1. Introduction

The main goal of this paper is to describe some methods of solution of nonlinear boundary value problems for ordinary or partial differential equations founded on a variational approach. The variational approach is quite powerful at two levels at least:

(i) The approximation by Galerkin type methods such as finite elements, spectral methods, etc...

(ii) The numerical solution of the approximate problems by efficient iterative methods.

To illustrate the above generalities, we shall discuss in Sections 2, 3, 4 the solution of nonlinear boundary value problems by least squares-conjugate gradient methods and apply them to quite classical nonlinear problems in Fluid Dynamics such as the Navier-Stokes equations for incompressible viscous fluids and the full potential equation modelling the transonic flows of compressible inviscid fluids.

In Sections 5, 6 we shall discuss the solution of a broad class of nonlinear variational problems by augmented Lagrangian methods and apply the corresponding techniques to the solution of some "hard" problems in finite elasticity.

The results of numerical experiments illustrate the possibilities of the methods discussed in this paper (more details about these methods can be found in [19], [20]).
2. Least squares solution of a nonlinear Dirichlet model problem

In order to introduce the methods that we shall apply in Sections 3, 4 to the solution of fluid dynamics problems, we shall consider the solution of a simple nonlinear Dirichlet problem by least squares and conjugate gradient methods. In Section 2.4, we shall briefly discuss the use of pseudo arc length continuation methods for solving nonlinear problems via least squares and conjugate gradient algorithms.

2.1. Formulation of the model problem. Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with a smooth boundary $\Gamma = \partial \Omega$; let $T$ be a nonlinear operator from $V = H^1_0(\Omega)$ to $V^* = H^{-1}(\Omega)$ ($H^{-1}(\Omega)$, the topological dual space of $H^1_0(\Omega)$). We consider the nonlinear Dirichlet problem

$$\begin{cases}
\text{Find } u \in H^1_0(\Omega) \text{ such that} \\
-\Delta u - T(u) = 0 \text{ in } \Omega
\end{cases} \quad (2.1)$$

and we observe that (2.1) implies

$$u = 0 \text{ on } \Gamma.$$ 

We do not discuss here the existence and uniqueness properties of the solutions of (2.1) since we do not want to be very specific about the operator $T$.

2.2. Least squares formulations of the model problem (2.1).

2.2.1. Generalities. We shall consider least squares formulations of the model problem (2.1). An obvious least squares formulation consists in saying that the required function $u$ minimizes the left-hand side of (2.1), in an $L^2(\Omega)$ least squares sense. That is,

$$\operatorname{Min}_{v \in V} \int_{\Omega} |\Delta v + T(v)|^2 \, dx, \quad (2.2)$$

where $V$ is a space of feasible functions. Let us introduce $\xi$ by

$$\begin{cases}
-\Delta \xi = -\Delta v - T(v) \text{ in } \Omega, \\
\xi = 0 \text{ on } \Gamma.
\end{cases} \quad (2.3)$$

Then (2.2) is equivalent to

$$\operatorname{Min}_{v \in V} \int_{\Omega} |\Delta \xi|^2 \, dx, \quad (2.4)$$
where $\xi$ is a (nonlinear) function of $v$, through (2.3). It is clear (see e.g. [10], [35]) that (2.3), (2.4) has the structure of an optimal control problem, where

(i) $v$ is the control vector,
(ii) $\xi$ is the state vector,
(iii) (2.3) is the state equation,
(iv) the functional occurring in (2.4) is the cost function.

Another least squares-optimal control formulation is

$$\min_{v \in V} \int_{\Omega} |\xi|^2 \, dx$$  \hspace{1cm} (2.5)

where $\xi$ still satisfies (2.3). This formulation has been used by Cea–Geymonat [11] to solve nonlinear partial differential problems (including the steady Navier–Stokes equations).

Actually, the two above least squares formulations may lead to a slow convergence, since the norm occurring in the cost functions is not appropriate for the state equation. An alternate choice, very well suited to nonlinear second order problems, will be discussed in the next section.

2.2.2. A $H^{-1}$-least squares formulation of (2.1). Let us recall some properties of $H^{-1}(\Omega)$, the topological dual space of $H^1_0(\Omega)$. If $L^2(\Omega)$ is identified to its dual, then

$$H^1_0(\Omega) \subset L^2(\Omega) \subset H^{-1}(\Omega);$$

moreover, $A(=\nabla^2)$ is an isomorphism from $H^1_0(\Omega)$ onto $H^{-1}(\Omega)$. In the sequel the duality pairing $\langle \cdot, \cdot \rangle$ between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$ is chosen in such a way that

$$\langle f, v \rangle = \int_{\Omega} fv \, dx \quad \forall f \in L^2(\Omega), \quad \forall v \in H^1_0(\Omega).$$  \hspace{1cm} (2.6)

The topology of $H^{-1}(\Omega)$ is defined by $\| \cdot \|_*$, where $\forall f \in H^{-1}(\Omega)$

$$\|f\|_* = \sup_{v \in H^1_0(\Omega) \setminus \{0\}} \frac{|\langle f, v \rangle|}{\|v\|_{H^1_0(\Omega)}}.$$  \hspace{1cm} (2.7)

A convenient least squares formulation to solve the model problem (2.1)

---

1 Convenient because the space $H^1_0(\Omega)$ in (2.8) is also the space in which we wish to solve (2.1) (as follows from the properties of $A$ and $T$).
seems to be

$$\min_{v \in H^1_0(\Omega)} \| \Delta v + T(v) \|_*.$$  \hspace{1cm} (2.8)

It is clear that if (2.1) has a solution, then this solution will be a solution of (2.8) for which the cost function vanishes. Let us introduce $\xi \in H^1_0(\Omega)$ by (2.3), so that (2.8) reduces to

$$\min_{v \in H^1_0(\Omega)} \| \Delta \xi \|_*$$  \hspace{1cm} (2.9)

where $\xi$ is a function of $v$ through (2.3).

Actually, it can be proved that if $\| \cdot \|_*$ is defined by (2.7) with $\langle \cdot, \cdot \rangle$ obeying (2.6), then

$$\| \Delta v \|_* = \| v \|_{H^1_0(\Omega)} = \left( \int_\Omega |\nabla v|^2 \, dx \right)^{1/2}, \quad \forall v \in H^1_0(\Omega).$$  \hspace{1cm} (2.10)

It follows from (2.10) that (2.9) may be reformulated as

$$\min_{v \in H^1_0(\Omega)} \int_\Omega |\nabla \xi|^2 \, dx$$  \hspace{1cm} (2.11)

where $\xi$ is a function of $v$ through (2.3); (2.11) has also the structure of an optimal control problem.

**Remark 2.1.** Nonlinear boundary value problems have been treated in [38] using a formulation closely related to (2.3), (2.11).

### 2.3. Conjugate gradient solution of the least squares problem (2.3), (2.11).

Let us define the function $J: H^1_0(\Omega) \to \mathbb{R}$ by

$$J(v) = \frac{1}{2} \int_\Omega |\nabla \xi|^2 \, dx$$  \hspace{1cm} (2.12)

where $\xi$ is a function of $v$ through (2.3); then we may also write (2.11) as

$$\begin{cases}
\text{Find } u \in H^1_0(\Omega) \text{ such that} \\
J(u) \leq J(v) \quad \forall v \in H^1_0(\Omega).
\end{cases}$$  \hspace{1cm} (2.13)

We shall use a conjugate gradient algorithm to solve (2.13). From among the possible conjugate gradient algorithms we have selected the Polak–Ribière version (see Polak [45]), since this algorithm produced the best
performances in the various experiments we did (good performance of the Polak–Ribière algorithm are discussed in [46]). Let us denote by \( J'(\cdot) \) the differential of \( J(\cdot) \); then the Polak–Ribière version of the conjugate gradient method, applied to the solution of (2.13) is

**Step 0: Initialization.**

\[
\begin{align*}
u^0 \in H^1_0(\Omega) & \text{ given}, \\
\text{compute } & g^0 \in H^1_0(\Omega) \text{ from} \\
& -\Delta g^0 = J'(u^0) \text{ in } \Omega, \quad g^0 = 0 \text{ on } \Gamma, \\
\text{and set} & \\
z^0 = g^0.
\end{align*}
\]

Then for \( n \geq 0 \), assuming \( u^n, g^n, z^n \) to be known, compute \( u^{n+1}, g^{n+1}, z^{n+1} \) by

**Step 1: Descent.**

\[
\begin{align*}u^{n+1} &= u^n - \lambda_n z^n, \\
\text{where } \lambda_n & \text{ is the solution of the one-dimensional minimization problem} \\
& \lambda_n \in \mathbb{R}, \quad J(u^n - \lambda_n z^n) \leq J(u^n - \lambda z^n) \quad \forall \lambda \in \mathbb{R}.
\end{align*}
\]

**Step 2: Construction of the new descent direction.**

Define \( g^{n+1} \in H^1_0(\Omega) \) by

\[
\begin{align*}-\Delta g^{n+1} &= J'(u^{n+1}) \text{ in } \Omega, \quad g^{n+1} = 0 \text{ on } \Gamma, \\
\text{then} & \\
\gamma_n &= \frac{\int_{\Omega} \nabla g^{n+1} \cdot \nabla (g^{n+1} - g^n) \, dx}{\int_{\Omega} |\nabla g^n|^2 \, dx}, \\
z^{n+1} &= g^{n+1} + \gamma_n z^n, \\
n &= n + 1, \quad \text{go to (2.17).} \quad \blacksquare
\end{align*}
\]

The two non-trivial steps of algorithm (2.14)-(2.22) are:

(i) The solution of the single variable minimization problem (2.18); the corresponding line search can be achieved by dichotomy or Fibonacci methods (see, for example, [6], [45]). We have to observe that each evalu-
ation of \( J(v) \) for a given argument \( v \) requires the solution of the linear Poisson problem (2.3) to obtain the corresponding \( \xi \).

(ii) The calculation of \( g^{n+1} \) from \( u^{n+1} \), which requires the solution of two linear Dirichlet problems (namely (2.3) with \( v = u^{n+1} \) and (2.19)).

**Remark 2.2.** As stopping criterion for algorithm (2.14)-(2.22) we should use

\[
J(u^n) \leq s \quad \text{or} \quad \| g^n \|_{H^1_0(\Omega)} \leq \varepsilon
\]

where \( s \) is a "small" positive number.

**Calculation of \( J'(u^n) \) and \( g^n \):** Owing to the importance of step (ii), let us describe in detail the calculation of \( J'(u^n) \) and \( g^n \):

Let \( w \in H^1_0(\Omega) \); then \( J'(v) \) may be defined by

\[
\langle J'(v), w \rangle = \lim_{t \to 0} \frac{J(v + tw) - J(v)}{t};
\]

we obtain from (2.3), (2.12), (2.23)

\[
\langle J'(v), w \rangle = \int_{\Omega} \nabla\xi \cdot \nabla \eta \, dx,
\]

where \( \eta \in H^1_0(\Omega) \) is the solution of

\[
\begin{cases}
\Delta \eta = \Delta w + T'(v) \cdot w \quad \text{in} \ \Omega, \\
\eta = 0 \quad \text{on} \ \Gamma;
\end{cases}
\]

(2.25) has the following variational formulation

\[
\begin{cases}
\int_{\Omega} \nabla \eta \cdot \nabla z \, dx = \int_{\Omega} \nabla w \cdot \nabla z \, dx - \langle T'(v) \cdot w, z \rangle \quad \forall z \in H^1_0(\Omega), \\
\eta \in H^1_0(\Omega).
\end{cases}
\]

Taking \( z = \xi \) in (2.26), we obtain from (2.24)

\[
\langle J'(v), w \rangle = \int_{\Omega} \nabla \xi \cdot \nabla w \, dx - \langle T'(v) \cdot w, \xi \rangle \quad \forall v, w \in H^1_0(\Omega).
\]

Therefore \( J'(v) \ (\in H^{-1}(\Omega)) \) may be identified with the linear functional on \( H^1_0(\Omega) \) defined by

\[
w \mapsto \int_{\Omega} \nabla \xi \cdot \nabla w \, dx - \langle T'(v) \cdot w, \xi \rangle.
\]
It follows from (2.19), (2.27), (2.28) that \( g^n \) is the solution of the following \textit{linear variational problem}:

\[
\begin{align*}
\text{Find } g^n \in H_0^1(\Omega) \text{ such that } & \forall w \in H_0^1(\Omega) \\
\int_{\Omega} \nabla g^n \cdot \nabla w \, dx = \int_{\Omega} \nabla \xi^n \cdot \nabla w \, dx - \langle T'(u^n) \cdot w, \xi^n \rangle 
\end{align*}
\]

(2.29)

where \( \xi^n \) is the solution of (2.3) corresponding to \( v = u^n \).

