CHAPTER V

Relaxation Methods and Applications[1]

1. Generalities

The key idea of *relaxation methods* is to reduce, using some iterative process, the solution of some problems posed in a product space $V = \prod_{i=1}^{N} V_i$ (minimization of functionals, solution of systems of equations and/or inequalities, etc.) to the solution of a sequence of subproblems of the same kind, but simpler, since they are posed in the V_i.

A typical example of such methods is given by the classical *point* or *block Gauss–Seidel methods* and their variants (S.O.R., S.S.O.R., etc.). For the solution of finite-dimensional linear systems by methods of this type, we refer to Varga [1], Forsythe and Wasow [1], D. Young [1] and the bibliographies therein. For the solution of systems of nonlinear equations, we refer to Ortega and Rheinboldt [1], Miellou [1], [2], and the bibliographies therein.

For the minimization of convex functionals by methods of this kind, let us mention S. Schecter [1], [2], [3], Cea [1], [2], A. Auslender [1], Cryer [1], [2], Cea and Glowinski [1], Glowinski [6], and the bibliographies therein. The above list is far from complete.

The basic knowledge of convex analysis required for a good understanding of this chapter may be found in Cea [1], Rockafellar [1], and Ekeland and Temam [1].

2. Some Basic Results of Convex Analysis

In this section we shall give, without proof, some classical results on the existence, uniqueness, and characterization of the solution of convex minimization problems. Let

(i) V be a real reflexive Banach space, V^* its dual space,
(ii) K be a nonempty closed convex subset of V,

[1] In this chapter we follow Cea and Glowinski [1] and Glowinski [6].

(iii) $J_0: V \to \mathbb{R}$, be a convex functional Frechet or Gateaux differentiable[2] on V,

(iv) $J_1: V \to \overline{\mathbb{R}}$, be[3] a proper l.s.c. convex functional.

We assume that $K \cap \mathrm{Dom}(J_1) \neq \varnothing$, where

$$\mathrm{Dom}(J_1) = \{v \mid v \in V, J_1(v) \in \mathbb{R}\}.$$

We define $J: V \to \overline{\mathbb{R}}$ by $J = J_0 + J_1$ and assume that

$$\lim_{\substack{\|v\| \to +\infty \\ v \in K}} J(v) = +\infty. \tag{2.1}$$

Under the above assumptions on K and J, we have the following fundamental theorem.

Theorem 2.1. *The minimization problem*

$$J(u) \leq J(v), \qquad \forall\, v \in K, \quad u \in K, \tag{2.2}$$

has a solution characterized by

$$\langle J_0'(u), v - u\rangle + J_1(v) - J_1(u) \geq 0, \qquad \forall\, v \in K, \quad u \in K. \tag{2.3}$$

This solution is unique if J is strictly convex.[4]

Remark 2.1. If K is bounded, then (2.1) may be omitted.

Remark 2.2. Problem (2.3) is a *variational inequality* (see Chapter I of this book).

Let us now recall some definitions about monotone operators.

Definition 2.1. Let $A: V \to V^*$. The operator A is said to be *monotone* if

$$\langle A(v) - A(u), v - u\rangle \geq 0, \qquad \forall\, u, v \in V,$$

and *strictly monotone* if it is monotone and

$$\langle A(v) - A(u), v - u\rangle > 0, \qquad \forall\, u, v \in V, \quad u \neq v.$$

---

[2] Let $F: V \to \mathbb{R}$; the Gateaux-differentiability property means that

$$\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{F(v + tw) - F(v)}{t} = \langle F'(v), w\rangle, \qquad \forall\, v, w \in V,$$

where $\langle \cdot, \cdot \rangle$ denotes the duality between V^* and V and $F'(v) \in V^*$; $F'(v)$ is said to be the Gateaux-derivative (or G-derivative) of F at v. Actually, we shall very often use the term gradient when referring to F'.

[3] $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

[4] i.e., $J(tv + (1-t)w) < tJ(v) + (1-t)J(w)$, $\forall\, t \in]0, 1[$, $\forall\, v, w \in \mathrm{Dom}(J)$, $v \neq w$.

We shall introduce the following proposition which will be very useful in the sequel of this chapter.

Proposition 2.1. *Let $F: V \to \mathbb{R}$ be G-differentiable. Then there is equivalence between the convexity of F (resp., the strict convexity of F) and the monotonicity (resp., the strict monotonicity) of F'.*

To prove Theorem 2.1 and Proposition 2.1, we should use the following:

Proposition 2.2. *If F is G-differentiable, then F is convex if and only if*

$$F(w) - F(v) \geq \langle F'(v), w - v \rangle, \qquad \forall\, v, w \in V. \tag{2.4}$$

3. Relaxation Methods for Convex Functionals: Finite-Dimensional Case

3.1. Statement of the minimization problem. Notations

With respect to Sec. 2, we assume that $V = \mathbb{R}^N$, with

$$v = \{v_1, \ldots, v_N\}, \qquad v_i \in \mathbb{R}, \quad 1 \leq i \leq N.$$

The following notation will be used in the sequel:

$$(u, v) = \sum_{i=1}^{N} u_i v_i, \qquad \|v\| = \sqrt{(v, v)}.$$

We also assume that

$$K = \{v \mid v \in \mathbb{R}^N, v_i \in K_i = [a_i, b_i], a_i \leq b_i, 1 \leq i \leq N\}, \tag{3.1}$$

where the a_i (resp., the b_i) may take the value $-\infty$ (resp., $+\infty$); K is obviously a nonempty closed convex subset of $\mathbb{R}^N$. Furthermore, we assume that

$$J(v) = J_0(v) + \sum_{i=1}^{N} j_i(v_i), \tag{3.2}$$

where $J_0 \in C^1(\mathbb{R}^N)$ and is strictly convex and where the $j_i \in C^0(\mathbb{R})$ and are convex, $1 \leq i \leq N$. We note that we are not assuming the differentiability of the j_i. Finally, we assume that

$$\lim_{\|v\| \to +\infty} J(v) = +\infty. \tag{3.3}$$

Then, under the above assumptions and from Theorem 2.1, it follows that the problem

$$J(u) \leq J(v) \qquad \forall\, v \in K, \quad u \in K, \tag{3.4}$$

has a unique solution characterized by

$$(J_0'(u), v - u) + \sum_{i=1}^{N} (j_i(v_i) - j_i(u_i)) \geq 0, \qquad \forall\, v \in K, \quad u \in K, \tag{3.5}$$

where $J_0'(v)$ denotes the gradient of J_0 at v.

3.2. Description of the relaxation algorithm

To solve (3.4) we can use the following relaxation algorithm:

$$u^0 \in K \text{ arbitrarily given}; \tag{3.6}$$

then, u^n being known, we compute u^{n+1}, component by component, by

$$\begin{aligned} &J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^{n+1}, u_{i+1}^n, \ldots) \\ &\quad \leq J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^n, \ldots), \qquad \forall\, v_i \in K_i, \quad u_i^{n+1} \in K_i, \end{aligned} \tag{3.7}$$

where $1 \leq i \leq N$.

From the above assumptions on K and J, and from Theorem 2.1, we obtain:

Proposition 3.1. *Each subproblem* (3.7) *has a unique solution which is characterized by*

$$\begin{aligned} &\frac{\partial J_0}{\partial v_i}(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots)(v_i - u_i^{n+1}) + j_i(v_i) - j_i(u_i^{n+1}) \geq 0, \\ &\qquad \forall\, v \in K_i, \quad u_i^{n+1} \in K_i. \end{aligned} \tag{3.8}$$

Remark 3.1. To compute u_i^{n+1}, we can proceed as follows. First we compute $\bar{u}_i^{n+1}$ by solving

$$\begin{aligned} &J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, \bar{u}_i^{n+1}, u_{i+1}^n, \ldots) \leq J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^n, \ldots), \\ &\qquad \forall\, v_i \in \mathbb{R}, \quad \bar{u}_i^{n+1} \in \mathbb{R}. \end{aligned} \tag{3.9}$$

Then we project $\bar{u}_i^{n+1}$ on K_i to obtain u_i^{n+1}; so

$$u_i^{n+1} = P_{K_i}(\bar{u}_i^{n+1}) = \operatorname{Max}(a_i, \operatorname{Min}(b_i, \bar{u}_i^{n+1})). \tag{3.10}$$

If j_i is differentiable, then $\bar{u}_i^{n+1}$ is the solution of the single-variable equation

$$\frac{\partial J_0}{\partial v_i}(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, \bar{u}_i^{n+1}, u_{i+1}^n, \ldots) + \frac{dj_i}{dv_i}(\bar{u}_i^{n+1}) = 0. \tag{3.11}$$

Equation (3.11) can be solved by various methods (see, for example, Householder [1]).

