elegant way to overcome this difficulty is to employ finite element approximations, as done in, e.g., [133].

**Remark 31.** A very detailed analysis of ADMM algorithms for the solution of image processing problems can be found in the chapter of this book by M. Burger, A. Sawatzky & G. Steidl (Chapter 10).

### 6.1.2 Total variation and the ROF model

One of the most popular variational models for image processing was proposed by Rudin, Osher, and Fatemi in their seminal work (ROF model) [144]. In [144], a denoised image is obtained by minimizing the following energy functional

$$
E(v) = \frac{1}{2} \int_{\Omega} |f - v|^2 \, dx + \eta \int_{\Omega} |\nabla v| \, dx,
$$

where: $dx = dx_1 dx_2$, $f : \Omega \rightarrow \mathbb{R}$ is a given noisy image defined on $\Omega$, $\int_{\Omega} |\nabla v| \, dx$ stands for the total variation of the trial function $v$ (see [157] and [169] for a definition of the notion of total variation), and $\eta > 0$ is a positive tuning parameter controlling how much noise will be removed. The remarkable feature of the ROF model lies in its effectiveness in preserving object edges while removing noise. In fact, the total variation regularizer has been widely employed to accomplish other image processing tasks such as deblurring, segmentation, and registration.

In order to incorporate more geometrical information into the regularizer, a number of higher order regularization models have been proposed and used for im-
age processing and computer vision problems. The ROF model has several unfavorable features. The main caveat is the stair-case effect, that is, the resulting cleaned image would present blocks even though the desired image may be smooth. Other undesirable properties include corner smearing and loss of image contrast. To remedy these drawbacks, a very rich list of results exists in the literature, see [2, 31, 120, 182, 185]. Despite the effectiveness of these models in removing the staircase effect, it is often a challenging issue to minimize the corresponding functionals. Note that if the functional $E$ contains second-order derivatives of $v$, the related Euler-Lagrange equation is a fourth-order linear or nonlinear partial differential equation.

6.1.3 Regularization using TV2

In [120], Lysaker et al. directly incorporated second order derivative information into the image denoising process, by proposing to minimize the following energy functional

$$E(v) = \frac{1}{2} \int_{\Omega} |f - v|^2 \, dx + \eta \int_{\Omega} \sqrt{(v_{x1x1})^2 + 2(v_{x1x2})^2 + (v_{x2x2})^2} \, dx$$  (132)

This higher order energy functional is much simpler than the Elastica regularizer that we shall introduce later. Numerically, this regularizer shows rather good performance with noise suppression and edge preservation. In the literature, there exists quite a number of related models, see [20, 24, 25, 26, 28, 39, 53, 58, 89, 99, 101, 134, 138, 147, 171, 146, 180]. The well-posedness of the variational problem associated with the energy functional in (132), and its gradient flow equation, have been studied in [88, 130]. High order models, such as the one associated with the energy in (132), have been discussed in, e.g., [15, 24, 32, 149, 176].

6.1.4 Regularization using the Euler’s Elastica energy

In order to ‘clean’ a given function $f : \Omega \rightarrow \mathbb{R}$, the Euler’s Elastica model relies on the minimization of the following energy functional

$$E(v) = \frac{1}{2} \int_{\Omega} |f - v|^2 \, dx + \int_{\Omega} \left[ a + b \left| \nabla \cdot \left( \frac{\nabla v}{|\nabla v|} \right) \right|^2 \right] |\nabla v| \, dx.$$  (133)

In (133), $a$ and $b$ are non-negative with $a + b > 0$. These two constants have to be chosen properly, depending of the application under consideration. The image processing model associated with the above energy functional comes from the Euler’s Elastica energy for curves (see [31, 124] for the derivation of this energy): indeed, for a given curve $\Gamma \subset \mathbb{R}^2$ with curvature $\kappa$, the Euler’s Elastica energy is defined (with obvious notation) by $\int_{\Gamma} (a + b\kappa^2) \, ds$. For a function $v$, the curvature of the
level curve $\Gamma_c := \{x|v(x) = c\}$ is $\kappa = \nabla \cdot \frac{\nabla v}{|\nabla v|}$ (if $\nabla v \neq 0$). Thus, the Euler's Elastica energy for the level curve $\Gamma_c$ is given by

$$l(c) = \int_{\Gamma_c} \left[ a + b \left| \nabla \cdot \frac{\nabla v}{|\nabla v|} \right|^2 \right] ds.$$ 

Summing up (integrating) the Euler's Elastica energy over all the level curves $\Gamma_c$, it follows from the co-area formula (see [168]) that the total Euler's Elastica energy is given by

$$\int_{-\infty}^{\infty} l(c) dc = \int_{-\infty}^{\infty} \int_{\Gamma_c} \left[ a + b \left| \nabla \cdot \frac{\nabla v}{|\nabla v|} \right|^2 \right] ds dc = \int_{\Omega} \left[ a + b \left| \nabla \cdot \frac{\nabla v}{|\nabla v|} \right|^2 \right] |\nabla v| dx.$$ 

### 6.1.5 Regularization using the image graph mean curvature

In [182], the authors proposed a variational image processing model making use of the mean curvature of the graph of function $f$, that is of the surface $\{x,y,z = f(x,y)\}$, to remove the noise. More specifically, the model considered in [182] employs the $L^1$ norm of the mean curvature of the above graph as a regularizer, the associated energy functional being defined by

$$E(v) = \frac{1}{2} \int_{\Omega} |f - v|^2 dx + \eta \int_{\Omega} \left| \nabla \cdot \frac{\nabla v}{\sqrt{1 + |\nabla v|^2}} \right| dx. \tag{134}$$ 

Above, $\eta (> 0)$ is a tuning parameter and the term $\frac{\nabla v}{\sqrt{1 + |\nabla v|^2}}$ is the mean curvature of the surface $\phi(x,y,z) = 0$ with $\phi(x,y,z) = u(x,y) - z$. Clearly, the model tries to fit the given noisy image surface $\{x,y,z = f(x,y)\}$ with a surface $\{x,y,z = u(x,y)\}$, $u$ being a minimizer of the $L^1$-mean curvature energy functional (134). This idea goes back to much earlier publications, [108] for example. The model can sweep noise while keeping object edges, and it also avoids the staircase effect. More importantly, as discussed in [185], the model is also capable of preserving image contrasts as well as object corners.

### 6.1.6 Interface problems: Chan-Vese segmentation model, labeling techniques, min-cut and continuous max-flow

In image processing, computer vision, etc., one encounters operations more complicated than denoising, segmentation being one of them. These applications require mathematical models more complicated (in some sense) than those considered in Sections 6.1.2 to 6.1.5, one of them being the Chan-Vese model introduced in [33]. Actually (as obvious from [33]), the snake and active contour model (ref. [106]) and
the Mumford-Shah model (ref. [132]) can be viewed as ancestors of the Chan-Vese model. Using the notation of [33], the Chan-Vese segmentation model relies on the minimization of the following energy functional:

$$E_{CV}(\phi, d_1, d_2) = \lambda_1 \int \Omega |f - d_1|^2 H(\phi) \, dx + \lambda_2 \int \Omega |f - d_2|^2 [1 - H(\phi)] \, dx \quad (135)$$

$$+ \mu \int \Omega |\nabla H(\phi)| \, dx + \nu \int \Omega H(\phi) \, dx,$$

where in (135): (i) \(\phi\) is a level set function whose zero level curves set represents the segmentation boundary. (ii) \(H(\cdot)\) is the Heaviside function. (iii) \(d_1\) and \(d_2\) are two real numbers. (iv) \(\lambda_1, \lambda_2\) and \(\mu\) (resp., \(\nu\)) are positive (resp., non-negative) tuning parameters (in many applications, one takes \(\lambda_1 = \lambda_2 = 1\)). The Euler-Lagrange equation associated with the minimization of the functional in (135) has been derived in [33]. In the above reference the associated gradient flow has been time-discretized by an explicit scheme to compute the solution of the above minimization problem (after an appropriate finite difference space discretization). Operator-splitting and ADMM can be used to develop algorithms with much faster convergence properties than the above explicit schemes; we will return on this issue in Section 6.2. Let us denote \(H(\phi)\) by \(v\); there is clearly equivalence between minimizing the functional defined by (135) and

$$\inf_{\{v, d_1, d_2\} \in \mathbb{V} \times \mathbb{R} \times \mathbb{R}} \left\{ \frac{\lambda_1}{2} \int \Omega |f - d_1|^2 \, v \, dx + \frac{\lambda_2}{2} \int \Omega |f - d_2|^2 [1 - v] \, dx \right.$$