\textbf{Remark 2.3.} It is clear from the above observations that an efficient \textit{Poisson solver} will be a basic tool for solving (2.1) (in fact a \textit{finite-dimensional approximation} of it) by the above conjugate gradient algorithm.

\textbf{Remark 2.4.} The fact that \( J'(v) \) is known through (2.27) is not at all a draw-back if a \textit{Galerkin} or a \textit{finite element method} is used to approximate (2.1). Indeed, we only need to know the value of \( \langle J'(v), w \rangle \) for \( w \) belonging to a basis of the \textit{finite-dimensional subspace} of \( H_0^1(\Omega) \) corresponding to the Galerkin or finite element approximation under consideration.

\textbf{Remark 2.5.} The above methodology extends easily to the solution of \textit{nonlinear elliptic systems} like

\[
\begin{align*}
- \Delta u_1 + u_1 \frac{\partial u_1}{\partial x_1} + u_2 \frac{\partial u_1}{\partial x_2} &= f_1 \text{ in } \Omega, \\
- \Delta u_2 + u_1 \frac{\partial u_2}{\partial x_1} + u_2 \frac{\partial u_2}{\partial x_2} &= f_2 \text{ in } \Omega, \\
u_1 = u_2 = 0 \text{ on } \Gamma,
\end{align*}
\]

(2.30)

where \( \Omega \) is a bounded domain of \( \mathbb{R}^2 \) and where \( f_1, f_2 \in H^{-1}(\Omega) \). Elliptic systems closely related to (2.30) occur in the solution of the \textit{time dependent Navier–Stokes equations} by \textit{alternating direction} methods (see Sec. 3).

\textbf{2.4. \textit{A nonlinear least squares approach to arc length continuation methods.}}

\textbf{2.4.1. Generalities. Synopsis.} We would like to show in this section that the above least squares methodology can be (slightly) modified in order to solve nonlinear problems by arc length continuation methods directly inspired from H.B. Keller [29], [30] (where the basic iterative methods are Newton's and quasi–Newton's instead of conjugate gradient).
As test problem, we have chosen a variant of the nonlinear Dirichlet problem (2.1); let us consider the following family of nonlinear Dirichlet problems (to be solved in $H_0^1(\Omega)$), parametrized by $\lambda \in \mathbb{R}$:

$$
\begin{align*}
- \Delta u &= \lambda T(u) \text{ in } \Omega, \\
u &= 0 \text{ on } \Gamma;
\end{align*}
$$

(2.31)

(2.1) corresponds to $\lambda = 1$.

2.4.2. Solution of (2.31) via arc length continuation methods. Following [29], [30] (for which we refer for justification (see also [43])), we associate to (2.31) a "continuation" equation; we have chosen (from among other possibilities)

$$
\int_\Omega |\nabla \dot{u}|^2 \, dx + \dot{\lambda}^2 = 1, \tag{2.32}
$$

where $\dot{u} = \partial u / \partial s$, $\dot{\lambda} = d\lambda / ds$, and where the curvilinear abscissa $s$ is defined by

$$
ds = \lambda \, ds + \int_0^s \nabla \dot{u} \cdot \nabla du \, ds, \tag{2.33}
$$
or equivalently by

$$(\delta s)^2 = (\delta \lambda)^2 + \int_\Omega \nabla u \cdot \nabla du \, dx. \tag{2.34}
$$

We are considering in fact a path in $H_0^1(\Omega) \times \mathbb{R}$ whose arc length is defined by (2.32)–(2.34). Then, in order to solve (2.31), we consider the family (parametrized by $s$) of nonlinear systems (2.31), (2.32). In practice we shall approximate (2.31), (2.32) by the following discrete family of nonlinear systems, where $\Delta s$ is an arc length step, positive or negative (possibly varying with $n$) and where $u^n \approx u(n \Delta s)$:

Take $u^0 = 0$, $\lambda^0 = 0$ and suppose that $\dot{\lambda}(0)$, $\dot{u}(0)$ are given (2.35)

(initialization (2.35) is justified by the fact that $u = 0$ is the unique solution of (2.31) if $\lambda = 0$; then for $n \geq 0$, assuming that $u^{n-1}$, $\lambda^{n-1}$, $u^n$, $\lambda^n$ are known, we obtain $\{u^{n+1}, \lambda^{n+1}\} \in H_0^1(\Omega) \times \mathbb{R}$ from the solution of

$$
\begin{align*}
- \Delta u^{n+1} &= \lambda^{n+1} T(u^{n+1}) \text{ in } \Omega, \\
u^{n+1} &= 0 \text{ on } \Gamma.
\end{align*}
$$

(2.36)
and
\[ \int_{\Omega} \nabla (u^1 - u^0) \cdot \nabla \hat{u}(0) \, dx + (\lambda^1 - \lambda^0) \hat{\lambda}(0) = \Delta s \quad \text{if } n = 0, \quad (2.37) \]
\[ \int_{\Omega} (\nabla u^{n+1} - u^n) \cdot \nabla \left( \frac{u^n - u^{n-1}}{\Delta s} \right) \, dx + (\lambda^{n+1} - \lambda^n) \left( \frac{\lambda^n - \lambda^{n-1}}{\Delta s} \right) = \Delta s \]
\[ \cdot \quad \text{if } n \geq 1; \quad (2.38) \]

obtaining \( \hat{u}(0) \) and \( \hat{\lambda}(0) \) is an easy task since we have (from (2.31))
\[ \begin{cases} - \Delta \hat{u}(0) = \hat{\lambda}(0) T(0) & \text{in } \Omega, \\ \hat{u}(0) = 0 & \text{on } \Gamma, \end{cases} \quad (2.39) \]

and therefore
\[ \hat{\lambda}^2(0)(1 + \int_{\Omega} |\nabla \hat{u}|^2 \, dx) = 1, \quad (2.40) \]
where \( \hat{u} \in H^1_0(\Omega) \) is the solution of
\[ \begin{cases} - \Delta \hat{u} = T(0) & \text{in } \Omega, \\ \hat{u} = 0 & \text{on } \Gamma \end{cases} \quad (2.41) \]

then clearly \( \hat{u}(0) = \hat{\lambda}(0) \hat{u} \).

Relations (2.36)–(2.38) look like a discretization scheme for solving the Cauchy problem for first order ordinary differential equations; from this analogy we can derive many other discretization schemes for the approximation of (2.31), (2.32) (Runge–Kutta, multisteps, etc.) and also methods for the automatic adjustment of \( \Delta s \).

2.4.3. **Nonlinear least squares and conjugate gradient solution of (2.36)–(2.38).** Without going into details (for which we refer to [21], [47]) we can solve (2.36)–(2.38) by a variant of algorithm (2.14)–(2.22) defined on the Hilbert space \( H^1_0(\Omega) \times \mathbb{R} \) equipped with the metric and inner product corresponding to
\[ \{ v, \mu \} \rightarrow \int_{\Omega} |\nabla v|^2 \, dx + \mu^2; \quad (2.42) \]

it is clear that other norms than (2.42) are possible, however, in all cases
the scaling of a conjugate gradient algorithm using a discrete variant of
\[
\begin{bmatrix}
  -A & 0 \\
  0 & 1 \\
\end{bmatrix}
\]
(or similar operators) (2.43)
will require an efficient solver and the conclusions of Sec. 2.3 (namely Remark 2.3) still hold.

**Remark 2.6.** To initialize the conjugate gradient algorithm solving (2.36)-(2.38) we have used \( \{2^n - \lambda^{-1}, 2u^n - u^{n-1}\} \) as initial guess to compute \( \{\lambda^{n+1}, u^{n+1}\} \) (this supposes \( \Delta s \) to be constant); with such a choice we obtain a much faster convergence than by taking \( \{\lambda^n, u^n\} \) as initial guess.

### 2.4.4. Applications

We describe in this section the solution of *nonlinear eigenvalue problems* by the continuation methods described above. The first problem is the *Bratu problem* and will be discussed in more details. The second one is the solution of the *Von Karman equations for nonlinear elastic plates* and will be discussed quite briefly.

**2.4.4.1. Formulation of the first test problem.** We shall apply the methods described in Sections 2.4.2, 2.4.3 to the solution of the following classical problem (known sometimes as the Bratu problem)

\[
\begin{cases}
  -Au = \lambda e^u & \text{in } \Omega, \\
  u = 0 & \text{on } \Gamma,
\end{cases}
\]

(2.44)

where \( \Omega \) is a bounded domain in \( \mathbb{R}^N \); we have to observe that unless \( N = 1 \), the mapping \( T \) defined by

\[ T(v) = e^v, \quad v \in H^1_0(\Omega), \]

is not continuous from \( H^1_0(\Omega) \) to \( H^{-1}(\Omega) \). We consider only the case where \( \lambda = 0 \), since if \( \lambda < 0 \), the operator \( v \to -\Delta v - \lambda e^v \) is *monotone* which implies that (2.44) has a unique solution.

If \( \lambda > 0 \), problem (2.44) has been considered by many authors; with regard to recent publications let us mention, among others, references [1], [15], [16], [39]-[41]. From the numerical point of view, problem (2.44) has been investigated in [13], [21], [31], [47] among others.

**2.4.4.2. Numerical solution of (2.44) by the methods of Sections 2.4.2, 2.4.3.** We have chosen the particular case of (2.44) where \( \Omega = [0,1] \times [0,1] \).
The practical application of the methods of Sections 2.4.2, 2.4.3 requires
the reduction of (2.44) to a *finite-dimensional problem*; this can done by
*finite elements* or (owing to the simplicity of $\Omega$) by *finite differences*.

Actually, the results presented here have been obtained using a finite
element method with piecewise linear approximations (see [19], [21],
[47] for more details). The continuation method (2.35)–(2.38) has been
applied with $\Delta s = 0.1$; we observe that $T(0) = 1$ in (2.41); algorithm
(2.35)–(2.38) ran "nicely" since an accurate *least squares solution* of the
nonlinear system (2.34), (2.38) required basically no more than 3 to 4
*conjugate gradient* iterations, even close to the turning point.

We have shown in Figure 2.1 the maximal value (reached at $x_1 = x_2 = 0.5$) of the computed solution $u_h$ as a function of $\lambda$; the computed
turning point is at $\lambda = 6.8591$ ...