3.3. A lemma on the monotonicity of the gradient of strictly convex functionals

To prove the convergence of algorithm (3.6), (3.7), we shall use the following:

Lemma 3.1. *Assume that $F \in C^1(\mathbb{R}^N)$ and is strictly convex. Then F is uniformly convex on the bounded sets of $\mathbb{R}^N$, i.e., $\forall\, M > 0$, $\exists\, \delta_M : [0, 2M] \to \mathbb{R}_+$, continuous, strictly increasing, and such that*

$$\delta_M(0) = 0, \tag{3.12}$$

$$(F'(v) - F'(u), v - u) \geq \delta_M(\|v - u\|)$$
$$\forall\, u, v \in \mathbb{R}^N, \quad \|u\| \leq M, \quad \|v\| \leq M, \tag{3.13}$$

and

$$F(v) \geq F(u) + (F'(u), v - u) + \tfrac{1}{2}\delta_M(\|v - u\|),$$
$$\forall\, u, v \in \mathbb{R}^N, \quad \|u\| \leq M, \quad \|v\| \leq M. \tag{3.14}$$

PROOF. Let $B_M = \{v \mid v \in \mathbb{R}^N, \|v\| \leq M\}$. For $\tau \in [0, 2M]$, we define δ_M^0 by

$$\delta_M^0(\tau) = \inf_{\substack{\|v-u\| = \tau \\ u, v \in B_M}} (F'(v) - F'(u), v - u). \tag{3.15}$$

From the definition of δ_M^0 it follows that

$$\delta_M^0(0) = 0 \tag{3.16}$$

and

$$(F'(v) - F'(u), v - u) \geq \delta_M^0(\|v - u\|), \qquad \forall\, u, v \in B_M. \tag{3.17}$$

Let $\tau_2 \in]0, 2M]$. From the continuity of $\{u, v\} \to (F'(v) - F'(u), v - u)$ and from the compactness of $B_M \times B_M$, it follows that there exists at least one pair $\{u_2, v_2\}$ realizing the minimum in (3.15). Then

$$\delta_M^0(\tau_2) = (F'(v_2) - F'(u_2), v_2 - u_2),$$

and

$$\delta_M^0(\tau_2) > 0$$

from the strict monotonicity of F' (cf. Sec. 2, Proposition 2.1). Let $\tau_1 \in]0, \tau_2[$. We define $w \in]u_2, v_2[$ by

$$w = u_2 + \frac{\tau_1}{\tau_2}(v_2 - u_2).$$

Since $0 < \tau_1/\tau_2 < 1$, from the strict monotonicity of F' it follows that

$$(F'(v_2) - F'(u_2), v_2 - u_2) > \left(F'\left(u_2 + \frac{\tau_1}{\tau_2}(v_2 - u_2)\right) - F'(u_2), v_2 - u_2\right).$$

This implies

$$(F'(v_2) - F'(u_2), v_2 - u_2) > \frac{\tau_2}{\tau_1}(F'(w) - F'(u_2), w - u_2) > (F'(w) - F'(u_2), w - u_2). \tag{3.18}$$

Since $\|w - u_2\| = \tau_1$, (3.18) in turn implies

$$\delta_M^0(\tau_2) > \delta_M^0(\tau_1).$$

Applying (3.17) to $\{u + t(v - u), u\}$, it follows that

$$(F'(u + t(v - u)), v - u) \geq (F'(u), v - u) + \frac{1}{t}\delta_M^0(t\|v - u\|), \qquad \forall\, t \in]0, 1]. \tag{3.19}$$

From the continuity of F', it easily follows that

$$\lim_{\tau \to 0+} \frac{1}{\tau}\delta_M^0(\tau) = 0. \tag{3.20}$$

Then from (3.20) it follows that (3.19) could be extended at $t = 0$. Integrating (3.19) on [0, 1], it follows that

$$F(v) - F(u) \geq (F'(u), v - u) + \int_0^1 \delta_M^0(t\|v - u\|)\frac{dt}{t}. \tag{3.21}$$

We also have

$$F(u) - F(v) \geq (F'(v), u - v) + \int_0^1 \delta_M^0(t\|v - u\|)\frac{dt}{t}. \tag{3.22}$$

Then, by summation of (3.21), (3.22), we obtain

$$\begin{aligned}(F'(v) - F'(u), v - u) &\geq 2\int_0^1 \delta_M^0(t\|v - u\|)\frac{dt}{t} \\ &= 2\int_0^{\|v-u\|} \delta_M^0(s)\frac{ds}{s}.\end{aligned} \tag{3.23}$$

Therefore the function δ_M defined by

$$\delta_M(\tau) = 2\int_0^\tau \delta_M^0(s)\frac{ds}{s} \tag{3.24}$$

has the required properties. Furthermore, (3.14) follows from (3.21) and from the definition of δ_M. □

Remark 3.2. The term *forcing function* is frequently used for functions such as δ_M (see Ortega and Rheinboldt [1]).

3.4. Convergence of algorithm (3.6), (3.7)

We have:

Theorem 3.1. *Under the above assumptions on K and J, the sequence $(u^n)_n$ defined* (3.6), (3.7) *converges, $\forall\, u^0 \in K$, to the solution u of* (3.4).

PROOF. For the sake of simplicity, we have split the proof into several steps.

Step 1. We shall prove that the sequence $J(u^n)$ is decreasing. We have

$$J(u^n) - J(u^{n+1}) = \sum_{i=1}^{N} (J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots) - J(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots)). \quad (3.25)$$

Since $u_i^n \in K_i$, it follows from (2.4), (3.8) that, $\forall\, i = 1, \ldots, N$, we have

$$\begin{aligned} &J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots) - J(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots) \\ &= J_0(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots) - J_0(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots) + j_i(u_i^n) - j_i(u_i^{n+1}) \quad (3.26) \\ &\geq \frac{\partial J_0}{\partial v_i}(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots)(u_i^n - u_i^{n+1}) + j_i(u_i^n) - j_i(u_i^{n+1}) \geq 0. \end{aligned}$$

Then (3.26) combined with (3.25) implies

$$J(u^n) \geq J(u^{n+1}), \qquad \forall\, n \geq 0. \quad (3.27)$$

Moreover, since J satisfies (3.3), it follows from (3.27) that there exists a constant M such that

$$\begin{aligned} &\|u\| \leq M, \|u^n\| \leq M, \qquad \forall\, n, \\ &\|\{u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots\}\| \leq M, \qquad \forall\, i = 1, \ldots, N, \quad \forall\, n. \end{aligned} \quad (3.28)$$

Step 2. From (3.8), (3.14), (3.25), (3.26), and (3.28), it follows that

$$J(u^n) - J(u^{n+1}) \geq \frac{1}{2} \sum_{i=1}^{N} \delta_M(|u_i^{n+1} - u_i^n|). \quad (3.29)$$

The sequence $J(u^n)$ is decreasing and bounded below by $J(u)$, where u is the solution of (3.4). Therefore the sequence $J(u^n)$ is convergent, and this implies

$$\lim_{n \to +\infty} (J(u^n) - J(u^{n+1})) = 0. \quad (3.30)$$

From (3.29), (3.30), and the properties of δ_M, it follows that

$$\lim_{n \to +\infty} (u^n - u^{n+1}) = 0. \quad (3.31)$$

Step 3. Let u be the solution of (3.4). Then it follows from (3.15), (3.28) that

$$(J_0'(u^{n+1}) - J_0'(u), u^{n+1} - u) \geq \delta_M(\|u^{n+1} - u\|),$$

which implies

$$(J_0'(u^{n+1}) - J_0'(u), u^{n+1} - u) + J_1(u^{n+1}) - J_1(u) \geq J_1(u^{n+1}) - J_1(u) + \delta_M(\|u^{n+1} - u\|). \quad (3.32)$$