$$+ \mu \int \Omega |\nabla v| \, dx + \nu \int \Omega v \, dx \right\}, \quad (136)$$

where \(\mathbb{V} = \{v \in L^\infty(\Omega), v(\cdot) \in \{0, 1\}, \text{a.e. in } \Omega, \nabla v \in L^1(\Omega)\}\). The model associated with (136) was proposed in [117] and referred as a binary level set based model. More generally, we can consider the minimization, over the above set \(V\), of energy functionals such as \(E_{\text{potts}}\) defined by

$$E_{\text{potts}}(v) = \int \Omega f_1 v \, dx + \int \Omega f_2 [1 - v] \, dx + \int \Omega g \nabla v \, dx, \quad (137)$$

where \(f_1\) and \(f_2\) are given functions indicating the possibility that a point belongs to phase 0 or to phase 1, and where \(g\) is a non-negative function, possibly constant; if \(d_1\) and \(d_2\) are fixed in (136), the Chan-Vese model becomes a particular case of the model associated with the functional \(E_{\text{potts}}\) defined by (137). It was recently observed (see [173, 175]) that minimizing \(E_{\text{potts}}\) over the above \(V\) is a (kind of) continuous min-cut problem, itself equivalent (by duality) to a max-flow problem. Indeed, let us consider the following continuous max-flow problem

$$\begin{align*}
\sup_{q_1, q_2} & \int \Omega q_1 \, dx \text{ subject to} \\
q_1 \leq f_1, & q_2 \leq f_2, \, |v| \leq g, \\
\nabla \cdot v = q_1 - q_2, & \text{in } \Omega, \, v \cdot n = 0 \text{ on } \Gamma (\, = \partial \Omega),
\end{align*} \quad (138)$$
where in (138): (i) \( \mathbf{v} = \{v_1, v_2\} \) and \( |\mathbf{v}| = \sqrt{v_1^2 + v_2^2} \), \( \mathbf{v} \) being the flow inside \( \Omega \). (ii) \( \mathbf{n} \) is the unit outward vector normal at \( \Gamma \). (iii) \( q_s \) (resp., \( q_t \)) represents a flow from a source (resp., to a sink). (iv) \( f_1 \) and \( f_2 \) are as in (137). We can also define \( |\mathbf{v}| \) by \( |\mathbf{v}| := |v_1| + |v_2| \); if we do so, the discretized max-flow problem can be solved by traditional graph cut methods. It follows from [175] that a dual of the max flow problem (138) reads as:

\[
\inf_{\mu \in \Lambda} \left[ \int_{\Omega} f_1(1 - \mu) \, dx + \int_{\Omega} f_2 \mu \, dx + \int_{\Omega} g |\nabla \mu| \, dx \right],
\]

(139)

where \( \Lambda = \{\mu | \mu \in L^\infty(\Omega), 0 \leq \mu(x) \leq 1, \text{a.e. in } \Omega \} \cap W^{1,1}(\Omega) \). We have recovered thus the functional \( E_{\text{potts}} \) from (137) and shown a link between the Chan-Vese model and the max-flow problem. The dual problem (139) is known as a (continuous) min-cut problem. Actually, Chan, Esedoglu and Nikolova have shown in [29] that there is equivalence between (139) and minimizing over \( V = \{v|v| \in L^\infty(\Omega), v(x) \in \{0, 1\}, \text{a.e. in } \Omega, \nabla v \in L^1(\Omega) \} \) the functional \( E_{\text{potts}} \) defined by (137), a most remarkable result indeed since problem (139) is a convex variational problem whose discrete variants can be solved by ADMM type algorithms (see [5, 6, 7, 8, 114, 173, 174, 175, 178] for more details and generalizations).

Remark 32. In (136), (138) and (139), it is on purpose that we used inf (resp., sup) instead of min (resp., max) since we have no guarantee that the minimizing sequences of the functionals under consideration will converge weakly in the space or set where the minimization takes place.

Remark 33. Suppose that in (138) we replace the constraint \( |\mathbf{v}| \leq g \) by \( |v_1| \leq g_1 \) and \( |v_2| \leq g_2 \), everything else being the same; then, the dual problem of the associated variant of (138) reads (with \( \Lambda \) as in (139)) as

\[
\inf_{\mu \in \Lambda} \left[ \int_{\Omega} f_1(1 - \mu) \, dx + \int_{\Omega} f_2 \mu \, dx + \int_{\Omega} \left( g_1 \left| \frac{\partial \mu}{\partial x_1} \right| + g_2 \left| \frac{\partial \mu}{\partial x_2} \right| \right) \, dx \right],
\]

clearly a close variant of (139). Similarly, if we replace in (138) the constraint \( |\mathbf{v}| \leq g \) by \( |v_1| + |v_2| \leq g \), we obtain (as expected) the following dual problem

\[
\inf_{\mu \in \Lambda} \left[ \int_{\Omega} f_1(1 - \mu) \, dx + \int_{\Omega} f_2 \mu \, dx + \int_{\Omega} g \sup \left( \left| \frac{\partial \mu}{\partial x_1} \right|, \left| \frac{\partial \mu}{\partial x_2} \right| \right) \, dx \right],
\]

the set \( \Lambda \) being as above.

### 6.1.7 Segmentation models with higher order regularization

As could have been expected, first order segmentation models have limitations (discussed in [132]). To give an example let us consider the situation depicted in Figure 8(a) where some parts of the four letters have been erased; albeit one can easily recognize the four letters, first order segmentation models such as Chan-Vese’s, might
often capture the existing boundary instead of restoring the missing ones, as illustrated in Figure 8(b). In *inpainting* problems (see [31, 124]), missing image information is also recovered, but within given regions assigned in advance. In contrast, one would like to have a segmentation model that can interpolate the missing boundaries automatically without specifying the region of interest. To this end, one may employ the Euler’s Elastica functional as a novel regularization term in the Chan-Vese’s model (135), in order to replace the weighted $TV$ term. Doing so we obtain the following energy functional (we assume $\nu = 0$, here):

$$E_{CVE}(\phi, d_1, d_2) = \lambda_1 \int_\Omega |f - d_1|^2 H(\phi) \, dx + \lambda_2 \int_\Omega |f - d_2|^2 [1 - H(\phi)] \, dx$$

$$+ \left[ a + b (\nabla \cdot \frac{\nabla \phi\vert}{\|\nabla \phi\|}) \right] \| \nabla H(\phi) \| \, dx$$

(140)

where $\lambda_1, \lambda_2, a$ and $b$ are positive parameters. If $\phi$ is the signed distance level set function, it can be proved that the last term in (140) is equal to the Euler’s elastica energy of the segmentation curve. This regularization was originally proposed and used in the celebrated paper on *segmentation with depth* by Nitzberg, Mumford and Shiota (ref. [135]). Actually, it has also been used in [31] (resp., [183, 184]) for the solution of the in-painting (resp., illusory contour) problem. In [146], linear programming was used to minimize (after space discretization) curvature dependent functionals, the functional defined by (140) being one of those considered in this article.

Remark 34. Observe that since (formally at least, but this can be justified using a well-chosen regularization of the Heaviside function, such as $\xi \rightarrow \frac{1}{2} \left[ 1 + \frac{\xi}{\sqrt{\varepsilon^2 + \xi^2}} \right]$)

$$\frac{\nabla \phi}{\|\nabla \phi\|} = \frac{\nabla H(\phi)}{\|\nabla H(\phi)\|},$$

only the sign $H(\phi)$ of the function $\phi$ is needed when solving the segmentation problem via the functional in (140). This property suggests, as done in [117], to use a binary level set representation via the introduction of the function $\nu = H(\phi)$. Such a change of function was also used in [29] for finding the global minimizer associated with the Chan-Vese’s model. More general binary level set
representations with global minimization techniques have been developed (see, e.g., [7, 173, 174, 175, 177]) using the relationships existing between graph cuts, binary labeling and continuous max flow problems. Since \( \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} = \nabla \cdot \frac{\nabla H(\phi)}{|\nabla H(\phi)|} \), one can rewrite the functional in (140) as

\[
E(v, d_1, d_2) = \lambda_1 \int_{\Omega} |f - d_1|^2 v \, dx + \lambda_2 \int_{\Omega} |f - d_2|^2 |1 - v| \, dx + \left[ a + b \left( \nabla \cdot \frac{\nabla v}{|\nabla v|} \right) \right] |\nabla v| \, dx
\]

with the values taken by \( v \) being either 0 or 1. Strictly speaking the mean curvature of the graph makes sense for “smooth” functions only; to fix this issue, one relaxes the above binary restriction by replacing it by \( 0 \leq v \leq 1 \), a less constraining condition indeed.