![Fig. 2.1](image)

2.4.4.3. *A second test problem*. The *least squares-continuation* methods
described in Sections 2.4.2, 2.4.3 have been applied to the solution of
nonlinear problems more complicated than (2.44); let us mention here
the Navier–Stokes equations for incompressible viscous fluids at high
Reynolds' number and also problems involving genuine bifurcation phenom-
انون like the *Von Karman equations* for plates. The details of these calcu-
lations can be found in [21], [47], [48]. We consider briefly here the Von Karman equations for thin clamped plates:

\[
\begin{align*}
\text{Find } u, \varphi \in H^2_0(\Omega) \text{ such that } \\
\Delta^2 u &= \lambda[\theta_0, u] + [\varphi, u] + f \text{ in } \Omega, \\
\Delta^2 \varphi &= -[u, u] \text{ in } \Omega,
\end{align*}
\]

where:

(i) $\Omega(= \mathbb{R}^2)$ is the two-dimensional spatial domain associated to the thin plate under consideration,

(ii) $\theta_0$ and $f$ are given functions, $f$ being the density of external forces normal to the plate,

(iii) $\lambda$ is a factor of proportionality for the external forces acting in the plane of the plate,

(iv) $u$ is the vertical displacement and $\varphi$ is the so-called Airy function,

(v) $[\cdot, \cdot]$ is defined by

\[
[v, w] = \frac{\partial^2 v}{\partial x_1^2} \frac{\partial^2 w}{\partial x_2^2} + \frac{\partial^2 v}{\partial x_2^2} \frac{\partial^2 w}{\partial x_1^2} - 2 \frac{\partial^2 v}{\partial x_1 \partial x_2} \frac{\partial^2 w}{\partial x_1 \partial x_2},
\]
3. Application to the solution of the Navier–Stokes equations for incompressible viscous fluids

We discuss briefly in this section the numerical solution of the Navier–
Stokes equations for incompressible viscous fluids. For more details, see [19], Chapter 7 and Appendix 3.

3.1. Formulation of the time dependent Navier–Stokes equations for incompressible viscous fluids. Let us consider a Newtonian incompressible viscous fluid. If \( \Omega \) and \( \Gamma \) denote the region of the flow (\( \Omega \subset \mathbb{R}^N \), \( N = 2, 3 \) in practice) and its boundary, respectively, then the flow is governed by the following Navier–Stokes equations

\[
\frac{\partial u}{\partial t} - \nu \Delta u + (u \cdot \nabla) u + \nabla p = f \quad \text{in} \quad \Omega, \\
\nabla \cdot u = 0 \quad \text{in} \quad \Omega \quad \text{(incompressibility condition)}.
\]  

(3.1)  

(3.2)

In (3.1), (3.2) we have

(a) \( \nabla = \left\{ \frac{\partial}{\partial x_i} \right\}_{i=1}^N \), \( \Delta = \nabla^2 = \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2} \),

(b) \( u = \left\{ u_i \right\}_{i=1}^N \) is the flow velocity,

(c) \( p \) is the pressure,

(d) \( \nu \) is the viscosity of the fluid,

(e) \( f \) is the density of external forces.

In (3.1), \( (u \cdot \nabla) u \) is a symbolic notation for the nonlinear (vector) term

\[
\left\{ \sum_{i=1}^N u_j \frac{\partial u_i}{\partial x_j} \right\}_{i=1}^N.
\]
Boundary conditions have to be added; for example, in the case of the airfoil $B$ of Figure 3.1, we have (since the fluid is viscous) the following adherence condition

$$u = 0 \quad \text{on } \partial B = I_B;$$

(3.3)

typical conditions at infinity are

$$u = u_\infty,$$

(3.4)

where $u_\infty$ is a constant vector (with regard to the space variables at least).

If $\Omega$ is a bounded region in $\mathbb{R}^N$, we may prescribe as boundary condition

$$u = g \quad \text{on } \Gamma,$$

(3.5)

where (by the incompressibility of the fluid) the given function $g$ has to satisfy

$$\int_{\Gamma} g \cdot n \, d\Gamma = 0,$$

(3.6)

where $n$ is the outward unit vector normal to $\Gamma$.

Finally, for time dependent problem (3.1), (3.2) an initial condition such as

$$u(x, 0) = u_0(x) \quad \text{a.e. on } \Omega,$$

(3.7)

with $u_0$ given, is usually prescribed.

Looking at the above equations, we observe three principal difficulties (even for flows with low Reynolds’ numbers in bounded regions $\Omega$), which are:

(i) The nonlinear term $(u \cdot \nabla)u$ in (3.1),

(ii) The incompressibility condition (3.2),

(iii) The fact that the solutions of the Navier–Stokes equations are vector-valued functions of $x, t$, whose components are coupled by $(u \cdot \nabla)u$ and by the incompressibility condition $\nabla \cdot u = 0$. 
Using convenient alternating direction methods for the time discretization of the Navier–Stokes equations, we shall be able to decouple the difficulties due to the nonlinearity and to incompressibility. For simplicity, we suppose from now on that $\Omega$ is bounded and that we have (3.5) as boundary conditions (with $g$ satisfying (3.6) and possibly depending on $t$).

3.2. Time discretization by alternating direction methods. Let $\Delta t (> 0)$ be a time discretization step and $\theta$ a parameter such that $0 < \theta < 1$.

3.2.1. A first alternating direction method. We first consider the following alternating direction method (of Peaceman–Rachford type):

$$u^0 = u_0,$$

then for $n \geq 0$, $u^n$ being known compute $\{u^{n+1/2}, p^{n+1/2}\}$ and $u^{n+1}$ by solving

$$\begin{align*}
\frac{u^{n+1/2} - u^n}{\Delta t/2} - \theta \nu \Delta u^{n+1/2} + \nabla p^{n+1/2} &= f^{n+1/2} + (1 - \theta) \nu \Delta u^n - (u^n \cdot \nabla) u^n \text{ in } \Omega, \\
\nabla \cdot u^{n+1/2} &= 0 \text{ in } \Omega, \\
u^{n+1/2} &= g^{n+1/2} \text{ on } \Gamma,
\end{align*}$$

and

$$\begin{align*}
\frac{u^{n+1} - u^{n+1/2}}{\Delta t/2} - (1 - \theta) \nu \Delta u^{n+1/2} + (u^{n+1} \cdot \nabla) u^{n+1} &= f^{n+1} + \theta \nu \Delta u^{n+1} - \nabla p^{n+1/2} \text{ in } \Omega, \\
u^{n+1} &= g^{n+1} \text{ on } \Gamma,
\end{align*}$$

respectively.

We use the notation $f^j(x) = f(x, j\Delta t)$, $g^j(x) = g(x, j\Delta t)$, and $u^j(x)$ is an approximation of $u(x, j\Delta t)$.

3.2.2. A second alternating direction method. We now consider the following alternating direction method (of Strang type):

$$u^0 = u_0,$$
then for $n \geq 0$ and starting from $u^n$ we solve

$$
\begin{align*}
\frac{u^{n+1/4} - u^n}{\Delta t/4} - \theta \nu \Delta u^{n+1/4} + \nabla p^{n+1/4} &= f^{n+1/4} + (1 - \theta) \nu \Delta u^n - (u^n \cdot \nabla) u^n \quad \text{in } \Omega, \\
\nabla \cdot u^{n+1/4} &= 0 \quad \text{in } \Omega, \quad u^{n+1/4} = g^{n+1/4} \quad \text{on } \Gamma,
\end{align*}
$$

(3.12)

$$
\begin{align*}
\frac{u^{n+3/4} - u^{n+1/4}}{\Delta t/2} - (1 - \theta) \nu \Delta u^{n+3/4} + (u^{n+3/4} \cdot \nabla) u^{n+3/4} &= f^{n+3/4} + \theta \nu \Delta u^{n+1/4} - \nabla u^{n+1/4} \quad \text{in } \Omega, \\
u \cdot u^{n+3/4} &= 0 \quad \text{in } \Omega, \quad u^{n+3/4} = g^{n+3/4} \quad \text{on } \Gamma,
\end{align*}
$$

(3.13)

$$
\begin{align*}
\frac{u^{n+1} - u^{n+3/4}}{\Delta t/4} - \theta \nu \Delta u^{n+1} + \nabla p^{n+1} &= f^{n+1} + (1 - \theta) \nu \Delta u^{n+3/4} - (u^{n+3/4} \cdot \nabla) u^{n+3/4} \quad \text{in } \Omega, \\
u \cdot u^{n+1} &= 0 \quad \text{in } \Omega, \quad u^{n+1} = g^{n+1} \quad \text{on } \Gamma.
\end{align*}
$$

(3.14)

3.2.3. Some comments and remarks concerning the alternating direction schemes (3.8)–(3.10) and (3.11)–(3.14). Using the two alternating schemes described in Sections 3.2.1, 3.2.2, we have been able to decouple nonlinearity and incompressibility in the Navier–Stokes equations. We shall describe in the following sections the specific treatment of the subproblems encountered at each step of (3.8)–(3.10) and (3.11)–(3.14); we shall first consider the case where the subproblems are still continuous in space (since the formalism of continuous problems is much simpler), and then the discrete case where a finite element method is used to approximate in space the Navier–Stokes equations.

Scheme (3.8)–(3.10) has a truncation error in $O(\Delta t)$; due to the symmetrization process that it involves, scheme (3.11)–(3.14) has a truncation error in $O(\Delta t^2)$.

We observe that $u^{n+1/2}$ and $u^{n+1/4}, u^{n+1}$ are obtained from the solution of linear problems ((3.9) and (3.12), (3.14), respectively) very close to the steady Stokes problem. Despite its greater complexity scheme (3.11)–(3.14) is almost as economical in use as scheme (3.8)–(3.10); this is mainly due to the fact that the "quasi" steady Stokes problems (3.9) and (3.12), (3.14) (in fact, convenient finite element approximation of them) can be solved by quite efficient solvers resulting in that most of the computer time used to solve a full alternating direction step ((3.9), (3.10) or (3.12)–
(3.14) is in fact used to solve the nonlinear subproblem ((3.10) or (3.13)). The good choice for $\theta$ is $\theta = 1/2$ (resp. $\theta = 1/3$) if one uses scheme (3.8)-(3.10) (resp. (3.11)-(3.14)); this follows from the fact that with the above choices for $\theta$, many computer subprograms can be used for both the linear and nonlinear subproblems, resulting therefore in quite substantial computer core memory savings.

Remark 3.1. A variant of scheme (3.8)-(3.10) is the following (it corresponds to $\theta = 1$):

$$u^0 = u_0,$$  

(3.15)

then, for $n \geq 0$ and starting from $u^n$,

$$
\begin{align*}
\frac{u^{n+1/2} - u^n}{\Delta t/2} - v Dw^{n+1/2} + \nabla p^{n+1/2} &= f^{n+1/2} - (u^n \cdot \nabla) u^n \quad \text{in } \Omega, \\
\nabla \cdot u^{n+1/2} &= 0 \quad \text{in } \Omega, \quad u^{n+1/2} = g^{n+1/2} \quad \text{on } \Gamma,
\end{align*}
$$  

(3.16)

$$
\begin{align*}
\frac{u^{n+1} - u^{n+1/2}}{\Delta t/2} + (u^{n+1/2} \cdot \nabla) u^{n+1} &= f^{n+1} + v D w^{n+1/2} \quad \text{in } \Omega, \\
u^{n+1} &= g^{n+1} \quad \text{on } \Gamma_{n+1/2},
\end{align*}
$$  

(3.17)

where

$$
\Gamma_{n+1/2} = \{ x \mid x \in \Gamma, \ g^{n+1/2}(x) \cdot n(x) < 0 \}.
$$

Both subproblems (3.16), (3.17) are linear; the first one is also a "quasi" steady Stokes problem and the second one, which is a first order system, can be solved by a method of characteristics.

A similar remark holds for scheme (3.11)-(3.14).

Such methods have been used by several authors, the space discretization having been done by finite element methods very close to those described in Section 3.5 of this paper (see [3], [28], [44] for a discussion of those characteristics-finite element methods for solving Navier–Stokes equations); these characteristics-finite element methods are slightly dissipative, which may be a drawback in some applications.