Since u is the solution of (3.4) and $u^{n+1} \in K$, we have (cf. (3.5))

$$(J_0'(u), u^{n+1} - u) + J_1(u^{n+1}) - J_1(u) \geq 0,$$

which, combined with (3.32), implies

$$(J_0'(u^{n+1}), u^{n+1} - u) + J_1(u^{n+1}) - J_1(u) \geq \delta_M(\|u^{n+1} - u\|). \tag{3.33}$$

Relation (3.33) implies

$$\begin{aligned}
&(J_0'(u^{n+1}), u^{n+1} - u) + J_1(u^{n+1}) - J_1(u) \\
&= \sum_{i=1}^{N} \left(\frac{\partial J_0}{\partial v_i}(u^{n+1}) - \frac{\partial J_0}{\partial v_i}(\hat{u}_i^{n+1}) \right)(u_i^{n+1} - u_i) \\
&\quad + \sum_{i=1}^{N} \left(\frac{\partial J_0}{\partial v_i}(\hat{u}_i^{n+1})(u_i^{n+1} - u_i) + j_i(u_i^{n+1}) - j_i(u_i) \right) \\
&\geq \delta_M(\|u^{n+1} - u\|),
\end{aligned} \tag{3.34}$$

where $\hat{u}_i^{n+1} = \{u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots\}$. Since $u_i \in K_i$, it follows from (3.8) that, $\forall\, i = 1, \ldots, N$,

$$\frac{\partial J_0}{\partial v_i}(\hat{u}_i^{n+1})(u_i^{n+1} - u_i) + j_i(u_i^{n+1}) - j_i(u_i) \leq 0. \tag{3.35}$$

Therefore (3.34) and (3.35) show that

$$\sum_{i=1}^{N} \left(\frac{\partial J_0}{\partial v_i}(u^{n+1}) - \frac{\partial J_0}{\partial v_i}(\hat{u}_i^{n+1}) \right)(u_i^{n+1} - u_i) \geq \delta_M(\|u^{n+1} - u\|). \tag{3.36}$$

Since $\|u^{n+1} - \hat{u}_i^{n+1}\| \leq \|u^{n+1} - u^n\|$, it follows from (3.31) that $\forall\, i = 1, \ldots, N$, we have

$$\lim_{n \to +\infty} (u^{n+1} - \hat{u}_i^{n+1}) = 0. \tag{3.37}$$

Since $J_0' \in C^0(\mathbb{R}^N)$, J_0' is uniformly continuous on the bounded subsets of $\mathbb{R}^N$. This property, combined with (3.37), implies, $\forall\, i = 1, \ldots, N$,

$$\lim_{n \to +\infty} \left\| \frac{\partial J_0}{\partial v_i}(u^{n+1}) - \frac{\partial J_0}{\partial v_i}(\hat{u}_i^{n+1}) \right\| = 0. \tag{3.38}$$

Therefore, from (3.28), (3.36), (3.38), and the properties of δ_M, it follows that

$$\lim_{n \to \infty} \|u^n - u\| = 0,$$

which completes the proof of the theorem. □

3.5. Various remarks

Remark 3.3. We assume that $K = \mathbb{R}^N$ and that $J \equiv J_0$ (i.e., $J_1 \equiv 0$), where

$J_0(v) = \frac{1}{2}(Av, v) - (b, v)$, where $b \in \mathbb{R}^N$ and

A is an $N \times N$ symmetrical positive-definite matrix.

The problem (3.4) associated with this choice of J and K obviously has an unique solution characterized (cf. (3.5)) by

$$Au = b. \tag{3.39}$$

If we apply the algorithm (3.6), (3.7) to this particular case, we obtain

$$u^0 \in \mathbb{R}^N, \text{ arbitrarily given;} \tag{3.40}$$

$$u_i^{n+1} = \frac{1}{a_{ii}}\left(b_i - \sum_{j<i} a_{ij}u_j^{n+1} - \sum_{j>i} a_{ij}u_j^n\right), \qquad 1 \leq i \leq N. \tag{3.41}$$

The algorithm (3.40), (3.41) is known as the *Gauss–Seidel method* for solving (3.39) (see, e.g., Varga [1] and D. Young [1]). Therefore, when A is symmetric and positive definite, optimization theory yields another proof of the convergence of the Gauss–Seidel method through Theorem 3.1.

Remark 3.4. *From the above remark it follows that the introduction of over-* or under-relaxation parameters could be effective for increasing the speed of convergence. This possibility will be discussed in the sequel of this chapter.

Let $F: V \to \overline{\mathbb{R}}$. We define

$$D(F) = \{v \mid v \in V, |F(v)| < +\infty\}. \tag{3.42}$$

If F is convex and proper, then $D(F)$ is a nonempty convex subset of V.

Remark 3.5. If in Sec. 3.1 we replace the conditions $j_i \in C^0(\mathbb{R})$ and j_i convex, $\forall\, i = 1, \ldots, N$, by

$$j_i: \mathbb{R} \to \overline{\mathbb{R}} \text{ is convex, proper, and l.s.c.}$$

and we assume $K_i \cap \mathrm{D}(j_i) \neq \varnothing$, $\forall\, i = 1, \ldots, N$, then the other assumptions being the same, (3.4) is still a well-posed problem and (3.5) still holds. Moreover, the algorithm (3.6), (3.7) could be used to solve (3.4), and the convergence result given by Theorem 3.1 would still hold.

Remark 3.6. We can complete Remark 3.5 in the following way. We take j_i as in Remark 3.5 and assume

$$J_0 \text{ strictly convex, proper, and l.s.c.,}$$

$D(J_0)$ is an open set of $\mathbb{R}^N$ and $J_0 \in C^1(D(J_0))$. Then, if $D(J) \cap K \neq \varnothing$ and if $\lim_{\|v\| \to +\infty} J(v) = +\infty$, problem (3.4) is well posed and (3.5) still holds. Moreover, algorithm (3.6), (3.7) could be used to solve (3.4).

Remark 3.7. A typical situation in which algorithm (3.6), (3.7) could be used is $K = \mathbb{R}^N$, J_0 as in Remark 3.3 and $J_1(v) = \sum_{i=1}^N \alpha_i |v_i|$, $\alpha_i \geq 0$, $\forall\, i = 1, \ldots, N$.

3.6. Some dangerous generalizations

In this section we would like to discuss some of the limitations of the relaxation methods.

3.6.1. *Relaxation and nondifferentiable functionals*

We consider $K = \mathbb{R}^2$ and $J \in C^0(\mathbb{R}^2)$, strictly convex, defined by

$$J(v) = \tfrac{1}{2}(v_1^2 + v_2^2) + |v_1 - v_2| - (v_1 + v_2);$$

J is nondifferentiable on the line $v_1 = v_2$. The unique solution of

$$J(u) = \operatorname*{Min}_{v \in \mathbb{R}^2} J(v) \tag{3.43}$$

is obviously $u = \{1, 1\}$.

Now, starting from $u^0 = \{0, 0\}$, let us apply (3.6), (3.7) to (3.43). Since u_1^1 is the solution of

$$\operatorname*{Min}_{v_1 \in \mathbb{R}} (\tfrac{1}{2}v_1^2 + |v_1| - v_1).$$

we have $u_1^1 = 0$. In a similar way we have $u_2^1 = 0$, so that

$$u^n = \{0, 0\}, \qquad \forall\, n.$$

From the above result it follows that the algorithm (3.6), (3.7) *does not converge to u.*

3.6.2. *Relaxation and nonfactorable convex sets*

In this section we assume that K is a closed convex subset of $\mathbb{R}^N$, such that

$$K \neq \prod_{i=1}^{N} K_i.$$

If all the other assumptions of Sec. 3.1 hold, we can generalize algorithm (3.6), (3.7) as follows:

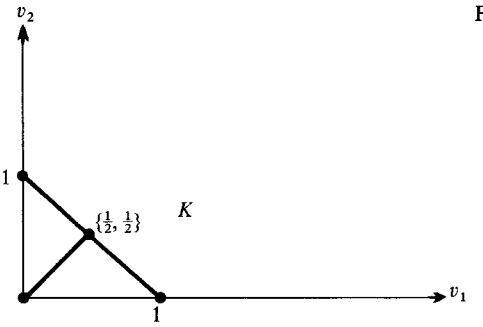
$$u^0 \in K, \tag{3.44}_1$$

$$\begin{aligned} &J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^{n+1}, u_{i+1}^n, \ldots) \leq J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^n, \ldots), \\ &\quad \forall \{u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^n, \ldots\} \in K, \\ &\quad \{u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots\} \in K; \qquad i = 1, \ldots, N. \end{aligned} \tag{3.44}_2$$

Since $u^0 \in K$, it is very easy to prove that (3.44) is a well-posed problem, $\forall\, n \geq 0$. A very simple example will show that $(3.44)_1, (3.44)_2$ generally does not converge if K is nonfactorable. We take $J = J_0$ defined by $J(v) = v_1^2 + v_2^2$ and K defined (see Fig. 3.1) by

$$K = \{v \mid v \in \mathbb{R}^2, v_1 \geq 0, v_2 \geq 0, v_1 + v_2 \geq 1\}.$$

Figure 3.1



The solution of

$$J(u) = \operatorname*{Min}_{v \in K} J(v) \tag{3.45}$$

is obviously $u = \{\frac{1}{2}, \frac{1}{2}\}$. Then, starting from $u^0 = \{0, 1\}$, let us apply the algorithm $(3.44)_1, (3.44)_2$ to (3.45). It is easy to see that $u^n = \{0, 1\} \neq u, \forall\, n \geq 0$. Therefore we do not have convergence to the solution.