6.2 Fast numerical algorithms for variational image processing models based on operator-splitting and augmented Lagrangian methods (ALM)

In this section, we will present operator-splitting and ALM based fast numerical algorithms, for the numerical treatment of variational image processing models.

6.2.1 Parallel splitting schemes for the ROF model

The first model that we are going to consider is the ROF model discussed in Section 6.1.2. The formal \textit{Euler-Lagrange equation} associated with the minimization of the (strictly convex) functional in (131) reads as

\[
-\eta \nabla \cdot \frac{\nabla u}{|\nabla u|} + u = f \text{ in } \Omega, \quad \frac{\nabla u}{|\nabla u|} \cdot n = 0 \text{ on } \partial \Omega. \tag{142}
\]

with \( n \) the outward unit vector normal at \( \partial \Omega \). In order to solve the nonlinear non-smooth elliptic equation (142) we associate with it an initial value problem and look for steady-state solutions. We consider thus

\[
\begin{cases}
\frac{\partial u}{\partial t} - \eta \nabla \cdot \frac{\nabla u}{|\nabla u|} + u = f \text{ in } \Omega \times (0, +\infty), \\
\frac{\nabla u}{|\nabla u|} \cdot n = 0 \text{ on } \partial \Omega \times (0, +\infty), \\
u(0) = u_0,
\end{cases} \tag{143}
\]
an obvious choice for $u_0$ in (143) being $u_0 = f$. Actually to overcome the difficulty associated with the non-smoothness of the elliptic operator in (142) and (143), we consider the following regularized variant of (143):

$$
\begin{aligned}
\frac{\partial u}{\partial t} - \eta \nabla \cdot \frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}} + u &= f \quad \text{in } \Omega \times (0, +\infty), \\
\frac{\partial u}{\partial n} &= 0 \quad \text{on } \partial \Omega \times (0, +\infty), \\
\left. u \right|_{t=0} &= u_0,
\end{aligned}
$$

(144)

with $\varepsilon$ a small positive number. The simplest time-stepping scheme we can think about to capture the steady state solution of (144) is clearly the forward-Euler scheme. Let $\triangle t (> 0)$ be a time-discretization step; applied to the solution of (144) the forward Euler scheme produces the following algorithm:

$$
u^n = u^n.
$$

(145)

For $n \geq 0$, $u^n \to u^{n+1}$ via

$$
\begin{aligned}
\frac{u^{n+1} - u^n}{\triangle t} - \eta \nabla \cdot \frac{\nabla u^n}{\sqrt{|\nabla u^n|^2 + \varepsilon^2}} + u^n &= f \quad \text{in } \Omega, \\
\frac{\partial u^{n+1}}{\partial n} &= 0 \quad \text{on } \partial \Omega.
\end{aligned}
$$

(146)

In practice, scheme (145)-(146) is applied to a discrete variant of (144) obtained by finite difference or finite element space discretization. Scheme (145)-(146) being explicit and the condition number of the operator in (146) rather large, its conditional stability requires small time steps leading to a slow convergence to a steady state solution. Suppose that $\Omega$ is the rectangle $(0, a) \times (0, b)$; in order to improve the speed of convergence to a steady state solution, we are going to apply to the solution of (144) the parallelizable operator-splitting scheme discussed in Section 2.8, taking advantage of the following decomposition of the operator in (144)

$$
-\eta \nabla \cdot \frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}} + u - f = A_1(u) + A_2(u),
$$

(147)

with

$$
\begin{aligned}
A_1(u) &= -\eta \frac{\partial}{\partial x_1} \left( \frac{\frac{\partial u}{\partial x_1}}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \right) + \frac{1}{2} (u - f), \\
A_2(u) &= -\eta \frac{\partial}{\partial x_2} \left( \frac{\frac{\partial u}{\partial x_2}}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \right) + \frac{1}{2} (u - f).
\end{aligned}
$$

(148)
Combining the scheme we mentioned just above with a semi-explicit time discretization of the nonlinear terms we obtain

\[ u^0 = u_0. \]  

For \( n \geq 0 \), \( u^n \rightarrow \{ u^{n+1/4}, u^{n+2/4} \} \rightarrow u^{n+1} \) via

\[
\begin{cases}
\frac{u^{n+1/4} - u^n}{2\Delta t} - \eta \frac{\partial}{\partial x_1} \left( \frac{\partial u^{n+1/4}}{\partial x_1} \right) + \frac{u^{n+1/4}}{2} = f \quad \text{in } \Omega, \\
\frac{\partial u^{n+1/4}}{\partial x_1}(0, x_2) = \frac{\partial u^{n+1/4}}{\partial x_1}(a, x_2) = 0 \quad \forall x_2 \in (0, b),
\end{cases}
\]

(150.1)

\[
\begin{cases}
\frac{u^{n+2/4} - u^n}{2\Delta t} - \eta \frac{\partial}{\partial x_2} \left( \frac{\partial u^{n+2/4}}{\partial x_2} \right) + \frac{u^{n+2/4}}{2} = f \quad \text{in } \Omega, \\
\frac{\partial u^{n+2/4}}{\partial x_2}(x_1, 0) = \frac{\partial u^{n+2/4}}{\partial x_2}(x_1, b) = 0 \quad \forall x_1 \in (0, a),
\end{cases}
\]

(150.2)

\[ u^{n+1} = \frac{1}{2}(u^{n+1/4} + u^{n+2/4}). \]  

(151)

Scheme (149)-(151) can accommodate large time steps implying a fast convergence to steady state solutions. It preserves also the symmetry of the images. Moreover, since in most applications \( \Omega \) is a rectangle with the image pixels uniformly distributed on it, it makes sense to use a finite difference discretization on a uniform Cartesian grid to approximate (150.1) and (150.2). For Dirichlet or Neumann boundary conditions, the finite difference discretization of (150.1) and (150.2) will produce two families of uncoupled tri-diagonal linear systems easily solvable (the good parallelization properties of the above scheme are quite obvious). The above operator-splitting scheme can be generalized to the numerical treatment of other variational models (such as Chan-Vese’s, and to models involving derivatives of order higher than one, as shown in, e.g., [89]). A closely related scheme is discussed in [167].

### 6.2.2 A split-Bregman method and related ADMM algorithm for the ROF model

In ref. [86], T. Goldstein and S. Osher proposed and tested a fast converging iterative method for the ROF model: this algorithm, of the split-Bregman type, is certainly one of the fastest numerical methods for the ROF model. It was quickly realized (see [156, 170, 172]) that the Bregman algorithm discussed in [86] is equivalent to an ADMM one. Here, we will explain the ideas in an informal way using the continuous model whose formalism is much simpler. As stated in Remark 29, to make
our discussion more rigorous mathematically, the functional spaces for which the continuous model makes sense have to be specified (here, they are of the \textit{bounded variation} type). This difficulty is one of the reasons explaining why some authors (as in [170]) consider discrete models, directly.

Let us denote \( \nabla u \) by \( p \); then, it is easy to see that (from (131)) the ROF model is equivalent to the following linearly constrained minimization problem:

\[
\{u, p\} = \arg \min_{v, q \in \mathbb{R}} \left\{ \eta \int_{\Omega} |q| \, dx + \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx \right\}. \tag{152}
\]

Clearly, problem (152) belongs to the family of variational problems discussed in Section 3.2, the associated augmented Lagrangian being defined (with \( r > 0 \)) by (see, e.g., [72] (Chapter 4)):

\[
\mathcal{L}_{rof}(v, q; \mu) = \eta \int_{\Omega} |q| \, dx + \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \frac{r}{2} \int_{\Omega} \|
abla v - q \|^2 \, dx + \int_{\Omega} \mu \cdot (\nabla v - q) \, dx. \tag{153}
\]

Above, \( u : \Omega \rightarrow \mathbb{R} \) denotes the restored image we are looking for, \( p = \nabla u \), \( \mu \) is a multiplier. Due to the strict convexity of the second term, the discrete analogues of the minimization problem (152) have a unique solution. Applying algorithm \textsc{Alg}2 of Section 3.2.2 to the solution of (152) we obtain the following

\textbf{Algorithm 6.1: An augmented Lagrangian method for the ROF model}

0. Initialization: \( \lambda^0 = 0 \), \( u^0 = f \).

For \( k = 0, 1, \ldots \), until convergence:

