JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 32, No. 3, NOVEMBER 1980

Auxiliary Problem Principle and Decomposition of Optimization Problems

G. COHEN[1]

Communicated by D. Q. Mayne

Abstract. The auxiliary problem principle allows one to find the solution of a problem (minimization problem, saddle-point problem, etc.) by solving a sequence of auxiliary problems. There is a wide range of possible choices for these problems, so that one can give special features to them in order to make them easier to solve. We introduced this principle in Ref. 1 and showed its relevance to decomposing a problem into subproblems and to coordinating the subproblems. Here, we derive several basic or abstract algorithms, already given in Ref. 1, and we study their convergence properties in the framework of infinite-dimensional convex programming.

Key Words. Convex programming, optimization algorithms, decomposition, coordination, large-scale systems.

1. Introduction

Motivated by such works as those of Arrow–Hurwicz (Ref. 1) in 1960, Takahara (Ref. 2) in 1964, Lasdon and associates (Refs. 3 and 4) in 1965, and Mesarovic and associates (Ref. 5) in 1970, a plentiful literature has been and is still currently devoted to what we may call two-level or *decomposition-coordination* algorithms for various optimization problems. Recently, we presented an attempt to provide a unified view of this field (Ref. 6). We considered two simple principles to start with and derived a few basic algorithms from them. Several examples were given to illustrate the assertion that most of the existing two-level algorithms can be considered as specializations of the previous basic algorithms to particular situations.

The main advantages of such an approach are the following. First, it reduces the so-called coordination principles or techniques to a minimum

[1] Maître de Recherche, Centre d'Automatique-Informatique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau, France; Conseiller Extérieur, IRIA, Laboria, Le Chesnay, France.

0022-3239/80/1100-0277$03.00/0

number of basic ones. Moreover, since classical algorithms (such as gradient, Newton-Raphson, Uzawa, etc.) can also be derived from the general formalism, we can bridge the gap between classical and decomposition algorithms. Although this abstract point of view, as is intrinsic in this kind of approach, may have the drawback of masking somewhat the practical issues of the decomposition-coordination approach (and thus may require some time for the reader to become familiar with it), this effort pays off in that many assumptions currently used by authors appear clearly unnecessary. In particular, we refer to separability assumptions, such as additive cost and additive constraints (see, e.g., Refs. 3 and 4). Hence, some coordination methods can be immediately extended to less restrictive situations. For example, the celebrated price coordination principle (Ref. 3) or interaction balance principle (Ref. 5) can be used in nonseparable problems as well. The last consequence of our systematic approach is that it gives guidelines for imagining new basic schemes, possibly more suited to particular problems.

Since all these issues have been discussed in Ref. 6, we shall not come back to them in the following sections. However, along the way, some of them will be briefly alluded to. Our main purpose here is to introduce once again the two principles that are the roots of all the basic algorithms and then to discuss these algorithms from the mathematical point of view, that is, to give the theorems of convergence and the proofs which were missing in Ref. 6.

The frame will be that of convex mathematical programming in possibly infinite-dimensional, reflexive Banach or Hilbert spaces. This includes most of the cases of interest in deterministic optimization, in particular optimal control problems of linear systems.[2] The restriction to the convex case allows us to state necessary and sufficient optimality conditions and also to obtain proofs of global convergence (i.e., starting from any initial guess). Of course, this does not prevent anybody from using the algorithms in practical nonconvex cases, since, if convergence results, then necessary conditions of overall optimality are generally met in the limit.

In the rest of the paper, the so-called auxiliary problem principle is introduced first, and a first family of algorithms is derived. It will be explained how they can be used for *parallel decomposition*, that is, when a problem is decomposed into subproblems which can be solved *independently* at each coordination step. More widespread and related to *sequential decomposition* is the relaxation principle. Combination of both principles yields relaxed algorithms. While, in this first part, only minimization problems on feasible subsets are considered, the second part extends the principles and algorithms to saddle-point problems and, more specifically, to

[2] Linearity is imposed by the convexity assumption.

minimization problems with explicit constraints (which, by duality, yield saddle-point problems for the Lagrangian functionals). All the proofs of convergence are collected in the Appendix. In the conclusions, we underline the basic assumption making the auxiliary problem principle, and thus parallel decomposition, possible. This suggests future directions of research when this assumption is not met.