Remark 3.8. Let's consider the "one-dimensional" torsion problem (3.46) of Chapter II, Sec. 3.7.1. In the sequel we assume that $f \in C^0[0, 1]$. Using the notation $v_h(x_i) = v_i$, the above problem is approximated by the following variant of (3.41) (Chapter II, Sec. 3.7.1):

$$\int_0^1 \frac{du_h}{dx}\left(\frac{dv_h}{dx} - \frac{du_h}{dx}\right) dx \geq h \sum_{i=1}^{N-1} f_i(v_i - u_i), \qquad \forall\, v_h \in K_h, \quad u_h \in K_h, \tag{3.46}$$

where $f_i = f(x_i)$, $i = 1, \ldots, N$.

Problem (3.46) is equivalent to the minimization problem

$$J_h(\hat{u}_h) \leq J_h(\hat{v}_h), \qquad \forall\, \hat{v}_h \in \hat{K}_h, \quad \hat{u}_h \in \hat{K}_h, \tag{3.47}$$

where

$$\hat{v}_h = \{v_0, \ldots, v_N\}, \tag{3.48}$$

$$J_h(\hat{v}_h) = \frac{h}{2} \sum_{i=0}^{N-1} \left| \frac{v_{i+1} - v_i}{h} \right|^2 - h \sum_{i=1}^{N-1} f_i v_i, \tag{3.49}$$

$$\hat{K}_h = \left\{ \hat{v}_h \,|\, \hat{v}_h \in \mathbb{R}^{N+1},\ v_0 = v_N = 0,\ \left| \frac{v_{i+1} - v_i}{h} \right| \leq 1, \forall\, i = 0, \ldots, N-1 \right\}; \tag{3.50}$$

$\hat{K}_h$ is a nonfactorable closed convex subset of $\mathbb{R}^{N+1}$. However, it is proved in G.L.T. [1, Chapter 3], [3, Chapter 3] that $(3.44)_1$, $(3.44)_2$ converges to the solution $\hat{u}_h$ of (3.47) if $f \geq 0$ on $[0, 1]$ and if we start with

$$\hat{u}_h^0 \leq \hat{u}_h \quad (\text{i.e., } u_i^0 \leq u_i, \forall\, i = 1, \ldots, N-1).$$

Since $f \geq 0$ implies $u_i \geq 0$, $\forall\, i = 1, \ldots, N-1$, an obvious choice for $\hat{u}_h^0$ is $\hat{u}_h^0 = 0$.

4. Block Relaxation Methods

In this section we take $V = \mathbb{R}^N$ such that

$$V = \prod_{i=1}^{q} V_i \quad \text{with } V_i = \mathbb{R}^{N_i}, \tag{4.1}$$

where

$$N_i \geq 1 \quad \text{and} \quad \sum_{i=1}^{q} N_i = N. \tag{4.2}$$

If $v \in V$, then $v = \{v_1, \ldots, v_q\}$, $v_i \in V_i$. We assume that $K = \prod_{i=1}^{q} K_i$, where

$$K_i \text{ is a closed convex subset of } V_i. \tag{4.3}$$

Then (4.3) implies that K is a closed convex subset of V. Finally we consider a convex functional $J: V \to \overline{\mathbb{R}}$ such that

$$D(J) \cap K \neq \varnothing, \qquad \lim_{\|v\| \to +\infty} J(v) = +\infty, \quad J = J_0 + J_1, \tag{4.4}$$

where

$$J_0 \in C^1(V), \quad J_0 \text{ strictly convex} \tag{4.5}$$

and

$$J_1(v) = \sum_{i=1}^{q} j_i(v_i), \tag{4.6}$$

where the j_i are convex, proper, and l.s.c. on V_i.

From Theorem 2.1 of Sec. 2 it follows that

$$J(u) \leq J(v), \qquad \forall\, v \in K, \quad u \in K, \tag{4.7}$$

has a unique solution (in $D(J) \cap K$) characterized by

$$(J_0'(u), v - u) + \sum_{i=1}^{q} (j_i(v_i) - j_i(u_i)) \geq 0, \qquad \forall\, v \in K, \quad u \in K. \tag{4.8}$$

The generalization of the point relaxation algorithm (3.6), (3.7) is obviously

$$u^0 \in K, \tag{4.9}$$

$$J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^{n+1}, u_{i+1}^{n}, \ldots) \leq J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^{n}, \ldots),$$
$$\forall\, v_i \in K_i, \quad u_i^{n+1} \in K_i; \quad i = 1, \ldots, q. \tag{4.10}$$

Algorithm (4.9), (4.10) is a block-relaxation algorithm. Using a variant of the proof of Theorem 3.1, it may easily be proved that the sequence u^n defined by (4.9), (4.10) converges to the unique solution of (4.2).

Remark 4.1. Most of the remarks we have made on algorithm (3.6), (3.7) still apply to (4.9), (4.10). Unfortunately, Remark 3.1 does not apply in general.

Remark 4.2. In Cea and Glowinski [1] and Glowinski [6], it is proved that (4.9), (4.10) could also be used in infinite-dimensional situations.[5] In this case the assumptions $J_0 \in C^1$ and strictly convex are not sufficient to insure the convergence of the relaxation algorithm. The basic reason for this is that closed bounded sets are generally noncompact in Banach spaces of infinite dimension. Using the same notation, in the two references cited above, it is proved that sufficient convergence conditions are

$$J_0 \text{ is } C^1 \text{ with } J_0' \text{ Lipschitz continuous on the bounded sets of } V; \tag{4.11}$$

$$J_0' \text{ locally uniformly convex, i.e.,}$$

$$\langle J_0'(v) - J_0'(u), v - u \rangle \geq \delta_M(\|v - u\|), \qquad \forall\, u, v, \quad \|u\| \leq M, \quad \|v\| \leq M, \tag{4.12}$$

where δ_M is the same as in Sec. 3.3.

5. Constrained Minimization of Quadratic Functionals in Hilbert Spaces by Under and Over-Relaxation Methods: Application

We follow Cea and Glowinski [1, Sec. 3].

5.1. Statement of the minimization problem

Let $V = \prod_{i=1}^{N} V_i$, where V_i is a real Hilbert space, $\forall\, i = 1, \ldots, N$. Norm and scalar product on V_i are, respectively, denoted by $\|\cdot\|_i$ and $((\cdot, \cdot))_i$. If $v \in V$, then

$$v = \{v_1, \ldots, v_N\} \quad \text{with } v_i \in V_i, \quad i = 1, \ldots, N.$$

On V we define a scalar product and a norm by

$$((u, v)) = \sum_{i=1}^{N} ((u_i, v_i))_i, \tag{5.1}$$

$$\|v\| = \left(\sum_{i=1}^{N} \|v_i\|_i^2 \right)^{1/2}; \tag{5.2}$$

[5] More precisely, in reflexive Banach spaces.

V is a Hilbert space for $((\cdot, \cdot))$ and $\|\cdot\|$. Let K be a closed convex subset of V such that

$$K = \prod_{i=1}^{N} K_i, \; K_i \neq \varnothing, \qquad \forall\, i = 1, \ldots, N,$$

where K_i is closed and convex in V_i.