1. Compute \( p^{k+1} \) from

\[
p^{k+1} = \arg \min_{q \in (L^2(\Omega))^2} \mathcal{L}_{rof}(u^k, q; \lambda^k). \tag{154}
\]

2. Compute \( u^{k+1} \) from

\[
u^{k+1} = \arg \min_{v \in H^1(\Omega)} \mathcal{L}_{rof}(v, p^{k+1}; \lambda^k). \tag{155}
\]

3. Update \( \lambda^k \) by

\[
\lambda^{k+1} = \lambda^k + r \langle \nabla u^{k+1} - p^{k+1} \rangle. \tag{156}
\]

It was observed in [156, 170] that this augmented Lagrangian algorithm is equivalent to the split-Bregman algorithm discussed in [86]. This equivalence is also explained in [172] for compressive sensing models. The minimization sub-problems (154) have closed form solutions which can be computed point-wise; solving them is thus quite easy. The minimization sub-problems (155) (in fact their discrete analogues) reduce to discrete well-posed linear Neumann problems; the associated ma-
trix being symmetric, positive definite and sparse, these discrete elliptic problems can be solved by a large variety of direct and iterative methods (among them: sparse Cholesky, multi-level, Gauss-Seidel, conjugate gradient, FFT, etc.; see [170, 172] for more details). The convergence of algorithm (154)-(156) is discussed in [170].

**Remark 35.** As described above, Algorithm 6.1 is largely formal since it operates in the space \( W = [H^1(\Omega) \times (L^2(\Omega))^2] \times (L^2(\Omega))^2 \), although the solution \( u \) of problem (131) may not have enough regularity to belong to \( H^1(\Omega) \). However, Algorithm 6.1 makes sense for the discrete analogues of problem (131) and space \( W \) obtained by finite difference or finite element approximation; for finite element approximations in particular, the formalisms of Algorithm 6.1 and of its discrete counterparts are nearly identical. The above observation applies to most of the ADMM algorithms described below (see Remark 36, for example).

**Remark 36.** As shown in, e.g., [109] (for image denoising applications), Algorithm 6.1 is easy to modify in order to handle those situations where the functional \( \int_{\Omega} |\nabla v| \, dx \) is replaced by \( \frac{1}{s} \int_{\Omega} |\nabla v|^s \, dx \) with \( 0 < s < 1 \), or by other non-convex functionals of \( |\nabla v| \); once discretized, these modifications of Algorithm 6.1 perform very well as shown in [109].

**Remark 37.** It is easy to extend algorithm (154)-(156) to the solution of the min-cut problem (139), since the additional constraint encountered in this last problem, namely \( 0 \leq \mu(x) \leq 1 \), a.e. in \( \Omega \), is (relatively) easy to treat; actually, this extension has been done in [22] (see also [4, 27], and Section 6.2.3, below, for a number of related new approaches). As shown in [170] (page 320), and [142, 143], it is also easy to extend algorithm (154)-(156) to those situations where one uses vector-TV regularization in order to process vector-valued data.

### 6.2.3 An augmented Lagrangian method for the continuous min-cut and max-flow problems

The continuous max-flow problems (138) and (139) are dual to each other in the sense that if the function \( \lambda \) is solution of (139), it is a *Lagrange multiplier* for the flow conservation equation in (138). We can solve both problems simultaneously using a primal-dual method à la ALG2 relying on the following *augmented Lagrangian* functional

\[
\mathcal{L}_c(q_1, q_2, v; \mu) = -\int_{\Omega} q_1 \, dx - \int_{\Omega} \mu(\nabla \cdot v - q_1 + q_2) \, dx + \frac{r}{2} \int_{\Omega} (\nabla \cdot v - q_1 + q_2)^2 \, dx,
\]

(157)

where in (157): \( r > 0 \), and \( q_1 \), \( q_2 \), and \( v \) verify, a.e. in \( \Omega \), \( q_1 \leq f_1 \), \( q_2 \leq f_2 \), \( |v| \leq g \); here \( |v| = \sqrt{v_1^2 + v_2^2} \) for all \( v \). Applying ALG2 to the computation of the saddle-points of \( \mathcal{L}_c \) over the set \( (Q_1 \times Q_2) \times L^2(\Omega) \), where \( Q_1 = \{ q \in L^2(\Omega), q \leq f_1 \} \), \( Q_2 = \{ q \in L^2(\Omega), q \leq f_2 \} \), and \( K = \{ v \in (L^2(\Omega))^2, \nabla \cdot v \in L^2(\Omega), v \cdot n = 0 \text{ on } \Gamma, |v| \leq g \} \), we obtain...
**Algorithm 6.2: An augmented Lagrangian method for the continuous max-flow problem**

0. Initialization: $\lambda^0 = 0$, $p_s^0 = f_1$, $p_t^0 = f_2$.

For $k = 0, 1, \ldots$, until convergence:

1. Compute $u^{k+1}$ from

   \[
   u^{k+1} = \arg \min_{v \in K} \mathcal{L}_c(p_s^k, p_t^k, v, \lambda^k).
   \]

   (158)

2. Compute $\{p_s^{k+1}, p_t^{k+1}\}$ from

   \[
   \{p_s^{k+1}, p_t^{k+1}\} = \arg \min_{\{q_s, q_t\} \in Q_1 \times Q_2} \mathcal{L}_c(q_s, q_t, u^{k+1}, \lambda^k).
   \]

   (159)

3. Update $\lambda^k$ by

   \[
   \lambda^{k+1} = \lambda^k - r(\nabla \cdot u^{k+1} - p_s^{k+1} + p_t^{k+1}).
   \]

   (160)

We observe that (159) has a closed form solution (and that $p_s^{k+1}$ and $p_t^{k+1}$ can be computed point-wise independently of each other). The sub-problem (158) is a simple variant of the dual of the ROF problem (that is, the unconstrained minimization of the functional in (131)). We just need to solve this problem approximately; indeed, in our implementations we just used few steps of a descent algorithm, followed by a projection on the convex set $\{v | v \in (L^2(\Omega))^2, |v| \leq g\}$ (see \[173\], \[175\] for more details on the solution of these sub-problems). The discrete variant of algorithm (158)-(160) that we implemented (via a finite difference discretization) proved being very robust with respect to initialization and to the value of the augmentation parameter $r$; it is also very efficient computationally.

**Remark 38.** As written, algorithm (158)-(160) is applicable only to the solution of two-phase flow problems. There are several ways to generalize this algorithm to models involving more than two phases, as shown in, e.g., \[5, 6, 7, 8, 173, 177\]. Also, we would like to emphasize the fact that the discrete analogue of algorithm (158)-(160) we implemented has good convergence properties no matter which of the following two norms we used for the flow constraint in (138) (see Remark 33 for the dual formulation associated with (162)):

\[
|v|_2 = \sqrt{v_1^2 + v_2^2}
\]

(161)

or

\[
|v|_1 = |v_1| + |v_2|.
\]

(162)

If one uses the meshes classically used in digital imaging, traditional graph cut methods (like those discussed in \[19\]) can be used to solve the discrete min-cut and max-flow problems if one uses the norm defined by (162) to bound $v$. On the other hand, the above mentioned graph cut methods cannot handle the norm defined by (161). It
is also known that the solutions of the discrete min-cut and max-flow problems suffer from the matrication error if the norm in (162) is used. Compared to graph cut methods, ADMM algorithms such as (158)-(160) can handle both norms without particular difficulty. Moreover, these augmented Lagrangian algorithms are easy to parallelize and to implement on GPUs; also, they use much less memory than traditional graph cut methods; this enables using these algorithms for high dimensional and large size images or data.

6.2.4 A split-Bregman method and related ADMM algorithm for a second order total variation model

Here, we will discuss the application of \textit{ALG2} (that is ADMM) to the solution of those image processing problems associated with the functional defined by (132) (also known as the \textit{TV2} model). The presentation follows [42, 73, 170], where the main ideas are: (i) Transfer the burden of nonlinearity from the Hessian

\[ \mathbf{D}^2 u = \left( \begin{array}{ccc} \partial^2 u / \partial x_1^2 & \partial^2 u / \partial x_1 \partial x_2 & \partial^2 u / \partial x_2^2 \\ \partial^2 u / \partial x_1 \partial x_2 & \partial^2 u / \partial x_2^2 & \partial^2 u / \partial x_1^2 \end{array} \right) \]


to an additional unknown \( p \), via the relation

\[ p = \mathbf{D}^2 u, \quad (163) \]

and (ii) Use a well-chosen augmented Lagrangian functional, associated with the linear relation (163). A similar idea has been (successfully) used in [42] for the augmented Lagrangian solution of the Dirichlet problem for the Monge-Ampère equation \( \det \mathbf{D}^2 u = f \) (see also Chapter 8 of this book).