2. Auxiliary Problem Principle. First Family of Basic Algorithms

2.1. Preliminaries. For convenience, we gather here some definitions, notation, terminology, and auxiliary results. Let $\mathcal{U}$ be a reflexive Banach space, which we sometimes assume to be a Hilbert space. $\mathcal{U}^*$ denotes the topological dual space of continuous linear functionals on $\mathcal{U}$. If $u \in \mathcal{U}$ and $p \in \mathcal{U}^*$, then $\langle p, u\rangle$ denotes the scalar which results from the action of p on u. If $\mathcal{U}$ is a Hilbert space, $\mathcal{U}^*$ is identified to $\mathcal{U}$ by a standard procedure, and $\langle \cdot , \cdot \rangle$ also denotes the scalar product in $\mathcal{U}$. $\mathcal{U}^f$ is a closed, convex subset of $\mathcal{U}$, and Π denotes the projection onto $\mathcal{U}^f$. Let J be a convex functional on $\mathcal{U}$.

Assumption (A). For a given $\mathcal{U}^f$, we say that Assumption (A) is met for J if, for all sequence $\{u^k \mid k \in \mathbb{N}, u^k \in \mathcal{U}^f\}$ such that $\|u^k\| \to +\infty$ (if any), then $J(u^k) \to +\infty$.

Of course, Assumption (A) is met for every J if $\mathcal{U}^f$ is bounded. The symbol $J'(u)$ denotes the *Gateaux derivative* (G-derivative) of J at u; that is (see Ref. 7), we assume that the limit of $(J(u+\epsilon h)-J(u))/\epsilon$, when $\epsilon \to 0$, exists and is equal to $\langle J'(u), h\rangle$ for all h, where $J'(u) \in \mathcal{U}^*$. If J is defined on a product space $\mathcal{U} \times \mathcal{V}$, $J'_u(u, v)$ denotes the partial derivative of J with respect to the first variable.

The derivative is *strongly monotone* with constant a, if $\exists a > 0$ such that,

$$\text{for all } u, w \in \mathcal{U}, \qquad \langle J'(u)-J'(w), u-w\rangle \geq a\|u-w\|^2. \tag{1}$$

The derivative is *Lipschitz* with constant A, if $\exists A > 0$ such that,

$$\text{for all } u, w \in \mathcal{U}, \qquad \|J'(u)-J'(w)\|_* \leq A\|u-w\|, \tag{2}$$

where $\|\cdot\|_*$ is the classical norm of operator in $\mathcal{U}^*$.

Lemma 2.1. If J' meets (1), then

$$\text{for all } u, w \in \mathcal{U}, \qquad J(w)-J(u) \geq \langle J'(u), w-u\rangle + \tfrac{1}{2}a\|u-w\|^2. \tag{3}$$

If J' meets (2), then

$$\text{for all } u, w \in \mathcal{U}, \qquad J(w) - J(u) \le \langle J'(u), w - u\rangle + \tfrac{1}{2}A\|u - w\|^2. \tag{4}$$

The proof is omitted. See Ref. 7 for a proof of (3).

Notice that, if (1) holds, then (3) implies that the sum of J and any Lipschitz functional on u meets Assumption (A).

We consider the so-called master problem [Problem (MP)]:

$$\text{(MP)} \qquad \min_{u \in \mathcal{U}^f} J(u) + J_1(u),$$

where J_1 is another lower semicontinuous convex functional which is not assumed to be differentiable. Notice that J is continuous, since we assume that it is G-differentiable. If Assumption (A) is met by $J + J_1$, then a solution u^* of Problem (MP) exists. Moreover, it is unique when (1) holds.