Let $J: V \to \mathbb{R}$ be defined by

$$J(v) = \tfrac{1}{2} a(v, v) - ((f, v)), \tag{5.3}$$

where the bilinear form $a(\cdot, \cdot)$ on $V \times V$, is continuous, symmetrical, and V-elliptic, i.e.,

$$\exists\, \alpha > 0 \text{ such that } a(v, v) \geq \alpha \|v\|^2, \qquad \forall\, v \in V,$$

and where $f \in V$.

Under the assumptions on V, K, and J, the optimization problem

$$J(u) \leq J(v), \qquad \forall\, v \in K, \quad u \in K \tag{5.4}$$

has a unique solution. This solution is characterized by

$$a(u, v - u) - ((f, v - u)) \geq 0, \qquad \forall\, v \in K, \quad u \in K. \tag{5.5}$$

5.2. Some preliminary results

From the properties of V it follows that

$$J(v) = J(v_1, \ldots, v_N) = \frac{1}{2} \sum_{1 \leq i, j \leq N} a_{ij}(v_i, v_j) - \sum_{i=1}^{N} ((f_i, v_i))_i, \tag{5.6}$$

where the a_{ij} are bilinear and continuous on $V_i \times V_j$ with $a_{ij} = a_{ji}^*$. The forms a_{ii} are V_i-elliptic (with the same constant α). Using the Riesz representation theorem, it is easily proved that there exists $A_{ij} \in \mathscr{L}(V_j, V_i)$ such that

$$a_{ij}(v_i, v_j) = ((v_i, A_{ij} v_j))_i, \tag{5.7}$$

$$A_{ij} = A_{ji}^*. \tag{5.8}$$

Moreover, the A_{ii} are self-adjoint and are isomorphisms from V_i to V_i. In the sequel it will be convenient to use the norm defined by a_{ii} on V_i, i.e.,

$$|||v_i|||_i^2 = a_{ii}(v_i, v_i) = ((A_{ii} v_i, v_i))_i, \qquad i = 1, \ldots, N. \tag{5.9}$$

The norms $\|\cdot\|_i$ and $|||\cdot|||_i$ are equivalent. The projection from V_i to K_i in the $|||\cdot|||_i$ norm will be denoted by P_i. Before giving the description of the iterative method, we shall prove some basic results on projections, useful in the sequel. Let:

(i) H be a real Hilbert space, with scalar product and norm denoted by $(\cdot, \cdot)$ and $\|\cdot\|$, respectively.

(ii) $b(\cdot, \cdot)$ be a bilinear form on H, continuous, symmetric, and H-elliptic (i.e., $\exists\, \beta > 0$ such that $b(v, v) \geq \beta \|v\|^2$, $\forall\, v \in H$).

Then from the Riesz representation theorem follows the existence of an isomorphism $B: H \to H$ such that

$$(Bu, v) = b(u, v), \qquad \forall\, u, v \in H,$$

$$B = B^*. \tag{5.10}$$

We denote by $[\cdot, \cdot]$ and $|\cdot|$ the scalar product on H and the norm on H, respectively, defined by

$$[u, v] = b(u, v), \qquad \forall\, u, v \in H, \tag{5.11}$$

$$|v|^2 = b(v, v), \qquad \forall\, v \in H. \tag{5.12}$$

The norms $|\cdot|$ and $\|\cdot\|$ are equivalent. Let

(iii) $C \neq \varnothing$ be a closed convex subset of H and π be the projector from $H \to C$ in the $|\cdot|$ norm.

(iv) $j: H \to \mathbb{R}$ be the functional defined by

$$j(v) = \tfrac{1}{2}b(v, v) - (g, v), \qquad \forall\, v \in H, \tag{5.13}$$

where $g \in H$.

Under the above assumptions, we have the following lemmas.

Lemma 5.1. *If u is the unique solution of*

$$j(u) \leq j(v), \qquad \forall\, v \in C, \quad u \in C, \tag{5.14}$$

then

$$u = \pi(B^{-1}g). \tag{5.15}$$

PROOF. The solution u of (5.14) is characterized by

$$(Bu - g, v - u) \geq 0, \qquad \forall\, v \in C, \quad u \in C. \tag{5.16}$$

From (5.16) it follows that

$$(B(v - u), B^{-1}g - u) = b(v - u, B^{-1}g - u) \leq 0, \qquad \forall\, v \in C, \quad u \in C, \tag{5.17}$$

and (5.17) characterizes u as the projection on C of $B^{-1}g$ in the norm $|\cdot|$. □

Lemma 5.2. *Let $u_0 \in C$ and let u_1 be defined by*

$$u_1 = \pi(u_0 + \omega(B^{-1}g - u_0)), \qquad \omega > 0. \tag{5.18}$$

Then

$$j(u_0) - j(u_1) \geq \frac{2 - \omega}{2\omega} |u_0 - u_1|^2. \tag{5.19}$$

PROOF. We have, $\forall\, v_1, v_2 \in H$,

$$j(v_1) - j(v_2) = \tfrac{1}{2}(|v_1 - B^{-1}g|^2 - |v_2 - B^{-1}g|^2). \tag{5.20}$$

Since $u_1 - B^{-1}g = u_0 - B^{-1}g + u_1 - u_0$, we have

$$|u_0 - B^{-1}g|^2 = |u_1 - B^{-1}g|^2 + 2[u_0 - B^{-1}g, u_0 - u_1] - |u_0 - u_1|^2. \tag{5.21}$$

From (5.18), and since $u_0 \in C$, it follows that

$$[u_0 + \omega(B^{-1}g - u_0) - u_1, u_0 - u_1] \leq 0.$$

This implies

$$|u_0 - u_1|^2 \leq \omega[u_0 - B^{-1}g, u_0 - u_1]. \tag{5.22}$$

Then (5.19) clearly follows from (5.20)–(5.22).

5.3. Description of the algorithm

For $i = 1, \ldots, N$, let the ω_i be positive numbers. Now let us consider the following algorithm:

$$u^0 = \{u_1^0, \ldots, u_N^0\} \textit{ arbitrarily given in } K; \tag{5.23}$$

u^n being known, we compute u^{n+1} by

$$u_i^{n+1} = P_i\left(u_i^n - \omega_i A_{ii}^{-1}\left(\sum_{j<i} A_{ij}u_j^{n+1} + \sum_{j\geq i} A_{ij}u_j^n - f_i\right)\right), \qquad i = 1, \ldots, N. \tag{5.24}$$

Remark 5.1. It follows from (5.24) that the computation of u_i^{n+1} can be achieved in three steps:

Step 1. On V_i we minimize the functional

$$v_i \to J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^n, \ldots, u_N^n).$$

Hence we obtain a solution

$$u_i^{n+1/3} = A_{ii}^{-1}\left(f_i - \sum_{j<i} A_{ij}u_j^{n+1} - \sum_{j>i} A_{ij}u_j^n\right).$$

Step 2. We compute $u_i^{n+2/3}$ by

$$u_i^{n+2/3} = u_i^n + \omega_i(u_i^{n+1/3} - u_i^n).$$

Step 3. At last we obtain u_i^{n+1} by

$$u_i^{n+1} = P_i(u_i^{n+2/3}).$$

We remark that if $\omega_i = 1, \forall\, i = 1, \ldots, N$, then algorithm (5.23), (5.24) reduces to the block algorithm (4.9), (4.10) (see Remark 4.2).

5.4. Convergence of algorithm (5.23), (5.24)

Proposition 5.1. *We have*

$$J(u^n) - J(u^{n+1}) \geq \sum_{i=1}^{N} \frac{2 - \omega_i}{2\omega_i} |||u_i^{n+1} - u_i^n|||_i^2. \tag{5.25}$$

PROOF. We have

$$J(u^n) - J(u^{n+1}) = \sum_{i=1}^{N} (J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n \ldots) - J(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots)). \tag{5.26}$$

Then (5.25) clearly follows from (5.24), and from the application of Lemma 5.2 at each of the differences of the right-hand side of (5.26).