Back to the \textit{TV2} model (132), let us recall that the related minimization problem reads as

\[ u = \arg \min_{v \in V} \left[ \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \eta \int_{\Omega} |\mathbf{D}^2 v| \, dx \right], \quad (164) \]

with \( V = \{ v \in L^2(\Omega), \mathbf{D}^2 v \in (L^1(\Omega))^2 \times 2 \} \) and \( |M| = \sqrt{\sum_{1 \leq i,j \leq 2} m_{ij}^2} \) denoting the Frobenius norm of matrix \( M \). Proceeding as in Section 3.2.2, we observe the equivalence between (164) and

\[ \{ u, \mathbf{D}^2 u \} = \arg \min_{\{ v, q \} \in W} \left[ \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \eta \int_{\Omega} |q| \, dx \right], \quad (165) \]

where

\[ W = \{ \{ v, q \} | v \in V, q \in (L^1(\Omega))^{d \times d}, \mathbf{D}^2 v - q = 0 \}, \]

an observation leading us to introduce the following augmented Lagrangian functional
\[ \mathcal{L}_{TV^2}(v, q; \mu) = \frac{1}{2} \int_{\Omega} |v - f|^2 dx + \eta \int_{\Omega} |q| dx 
+ \frac{r}{2} \int_{\Omega} |D^2 v - q|^2 dx + \int_{\Omega} \mu : (D^2 v - q) dx, \]  \tag{166}

where, in (166), \( r > 0 \), and (with obvious notation) \( S : T = \sum_{i,j \leq 2} s_{ij} t_{ij} \). Applying the methods discussed in Section 3.2.2 to the solution of the minimization problem (164) we obtain the following

**Algorithm 6.3: An augmented Lagrangian method for the TV2 model**

0. Initialization: \( \lambda^0 = 0, u^0 = f \).

For \( k = 0, 1, \ldots \), until convergence:

1. Compute \( p^{k+1} \) from
   \[ p^{k+1} = \arg \min_{q \in (L^2(\Omega))^2 \times 2} \mathcal{L}_{TV^2}(u^k, q; \lambda^k). \]  \tag{167}

2. Compute \( u^{k+1} \) from
   \[ u^{k+1} = \arg \min_{v \in H^2(\Omega)} \mathcal{L}_{TV^2}(v, p^{k+1}; \lambda^k). \]  \tag{168}

3. Update \( \lambda^k \) by
   \[ \lambda^{k+1} = \lambda^k + r(D^2 u^{k+1} - p^{k+1}). \]  \tag{169}

As with Algorithm 6.1 (that is (154)-(156)), the sub-problems (167) have closed-form solutions which can be computed point-wise. On the other hand, the sub-problems (168) reduce to linear bi-harmonic problems for the elliptic operator \( I + r\nabla^4 \); if properly discretized on a uniform grid (typically by finite differences), the discrete analogues of these bi-harmonic problems can be solved by FFT or by iterative methods (see [170] (page 324) for details).

**Remark 39.** Obviously, Remark 35 applies also to Algorithm 6.3, with \( H^2(\Omega) \) playing here the role of \( H^1(\Omega) \) there.

### 6.2.5 An augmented Lagrangian method for the Euler’s Elastica model

The energy functional defined by (133), namely

\[ E(v) = \frac{1}{2} \int_{\Omega} |f - v|^2 dx + \int_{\Omega} \left[ a + b \left| \nabla \frac{\nabla v}{|\nabla v|} \right|^2 \right] |\nabla v| dx, \]

makes no sense on the subset of \( \Omega \) where \( \nabla v \) vanishes. Following an approach very common in visco-plasticity (see, e.g., [66, 83]) one make things more rigorous by defining (following [154]) the energy functional by
\[ E(v, m) = \frac{1}{2} \int_{\Omega} |f - v|^2 \, dx + \int_{\Omega} \left[ a + b |\nabla v|^2\right] |\nabla v| \, dx \tag{170} \]

the functions \(v\) and \(m\) in (170) verifying
\[ |\nabla v| = m \cdot \nabla v, \quad |m| \leq 1. \tag{171} \]

The related minimization problem reads as
\[
\begin{cases}
\{u, n\} = \arg \min_{\{v, m\}} E(v, m), \\
\text{with } \{v, m\} \text{ verifying (171)}.
\end{cases}
\tag{172}
\]

Introducing the vector-valued function \(p\) verifying \(p = \nabla u\), we clearly have equivalence between (172) and
\[
\begin{cases}
\{u, p, n\} = \arg \min_{\{v, q, m\}} \left[ \frac{1}{2} \int_{\Omega} |f - v|^2 \, dx + \int_{\Omega} \left[ a + b |\nabla v|^2\right] |q| \, dx \right], \\
\text{with } \{v, q, m\} \text{ verifying } q = \nabla v, \quad |q| = m \cdot q, \quad |m| \leq 1.
\end{cases}
\tag{173}
\]

Following [154], we associate with the minimization problem (173) the following augmented Lagrangian functional
\[
\mathcal{L}_{elast}\{v, q, m; \mu_1, \mu_2\} = \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \int_{\Omega} \left[ a + b |\nabla v|^2\right] |q| \, dx \\
+ \frac{r_1}{2} \int_{\Omega} |\nabla v - q|^2 \, dx + r_2 \int_{\Omega} (|q| - q \cdot m) \, dx + \int_{\Omega} \mu_1 \cdot (\nabla v - q) \, dx + \int_{\Omega} \mu_2 (|q| - q \cdot m) \, dx, \tag{174}
\]

with \(r_1\) and \(r_2\) both positive. Suppose that in (174) the vector-valued function \(m\) belongs to \(M\), the closed convex set of \((L^2(\Omega))^2\) defined by
\[
M = \{m | m \in (L^2(\Omega))^2, |m(x)| \leq 1, \text{a.e. in } \Omega\};
\]

we have then \(|q| - q \cdot m \geq 0\), implying (since \(|q| - q \cdot m = ||q| - q \cdot m||\)) that the variant of ALG2 described just below will force the condition \(|q| - q \cdot m = 0\) in the sense of \(L^1(\Omega)\). This variant of ALG2 reads as follows when applied to the solution of problem (172) (below, \(H(\Omega; \text{div}) = \{v \in (L^2(\Omega))^2, \nabla \cdot v \in L^2(\Omega)\}\)):

**Algorithm 6.4:** An augmented Lagrangian method for the Euler’s Elastica model

0. Initialization: \(\lambda_1^0 = 0, \lambda_2^0 = 0, u^0 = f, n^0 = 0\).

For \(k = 0, 1, \ldots\), until convergence:

1. Compute \(p^{k+1}\) from
\[
p^{k+1} = \arg \min_{q \in (L^2(\Omega))^2} \mathcal{L}_{elast}(u^k, q, n^k; \lambda_1^k, \lambda_2^k). \tag{175}\]
2. Compute \( n^{k+1} \) from
\[
n^{k+1} = \arg \min_{m \in H(\Omega; \text{div}) \cap M} \mathcal{L}_{\text{elas}}(u^k, p^k, m; \lambda^k_1, \lambda^k_2).
\] (176)

3. Compute \( u^{k+1} \) from
\[
u^{k+1} = \arg \min_{v \in H^1(\Omega)} \mathcal{L}_{\text{elas}}(v, p^k, n^k; \lambda^k_1, \lambda^k_2).
\] (177)

4. Update \( \{ \lambda^k_1, \lambda^k_2 \} \) by
\[
\begin{align*}
\lambda^{k+1}_1 &= \lambda^k_1 + r_1 (\nabla u^{k+1} - p^{k+1}), \\
\lambda^{k+1}_2 &= \lambda^k_2 + r_2 (p^{k+1} - p^{k+1} - n^{k+1}).
\end{align*}
\] (178)

Below, we will give some details and comments about the solution of the sub-problems encountered when applying algorithm (175)-(178); implementation issues will be also addressed. Further information is provided in [154].