2.2. Auxiliary Problem Principle

Lemma 2.2. Let G be a convex, G-differentiable functional, let ϵ be a positive constant, and let u^* be a solution of

$$\min_{u \in \mathcal{U}^f} G(u) + \epsilon J_1(u). \tag{5}$$

Assume that

$$G'(u^*) = \epsilon J'(u^*). \tag{6}$$

Then, u^* solves Problem (MP).

Proof. A necessary and sufficient condition for a solution u^* of Problem (MP) is that (see Ref. 7)

$$u^* \in \mathcal{U}^f \quad \text{and, for all } u \in \mathcal{U}^f, \qquad \langle J'(u^*), u - u^*\rangle + J_1(u) - J_1(u^*) \ge 0. \tag{7}$$

One can write the same kind of condition for Problem (5) and derive (7) using (6) and $\epsilon > 0$. □

For any $v \in \mathcal{U}^f$, any $\epsilon > 0$, and any convex, G-differentiable functional K, consider the following functional:

$$G^v : u \mapsto K(u) + \langle \epsilon J'(v) - K'(v), u\rangle. \tag{8}$$

We check that

$$(G^v)'(v) = \epsilon J'(v), \tag{9}$$

which looks like (6). Hence, if v happens to be a solution of (5), with G^v

replacing G, then v is a solution of Problem (MP). This suggests the following algorithm of the fixed-point type.

Let a sequence of functionals $\{K^k, k \in \mathbb{N}\}$ and positive numbers $\{\epsilon^k, k \in \mathbb{N}\}$ be chosen.

Algorithm 2.1

(i) Choose $u^0 \in \mathcal{U}^f$. Set $k = 0$.

(ii) Solve the auxiliary problem

$$(\mathrm{AP}^k) \quad \min_{u \in \mathcal{U}^f} K^k(u) + \langle \epsilon^k J'(u^k) - (K^k)'(u^k), u \rangle + \epsilon^k J_1(u).$$

Let u^{k+1} be a solution.

(iii) Stop if $\|u^k - u^{k+1}\|$ or $|(J+J_1)(u^k) - (J+J_1)(u^{k+1})|$ is below some desired threshold. Otherwise, make $k \leftarrow k+1$, and return to step (ii).

Notice that, if $u^0 \notin \mathcal{U}^f$, then u^1 can be relabeled u^0. The existence and even uniqueness of a solution to Problem (AP^k) can be ensured by a proper choice of K^k. Moreover, we can take advantage of this choice to obtain a good numerical conditioning of Problem (AP^K); that is, $(K^k)'$ can be strongly enough monotone. Finally, assuming a decomposition

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_N, \qquad \mathcal{U}^f = \mathcal{U}_1^f \times \cdots \times \mathcal{U}_N^f, \tag{10}$$

and provided that J_1, the nondifferentiable part, be additive with respect to this decomposition, we can also choose K^k additive, so that Problem (AP^k) splits into N independent problems. As one sees, we can give to Problem (AP^k) many special features by choosing K^k. This is why we call it a *core*. Actually, this choice is subjected to rather mild conditions as indicated in the theorem below. In Ref. 6, we gave some references where the auxiliary problem principle, which is not really new, appeared in less general forms than that presented here.

Theorem 2.1. We assume the following:

(i) Assumption (A) is met for $J + J_1$;

(ii) J is convex, with a G-derivative Lipschitz with constant A [see (2)] on $\mathcal{U}^f$;

(iii) J_1 is convex, lower semicontinuous, and such that $J_1(u) + m\|u\|^r$ is bounded from below on $\mathcal{U}^f$, for some $m > 0$ and $r < 2$;

(iv) the functional K^k are convex, with G-derivatives strongly monotone, with constant b^k and Lipschitz with constant B^k on $\mathcal{U}^f$.