Proposition 5.2. *If* $0 < \omega_i < 2, \forall\, i = 1, \ldots, N$, *we have*

$$J(u^n) \geq J(u^{n+1}), \qquad \forall\, n$$

and

$$\lim_{n \to +\infty} (u^n - u^{n+1}) = 0 \text{ strongly in } V. \tag{5.27}$$

PROOF. Since $0 < \omega_i < 2$ implies $(2 - \omega_i)/\omega_i > 0, \forall\, i = 1, \ldots, N$, it follows from (5.25) that $J(u^n)$ is a decreasing sequence. Since $J(u^n)$ is bounded below by $J(u)$, where u is the solution of (5.4), $J(u^n)$ is convergent. This implies

$$\lim_{n \to +\infty} (J(u^n) - J(u^{n+1})) = 0. \tag{5.28}$$

Then, from (5.25), (5.27) and from $(2 - \omega_i)/\omega_i > 0, \forall\, i = 1, \ldots, N$, it clearly follows that

$$\lim_{n \to +\infty} \|u_i^{n+1} - u_i^n\|_i = 0, \qquad \forall\, i = 1, \ldots, N.$$

This implies (5.27).

From these two propositions we deduce:

Theorem 5.1. *If* $0 < \omega_i < 2, \forall\, i = 1, \ldots, N$, *then the sequence* u^n *defined by* (5.23), (5.24) *satisfies*

$$\lim_{n \to +\infty} u^n = u,$$

where u *is the solution of* (5.4).

PROOF. The V-ellipticity of $a(\cdot, \cdot)$ implies

$$a(u^{n+1} - u, u^{n+1} - u) \geq \alpha \|u^{n+1} - u\|^2. \tag{5.29}$$

From (5.29) it follows that

$$a(u^{n+1}, u^{n+1} - u) - ((f, u^{n+1} - u)) \geq a(u, u^{n+1} - u) - ((f, u^{n+1} - u)) + \alpha \|u^{n+1} + u\|^2. \tag{5.30}$$

Since u is the solution of (5.4), and since $u^{n+1} \in K$, we have

$$a(u, u^{n+1} - u) - ((f, u^{n+1} - u)) \geq 0,$$

which, combined with (5.30), implies

$$a(u^{n+1}, u^{n+1} - u) - ((f, u^{n+1} - u)) \geq \alpha \|u^{n+1} - u\|^2. \tag{5.31}$$

The left-hand side of (5.31) could be written as follows:

$$a(u^{n+1}, u^{n+1} - u) - ((f, u^{n+1} - u)) = \sum_{i=1}^{N} \left(\left(\sum_{j=1}^{N} A_{ij} u_j^{n+1} - f_i, u_i^{n+1} - u_i \right) \right)_i. \tag{5.32}$$

Let $\bar{u}_i^{n+1}$ be the vector of K_i for which the functional

$$v_i \to J(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^{n}, \ldots)$$

attains its minimum on K_i. From Lemma 5.1 it follows that

$$\bar{u}_i^{n+1} = P_i \left(A_{ii}^{-1} \left(f_i - \sum_{j<i} A_{ij} u_j^{n+1} - \sum_{j>i} A_{ij} u_j^{n} \right) \right). \tag{5.33}$$

Moreover, from the usual characterization of the minimum we have

$$\left(\left(A_{ii} \bar{u}_i^{n+1} + \sum_{j<i} A_{ij} u_j^{n+1} + \sum_{j>i} A_{ij} u_j^{n} - f_i, v_i - \bar{u}_i^{n+1} \right) \right)_i \geq 0, \qquad \forall\, v_i \in K_i. \tag{5.34}$$

It follows from (5.32) that

$$\begin{aligned}
&a(u^{n+1}, u^{n+1} - u) - ((f, u^{n+1} - u)) \\
&= \sum_{i=1}^{N} ((A_{ii}(u_i^{n+1} - \bar{u}_i^{n+1}), u_i^{n+1} - u_i))_i \\
&\quad + \sum_{i=1}^{N} \left(\left(\sum_{j>i} A_{ij} \, (u_j^{n+1} - u_j^{n}), u_i^{n+1} - u_i \right) \right)_i \\
&\quad + \sum_{i=1}^{N} \left(\left(\sum_{j<i} A_{ij} u_j^{n+1} + A_{ii} \bar{u}_i^{n+1} + \sum_{j>i} A_{ij} u_j^{n} - f_i, u_i^{n+1} - \bar{u}_i^{n+1} \right) \right)_i \\
&\quad + \sum_{i=1}^{N} \left(\left(\sum_{j<i} A_{ij} u_j^{n+1} + A_{ii} \bar{u}_i^{n+1} + \sum_{j>i} A_{ij} u_j^{n} - f_i, \bar{u}_i^{n+1} - u_i \right) \right)_i.
\end{aligned} \tag{5.35}$$

Since $u_i \in K_i$, (5.34) implies that the last term on the right-hand side of (5.35) is ≤ 0. Therefore (5.31), (5.35) imply

$$\begin{aligned}
&\sum_{i=1}^{N} ((A_{ii}(u_i^{n+1} - \bar{u}_i^{n+1}), u_i^{n+1} - u_i))_i + \sum_{i=1}^{N} \left(\left(\sum_{j<i} A_{ij} u_j^{n+1} + A_{ii} \bar{u}_i^{n+1} \right. \right. \\
&\quad \left. \left. + \sum_{j>i} A_{ij} u_j^{n} - f_i, u_i^{n+1} - \bar{u}_i^{n+1} \right) \right)_i \\
&\quad + \sum_{i=1}^{N} \left(\left(\sum_{j>i} A_{ij} (u_j^{n+1} - u_j^{n}), u_i^{n+1} - u_i \right) \right)_i \geq \alpha \|u^{n+1} - u\|^2.
\end{aligned} \tag{5.36}$$

From (5.36) it follows that to prove the strong convergence of u^n to u, it suffices to prove that the left-hand side of (5.36) converges to zero. Since the A_{ij} are linear and continuous, the u_i^n, $\bar{u}_i^n$ are bounded uniformly in i and n, and $\lim_{n\to+\infty} \|u^{n+1} - u^n\| = 0$, it follows from (5.36) that $\lim_{n\to+\infty} \|\bar{u}^n - u^n\| = 0$ (where $\bar{u}^n = \{\bar{u}_1^n, \ldots, \bar{u}_i^n, \ldots, \bar{u}_N^n\}$) will imply convergence. Let us prove the last property. Since ω_i is >0, from (5.34) it follows that

$$\bar{u}_i^{n+1} = P_i\left(\bar{u}_i^{n+1} - \omega_i A_{ii}^{-1}\left(\sum_{j<i} A_{ij}u_j^{n+1} + A_{ij}\bar{u}_i^{n+1} + \sum_{j>i} A_{ij}u_j^n - f_i\right)\right). \tag{5.37}$$

From (5.24), (5.37), and since P_i is a contraction, it follows that

$$|||\bar{u}_i^{n+1} - u_i^{n+1}|||_i \leq |1 - \omega_i|\,|||\bar{u}_i^{n+1} - u_i^n|||_i, \qquad \forall\, i = 1, 2, \ldots, N. \tag{5.38}$$

Since $0 < \omega_i < 2$ implies $0 < |1 - \omega_i| < 1$, it follows from (5.38) that

$$|||\bar{u}_i^{n+1} - u_i^{n+1}|||_i \leq |||\bar{u}_i^{n+1} - u_i^n|||_i, \qquad \forall\, i = 1, \ldots, N. \tag{5.39}$$

From the triangle inequality and (5.38), (5.39), it follows that

$$\begin{aligned} |||u_i^n - u_i^{n+1}|||_i &\geq |||u_i^n - \bar{u}_i^{n+1}|||_i - |||\bar{u}_i^{n+1} - u_i^{n+1}|||_i \\ &\geq (1 - |1 - \omega_i|)|||\bar{u}_i^{n+1} - u_i^n|||_i \\ &\geq (1 - |1 - \omega_i|)|||\bar{u}_i^{n+1} - u_i^{n+1}|||_i, \end{aligned} \tag{5.40}$$

where $0 < 1 - |1 - \omega_i| < 1$.

From Proposition 5.2 we have $\lim_{n\to+\infty} \|u^{n+1} - u^n\| = 0$; this combined with (5.40) implies $\lim_{n\to\infty} \|\bar{u}^n - u^n\| = 0$, which completes the proof of the theorem. □

Remark 5.2. The above theorem generalizes Cryer [2], and also generalizes a classical result in finite dimensions (without constraints) for which we refer to R. S. Varga [1] and D. M. Young [1].