- The minimization sub-problem (175) has a unique closed-form solution which can be computed point-wise.
- The minimization sub-problem (176) is equivalent to the following elliptic variational inequality
\[
\begin{align*}
\begin{cases}
 n^{k+1} & \in H(\Omega; \text{div}) \cap M, \\
 b \int_\Omega |p^{k+1}| \nabla \cdot n^{k+1} \cdot (m - n^{k+1}) \, dx & \geq \int_\Omega (r_2 + \lambda^k_2) p^{k+1} \cdot (m - n^{k+1}) \, dx, \\
 \forall m & \in H(\Omega; \text{div}) \cap M.
\end{cases}
\end{align*}
\] (179)

We observe that the bilinear functional in the left-hand side of (179) is symmetric and positive semi-definite (indeed, \( \int_\Omega |p^{k+1}| (\nabla \cdot m)^2 \, dx = 0 \) if \( m = \nabla \times z \). However, the boundedness of \( M \) implies that the variational problem (176), (179) has solutions. For the solution of the discrete analogues of the above problem we advocate using few iterations of those relaxation methods with projection discussed in, e.g., [66, 76] (other methods are possible as shown in [154]).

- The minimization sub-problem (177) has a unique solution characterized by
\[
\begin{align*}
u^{k+1} & \in H^1(\Omega), \\
\int_\Omega u^{k+1} v \, dx + r_1 \int_\Omega \nabla u^{k+1} \cdot \nabla v \, dx & = \int_\Omega f v \, dx + \int_\Omega (r_1 p^{k+1} - \lambda^k_1) \cdot \nabla v \, dx, \\
\forall v & \in H^1(\Omega).
\end{align*}
\] (180)

Actually, (180) is nothing but a variational formulation of the following Neumann problem
Remark 40. In [154], an alternative method for the solution of the Euler's Elastica problem (172) is also considered. It relies on the equivalence between (172) and an augmented Lagrangian associated with (182) is clearly the one defined by

\[
\begin{align*}
\mathcal{L}_{\text{elas}} \left( \nu, q, m_1, m_2; \mu_1, \mu_2, \mu_3 \right) &= \frac{1}{2} \int_{\Omega} |\nu - f|^2 dx + \int_{\Omega} \left[ a + b |\nabla \cdot m_1|^2 \right] |q| dx \\
&+ \frac{r_1}{2} \int_{\Omega} |\nabla \nu - q|^2 dx + r_2 \int_{\Omega} (|q| - q \cdot m_2) dx + r_3 \int_{\Omega} |m_1 - m_2|^2 dx \\
&+ \frac{1}{2} \int_{\Omega} |\nabla \nu - q|^2 dx + \int_{\Omega} \mu_2 (|q| - q \cdot m_2) dx + \int_{\Omega} \mu_3 \cdot (m_1 - m_2) dx, 
\end{align*}
\]

where, in (181), \( \nu \) denotes the outward unit vector normal at the boundary \( \partial \Omega \) of \( \Omega \). The numerical solution of linear elliptic problems such as (181) is routine nowadays; after an appropriate space discretization it can be achieved by a large variety of direct and iterative methods (sparse Cholesky, FFT, relaxation, multilevel, etc.).

- Since the energy functional associated with the Euler's Elastica is non-convex (see (170)) the augmentation parameters \( r_1 \) and \( r_2 \) have to be chosen large enough to guarantee the convergence of algorithm (175)-(179). Actually, the tuning of \( r_1 \) and \( r_2 \) is a delicate issue in itself and we can expect (as shown for example in [133], for a problem involving three augmentation parameters) the optimal values of these parameters to be of different orders of magnitude with respect to the space discretization \( h \).

- Another solution method for the Euler's Elastica is discussed in [21]. It relies on tractable convex relaxation in higher dimension.

Remark 41. As shown in [186], the methodology we employed to solve the minimization problem (172) can be easily modified in order to handle the Chan-Vese Elastica model.
6.2.6 An augmented Lagrangian method for the $L^1$-mean curvature model

In this section, we follow closely the presentation used in [185]. The rational of the $L^1$-mean curvature model has been given in Section 6.1.5, leading one to consider the following minimization problem

$$ u = \arg \min_{v \in V} \left[ \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \eta \int_{\Omega} \frac{\nabla v \cdot \nabla v}{\sqrt{1 + |\nabla v|^2}} \, dx \right], \quad (184) $$

where $\nabla = \{\partial / \partial x_i\}_{i=1}^2$. In (184), the choice of $V$ is a delicate theoretical issue; indeed the safest way to proceed would be to take $V = H^2(\Omega)$ in (184), and to replace min by inf (a kind of justification for this approach can be found in [133]).

Let observe (as in [185], where a slightly different notation is used) that

$$ \nabla \cdot \nabla v \sqrt{1 + |\nabla v|^2} = \nabla_3 \cdot \{\nabla v, -1\} \left| \{\nabla v, -1\} \right|, \quad (185) $$

where, in (185), $\nabla_3 = \{\partial / \partial x_1, \partial / \partial x_2, 0\}$, and where $\{\nabla v, -1\}$ denotes the 3-dimensional vector-valued function $\{\partial \nabla v / \partial x_1, \partial \nabla v / \partial x_2, -1\}$. In order to simplify (in some sense) the nonlinear structure of the minimization problem (184), we associate new unknown functions with its solution $u$, namely $p$, $n$ and $\psi$ verifying

$$ \begin{cases} p = \{\nabla u, -1\}, \\ n = \frac{p}{|p|}, \text{ or equivalently here } |p| - p \cdot n = 0, |n| \leq 1, \\ \psi = \nabla_3 \cdot n. \end{cases} \quad (186) $$

From (185) and (186), there is clearly equivalence between (184) and

$$ \begin{cases} \{u, \psi, p, n\} = \arg \min_{\{v, \varphi, q, m\}} \left[ \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \eta \int_{\Omega} |\varphi| \, dx \right], \\ \text{with } \{v, \varphi, q, m\} \text{ verifying } q = \{\nabla v, -1\}, |q| - q \cdot m = 0, |m| \leq 1, \nabla_3 \cdot m = \varphi. \end{cases} \quad (187) $$

In order to solve the minimization problem (184), taking advantage of its equivalence with (187), we introduce the following augmented Lagrangian functional
\[ L_{MC}(v, \varphi, q, z, m; \mu_1, \mu_2, \mu_3, \mu_4) = \frac{1}{2} \int_{\Omega} |v - f|^2 \, dx + \eta \int_{\Omega} |\varphi| \, dx \\
+ \frac{r_1}{2} \int_{\Omega} (|q| - q \cdot z) \, dx + \int_{\Omega} |\mu_1| (|q| - q \cdot z) \, dx \\
+ \frac{r_2}{2} \int_{\Omega} (\{\nabla v, -1\} - q)^2 \, dx + \int_{\Omega} \mu_2 : (\{\nabla v, -1\} - q) \, dx \]

(188)

The additional vector-valued function \( z \) has been introduced in order to decouple \( \nabla_3 \cdot m \) from the nonlinear relations verified by \( m \) in (187). Following [185], and taking (187) and (188) into account, we advocate the following algorithm for the solution of problem (184):

**Algorithm 6.5**: An augmented Lagrangian method for the \( L^1 \)-mean curvature model

0. Initialization: \( \lambda_1^0 = 0, \lambda_2^0 = 0, \lambda_3^0 = 0, \lambda_4^0 = 0, u^0 = f, p^0 = \{\nabla u^0, -1\}, n^0 = y^0 = \frac{p^0}{|p^0|}, \psi^0 = \nabla_3 \cdot n^0. \)

For \( k = 0, 1, \ldots \), until convergence:

1. Compute \( u^{k+1} \) from

\[ u^{k+1} = \text{arg min}_{v \in H^1(\Omega)} L_{MC}(v, \psi^k, p^k, y^k, n^k; \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k). \]  

(189)

2. Compute \( \psi^{k+1} \) from

\[ \psi^{k+1} = \text{arg min}_{\varphi \in L^2(\Omega)} L_{MC}(u^{k+1}, \varphi, p^k, y^k, n^k; \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k). \]

(190)

3. Compute \( p^{k+1} \) from

\[ p^{k+1} = \text{arg min}_{q \in L^2(\Omega)} L_{MC}(u^{k+1}, \psi^{k+1}, q, y^k, n^k; \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k). \]

(191)

4. Compute \( y^{k+1} \) from

\[ y^{k+1} = \text{arg min}_{z \in Z} L_{MC}(u^{k+1}, \psi^{k+1}, p^{k+1}, z, n^k; \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k). \]

(192)

5. Compute \( n^{k+1} \) from

\[ n^{k+1} = \text{arg min}_{m \in M} L_{MC}(u^{k+1}, \psi^{k+1}, p^{k+1}, y^{k+1}, m; \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k). \]