Moreover, $\exists b > 0$, $\exists B > 0$, such that,

$$\text{for all } k \in \mathbb{N}, \qquad b^k \geq b, \qquad B^k \leq B. \tag{11}$$

Then, a solution u^* of Problem (MP) exists, as well as a unique solution u^{k+1} of Problem (AP^k) for all k. If the ϵ^k are such that

$$\alpha \le \epsilon^k \le 2b^k/(A+\beta), \quad \text{for some } \alpha > 0 \text{ and } \beta > 0, \tag{12}$$

then the sequence $\{(J+J_1)(u^k)\}$ is strictly decreasing (unless $u^k = u^*$ for some k), and it converges toward $(J+J_1)(u^*)$. Every weak cluster point of the sequence $\{u^k\}$ (at least one exists) is a solution of Problem (MP). Hence, the sequence $\{u^k\}$ weakly converges to u^* if this is unique.

Moreover, assume that

(v) J' is strongly monotone with constant a on $\mathcal{U}^f$.

Then, $\{u^k\}$ converges *strongly* toward the unique u^*, and we have the a posteriori error estimation

$$\|u^{k+1} - u^*\| \le (1/a)(B^k/\epsilon^k + A)\|u^{k+1} - u^k\|. \tag{13}$$

2.3. Other Algorithms and Remarks. A first way of modifying Algorithm 2.1 is the following. If the core $K^k(\cdot)$ is changed into $K^k(\cdot) + \gamma^k\|\cdot\|^2$, $\gamma^k > 0$, and ϵ^k is taken equal to 1, then one gets an additional term $\gamma^k\|u - u^k\|^2$ in Problem (AP^k). The convergence can then be obtained with γ^k large enough.

Another way, still with $\varepsilon^k = 1$, is the following algorithm.

Algorithm 2.2. Here, Step (ii) is replaced by the follow up steps.

(ii-a) Solve Problem (AP^k), with $\varepsilon^k = 1$. Let $\hat{u}^{k+1}$ be a solution.

(ii-b) Set

$$u^{k+1} = \rho^k \hat{u}^{k+1} + (1-\rho^k)u^k, \tag{14}$$

with $\rho^k > 0$ and such that $u^{k+1} \in \mathcal{U}^f$ ($\rho^k \le 1$, if necessary).

When $\rho^k \le 1$, this is called *under-relaxation*; and, when $\rho^k \ge 1$, this is called *over-relaxation* (see Ref. 8). Algorithm 2.2 generates sequences generally different from those of Algorithm 2.1, but we have the following theorem.

Theorem 2.2. With Assumptions (i) to (v) of Theorem 2.1, the same conclusions hold for Algorithm 2.2, if the sequence $\{\rho^k\}$ meets the condition (12) when J_1 is not present and, in addition, if $\rho^k \le 1$ when J_1 is present. The a posteriori error estimation is now

$$\|\hat{u}^{k+1} - u^*\| \le (1/a)(B^k + A)\|\hat{u}^{k+1} - u^k\|. \tag{15}$$

Remark 2.1. To see that ρ^k must not be taken greater than 1 when the nondifferentiable part J_1 is present, consider the case when J_1 is the indicator function of a feasible convex set (Rockafellar, Ref. 9).

As previously indicated, our main motivation for considering the auxiliary problem principle and the resulting algorithms is in the case when a decomposition such as (10) is given. Moreover, we must assume that the nondifferentiable part J_1, when it exists, is additive with respect to (10). Then, with an additive core, Problem (AP^k) turns out to be made up of independent subproblems that may be solved in parallel if one has N processors at his disposal. Several examples are given in Ref. 6, with systematic ways of deriving additive cores from J.

We shall limit ourselves here to the case when $\mathcal{U}$ is a Hilbert space and the core $K^k(\cdot)$ is $\frac{1}{2}\|\cdot\|^2$. Then, Algorithm 2.1 yields, assuming $J_1 \equiv 0$,

$$u^{k+1} = \Pi(u^k - \epsilon^k J'(u^k)), \tag{16}$$

namely a projected-gradient algorithm. Also in Ref. 6, we recovered the Newton-Raphson algorithm for a twice-differentiable functional J.