Remark 5.3. If $\omega_i > 1$ (resp., $\omega_i = 1$, $\omega_i < 1$), $\forall\, i = 1, \ldots, N$, then algorithm (5.23), (5.24) is an over-relaxation algorithm (resp., relaxation, under-relaxation algorithm) with projection.

5.5. Application to an elastic-plastic torsion problem

5.5.1. *Statement of the continuous problem*

We consider the elastic-plastic torsion problem described in Chapter II, Sec. 3 (we use the notation of Chapter II, Sec. 3):

$$a(u, v - u) \geq C \int_\Omega (v - u)\, dx, \qquad \forall\, v \in K, \quad u \in K, \tag{5.41}$$

where $a(u, v) = \int_\Omega \nabla u \cdot \nabla v\, dx$, and where

$$K = \{v \mid v \in H_0^1(\Omega), |\nabla v(x)| \leq 1 \text{ a.e.}\}.$$

From the equivalence result of Brezis and Sibony [1] mentioned in Chapter II, Sec. 3.4, it follows that the solution of (5.41) is also the solution of

$$a(u, v-u) \geq C \int_{\Omega} (v-u)\, dx, \qquad \forall\, v \in \hat{K},$$

$$u \in \hat{K} = \{v \mid v \in H_0^1(\Omega),\ |v(x)| \leq d(x, \Gamma) \text{ a.e.}\}. \tag{5.42}$$

In this section we consider only the numerical analysis of (5.42).

5.5.2. *A finite element approximation of* (5.42)

We consider piecewise linear approximations only. As mentioned in Chapter II, Sec. 3.4, (5.42) is a variant of the obstacle problem we considered in Chapter II, Sec. 2.

We assume that Ω is a polygonal domain and that $\mathcal{T}_h$ is a standard triangulation of Ω (see Chapter II). From Chapter II, Sec. 2 it follows that $H_0^1(\Omega)$ and $\hat{K}$ may be, approximated respectively, by[6]

$$V_h = \{v_h \mid v_h \in C^0(\bar{\Omega}),\ v_h = 0 \text{ on } \Gamma,\ v_h|_T \in P_1,\ \forall\, T \in \mathcal{T}_h\},$$
$$\hat{K}_h = \{v_h \mid v_h \in V_h,\ |v_h(x)| \leq d(x, \Gamma),\ \forall\, x \in \mathring{\Sigma}_h\}.$$

Then (5.42) is approximated by

$$a(u_h, v_h - u_h) \geq C \int_{\Omega} (v_h - u_h)\, dx, \qquad \forall\, v_h \in \hat{K}_h,\quad u_h \in \hat{K}_h. \tag{5.43}$$

From the results of Chapters I and II it easily follows that:

Proposition 5.3. *The approximate problem* (5.43) *has a unique solution. This solution is also the unique solution of the minimization problem*

$$J(u_h) \leq J(v_h), \qquad \forall\, v_h \in \hat{K}_h,\quad u_h \in \hat{K}_h, \tag{5.44}$$

where $J(v) = \frac{1}{2}a(v, v) - C \int_{\Omega} v\, dx$.

Moreover, using the methods of Chapter II, we may prove the following:

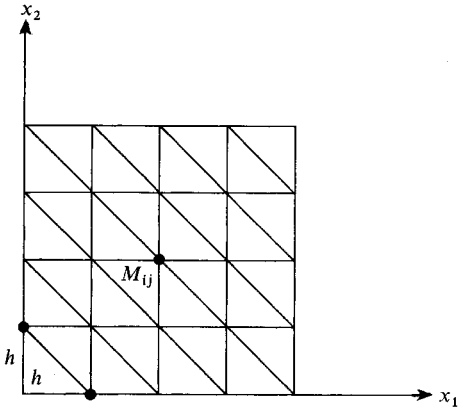
Theorem 5.2. *Suppose the angles of* $\mathcal{T}_h$ *are uniformly bounded below by* $\theta_0 > 0$ *as* $h \to 0$. *Then*

$$\lim_{h \to 0} \|u_h - u\|_{H_0^1(\Omega)} = 0,$$

where u *and* u_h *are, respectively, the solutions of* (5.42) *and* (5.43).

[6] $\mathring{\Sigma}_h$: set of the interior nodes of $\mathcal{T}_h$.

Figure 5.1



5.5.3. *Solution of the approximate problem*

The natural unknowns of the approximate problems (5.43) and (5.44) are $u_h(P)$, $P \in \mathring{\Sigma}_h$. Since $v_h \to J(v_h)$ is quadratic with respect to $v_h(P)$, it follows from the structure of $\hat{K}_h$ that algorithm (5.23), (5.24) could be used to solve (5.43), (5.44).

5.5.4. *Application to the case* $\Omega =]0, 1[\times]0, 1[$

When $\Omega =]0, 1[\times]0, 1[$, or, more generally, a rectangle, it is convenient to use a finite element approximation which is actually equivalent to a finite difference approximation. Let N be a positive integer and $h = 1/N$. On Ω we use the triangulation $\mathcal{T}_h$ of Fig. 5.1. We have

$$\Sigma_h = \{M_{ij} | M_{ij} = \{ih, jh\},\ i, j \text{ integers},\ 0 \leq i, j \leq N\},$$

$$\mathring{\Sigma}_h = \{M_{ij} | M_{ij} \in \Sigma_h,\ 1 \leq i, j \leq N - 1\}.$$

By analogy with the finite difference notation, it is convenient to denote $v_h(M_{ij})$ by v_{ij}. Then the approximate problems (5.43), (5.44) are easily expressed in function of v_{ij}, since

$$J(v_h) = \frac{h^2}{4} \sum_{0 \leq i, j \leq N} \left(\left(\frac{v_{i+1\,j} - v_{ij}}{h} \right)^2 + \left(\frac{v_{i-1\,j} - v_{ij}}{h} \right)^2 + \left(\frac{v_{ij+1} - v_{ij}}{h} \right)^2 + \left(\frac{v_{ij-1} - v_{ij}}{h} \right)^2 \right) - h^2 C \sum_{1 \leq i, j \leq N-1} v_{ij},$$

where $v_{kl} = 0$ if $M_{kl} \notin \mathring{\Sigma}_h$,

and since

$$\hat{K}_h = \{v_h | v_h \in V_h,\ |v_{ij}| \leq d_{ij},\ 1 \leq i, j \leq N - 1\},$$

where $d_{ij} = d(M_{ij}, \Gamma)$.

Algorithm (5.23), (5.24) has been used with $\omega_{ij} = \omega$, $\forall\, 1 \leq i, j \leq N-1$. Therefore the explicit form of (5.23), (5.24) is

$$u_h^0 \in \hat{K}_h \text{ arbitrary given,} \tag{5.45}$$

and for $1 \leq i, j \leq N-1$,

$$u_{ij}^{n+1/2} = (1-\omega)u_{ij}^n + \frac{\omega}{4}(u_{i+1j}^n + u_{ij+1}^n + u_{i-1j}^{n+1} + u_{ij-1}^{n+1} + h^2 C) \tag{5.46}$$

$$u_{ij}^{n+1} = \max(-d_{ij}, \min(u_{ij}^{n+1/2}, d_{ij})), \tag{5.47}$$

where $u_{kl}^m = 0$ if $M_{kl} \notin \mathring{\Sigma}_h$.

From Theorem 5.1 it follows that u_h converges to the solution of (5.43), (5.44) if $0 < \omega < 2$.

Numerical experiments have been made with $C = 10$, $h = \frac{1}{40}$, $u_h^0 = 0$. The *stopping criterion* used was

$$\sum_{1 \leq i, j \leq N-1} |u_{ij}^{n+1} - u_{ij}^n| < 10^{-4}. \tag{5.48}$$

The number of iterations necessary to obtain the convergence is given below as a function of ω:

ω	1	1.4	1.5	1.6	1.7	1.8	1.9	2
Number of Iterations	290	138	118	98	93	108	188	

The C.P.U. time on an IBM 360/91 was 7.33 sec for $\omega = 1$ and 2.51 sec for $\omega = 1.7$.

Figure 5.2 shows the behavior of

$$R^n = \sum_{1 \leq i, j \leq N-1} |u_{ij}^{n+1} - u_{ij}^n|$$

as a function of n (when $\omega = 1.7$).