Remark 3.2. In order to improve the well-posedness properties of the nonlinear steps (3.10) and (3.13) (and also to simplify convergence proofs)
one may replace the original nonlinear term $B(u) = (u \cdot \nabla)u$ by

$$\tilde{B}(u) = (u \cdot \nabla)u + \frac{1}{2}u(\nabla \cdot u),$$

following the lines of [52]; it is clear that $B(u) = \tilde{B}(u)$ if $\nabla \cdot u = 0$. Actually, the good property of $\tilde{B}$ is that

$$\int_{\Omega} \tilde{B}(v) \cdot v \, dx = 0 \quad \forall v \text{ sufficiently smooth}$$

such that $v = 0$ on $\Gamma$, even if $\nabla \cdot v \neq 0$ (see [52] for more details).

The numerical results obtained using either $B$ or $\tilde{B}$ are practically identical, provided that $\Delta t$ is "reasonably" small.

3.3. Least squares-conjugate gradient solution of the nonlinear subproblems.

3.3.1. Classical and variational formulations. At each full step of the alternating direction methods (3.8)-(3.10) and (3.11)-(3.14) we have to solve a nonlinear elliptic system of the following type:

$$\begin{cases}
\alpha u - v \Delta u + (u \cdot \nabla)u = f & \text{in } \Omega, \\
u = g & \text{on } \Gamma,
\end{cases}$$

(3.18)

where $\alpha$ and $v$ are two positive parameters and where $f$ and $g$ are two given functions defined on $\Omega$ and $\Gamma$, respectively. We shall not discuss here the existence and uniqueness of solutions for problem (3.18). We introduce the following spaces of Sobolev's type

$$V_0 = (H^1_0(\Omega))^N,$$

(3.19)

$$V_g = \{v \mid v \in (H^1(\Omega))^N, v = g \text{ on } \Gamma\};$$

(3.20)

if $g$ is sufficiently smooth, then $V_g \neq \emptyset$.

We use in the following the notation

$$u \cdot v = \sum_{i=1}^{N} u_i v_i, \quad \nabla u \cdot \nabla v = \sum_{i=1}^{N} \nabla u_i \cdot \nabla v_i = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial u_i}{\partial x_j} \frac{\partial v_i}{\partial x_j}.$$
From Green's formula we have, for sufficiently smooth functions \( u \) and \( v \), belonging to \( (H^1(\Omega))^N \) and \( V_0 \), respectively,

\[
- \int_{\Omega} \Delta u \cdot v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx. \tag{3.21}
\]

If \( u \) is a solution of (3.18) it is also a solution of the nonlinear variational problem

\[
\begin{align*}
\begin{cases}
\text{Find } u \in V_g \text{ such that } \\
\alpha \int_{\Omega} u \cdot v \, dx + \nu \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Omega} ((u \cdot \nabla) u) \cdot v \, dx = \int_{\Omega} f \cdot v \, dx \\
\forall v \in V_0,
\end{cases}
\end{align*}
\]

and conversely. We observe that (3.18), (3.22) is not equivalent to a problem of the calculus of variations since there is no functional of \( v \) with \( (v \cdot \nabla) v \) as differential. Using, however, a nonlinear least-squares formulation like those discussed in Section 2, we shall be able to solve (3.18), (3.22) by efficient methods from nonlinear programming, like conjugate gradient, for example. The finite element approximation of (3.18), (3.22) is briefly discussed in Section 3.5.

3.3.2. Least-squares formulation of (3.18), (3.22). Let \( v \in V_g \); from \( v \) we define \( y = y(v) \in V_0 \) as the solution of

\[
\begin{align*}
\begin{cases}
\alpha y - \nu \Delta y = \alpha v - \alpha \Delta v + (v \cdot \nabla)v - f \quad \text{in } \Omega, \\
y = 0 \quad \text{on } \Gamma.
\end{cases}
\end{align*} \tag{3.23}
\]

We observe that \( y \) is obtained from \( v \) via the solution of \( N \) uncoupled linear Poisson problems (one for each component of \( y \)); using (3.21), it can be shown that (3.23) is actually equivalent to the linear variational problem

\[
\begin{align*}
\begin{cases}
\text{Find } y \in V_0 \text{ such that } \\
\alpha \int_{\Omega} y \cdot z \, dx + \nu \int_{\Omega} \nabla y \cdot \nabla z \, dx \\
= \alpha \int_{\Omega} v \cdot z \, dx + \nu \int_{\Omega} \nabla v \cdot \nabla z \, dx + \int_{\Omega} ((v \cdot \nabla)v) \cdot z \, dx - \int_{\Omega} f \cdot z \, dx,
\end{cases}
\end{align*} \tag{3.24}
\]

which has a unique solution. Suppose now that \( v \) is a solution of (3.18), (3.22); the corresponding \( y \) (obtained through the solution of (3.23), (3.24)) is clearly \( 0 \). From these observations it is quite natural to introduce
the following **nonlinear least squares** formulation of problem (3.18), (3.22) (which is a straightforward variant of the one discussed in Section 2.2.2 for the solution of problem (2.1)):

\[
\begin{align*}
\text{Find } u \in V_g & \text{ such that } \\
J(u) & \leq J(v) \quad \forall v \in V_g,
\end{align*}
\]

\[\text{(3.25)}\]

where \( J: (H^1(\Omega))^N \to \mathbb{R} \) is the functional of \( v \) defined by

\[
J(v) = \frac{1}{2} \int_{\Omega} \{a|y|^2 + v|\nabla y|^2\} \, dx,
\]

with \( y \) defined from \( v \) by the solution of the linear problem (3.23), (3.24).

### 3.3.3. Conjugate gradient solution of the least-squares problem (3.25).

We use the following generalization of algorithm (2.14)-(2.22) discussed in Section 2.3:

**Step 0: Initialization.**

\[
u^0 \in V_g \text{ given;}
\]

\[\text{(3.26)}\]

we define \( g^0, w^0 \in V_0 \) by

\[
a \int g^0 \cdot z \, dx + v \int \nabla g^0 \cdot \nabla z \, dx = \langle J'(u^0), z \rangle \quad \forall z \in V_0, \quad g^0 \in V_0,
\]

\[\text{(3.27)}\]

\[
w^0 = g^0,
\]

\[\text{(3.28)}\]

respectively. ■

Then, for \( n \geq 0 \), assuming that \( u^n, g^n, w^n \) are known, we obtain \( u^{n+1}, g^{n+1}, w^{n+1} \) by

**Step 1: Descent.**

\[
\begin{align*}
\text{Find } \lambda^n \in \mathbb{R} & \text{ such that } \\
J(u^n - \lambda^nw^n) & \leq J(u^n - \lambda w^n) \quad \forall \lambda \in \mathbb{R},
\end{align*}
\]

\[\text{(3.29)}\]

\[
u^{n+1} = u^n - \lambda^nw^n.
\]

\[\text{(3.30)}\]
Step 2: Calculation of the new descent direction.

Find \( g^{n+1} \in V_0 \) such that

\[
\alpha \int_D g^{n+1} \cdot z \, dx + \nu \int_D \nabla g^{n+1} \cdot \nabla z \, dx = \langle J'(u^{n+1}), z \rangle \quad \forall z \in V_0,
\]

(3.31)

\[
\gamma_n = \frac{\alpha \int_D g^{n+1} \cdot (g^{n+1}-g^n) \, dx + \nu \int_D \nabla g^{n+1} \cdot (\nabla g^{n+1}-\nabla g^n) \, dx}{\alpha \int_D \vert g^n \vert^2 \, dx + \nu \int_D \vert \nabla g^n \vert^2 \, dx},
\]

(3.32)

\[
\omega^{n+1} = g^{n+1} + \gamma_n \omega^n,
\]

(3.33)

\( n = n+1 \), go to (3.29).

As we shall see below, applying algorithm (3.26–(3.33) to solve (3.25) requires, at each iteration, the solution of several Dirichlet problems for the elliptic operator \( aI - \nu A \).

Calculation of \( J' \): By a method similar to the one used in Section 2.3 we can prove that \( J'(v) \) can be identified with the linear functional from \( V_0 \) to \( \mathbb{R} \) defined by

\[
\langle J'(v), z \rangle = \alpha \int_D y \cdot z \, dx + \nu \int_D \nabla y \cdot \nabla z \, dx + \int_D y \cdot (z \cdot \nabla) v \, dx + \int_D y \cdot (v \cdot \nabla) z \, dx \quad \forall z \in V_0,
\]

(3.34)

with \( y \) being the solution of (3.23), (3.24); it has therefore a purely integral representation, a property of major importance in view of finite element implementations of algorithm (3.26–(3.33).

From the above results, to obtain \( \langle J'(u^{n+1}), z \rangle \) we proceed as follows:

(i) We compute \( y^{n+1} \) from \( u^{n+1} \) through the solution of (3.23), (3.24) with \( v = u^{n+1} \).

(ii) We obtain \( \langle J'(u^{n+1}), z \rangle \) by taking \( v = u^{n+1} \) and \( y = y^{n+1} \) in (3.34).

Further comments on algorithm (3.26–(3.33). Each step of algorithm (3.26–(3.33) requires the solution of several Dirichlet systems for the operator \( aI - \nu A \); more precisely, we have to solve the following systems:

(i) System (3.23), (3.24) to obtain \( y^{n+1} \) from \( u^{n+1} \);

(ii) System (3.31) to obtain \( g^{n+1} \) from \( u^{n+1}, y^{n+1} \).
(iii) Two systems to obtain the coefficients of the quartic polynomial \( \lambda \rightarrow J(u^n - \lambda u^n) \).

Thus we have to solve 4 Dirichlet systems at each iteration (or equivalently 4N scalar Dirichlet problems for \( aI - \nu \Delta \) at each iteration); from these observations it appears clearly that the practical implementation of algorithm (3.26)-(3.33) will require an efficient (direct or iterative) elliptic solver (in fact 3N problems suffice).

The solution of the one-dimensional problem (3.29) can be done very efficiently since it is equivalent to finding the roots of a single variable cubic polynomial whose coefficients are known.

As a last comment, we would like to mention that algorithm (3.26)-(3.33) (in fact, its finite element variants) is quite efficient; when used in combination with the alternating direction methods of Section 3.2, to solve the test problems of Section 3.6, three iterations suffice to reduce the value of the cost function \( J \) by a factor of \( 10^4 \) to \( 10^6 \).

**Remark 3.3.** The above method can be applied also to the variant of problem (3.18), obtained by replacing \( (u- V)u \) by \( (u- V)u + \frac{1}{2}(V \cdot u)u \) in (3.18) (cf. Remark 3.2), i.e., to the nonlinear Dirichlet problem

\[
au - \nu \Delta u + (u \cdot \nabla) u + \frac{1}{2}(\nabla \cdot u)u = f \quad \text{in } \Omega,
\]

\[u = g \quad \text{on } \Gamma.
\]

### 3.4. Solution of the "quasi" Stokes linear subproblems.

#### 3.4.1. Formulation. Synopsis.

At each full step of the alternating direction methods (3.8)-(3.10) and (3.11)-(3.14) we have to solve a linear problem of the following type:

\[
\begin{cases}
au - \nu \Delta u + \nabla \rho = f & \text{in } \Omega, \\
\nabla \cdot u = 0 & \text{in } \Omega, \\
u = g \text{ on } \Gamma \text{ (with } \int_{\Gamma} g \cdot n d\Gamma = 0),
\end{cases}
\]

(3.35)

where \( a \) and \( \nu \) are two positive parameters and \( f \) and \( g \) are two given functions defined on \( \Omega \) and \( \Gamma \), respectively.

We recall (see, e.g., [18], [36], [50], [52]) that if \( f \) and \( g \) are sufficiently smooth, then problem (3.35) has a unique solution in \( V_g \times (L^2(\Omega)/\mathbb{R}) \) (with \( V_g \) still defined by (3.20); \( \rho \in L^2(\Omega)/\mathbb{R} \) means that \( \rho \) is defined only up to an arbitrary constant). We shall briefly discuss in the following sections
solution of some iterative methods for solving (3.35), quite easy to implement using finite element methods (other methods are discussed in [19], Chapter 7).