(193)

6. Update \( \{\lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k\} \) by
\begin{equation}
\begin{aligned}
\lambda_{1}^{k+1} &= \lambda_{1}^{k} + r_{1}(|p_{1}^{k+1}| - p_{1}^{k+1} \cdot y^{k+1}), \\
\lambda_{2}^{k+1} &= \lambda_{2}^{k} + r_{2}(\{\nabla u_{1}^{k+1}, -1\} - p_{2}^{k+1}), \\
\lambda_{3}^{k+1} &= \lambda_{3}^{k} + r_{3}(\psi^{k+1} - \left(\frac{\partial n_{1}^{k+1}}{\partial x_{1}} + \frac{\partial n_{2}^{k+1}}{\partial x_{2}}\right)), \\
\lambda_{4}^{k+1} &= \lambda_{4}^{k} + r_{4}(y^{k+1} - n^{k+1}).
\end{aligned}
\end{equation}

In (189)-(194), the sets \( Z \) and \( M \) are defined by
\[ Z = \{ z | z \in (L^{2}(\Omega))^{3}, |z(x)| \leq 1, \text{ a.e. in } \Omega \}, \]
and
\[ M = \{ m | m \in (L^{2}(\Omega))^{3}, \frac{\partial m_{1}}{\partial x_{1}} + \frac{\partial m_{2}}{\partial x_{2}} \in L^{2}(\Omega) \}, \]
respectively.

We observe that that the minimization sub-problems (190), (191) and (192) have closed form solutions which can be computed point-wise. On the other hand, the Euler-Lagrange equations of the sub-problems (189) and (193) are well-posed linear elliptic equations with constant coefficients; fast solvers exist for the solution of the discrete analogues of these elliptic problems (see [185] for details and the results of numerical experiments validating the above algorithm). An important issue is the tuning of the augmentation parameters \( r_{1}, r_{2}, r_{3} \) and \( r_{4} \); the comments we did in Section 6.2.5, concerning the adjustment of \( r_{1} \) and \( r_{2} \) in algorithm (176)-(178), still apply here.

Remark 42. Another augmented Lagrangian based solution method for the \( L^{1} \)-mean curvature problem (184) is discussed and numerically tested in ref. [133]. The related ADMM algorithm involves only three Lagrange multipliers, and three augmentation parameters. Moreover, the various vector-valued functions encountered in the approach discussed in [133] map \( \Omega \) into \( \mathbb{R}^{2} \) (instead of \( \mathbb{R}^{3} \), as it is the case for algorithm (189)-(194)).

7 Further comments and complements

There is much more to say about operator-splitting and ADMM algorithms; fortunately, many of these issues and topics we left behind, or said very little about, are developed in the other chapters of this book. There are however some issues we would like to briefly-comment to conclude this chapter, namely:

(i) The convergence of operator-splitting methods and ADMM algorithms, when applied to the solution of problems involving non-monotone operators and/or non-convex functionals.
(ii) The choice of the augmentation parameters and their dynamical adjustment when applying ADMM algorithms.
(iii) The derivation of operator-splitting schemes of high (or higher) orders of accuracy.

(iv) Trying to understand why the Douglas-Rachford scheme is more robust than the Peaceman-Rachford one, using simple model problems to clarify this issue.

(v) Very few problems have generated as many operator-splitting based solution methods than the Navier-Stokes equations modelling viscous fluid flows. From this fact, providing the reader with a significant number of related references is a must in a book like this one. These references will conclude this chapter.

Concerning the first issue, to the best of our knowledge, there is no general theory concerning the convergence of operator-splitting methods and ADMM algorithms when the problem under consideration involves at least one non-monotone operator and/or a non-convex functional. Actually, one can find in the literature convergence results for some problems lacking monotonicity and/or convexity, but, most often, the proofs of these results are very specific of the problem under consideration, and therefore are not easy to generalize to other situations. However, some recent results obtained by R. Luke [96, 115] and W. Yin [166], and collaborators, suggest that a fairly general theory is not out of reach. However, we think that there always will be situations where one not will be able to prove the convergence of operator-splitting methods and ADMM algorithms. This is not surprising since these methods and algorithms have been quite successful at solving problems for which the existence of solutions has not been proved.

The second issue, concerning the choice and the dynamical adaptation of the augmentation parameters is another complicated one, particularly for those non-convex and non-monotone situations involving more than one of such parameters. Indeed, numerical experiments have shown that the optimal values of these parameters may have several orders of magnitude (as shown in, e.g., [80] and [133]), and, from the possible existence of multiple solutions, that bifurcations can take place depending also of the values of these parameters (and of the algorithm initialization). However, for particular problems, heuristics have been found, significantly improving the speed of convergence of these ADMM algorithms (see, e.g., [46]).

In order to address the high (or higher) orders of accuracy issue (our third issue) we return to Section 2.3 of this chapter (the one dedicated to the Strang symmetrized operator-splitting scheme), and consider the following initial value problem

\[
\begin{aligned}
\frac{dX}{dt} + (A + B)X &= 0 \text{ on } (0, T), \\
X(0) &= X_0,
\end{aligned}
\]

where \(A\) and \(B\) are linear operators independent of \(t\). When applied to the solution of the initial value problem (195), the Strang symmetrized scheme (7)-(10) can be written in the following more compact form

\[
\begin{aligned}
X^0 &= X_0, \\
X^{n+1} &= e^{-A\Delta t/2}e^{-B\Delta t}e^{-A\Delta t/2}X^n, \quad \forall n \geq 0.
\end{aligned}
\]
The relation
\[ e^{-[A+B]△t} - e^{-A△t/2} e^{-B△t} e^{-A△t/2} = O(△t^3), \]
shows that scheme (196) is second order accurate (and exact if \(AB = BA\)). For those situations requiring an order of accuracy higher than two, several options do exist, the best known being:

(a) The 4th order Strang-Richardson scheme discussed in [49, 50, 48] where it is applied (among other problems) to the numerical solution of real-valued or vector-valued reaction-diffusion equations such as

\[ \frac{∂u}{∂t} - M∇^2 u + F(u) = 0, \]

where \(u(x,t) \in \mathbb{R}^d\), \(∇^2\) denotes the Laplace operator, \(M\) is a \(d \times d\) symmetric definite matrix, and \(F\) is a smooth mapping from \(\mathbb{R}^d\) into \(\mathbb{R}^d\).

(b) The exponential operator-splitting schemes. Actually, the Lie and Strang splitting schemes belong to this family of time discretization methods, whose origin (concerning schemes of order higher than two) is not easy to track back, early significant publications being [150, 151] (see also the references therein and those in [161], and in Google Scholar). Arbitrary high accuracy can be obtained with these methods, the price to pay being their reduced stability (compared to the Strang scheme, for example).

The best way to introduce the Strang-Richardson scheme is to start, one more time, from the simple initial value problem (195). Applied to the solution of (195), the Strang-Richardson scheme reads as

\[
X^0 = X_0,
\]

\[
X^{n+1} = \frac{1}{3} \left[ 4 e^{-A△t/4} e^{-B△t/2} e^{-A△t/2} e^{-B△t/2} e^{-A△t/4} - e^{-A△t/2} e^{-B△t} e^{-A△t/2} \right] X^n, \forall n ≥ 0.
\]

A more practical equivalent formulation of the symmetrized scheme (197) can be found in the Chapter 6 of [70]; it avoids the use of matrix exponentials and can be generalized easily to nonlinear problems (it requires the solution of eight sub-initial value problems per time step). Scheme (197) is fourth order accurate but not as stable as the original Strang scheme (scheme (196)). Also, its application to decompositions involving more than two operators becomes a bit complicated to say the least (higher order methods of the same type are discussed in [85]).

In a similar fashion, we consider again the initial value problem (195) to introduce exponential splitting methods. Applied to the solution of (195) the typical exponential operator-splitting scheme reads as follows:

\[
X^0 = X_0,
\]

\[
X^{n+1} = \left( \prod_{j=1}^{J} e^{-b_j B△t} e^{-a_j A△t} \right) X^n, \forall n ≥ 0.
\]
where \( a_j, b_j \in \mathbb{R} \), for \( 1 \leq j \leq J \). The Strang symmetrized scheme (196) is a particular case of (198) (corresponding to \( J = 2, b_1 = 0, a_1 = 1/2, b_2 = 1, a_2 = 1/2 \)). By an appropriate choice of \( J \), and of the coefficients \( a_j \) and \( b_j \), scheme (198) can be made of order higher than two (as shown in, e.g., [16]), the price to pay being that some of the coefficients \( a_j, b_j \) are negative making the scheme inappropriate to those situations where some of the operators are dissipative. On the other hand, these higher order schemes produce spectacular results when applied to reversible systems, like those associated with some linear and nonlinear Schrödinger operators, as shown in, e.g., [51, 161]. Their generalization to those (fairly common) situations involving more than two operators is rather complicated, although theoretically doable.