3. Relaxation and Relaxed Algorithms

Relaxation is a classical principle in optimization to convert a problem into a sequence of subproblems solved sequentially (that is, each partial solution is immediately used in the next subproblem). Assuming (10), at step tN + i, one solves:

$$\min_{u_i \in \mathcal{U}_i^f} J(u_1^{t+1}, \ldots, u_{i-1}^{t+1}, u_i, u_{i+1}^t, \ldots, u_N^t), \tag{17}$$

yielding u_i^{t+1}.

This principle can be combined with the auxiliary problem principle in the following way. For the sake of notational simplicity, we assume that $J_1 \equiv 0$, that K and ϵ are chosen independent of k, and that $\mathcal{U}$ is decomposed into only two components, which we now call $\mathcal{U}$ and $\mathcal{V}$. That is, Problem (MP) is now the following Problem (MP′):

(MP′) $\min J(u, v)$, with $u \in \mathcal{U}^f \subset \mathcal{U}$, $v \in \mathcal{V}^f \subset \mathcal{V}$.

Algorithm 3.1

(i) Choose $(u^0, v^0) \in \mathcal{U}^f \times \mathcal{V}^f$. Set $k = 0$.

(ii) Solve the following auxiliary problem:

(AP_u^k) $\min_{u \in \mathcal{U}^f} K(u, v^k) + \langle \epsilon_1 J'_u(u^k, v^k) - K'_u(u^k, v^k), u \rangle.$

Let u^{k+1} be a solution.

(iii) Solve the following auxiliary problem:

$$(\mathrm{AP}_v^k) \quad \min_{v \in \mathcal{V}^f} K(u^{k+1}, v) + \langle \epsilon_2 J'_v(u^{k+1}, v^k) - K'_v(u^{k+1}, v^k), v \rangle.$$

Let v^{k+1} be a solution.

(iv) Stop if some degree of accuracy is reached. Otherwise, make $k \leftarrow k+1$, and return to step (ii).

Notice that we can choose ϵ_1 different from ϵ_2, since Lemma 2.2 can be generalized to the situation (10) when (6) holds for each partial derivative with a different ϵ_i.

Theorem 3.1. We assume that J is jointly convex in (u, v), and its G-derivative is strongly monotone with constant a on $\mathcal{U}^f \times \mathcal{V}^f$. The mappings $u \mapsto J'_u(u, v)$, $v \mapsto J'_v(u, v)$ are Lipschitz with respective constant A_1 (independent of v), A_{12}, and A_2 (independent of u). The restricted mappings $u \mapsto K(u, v)$ and $v \mapsto K(u, v)$ are separately convex, respectively, in u for all v and in v for all u; their G-derivatives are strongly monotone with constants b_1 and b_2, respectively, and Lipschitz with constants B_1 and B_2. Then, if

$$0 < \epsilon_i < 2b_i/A_i, \qquad \text{for } i = 1, 2, \tag{18}$$

the sequence $\{u^k, v^k\}$ generated by Algorithm 3.1 converges toward the unique solution (u^*, v^*) of (MP'), and $\{J(u^k, v^k)\}$ monotonously decreases.

Notice that we have no assumption on K jointly in (u, v), whereas we assume J to be jointly convex. Assume now that J meets Assumption (ii) of Theorem 2.1 and K meets Assumption (iv) of the same theorem. Then, we clearly have that

$$A \geq \max(A_1, A_2) \qquad \text{and} \qquad b \leq \min(b_1, b_2).$$

Hence, the upper bound of ϵ in Theorem 2.1 is smaller than the upper bounds of ϵ_1 and ϵ_2 in (18). This gives a heuristic comparison between Algorithm 2.1 and its relaxed version, Algorithm 3.1. It is well known that the latter requires less steps to achieve a desired accuracy. However, in this case, the partial minimizations are sequential, so that no parallel computations are possible as it is with Algorithm 2.1 when, for example, we consider the additive core derived from the core K used in Algorithm 3.1 via the rule

$$K^k : (u, v) \mapsto K(u, v^k) + K(u^k, v). \tag{19}$$

Thus, time savings may occur if a multiprocessor is used for the implementation of Algorithm 2.1, despite its slower convergence. Further aspects of parallel and sequential decomposition algorithms are discussed in Ref. 6.