3.4.2. Gradient and conjugate gradient methods for solving (3.35). A standard method to solve (3.35) is defined as follows:

\[ p^0 \in L^2(\Omega), \text{ given,} \]  

then, for \( n \geq 0 \), define \( u^n \) and \( p^{n+1} \) from \( p^n \) by

\[
\begin{align*}
\alpha u^n - \nu \Delta u^n &= f - \nabla p^n \quad \text{in } \Omega, \\
u^n &= g \quad \text{on } \Gamma,
\end{align*}
\]

(3.37)

\[ p^{n+1} = p^n - \varrho \nabla \cdot u^n. \]  

(3.38)

Concerning the convergence of (3.36)-(3.38), one can prove (see e.g. [19], Chapter 7, [18], [52]) the following

**Proposition 3.1.** Suppose that

\[ 0 < \varrho < 2\nu / N; \]

(3.39)

we then have

\[
\lim_{n \to +\infty} \{u^n, p^n\} = \{u, p_0\} \quad \text{strongly in } (H^1(\Omega))^N \times L^2(\Omega),
\]

(3.40)

where \( \{u, p_0\} \) is the solution of (3.35) such that

\[
\int_{\Omega} p_0 \, dx = \int_{\Omega} p^0 \, dx
\]

(3.41)

(actually the convergence is linear).

**Remark 3.4.** When applying algorithm (3.36)-(3.38) to solve the "quasi" Stokes problem (3.35) we have to solve at each iteration \( N \) uncoupled scalar Dirichlet problems for \( \alpha I - \nu \Delta \), to obtain \( u^n \) from \( p^n \). We see again (as in Sec. 3.3) the importance of having efficient Dirichlet solvers for \( \alpha I - \nu \Delta \).

**Remark 3.5.** Instead of algorithm (3.36)-(3.38) we should rather use in practice the following conjugate gradient variant of it, whose convergence is much faster in most cases and which is, in addition, no more costly to implement:
Description of the conjugate gradient algorithm:

Step 0: Initialization.

\( p^0 \in L^2(\Omega) \) given arbitrarily, \( (3.42) \)

solve

\[
\begin{align*}
\alpha u^0 - \nu \Delta u^0 &= f - \nabla p^0 \quad \text{in } \Omega, \\
 u^0 &= g \quad \text{on } \Gamma,
\end{align*}
\]

and set

\[
\begin{align*}
g^0 &= \nabla \cdot u^0, \\
w^0 &= g^0.
\end{align*}
\]

Then for \( n \geq 0 \), we obtain \( p^{n+1}, g^{n+1}, w^{n+1} \) from \( p^n, g^n, w^n \) by

Step 1: Descent. Compute \( \chi^n \in (H^1_0(\Omega))^N \) as the solution of

\[
\begin{align*}
\alpha \chi^n - \nu \Delta \chi^n &= -\nabla w^n \quad \text{in } \Omega, \\
 \chi^n &= 0 \quad \text{on } \Gamma,
\end{align*}
\]

then set

\[
\bar{e}_n = \frac{\int_{\Omega} w^n g^n \, dx}{\int_{\Omega} \nabla \cdot \chi^n w^n \, dx} = \frac{\int_{\Omega} |g^n|^2 \, dx}{\int_{\Omega} \nabla \cdot \chi^n \nabla w^n \, dx},
\]

and finally

\[
p^{n+1} = p^n - \bar{e}_n w^n.
\]

Step 2: Calculation of the new descent direction.

\[
g^{n+1} = g^n - \bar{e}_n \nabla \cdot \chi^n,
\]

\[
\gamma_n = \frac{\|g^{n+1}\|_{L^2(\Omega)}}{\|g^n\|_{L^2(\Omega)}} \gamma_n,
\]

\[
w^{n+1} = g^{n+1} + \gamma_n w^n.
\]

Then take \( n = n + 1 \) and go to \( (3.46) \).

Once the convergence of \( (3.42)-(3.51) \) to \( p_0 \) (that pressure solution such that \( \int_{\Omega} p_0 \, dx = \int_{\Omega} p^0 \, dx \)) has been settled we compute \( u \) from \( p_0 \) by the
solution of the Dirichlet system

\[
\begin{align*}
au - vA u &= f - \nabla p_0 \quad \text{in } \Omega, \\
u &= g \quad \text{on } \Gamma.
\end{align*}
\]

3.4.3. Another iterative method for solving (3.35). This method is in fact a penalization variant of algorithm (3.36)-(3.38) and is defined as follows (with \( r \) a positive parameter):

\begin{equation}
p^0 \in L^2(\Omega) \quad \text{given},
\end{equation}

then, for \( n \geq 0 \), define \( u^n \) and \( p^{n+1} \) from \( p^n \) by

\begin{equation}
\begin{align*}
au^n - vA u^n - r \nabla (\nabla \cdot u^n) &= f - \nabla p^n \quad \text{in } \Omega, \\
u^n &= g \quad \text{on } \Gamma, \\
p^{n+1} &= p^n - \varrho \nabla \cdot u^n.
\end{align*}
\end{equation}

**Proposition 3.2.** Suppose that

\begin{equation}
0 < \varrho < 2(r + v/N);
\end{equation}

then the convergence result (3.40) still holds for \( \{u^n, p^n\} \).

For a proof see e.g. [19], Chapter 7.

**Remark 3.6 (About the choice of \( \varrho \) and \( r \).** In practice, we should use \( \varrho = r \), since it can be proved that in that case the convergence ratio of algorithm (3.52)-(3.54) is \( O(r^{-1}) \), for large value of \( r \). In many applications, taking \( r = 10^4 v \), we have a practical convergence of algorithm (3.52)-(3.54) in 3 to 4 iterations. There is, however, a practical upper bound for \( r \): this follows from the fact that for too large values of \( r \), problem (3.53) will be ill-conditioned and its practical solution sensitive to round off errors.

**Remark 3.7.** Problem (3.53) is more complicated to solve in practice than problem (3.37) since the components of \( u^n \) are coupled by the linear term \( \nabla (\nabla \cdot u^n) \). Actually the partial differential elliptic operator in the left-hand side of (3.53) is very close to the linear elasticity operator, and close variants of it occur naturally in compressible and/or turbulent viscous flow problems.

**Remark 3.8.** Other techniques for solving (3.35) are discussed in [19], Chapter 7.
3.5. Finite element approximation of the time dependent Navier–Stokes equations.

3.5.1. Generalities. Synopsis. We shall briefly discuss in this section a specific finite element approximation for the time dependent Navier–Stokes equations. Actually this method which leads to continuous approximations for both pressure and velocity is fairly simple and has been known for years; it has been advocated by Hood–Taylor [51], and also by other authors. Other finite element approximations of the Navier–Stokes equations can be found in e.g. [18]–[20], [52], [53]. A most important reference for the theoretical study of the convergence of the approximate solution of the time dependent Navier–Stokes equations is Heywood–Rannacher [26].

3.5.2. Basic hypotheses. Fundamental discrete spaces. We suppose for simplicity that $\Omega$ is a bounded domain in $\mathbb{R}^2$. With $\mathcal{T}_h$ a standard finite element triangulation of $\Omega$, and $h$ the maximal length of the edges of the triangles of $\mathcal{T}_h$, we introduce the following discrete spaces (with $P_k =$ space of the polynomials in two variables of degree $\leq k$)

$$H^1_h = \{q_h | q_h \in C^0(\bar{\Omega}), q_h|_\Gamma \in P_1 \ \forall T \in \mathcal{T}_h \},$$

$$V_h = \{v_h | v_h \in C^0(\bar{\Omega}) \times C^0(\bar{\Omega}), v_h|_\Gamma \in P_2 \times P_2 \ \forall T \in \mathcal{T}_h \},$$

$$V_{oh} = V_0 \cap V_h = \{v_h | v_h \in V_h, v_h = 0 \ \text{on} \ \Gamma \}.$$ (3.56) (3.57) (3.58)

3.5.3. Space discretization of the time dependent Navier–Stokes equations. Using the above spaces $H^1_h, V_h, V_{oh}$ we approximate the time dependent Navier–Stokes equations as follows:

**Find** $\{u_h(t), p_h(t)\} \in V_h \times H^1_h \ \forall t \geq 0$, such that

$$\int_{\Omega} \frac{\partial u_h}{\partial t} \cdot v_h \, dx + \nu \int_{\Omega} \nabla u_h \cdot \nabla v_h \, dx + \int_{\Omega} (u_h \cdot \nabla) u_h \cdot v_h \, dx + \int_{\Omega} \nabla p_h \cdot v_h \, dx$$

$$= \int_{\Omega} f_h \cdot v_h \, dx \ \forall v_h \in V_{oh},$$ (3.59)

$$\int_{\Omega} \nabla \cdot u_h q_h \, dx = 0 \ \forall q_h \in H^1_h,$$ (3.60)

$$u_h = g_h \ \text{on} \ \Gamma,$$ (3.61)

$$u_h(x, 0) = u_{oh}(x)(u_{oh} \in V_h);$$ (3.62)
in (3.59)–(3.62), \( f_h \) and \( u_{oh} \) are convenient approximations of \( f \) and \( u_0 \), respectively, and \( g_h \) is an approximation of \( g \) such that \[ \int_D g_h \cdot n \, dI = 0 \] (for the construction of \( g_h \) see [19], Appendix 3, or [25]).

We have thus reduced the solution of the time dependent Navier-Stokes equations to that of a nonlinear system of algebraic and ordinary differential equations. We observe that the incompressibility condition is only approximately satisfied. The time discretization of (3.59)–(3.62) is discussed in Section 3.5.4 below.

3.5.4. Time discretization of (3.59)–(3.62) by alternating direction methods.

We now consider a fully discrete version of the scheme (3.8)–(3.10) discussed in Sec. 3.2.1; it is defined as follows (with \( \Delta t \) and \( \theta \) as in Sec. 2):

\[ u_h^0 = u_{oh}, \quad (3.63) \]

then, for \( n \geq 0 \), compute (from \( u_h^n \)) \( \{u_h^{n+1/2}, p_h^{n+1/2}\} \in V_h \times H_h \), and \( u_h^{n+1} \in V_h \), by solving

\[ \int_D \frac{u_h^{n+1/2} - u_h^n}{\Delta t/2} \cdot v_h \, dx + \nu \int_D \nabla u_h^{n+1/2} \cdot \nabla v_h \, dx + \int_D \nabla p_h^{n+1/2} \cdot v_h \, dx \]

\[ = \int_D f_h^{n+1/2} \cdot v_h \, dx -(1-\theta) \nu \int_D \nabla u_h^n \cdot \nabla v_h \, dx - \int_D (u_h^n \cdot \nabla) u_h^n \cdot v_h \, dx \quad \forall v_h \in V_{oh}, \quad (3.64) \]

\[ \int_D \nabla \cdot u_h^{n+1/2} g_h \, dx = 0 \quad \forall g_h \in H_h^1, \quad (3.65) \]

\[ u_h^{n+1/2} \in V_h, \quad p_h^{n+1/2} \in H_h^1, \quad u_h^{n+1/2} = g_h^{n+1/2} \text{ on } \Gamma, \quad (3.66) \]

and then

\[ \int_D \frac{u_h^{n+1} - u_h^{n+1/2}}{\Delta t/2} \cdot v_h \, dx + (1-\theta) \nu \int_D \nabla u_h^{n+1} \cdot \nabla v_h \, dx + \int_D (u_h^{n+1} \cdot \nabla) u_h^{n+1} \cdot v_h \, dx \]

\[ = \int_D f_h^{n+1} \cdot v_h \, dx - \theta \nu \int_D \nabla u_h^{n+1/2} \cdot \nabla v_h \, dx - \int_D \nabla p_h^{n+1/2} \cdot v_h \, dx \quad \forall v_h \in V_{oh}, \quad (3.67) \]

\[ u_h^{n+1} \in V_h, \quad u_h^{n+1} = g_h^{n+1} \text{ on } \Gamma. \quad (3.68) \]
Obtaining the fully discrete analogue of scheme (3.11)-(3.14) is straightforward. Solving the linear and nonlinear subproblems encountered at each step of (3.63)-(3.68) can be done by the discrete analogues of the methods discussed in Sections 3.3, 3.4; for more details see [19], Chapter 7 and Appendix 3, where the interest of efficient Poisson solvers as basic tools appears clearly.