In the Fig. 5.3 we can see the elastic part of Ω (in white) and the plastic parts of Ω (striped area). In the plastic parts we have $|\nabla u| = 1$ and $|u(x)| = d(x, \Gamma)$. In the elastic part we have $-\Delta u = C$ and $|u(x)| < d(x, \Gamma)$.

Remark 5.4. The optimal value of ω is very close to the optimal value of ω corresponding to the solution, by S.O.R., of the Dirichlet problem $-\Delta u = C$ in the elastic part of Ω (with the same discretization step h). This follows directly from the fact that the plastic part of Ω[7] is very quickly obtained. From

[7] It is the part of Ω in which the constraints $|u| \leq d(x, \Gamma)$ are active (i.e., $|u| = d(x, \Gamma)$).

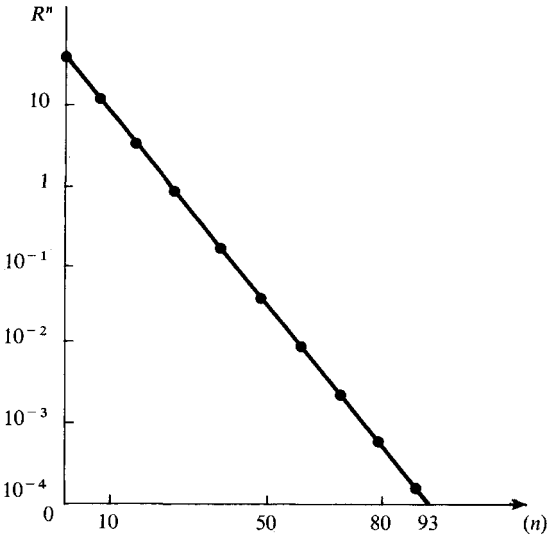


Figure 5.2

this fact it follows that the computation time is mainly used to solve $-\Delta u = C$ in the elastic part of Ω. Moreover, the plastic part increases with C; this fact explains that the optimal value of ω and the number of iterations needed to obtain the convergence are, for a given h, decreasing functions of C (as is the computational time).

Remark 5.5. In Cryer [1], [2] we can find a discussion on the choice of ω, when using point S.O.R. with projection to minimize quadratic functionals on a product of intervals.

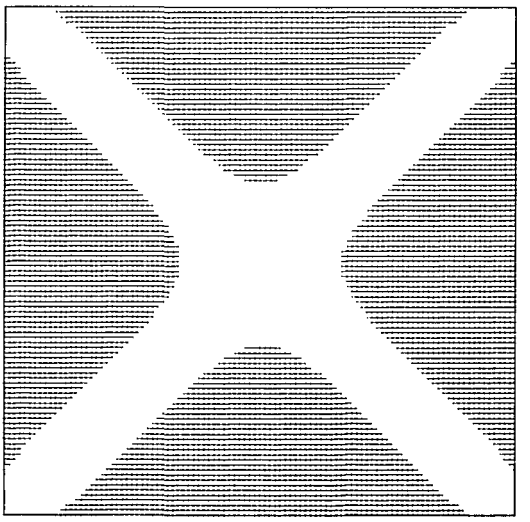


Figure 5.3

6. Solution of Systems of Nonlinear Equations by Relaxation Methods

6.1. Statement of the problem

In this section we very briefly describe some relaxation methods for solving systems of nonlinear equation such that

$$\begin{aligned} f_1(u_1, u_2, \ldots, u_N) &= 0, \\ &\vdots \\ f_i(u_1, u_2, \ldots, u_N) &= 0, \\ &\vdots \\ f_N(u_1, u_2, \ldots, u_N) &= 0, \end{aligned} \tag{6.1}$$

where $f_i: \mathbb{R}^N \to \mathbb{R}$. We define $F: \mathbb{R}^N \to \mathbb{R}^N$ by $F = \{f_1, \ldots, f_N\}$.

6.2. A first algorithm

We consider the following algorithm:

$$u^0 \text{ given}; \tag{6.2}$$

u^n being known, we compute u^{n+1} by

$$\begin{aligned} f_i(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^{n+1/2}, u_{i+1}^n, \ldots) &= 0, \\ u_i^{n+1} &= u_i^n + \omega(u_i^{n+1/2} - u_i^n), \\ 1 \leq i &\leq N. \end{aligned} \tag{6.3}$$

If $F = \nabla J$, where $J: \mathbb{R}^N \to \mathbb{R}$ is a strictly convex C^1 function such that $\lim_{\|v\| \to +\infty} J(v) = +\infty$, then (6.3) has a unique solution (see Sec. 2). This solution is also the unique solution of

$$J(u) \leq J(v), \qquad \forall\, v \in \mathbb{R}^N, \quad u \in \mathbb{R}^N. \tag{6.4}$$

Moreover, if $\omega = 1$, it follows from Theorem 3.1 that algorithm (6.2), (6.3) converges to the solution u of (6.1), (6.4).

If $F = \nabla J$, $\omega \neq 1$, we refer to S. Schechter [1], [2], [3]. In these papers it is proved that under the hypothesis

(i) $J \in C^2(\mathbb{R}^N)$,
(ii) $(J'(w) - J'(v), w - v) \geq \alpha \|w - v\|^2$, $\forall\, v, w$; $\alpha > 0$,
(iii) $0 < \omega < \omega_M$,

algorithm (6.2), (6.3) converges to the solution of (6.1), (6.4). Moreover, estimates of ω_M and of the optimal value of ω are given.

For the convergence of (6.2), (6.3) when $F \neq \nabla J$, we refer to Ortega and Rheinboldt [1], Miellou [1], [2], and the bibliography therein.

6.3. A second algorithm

This algorithm is given by

$$u^0 \text{ given}; \tag{6.5}$$

u^n *being known, we compute* u^{n+1} *by*

$$f_i(u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots) = (1 - \omega) f_i(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots),$$
$$1 \leq i \leq N. \tag{6.6}$$

To our knowledge the convergence of (6.5), (6.6) for $\omega \neq 1$ and F nonlinear has not yet been considered.

Remark 6.1. Algorithms (6.2), (6.3) and (6.5), (6.6) are identical if $\omega = 1$ and/or F is linear.

Remark 6.2. In many applications, from the numerical experiments it appears that (6.5), (6.6) is faster than (6.2), (6.3), ω having its (experimental) optimal value in both cases. Intuitively this seems to be related to the fact that (6.5), (6.6) is "more implicit" than (6.2), (6.3). For instance, (6.5), (6.6) could easily be used if F is only defined on a subset D on $\mathbb{R}^N$; in such a situation, when using (6.2), (6.3) with $\omega > 1$, it could happen that $\{u_1^{n+1}, \ldots, u_i^{n+1}, u_{i+1}^n, \ldots\} \notin D$.

6.4. A third algorithm

In this section we assume that $F \in C^1(\mathbb{R}^N)$. A natural method for computing $u_i^{n+1/2}$ in (6.3) or u_i^{n+1} in (6.6) is Newton's method. We recall that Newton's method applied to the solution of the single-variable equation

$$f(x) = 0$$

is basically:

$$x^0 \text{ given}; \tag{6.7}$$

$$x^{m+1} = x^m - \frac{f(x^m)}{f'(x^m)}. \tag{6.8}$$

In the computation of $u_i^{n+1/2}$ in (6.3) or u_i^{n+1} in (6.6) by (6.7), (6.8), the obvious starting value is u_i^n. Then obvious variants of (6.2), (6.3) and (6.5), (6.6) are obtained if we run only one Newton iteration. Actually, in such a case, (6.2), (6.3) and (6.5), (6.6) reduce to the same algorithm, which is

$$u^0 \text{ given}; \tag{6.9}$$

$$u_i^{n+1} = u_i^n - \omega \frac{f_i(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots)}{(\partial f_i / \partial v_i)(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^n, \ldots)}, \qquad 1 \leq i \leq N. \tag{6.10}$$

In S. Schecter, *loc. cit.*, the convergence of (6.9), (6.10) is proved, if $F = \nabla J$, under the same assumptions as in Sec. 6.2 for algorithm (6.2), (6.3) (with a different ω_M in general).

Remark 6.3. In Glowinski and Marrocco [1], [2] and Concus [1], we can find comparisons between the above methods when applied to the numerical solution of the nonlinear elliptic equation modeling the magnetic state of ferromagnetic media (see also Winslow [1]). Applications of the first algorithm for solving minimal surface problems may be found in Jouron [1].