Concerning the fourth issue, the Peaceman-Rachford and Douglas-Rachford schemes have been briefly discussed in Sections 2.4 and 2.5, respectively. In order to have a better idea of their accuracy and stability properties, we will consider the particular situation where, in problem (14), \( \phi_0 \in \mathbb{R}^d, T = +\infty \), and where \( A_1 \) (resp., \( A_2 \)) is given by \( A_1 = \alpha A \) (resp., \( A_2 = \beta A \)), \( A \) being a real symmetric positive definite \( d \times d \) matrix, and \( \alpha, \beta \) verifying \( 0 \leq \alpha, \beta \leq 1 \) and \( \alpha + \beta = 1 \). The exact solution of the associated problem (14) reads as
\[
\phi(t) = e^{-At}\phi_0, \quad \forall t \geq 0,
\]
which implies (by projection on an orthonormal basis of eigenvectors of matrix \( A \), and with obvious notation)
\[
\phi(t) = e^{-\lambda_i t}\phi_0, \quad \forall t \geq 0, \quad \forall i = 1, \ldots, d,
\]
where \( 0 < \lambda_1 \leq \cdot \cdot \cdot \leq \lambda_i \leq \cdot \cdot \cdot \leq \lambda_d \), the \( \lambda_i \)'s being the eigenvalues of matrix \( A \). Applying the Peaceman-Rachford scheme (15) to the particular problem (14) defined above, we obtain the following discrete analogue of (199):
\[
\phi^n_i(t) = (R_1(\lambda_i t))^n\phi_0, \quad \forall n \geq 0, \quad \forall i = 1, \ldots, d,
\]
(200)

\( R_1 \) being the rational function defined by
\[
R_1(\xi) = \frac{(1 - \frac{\alpha}{2}\xi)}{(1 + \frac{\alpha}{2}\xi)} \frac{(1 - \frac{\beta}{2}\xi)}{(1 + \frac{\beta}{2}\xi)}.
\]
(201)

Since \( |R_1(\xi)| < 1, \forall \xi > 0 \), the Peaceman-Rachford scheme (15) is unconditionally stable in the particular case considered here. However, the property \( \lim_{\xi \to +\infty} R_1(\xi) = 1 \) shows that the above scheme is not stiff A-stable, making it not a first choice scheme to capture steady state solutions or to simulate fast transient phenomena. Actually, the stability drawback we just mentioned is not specific to the particular case we are considering, but seems to hold in general for scheme (15). Incidentally, the relation
\[
R_1(\xi) - e^{-\xi} = O(\xi^3)
\]
in the neighborhood of \( \xi = 0 \)
implies that in the particular case under consideration (where $A_1$ and $A_2$ commute) scheme (15) is second order accurate. Applying now the Douglas-Rachford scheme (17) to the same particular case of problem (14), we obtain

$$\phi^{n+1} = (I + \alpha \Delta tA)^{-1}(I + \beta \Delta tA)^{-1}(I + \alpha \beta (\Delta t)^2 A^2)\phi^n, \forall n \geq 0,$$

which implies

$$\phi^n = (I + \alpha \Delta tA)^{-n}(I + \beta \Delta tA)^{-n}(I + \alpha \beta (\Delta t)^2 A^2)^n\phi_0, \forall n \geq 0.$$  \hspace{1cm} (202)

By projection of (202) on an orthonormal basis of $\mathbb{R}^d$ consisting of eigenvectors of $A$, we obtain the following variant of (200):

$$\phi^n_i(t) = (R_2(\lambda_i \Delta t))^n \phi_0_i, \forall n \geq 0, \forall i = 1, \ldots, d, \hspace{1cm} (203)$$

$R_2$ being the rational function defined by

$$R_2(\xi) = \frac{1 + \alpha \beta \xi^2}{(1 + \alpha \xi)(1 + \beta \xi)}.$$  \hspace{1cm} (204)

Since $0 < R_2(\xi) < 1, \forall \xi > 0$, the Douglas-Rachford scheme (17) is unconditionally stable in the particular case considered here. However, the property $\lim_{\xi \to +\infty} R_2(\xi) = 1$ shows that the above scheme is not stiff $A$-stable, making it not a first choice scheme to capture steady state solutions or to simulate fast transient phenomena. Actually, the stability drawback we just mentioned is not specific to the particular case we are considering, but seems to hold in general for scheme (17). Concerning the accuracy of scheme (17), we observe that in the neighborhood of $\xi = 0$, we have

$$R_2(\xi) = 1 - \xi + \xi^2 + O(\xi^3),$$

which implies, by comparison with $e^{-\xi} = 1 - \xi + \frac{\xi^2}{2} + O(\xi^3)$, that scheme (17) is no better than first order accurate in the particular case we are considering. Since this particular case is the most favorable one can think about, one expects the Douglas-Rachford scheme (17) to be generically first order accurate, a prediction supported by the results of various numerical experiments. It is worth mentioning that in order to improve the accuracy of the Douglas-Rachford scheme (17), J. Douglas & S. Kim introduced in the late 90s-early 2000s ([56]), the following variant of the above scheme

$$\phi^0 = \phi_0.$$  \hspace{1cm} (205)

For $n \geq 0$, $\phi^n \to \hat{\phi}^{n+1} \to \phi^{n+1}$ as follows:

Solve

$$\frac{\hat{\phi}^{n+1} - \phi^n}{\Delta t} + A_1 \left( \frac{\hat{\phi}^{n+1} + \phi^n}{2}, t^{n+1/2} \right) + A_2(\phi^n, t^n) = 0,$$  \hspace{1cm} (206)

and
\[
\frac{\phi^{n+1} - \phi^n}{\Delta t} + A_1 \left( \frac{\phi^{n+1} + \phi^n}{2}, t^{n+1/2} \right) + A_2 \left( \frac{\phi^{n+1} + \phi^n}{2}, t^{n+1/2} \right) = 0. \tag{207}
\]

The Douglas-Kim scheme (205)-(207) is clearly inspired from the Crank-Nicolson scheme. Scheme (205)-(207) is second order accurate if the operators \(A_1\) and \(A_2\) are sufficiently smooth, the price to pay for this accuracy enhancement being a reduction of stability and robustness compared to the original Douglas-Rachford scheme (17).

At those wondering how to choose between Peaceman-Rachford and Douglas-Rachford schemes we will say that on the basis of many numerical experiments, it seems that the second scheme is more robust and faster for those situations where one of the operators is non-smooth (multivalued or singular, for example), particularly if one is interested at capturing steady state solutions. Actually, this behavior is consistent with the fact that the rational function \(R_1\) associated with the Peaceman-Rachford scheme (the one defined by (201)), may change sign when \(\xi\) varies on \((0, +\infty)\), unlike the rational function \(R_2\) defined by (204) (the one associated with the Douglas- Rachford scheme) which stays positive on the above interval. These sign changes, suggest a more oscillatory behavior for the associated scheme if fast transients take place, or if one tries to capture steady state solutions starting far away from these solutions.

As a final comment on ADI methods we have to mention that one their main contributors (if not the main one), beyond their founders (J. Douglas, H. Rachford and D. Peaceman), is definitely E. Wachpress: His wonderful book The ADI Model Problem [164] is an invaluable source of information and references on the Peaceman-Rachford and Douglas-Rachford methods, from the theoretical and practical points of view.

As a conclusion, let observe that the Navier-Stokes equations modelling the flow of viscous fluids have been mentioned quite a few times in this chapter (Section 4 in particular), and in other chapters of this book. There is no doubt that very few partial differential equation problems have motivated such a large number of operator-splitting based solution methods. Focusing on those publications with which we have some familiarity, let us mention: [11, 12, 13, 23, 35, 43, 47, 70, 72, 73, 90, 91, 92, 94, 105, 107, 111, 112, 116, 122, 123, 158, 159, 160] (see also the references therein, Google Scholar, and Chapters 21, 22 and 23 of this book).

References


171. Zhu, W., Tai, X.C., Chan, T.F.: Image segmentation using Euler’s elastica as the regulariza-


175. Zhu, W., Tai, X.C., Chan, T.F.: Image segmentation using Euler’s elastica as the regulariza-