Modifying (3.64)-(3.67), to take into account the augmented nonlinear operator $\tilde{B}$ introduced in Section 3.2.3, Remark 3.2, is quite easy, but as mentioned before, it has no practical influence on the numerical results that we obtained.

3.5.5. **Numerical experiments.** We illustrate the numerical techniques discussed in the above sections by presenting the results of numerical experiments where these techniques have been used to simulate several incompressible viscous flows modeled by the Navier–Stokes equations.

3.5.5.1. **Flow in a channel with a step.** The first numerical experiment that we have done concerns a Navier–Stokes flow in a channel with a step at $Re = 191$; the characteristics length used to compute the Reynolds number is the height of the step. Poiseuille’s profiles of velocity have been prescribed upstream and quite far downstream. The corresponding streamlines are shown in Fig. 3.2; we clearly see in Fig. 3.2 a thin separation layer starting slightly below the upper corner of the step, and separating a recirculation region from a region where the flow is quasi potential. The results obtained for this test problem are in very good agreement with those obtained by several authors using different methods (see [19] and [42]).

---

Fig. 3.2 The streamlines shown are those for which the streamfunction assumes values $(n/15)$, for integers $n$ between $-5$ and $+15$. The stepped (lower) boundary of the channel corresponds to $n = 0$. 
3.5.5.2. *Flow around and inside a nozzle.* This experiment concerns an unsteady flow around and inside a nozzle at high incidence and at $Re = 750$ (the characteristic length being the distance between the nozzle walls). We have shown in Fig. 3.3 a part of the finite element triangulation used for the computation and in Figs. 3.4–3.7 the streamlines at $t = 0, 0.2, 0.4, 0.6$, respectively, showing clearly the creation and the motion of eddies, inside and behind the nozzle.
4. Application to the calculation of potential transonic flows for compressible inviscid fluids

4.1. Generalities. The physical problem. The numerical simulation of transonic potential flows of compressible inviscid fluids is a non-trivial problem since

(1) The equations governing these flows are nonlinear and of changing type (elliptic in the subsonic region of the flow, hyperbolic in the supersonic region);

(2) Shocks may exist corresponding to discontinuities of velocity, pressure and density;

(3) An entropy condition must be included in order to eliminate rarefaction shocks since they correspond to unphysical situations.

We suppose in the following that the fluids to be considered are compressible and inviscid and that their flows are potential and therefore quasi-isentropic, with weak shocks only; in fact, this is only an approximation since usually a flow is no longer potential after a shock (cf. [32]). In the case of flows past bodies we shall suppose that these bodies are sufficiently thin and parallel to the main flow in order not to create a wake in the outflow.

4.2. Mathematical formulation. Let \( \Omega \) be the region of the flow and \( \Gamma \) its boundary; it follows from [32] that the flow is governed by

\[
\nabla \cdot \rho \mathbf{u} = 0 \quad \text{in} \quad \Omega,
\]

where

\[
\rho = \rho_0 \left(1 - \frac{(\gamma - 1)}{(\gamma + 1)} \frac{|\mathbf{u}|^2}{c_s^2}\right)^{1/(\gamma - 1)},
\]

\[
\mathbf{u} = \nabla \varphi.
\]

In the above relations \( \varphi \) is the velocity potential, \( \rho \) is the density of the fluid, \( \gamma \) (= 1.4 in air) is the ratio of specific heats and \( c_s \) is the critical velocity.

For an airfoil \( B \) (see Fig. 4.1) we assume that the flow is uniform on \( \Gamma_\infty \) and tangential at \( \Gamma_B \). We then have

\[
\frac{\partial \varphi}{\partial n} = u_\infty \cdot \mathbf{u} \quad \text{on} \quad \Gamma_\infty, \quad \frac{\partial \varphi}{\partial n} = 0 \quad \text{on} \quad \Gamma_B.
\]
Since only Neumann boundary conditions are involved the potential is determined up to an arbitrary constant. To remedy this we should prescribe the value of \( \varphi \) at some point within \( \Omega \cap \Gamma_B \) and, for example,

\[
\varphi = 0 \quad \text{at the trailing edge (T.E.) of } B. \quad (4.5)
\]

In addition to (4.4), (4.5) another condition known as the Kutta–Joukowsky condition has to be prescribed; it requires "some" continuity of \( u \), even at the corners, and is particularly important for lifting bodies.

Since this condition is not specific of transonic flows (it occurs also for compressible inviscid subsonic flows and incompressible inviscid flows), it will not be discussed here (see [7], [8] for the numerical treatment of the Kutta–Joukowsky condition).

Another most important feature of inviscid transonic flows is the existence of shocks; across a shock the flow must satisfy the Rankine–Hugoniot conditions

\[
(q u \cdot n)_+ = (q u \cdot n)_- \quad \text{(where } n \text{ is normal at the shock line or surface)},
\]

\[
\text{the tangential component of the velocity is continuous.} \quad (4.6)
\]

As regards the entropy condition, it can be formulated as follows:

\[
\text{Following the flow, we cannot have a positive variation of velocity through a shock since this would imply a negative variation of entropy which is an unphysical phenomenon.} \quad (4.8)
\]
4.3. Least-squares formulation of the continuous problem. We will not consider here the practical implementation of (4.8) (it will be discussed briefly in Sec. 4.4); we consider only the variational formulation of (4.1)–(4.4), (4.6), (4.7) and an associated nonlinear least squares formulation.

4.3.1. A variational formulation of the continuity equation. We consider for simplicity the situation in Fig. 4.2 which shows a symmetric flow, subsonic at infinity, around a symmetric airfoil; thus the Kutta-Joukowski condition is automatically satisfied.

For practical purposes (other approaches are also possible) we imbed the airfoil in a “large” domain; using the notation of Section 4.2, the continuity equation and the boundary conditions are

\[ \nabla \cdot (\rho \varphi \nabla \varphi) = 0 \quad \text{in} \ \Omega \]  

(4.9)

with

\[ \rho(\varphi) = \rho_0 \left(1 - \frac{\gamma - 1}{\gamma + 1} \frac{|
abla \varphi|^2}{\rho_0^2} \right)^{\gamma/(\gamma - 1)} \]  

(4.10)

and

\[ \rho \frac{\partial \varphi}{\partial n} = 0 \quad \text{on} \ \Gamma_B, \quad \rho \frac{\partial \varphi}{\partial n} = \rho_\infty \mathbf{u}_\infty \cdot \mathbf{n} \quad \text{on} \ \Gamma_\infty. \]  

(4.11)

Define \( g \) on the set \( \Gamma(= \Gamma_B \cup \Gamma_\infty) \) by

\[ g = 0 \quad \text{on} \ \Gamma_B, \quad g = \rho_\infty \mathbf{u}_\infty \cdot \mathbf{n} \quad \text{on} \ \Gamma_\infty. \]  

(4.12)

Clearly, we have

\[ \rho \frac{\partial \varphi}{\partial n} = g \quad \text{and} \quad \int_{\Gamma} g \, d\Gamma = 0. \]  

(4.13)
An equivalent variational formulation is
\[ \int_Q q(\varphi) \nabla \varphi \cdot \nabla v \, dx = \int_I g(v) \, d\Gamma, \quad \forall v \in H^1(\Omega), \quad \varphi \in W^{1,\infty}(\Omega)/\mathbb{R}. \] (4.14)

The space \( W^{1,\infty}(\Omega) \) is a natural choice for \( \varphi \) since physical flows require (among other properties) a positive density \( q \); therefore, in view of (4.10), \( \varphi \) must satisfy
\[ |\nabla \varphi| \leq \delta < \left( \frac{\gamma + 1}{\gamma - 1} \right)^{1/2} c_* \text{ a.e. in } \Omega. \]

4.3.2. A least squares formulation of (4.14). For a genuine transonic flow, problem (4.14) is not equivalent to a standard problem of the calculus of variations (as it would be for purely subsonic flows); to remedy this situation and — in some sense — convexify the problem under consideration, we introduce a nonlinear least squares formulation of (4.14) as follows. Let \( X \) be a set of feasible solutions; the least squares problem is then
\[ \min_{\xi \in X} J(\xi) \] (4.15)
with
\[ J(\xi) = \frac{1}{2} \int_{\Omega} |\nabla y(\xi)|^2 \, dx \] (4.16)
where, in (4.16), \( y(\xi) (= y) \) is a solution of
\[ \begin{cases} \text{Find } y \in H^1(\Omega)/\mathbb{R} \text{ such that} \\ \int_{\Omega} \nabla y \cdot \nabla v \, dx = \int_{\Omega} q(\xi) \nabla y \cdot \nabla v \, dx - \int_I g(v) \, d\Gamma, \quad \forall v \in H^1(\Omega). \end{cases} \] (4.17)

If (4.14) has solutions, these solve (4.15) and give the value zero to the objective function \( J \).

4.4. Finite element approximation. We consider here only two-dimensional problems but the methods described have been applied to three-dimensional problems.

4.4.1. Finite element approximation of (4.14). We still consider the nonlifting situation of Section 4.3.1; once the flow region has been embedded in a large domain \( \Omega \), we approximate this latter domain by a polygonal
domain \( \Omega_h \); with \( \mathcal{T}_h \) a standard triangulation of \( \Omega_h \), we approximate \( H^1(\Omega) \) (and in fact \( W^{1,p}(\Omega), \forall p \geq 1 \)) by

\[
H^1_h = \{ v_h \mid v_h \in C^0(\overline{\Omega}_h), v_h|_T \in P_1 \ \forall T \in \mathcal{T}_h \} \tag{4.18}
\]

where \( P_1 \) is the space of polynomials in two variables of degree \( \leq 1 \). We prescribe the value for the potential at T.E.; this leads to

\[
V_h = \{ v_h \in H^1_h, v_h(T.E.) = 0 \}. \tag{4.19}
\]

Clearly,

\[
\dim V_h^1 = 1 + \dim V_h = \text{number of vertices of } \mathcal{T}_h. \tag{4.20}
\]

We then approximate the variational equation (4.14) (dropping \( h \) in \( \Omega_h \) and \( I_h \)) by

\[
\begin{align*}
&\text{Find } \varphi_h \in V_h \text{ such that} \\
&\int_{\Omega} g(\varphi_h) \nabla \varphi_h \cdot \nabla v_h \, dx = \int_{I} g_h v_h \, d\Gamma \quad \forall v_h \in V_h
\end{align*} \tag{4.21}
\]

where \( g_h \) is an approximation of the function \( g \) of (4.13). Let \( \mathcal{B}_h = \{ w_i \}_{i=1}^{N_h} \) be a vector basis of \( V_h \). Then (4.21) is equivalent to the nonlinear finite-dimensional system

\[
\begin{align*}
\varphi_h &= \sum_{j=1}^{N_h} \varphi_j w_j, \\
\int_{\Omega} g(\varphi_h) \nabla \varphi_h \cdot \nabla w_i \, dx &= \int_{I} g_h w_i \, d\Gamma \quad \forall i = 1, \ldots, N_h. \tag{4.22}
\end{align*}
\]

With the above choice for \( H^1_h \) and \( V_h \), there is no problem of numerical integration since, in (4.21) and (4.22), \( \nabla \varphi_h, \nabla v_h \) (and therefore \( g(\varphi_h) \)) are piecewise constant.

4.4.2. Numerical implementation of the entropy condition. The numerical implementation of the entropy condition (4.8), in order to eliminate rarefaction shocks, is a non-trivial matter. Without going into details, we should mention that methods founded on the upwinding of the density have been implemented, producing rather good numerical results (see [9], [27] and also [19], Chapter 7, for technical details and further references).
The method for upwinding the density discussed in [9] leads to the following equation:

\[
\begin{aligned}
\text{Find } \varphi_h \in V_h \text{ such that }
\int_{\Omega} c(\varphi_h) \nabla \varphi_h \cdot \nabla v_h \, dx + \int_{\Omega} R_h(\varphi_h) v_h \, dx = \int_{\Gamma} g_h v_h \, d\Gamma \quad \forall v_h \in V_h,
\end{aligned}
\]

in which \( R_h \) can be viewed as an artificial viscosity operator (see [9] for a full description of \( R_h \)).

The solution of (4.23) by nonlinear least squares methods is achieved by the following variant of (4.15):

\[
\begin{aligned}
\min_{\xi_h \in V_h} J_h(\xi_h)
\end{aligned}
\]

Fig. 4.3. \( \alpha = 6^\circ, M_\infty = .6 \)
with

\[ J_h(\xi_h) = \frac{1}{2} \int_{\Omega} |\nabla y_h|^2 \, dx \]  \hspace{1cm} (4.25)

where, in (4.25), \( y_h \) is a solution of

\[
\begin{align*}
\int_{\Omega} \nabla y_h \cdot \nabla v_h \, dx &= \int_{\Omega} \eta(\xi_h) \nabla \xi_h \cdot \nabla v_h \, dx + \int_{\partial \Omega} R_h(\xi_h) v_h \, dx - \int_{\Gamma} g_h v_h \, d\Gamma.
\end{align*}
\]  \hspace{1cm} (4.26)

The solution of (4.24)–(4.26) by a conjugate gradient algorithm (fairly close to algorithm (2.14)–(2.22)) is discussed in [19], Chapter 7, and [9].

Fig. 4.4. \( a = 6^\circ, M_\infty = .6 \)

4.5. Numerical experiments.

4.5.1. Flows around a NACA 0012 airfoil. Figures 4.3–4.5 show the pressure distribution and isomach lines for flows around a NACA 0012 airfoil at various \( M_\infty \) and angles of attack.
Fig. 4.5. $a = 0^\circ$, $M_\infty = .85$
Fig. 4.6. $\alpha = 0^\circ$, $M_\infty = 0.90$
These results are in good agreement with those obtained by other authors by quite different methods (mostly finite difference methods); see [19] and [9] for further references.

4.5.2. Flows around a NACA 64006 airfoil. We have shown in Figs. 4.6, 4.7 the Mach distribution corresponding to flows around a NACA 64006 airfoil for $M_\infty = .89$ and the angle of attack $\alpha = 0^\circ$. The flow on Fig.

Fig. 4.6

42 — Proceedings..., t. II
4.6 (resp. 4.7) is symmetrical (resp. non-symmetrical); thus we have (at least) three solutions to the same problem (the third one is obtained from the one of Fig. 4.7 by symmetry with respect to the symmetry axis of the airfoil); these three solutions satisfy the continuity equation, the Rankine–Hugoniot, Kutta–Joukowski and entropy conditions. Actually, the symmetrical one seems to be instable with regards to small nonsymmetric perturbations; for more details, see e.g. [49].
5. Decomposition methods for variational problems by augmented Lagrangians. An application in finite elasticity

5.1. Generalities. The main goal of this section is to give a brief account of solution methods for variational problems when some decomposition property holds; introducing a convenient augmented Lagrangian, we obtain solution methods, taking advantage of the special structure of the problem under consideration; these methods are described and discussed in Section 5.2. An application in Finite Elasticity is considered in Section 5.3.

For more details and further references, see [20] and also [19], Chapter 6.


5.2.1. A family of variational problems. In the sequel we consider real Hilbert spaces only; let $V$ and $H$ be two such vector spaces, equipped with the norms and scalar products $\|\cdot\|$, $(\cdot, \cdot)$ and $|\cdot|$, $(\cdot, \cdot)$, respectively. Let $B \in \mathcal{L}(V, H)$ and $F, G$ be two functionals convex, proper, l.s.c., from $H$ and $V$ to $\mathbb{R} \cup \{+\infty\}$, respectively. We suppose that

$$\text{dom}(G) \cap \text{dom}(F \circ B) \neq \emptyset,$$  \hfill (5.1)

where

$$\text{dom}(G) = \{v \mid v \in V, -\infty < G(v) < +\infty\}$$

and a similar definition for dom($F \circ B$). We associate with $V, H, B, F, G$ the following minimization problem:

$$\begin{align*}
\text{Find } u \in V \text{ such that } \\
J(u) \leq J(v) \quad \forall v \in V
\end{align*}$$

(P)

where $J: V \rightarrow \mathbb{R}$ is defined by

$$J(v) = F(Bv) + G(v).$$ \hfill (5.2)

Since $J(\cdot)$ has a special structure, it is natural to look for methods taking advantage of this fact.

Remark 5.1. Most of the following considerations can be applied to the solution of variational problems such as

$$f \in B' A_1(Bu) + A_2(u),$$ \hfill (5.3)
where \( f \in V' \) (the dual space of \( V \)) and where \( A_1, A_2 \) are monotone operators from \( H \) to \( H' \) (dual space of \( H \)) and from \( V \) to \( V' \), respectively. In general, \( A = B' \circ A_1 \circ B + A_2 \) is not the gradient (or subgradient) of a functional \( (B' \) is the transposed operator of \( B \)).

5.2.2. A decomposition principle. We define a set \( W \subset V \times H \) by

\[
W = \{ (v, q) \in V \times H, Bv - q = 0 \}.
\]  

Problem (P) is equivalent to

\[
\begin{align*}
\text{Find } \{ u, p \} \in W \text{ such that } \\
\text{subject to } j(u, p) \leq j(v, q) \quad \forall (v, q) \in W,
\end{align*}
\]

with

\[
j(v, q) = F(q) + G(v).
\]  

Remark 5.2. Problems (P) and (II) are equivalent, but (II) has, in some sense, a simpler structure than (P), despite the fact that it contains an extra variable. This is due to the fact that the linear relation

\[
Bv - q = 0
\]

can be efficiently treated, using simultaneously — via an appropriate augmented Lagrangian — the penalty and Lagrange multiplier methods of solution.

5.2.3. An augmented Lagrangian associated to (II). Let \( r > 0 \); we define

\[
\mathcal{L}_r : V \times H \times H \to \mathbb{R}
\]

by

\[
\mathcal{L}_r(v, q, \mu) = F(q) + G(v) + \frac{r}{2} |Bv - q|^2 + (\mu, Bv - q).
\]

We easily prove that if \( \{ (u, p), \lambda \} \) is a saddle point of \( \mathcal{L}_r \) over \( (V \times H) \times H \), then \( \{ u, p \} \) is a solution of (II), i.e. \( u \) is a solution of (P) (with \( p = Bu \)).

5.2.4. A first algorithm for solving (P). To solve (P) and (II) we look for saddle points of \( \mathcal{L}_r \) using duality algorithms like those discussed in [24]; such an algorithm is defined as follows:

\[
\lambda^0 \in H \text{ given},
\]  

\[
\lambda^1 = \lambda^0 - B^* \mu^0,
\]

\[
\mu^1 = \mu^0 + B\lambda^1 - r \lambda^1,
\]

\[
\mathcal{L}_r(u^1, p^1, \lambda^1) = \min_{(v, q)} \mathcal{L}_r(v, q, \lambda^1),
\]

\[
\text{subject to } j(u^1, p^1) \leq j(v, q) \quad \forall (v, q) \in W.
\]
then for \( n \geq 0 \), \( \lambda^n \) being known, we compute \( u^n, p^n \) and \( \lambda^{n+1} \) by

\[
\begin{align*}
\text{Find } \{u^n, p^n\} &\in V \times H \text{ such that } \\
\mathcal{L}_r(u^n, p^n, \lambda^n) &\leq \mathcal{L}_r(v, q, \lambda^n) \quad \forall \{v, q\} \in V \times H, \\
\lambda^{n+1} &= \lambda^n + \varepsilon (Bu^n - p^n).
\end{align*}
\] (5.9)

We have proved in [20], Chapter 3, that if \( F, B, G \) satisfy quite reasonable hypotheses and if

\[ 0 < \varepsilon < 2r, \] (5.11)

then we have

\[
\begin{align*}
limit_{n \to +\infty} u^n &= u \quad \text{strongly in } V, \\
limit_{n \to +\infty} p^n &= p (= Bu) \quad \text{strongly in } H, \\
limit_{n \to +\infty} \lambda^n &= \lambda \quad \text{weakly in } H,
\end{align*}
\] (5.12-5.14)

where \( u \) is the solution of (P), and where \( \lambda \) is such that \( \{u, p\}, \lambda \) is a saddle point of \( \mathcal{L}_r \) on \( (V \times H) \times H \).

Remark 5.3. To solve (5.9), we can use block-relaxation algorithms like those discussed in [12] (see [20] for more details); if we use these relaxation methods to solve (5.9) and limit ourselves to only one inner iteration, we obtain the algorithm described in Sec. 5.2.5.

5.2.5. A second algorithm for solving (P). It is defined by

\[
\{u^{-1}, \lambda^0\} \in V \times H \quad \text{given},
\] (5.15)

then, for \( n \geq 0 \), \( u^{n-1} \) and \( \lambda^n \) being known, we compute \( p^n, u^n, \lambda^{n+1} \) by

\[
\begin{align*}
\mathcal{L}_r(u^{n-1}, p^n, \lambda^n) &\leq \mathcal{L}_r(u^{n-1}, q, \lambda^n) \quad \forall q \in H, \quad p^n \in H, \\
\mathcal{L}_r(u^n, p^n, \lambda^n) &\leq \mathcal{L}_r(v, p^n, \lambda^n) \quad \forall v \in V, \quad u^n \in V, \\
\lambda^{n+1} &= \lambda^n + \varepsilon (Bu^n - p^n).
\end{align*}
\] (5.16-5.18)

Remark 5.4. Several variants of (5.15)-(5.18) can be derived; we may for example

(i) interchange the role of \( q \) and \( v \) (see also Remark 5.5),
(ii) update $\lambda^n$ between steps (5.16), (5.17); doing so we obtain the following variant of (5.15)-(5.18) (due to Gabay [17]):

$$\{u^{-1}, \lambda^0\} \text{ given in } V \times H, \tag{5.19}$$

then for $n \geq 0$, $u^{n-1}$ and $\lambda^n$ being given, we compute $p^n$, $\lambda^{n+1/2}$, $u^n$, $\lambda^{n+1}$ by

$$\mathcal{L}_r(u^{n-1}, p^n, \lambda^n) \leq \mathcal{L}_r(u^{n-1}, q, \lambda^n) \quad \forall q \in H, \quad p^n \in H, \tag{5.20}$$

$$\lambda^{n+1/2} = \lambda^n + (Bu^{n-1} - p^n), \tag{5.21}$$

$$\mathcal{L}_r(u^n, p^n, \lambda^{n+1/2}) \leq \mathcal{L}_r(v, p^n, \lambda^{n+1/2}) \quad \forall v \in V, \quad u^n \in V, \tag{5.22}$$

$$\lambda^{n+1} = \lambda^{n+1/2} + \varepsilon (Bu^n - p^n); \tag{5.23}$$

$q$ and $v$ play a more symmetrical role in (5.19)–(5.23) than in (5.15)–(5.18).

**Remark 5.5.** If one uses (5.15)–(5.18), it is suggested to solve in the second step the problem with the best ellipticity properties (cf. [20], Chapter 3, for the justification of such a choice). ■

As regards the convergence of (5.15)–(5.18), one proves in [20], Chapter 3, that the convergence results (5.12)–(5.14) hold if

$$0 < \varepsilon < \frac{1 + \sqrt{5}}{2} r. \tag{5.24}$$

5.2.6. **Comments on the choice of $\varepsilon$ and $r$.** For a given $r$, it follows from various numerical experiments we have done that the optimal choice for $\varepsilon$ is close to $r$. The choice of $r$ is more complicated; theoretically, the speed of convergence of (5.8)–(5.10) increases with $r$, but the conditioning of problem (5.9) deteriorates as $r$ increase.

If one uses algorithms (5.15)–(5.19) and (5.19)–(5.23) with $\varepsilon = r$, the optimal choice for $r$ is again a problem difficult to analyse.

5.2.7. **Relations with alternating direction methods.** Algorithms (5.15)–(5.18) and (5.19)–(5.23) are closely related to alternating direction methods, as shown in [20], Chapters 8 and 9; for the convergence properties of these alternating direction methods and relations with the numerical integration of time dependent problems see [17], [37].

5.3. **Application in finite elasticity.** We apply the decomposition methods of Section 5.2 to the numerical solution of nonlinear problems in finite elasticity dealing with incompressible materials of the Mooney–Rivlin type (we follow here [20], Chapter 8, and [22]).
5.3.1. Formulation of the elasticity problem. A fundamental problem in finite elasticity is the calculation of the deformations and displacements of a solid body made of an homogeneous, isotropic, hyperelastic and incompressible material submitted to volume forces \(f_0\) (\(f_0\) is the density of the material) and superficial forces \(S_0\). Using a Lagrangian formulation, the functional of energy associated with a displacement field \(v\) is given by

\[
\Pi(v) = \int_{\Omega} f_0 (\sigma(v) - f_0 v) \, dx - \int_{\partial\Omega_2} S_0 \cdot v \, d\Gamma, \tag{5.25}
\]

where \(\Omega\) is the domain in \(\mathbb{R}^N\) corresponding to the reference configuration; \(\partial\Omega = \partial\Omega_1 \cap \partial\Omega_2\) is the boundary of \(\Omega\). The body being fixed along \(\partial\Omega_1\), we have denoted by \(\sigma(v)\) the stored energy functional (per unit mass). For a Mooney-Rivlin material we have

\[
\sigma(v) = E_1 (I_1 - 2) \quad \text{if} \quad N = 2, \tag{5.26}
\]

\[
\sigma(v) = E_1 (I_1 - 3) + E_2 (I_2 - 3) \quad \text{if} \quad N = 3, \tag{5.27}
\]

where \(I_i\) is the \(i\)th invariant of the tensor \(FF^t\), where

\[
F = J + \nabla v, \tag{5.28}
\]

and where \(E_1, E_2\) are positive coefficients, material dependent.

The displacement \(v\) satisfies the incompressibility condition

\[
det F(v) = 1 \quad \text{a.e. on } \Omega. \tag{5.29}
\]

Remark 5.6. We have supposed in (5.25) that \(f\) and \(S_0\) are independent of \(v\) (dead load hypothesis); actually we refer to [22], [33], [34] where the algorithms to be described below are generalized to problems for which the above hypothesis is not satisfied.

It is reasonable to suppose that the displacements \(u\) corresponding to the stable equilibria satisfy

\[
u \text{ is a local minimizer over } K \text{ of the functional } \Pi, \tag{5.30}
\]

where, for a Mooney-Rivlin incompressible material, we have

\[
K = \{ v \in (H^1(\Omega))^N | \ v = 0 \text{ on } \partial\Omega_1, \ det \ F(v) = 1 \}
\]

\[
a.e., \ F^{-1}(v) \in (L^2(\Omega))^{N \times N} \}
\]

and where \(\Pi\) is defined by (5.25)-(5.27).

The existence of solutions for (5.30), (5.31) is proved in [2].
We can give also a formulation founded on the following augmented Lagrangian (associated with the linear relation (5.28)), with \( r > 0 \):

\[
\mathcal{L}_r(v, G, \mu) = H(v) + \frac{r}{2} \|\nabla v + J - G\|_{L^2}^2 - \int_\Omega \mu \cdot (\nabla v + J - G) \, dv. 
\]  

(5.32)

This leads to the following formulation of the elastostatic problem:

\[ \text{Find } \{u, F, \lambda\} \in W = X \times Y \times (L^2(\Omega))^{N \times N}, \text{ a stationary point of } \mathcal{L}_r \text{ over } W, \]  

(5.33)

where

\[ Y = \{G \mid G \in (L^2(\Omega))^{N \times N}, G^{-1} \in (L^2(\Omega))^{N \times N}, \det G = 1 \text{ a.e.} \}. \]

The relations between formulations (5.30), (5.31) and (5.33) are discussed in [22], [33], [34] (together with other formulations).

5.3.2. Solution of problem (5.30), (5.31).

5.3.2.1. A first algorithm. It corresponds to (5.8)-(5.10) of Section 5.2 and is defined by:

\[ \lambda^0 \text{ is given in } (L^2(\Omega))^{N \times N}, \]  

(5.34)

then for \( n \geq 0 \), \( \lambda^n \) being known, we obtain \( u^n, F^n \) and \( \lambda^{n+1} \) from the solution of

\[
\begin{cases} 
\mathcal{L}_r(u^n, F^n, \lambda^n) \leq \mathcal{L}_r(v, G, \lambda^n) & \forall \{v, G\} \in X \times Y, \\
\{u^n, F^n\} \in X \times Y, \\
\lambda^{n+1} = \lambda^n - \varepsilon (\nabla u^n + J - F^n), \quad \varepsilon > 0.
\end{cases} 
\]  

(5.35)

(5.36)

Remark 5.7. Problem (5.35) is equivalent to the nonlinear system

\[ \mathcal{L}_r(u^n, F^n, \lambda^n) \leq \mathcal{L}_r(u^n, G, \lambda^n) \quad \forall G \in Y, F^n \in Y, \]  

(5.37)

\[ \partial_u \mathcal{L}_r(u^n, F^n, \lambda^n) \cdot v = 0 \quad \forall v \in X, u^n \in X, \]  

(5.38)

whose solution using block relaxation methods leads to the algorithm thereafter.
5.3.2.2. A second algorithm. It corresponds to (5.15)-(5.18) of Section 5.2 and is defined by:

\[ u^{-1} \text{ is given in } X, \lambda^0 \text{ is given in } (L^2(\Omega))^{N \times N}, \]

(5.39)

then for \( n \geq 0 \), \( u^{n-1} \) and \( \lambda^n \) being known, we obtain \( F^m, u^n \) and \( \lambda^{n+1} \) by the solution of

\[ \mathcal{L}_r(u^{n-1}, F^m, \lambda^n) \leq \mathcal{L}_r(u^{n-1}, F, \lambda^n) \quad \forall G \in Y, F^m \in Y, \]  
\[ \partial_r \mathcal{L}_r(u^n, F^m, \lambda^n) \cdot v = 0 \quad \forall v \in X, u^n \in X, \]
\[ \lambda^{n+1} = \lambda^n - \varrho (\nabla u^n + J - F^m), \quad \varrho > 0. \]

(5.40)  
(5.41)  
(5.42)

Problem (5.41), which is equivalent to

\[
\begin{aligned}
\text{Find } u^n \in X \text{ such that } \\
\mathcal{L}_r(u^n, F^m, \lambda^n) \leq \mathcal{L}_r(v, F^m, \lambda^n) \quad \forall v \in X,
\end{aligned}
\]

(5.43)

is in fact an unconstrained minimization problem whose solution is rather easy, particularly if \( r \) is sufficiently large; if \( N = 2 \) the functional in (5.43) is quadratic, and solving (5.41), (5.43) is equivalent to solving a linear problem for a second order partial differential operator (close to the linear elasticity operator), independent of \( n \), and whose finite-dimensional variants are linear systems for symmetric, positive-definite matrices, independent of \( n \). Problem (5.40) is more complicated (apparently, at least); if \( N = 2 \), (5.40) leads to

\[
\begin{aligned}
\text{Find } F \in (L^2(\Omega))^4 \text{ such that } F_{11}F_{22} - F_{12}F_{21} = 1 \text{ a.e.} \\
\text{and minimizing the functional } \\
G \rightarrow \int \left[rG_{ij}^2 - 2(r(u_{ij} + \delta_{ij}) - \lambda_{ij})G_{ij}\right] dx \\
\text{on the set of the } G \in (L^2(\Omega))^4 \text{ such that } G_{11}G_{22} - G_{12}G_{21} = 1 \text{ a.e.;}
\end{aligned}
\]

(5.44)

in (5.44) \( n \) is omitted, and the summation convention of repeated indices is used, \( u_{ij} = \partial u_{i}/\partial x_j \) and \( \delta_{ij} \) is the Kronecker symbol. Since (5.44) does not contain any derivative of \( G \) and \( F \), we can solve this latter problem pointwise as shown in [20], Chapter 8, [22], [33], [34], using a diagonalization of the constraint \( F_{11}F_{22} - F_{12}F_{21} = 1 \), via the transformation

\[ b_1 = (F_{11} + F_{22})/\sqrt{2}, \quad b_2 = (F_{11} - F_{22})/\sqrt{2}, \]
\[ b_3 = (F_{12} + F_{21})/\sqrt{2}, \quad b_4 = (F_{12} - F_{21})/\sqrt{2}. \]

(5.45)
5.3.3. A numerical experiment. We suppose \( N = 2 \); we approximate (5.30), (5.31) (and (5.33)) using a finite element method. We have used rectangular finite elements \( K \in Q_h \), where \( Q_h \) is a quadrangulation of \( \Omega \). We approximate then the displacement \( v \) by \( v_h \in C^0(\overline{\Omega}) \times C^0(\overline{\Omega}) \), such that

\[
v_h|_K \in Q_1 \times Q_1 \quad \forall K \in Q_h,
\]

where

\[
Q_1 = \{ q \mid q(x_1, x_2) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 \};
\]

we require the incompressibility condition (5.29) at the center of each elementary rectangle \( K \in Q_h \) (which is equivalent to require it in an averaging sense). The convergence of the approximate solutions, as \( h \to 0 \) is a very difficult problem, discussed e.g. in [33], [34].

In the following numerical experiments \( \Omega \) is a two-dimensional bar containing a (non-propagating) crack. We have shown in Fig. 5.1 the right half part of the bar, the crack and the quadrangulation \( Q_h \) (or rather the right half part of it). We suppose that in (5.25), (5.26) we have \( c_0 = 1 \), \( E_1 = 1 \), \( \partial \Omega_2 = \partial \Omega \) and that \( S_0 \) corresponds to horizontal stretching forces whose density modulus is 2, these forces being applied at the extremities
of the bar. Under the action of these forces we have a stretching phenomenon and we have shown in Figure 5.2 the equilibrium configuration, computed by a discretized variant of algorithm (5.34)-(5.36), with \( q = r = 10 \); the convergence is obtained in 20 iterations corresponding to a running time of 5 seconds on CDC 6400. We should observe with interest the behavior of the crack.

One can find in [22], [23], [33], [34] numerical experiments for other two-dimensional problems and also for axisymmetric and three-dimensional problems.

6. Conclusion

We have shown in this paper that variational methods can be applied to the numerical solution of large classes of nonlinear problems governed by partial differential equations, even for situations which are not equivalent, in a strict sense, to a problem of the calculus of variations. For a more complete discussion concerning these methods and their numerical applications, see, e.g. [19] and [20].

References


[47] Reinhart L., *Sur la résolution numérique de problèmes aux limites non linéaires*
par des méthodes de continuation, Thèse de 3ème cycle, Université Pierre et Marie Curie, Paris, 1980.


UNIVERSITÉ PIERRE ET MARIE CURIE,
4 PLACE JUSSIEU – 75230 PARIS CEDEX 05 AND INRIA