4. Auxiliary Problem Principle in Saddle-Point Problems

Let Φ be a functional on $\mathscr{U} \times \mathscr{P}$ (two reflexive Banach or Hilbert spaces), such that $u \mapsto \Phi(u, p)$ is lower semicontinuous and convex for all p and $p \mapsto \Phi(u, p)$ is upper semicontinuous and concave for all u.

Let $\mathscr{U}^f \subset \mathscr{U}$ and $\mathscr{P}^f \subset \mathscr{P}$ be closed, convex subsets. The master problem now is to find $(u^*, p^*) \in \mathscr{U}^f \times \mathscr{P}^f$, such that

$$\text{for all } u \in \mathscr{U}^f, p \in \mathscr{P}^f, \Phi(u^*, p) \leq \Phi(u^*, p^*) \leq \Phi(u, p^*). \tag{20}$$

A solution exists under mild additional assumptions (see Ref. 9). Assume now that G-derivatives with respect to u and p (separately) exist. Then, the following variational inequalities are equivalent to (20):

$$\text{for all } u \in \mathscr{U}^f, \quad \langle \Phi'_u(u^*, p^*), \quad u - u^* \rangle \geq 0,$$
$$\text{for all } p \in \mathscr{P}^f, \quad \langle \Phi'_p(u^*, p^*), \quad p - p^* \rangle \leq 0.$$

We generalize Lemma 2.2 as follows.

Lemma 4.1. Let Γ be a functional of the same kind as Φ. Let (u^*, p^*) be a saddle point of Γ on $\mathscr{U}^f \times \mathscr{P}^f$, and assume that there exists $\epsilon > 0$ and $\rho > 0$, such that

$$\Gamma'_u(u^*, p^*) = \epsilon \Phi'_u(u^*, p^*), \qquad \Gamma'_p(u^*, p^*) = \rho \Phi'_p(u^*, p^*).$$

Then, (u^*, p^*) is also a saddle point of Φ on the same feasible sets.

The proof is straightforward. Notice that we could have added a nondifferentiable part $\Phi_1(u) + \Phi_2(p)$ to Φ and, consequently, $\epsilon\Phi_1(u) + \rho\Phi_2(p)$ to Γ. Several other tricks of this kind could also have been considered. In one instance hereafter, we shall use an additional $\Phi_1(u, p)$ differentiable in p, but not in u. As previously, Lemma 4.1 suggests an iterative algorithm by choosing a core $\Psi(u, p)$, that may depend on the step k, and by adding linear modifications in u and p to it in order to force the solutions of the successive auxiliary problems to converge toward that of (20). However, due to the presence of two variables, namely u and p, we may or may not use the relaxation principle. In the former case, we shall obtain successive minimization and maximization problems, while in the latter case we shall obtain a sequence of saddle-point problems. We begin with the latter case.

Algorithm 4.1

(i) Choose $(u^0, p^0) \in \mathscr{U}^f \times \mathscr{P}^f$. Set $k = 0$.

(ii) Solve the following auxiliary problem: find the saddle-point (u^{k+1}, p^{k+1}) of the following functional on $\mathcal{U}^f \times \mathcal{P}^f$:

$$(u, p) \mapsto \Psi(u, p) + \langle \epsilon \Phi'_u(u^k, p^k) - \Psi'_u(u^k, p^k), u \rangle + \langle \rho \Phi'_p(u^k, p^k) - \Psi'_p(u^k, p^k), p \rangle.$$

(iii) Stop if some degree of accuracy is reached. Otherwise, make $k \leftarrow k+1$, and return to Step (ii).

As Ψ, the quantities ϵ and ρ might depend on k. If $\mathcal{U}$ and $\mathcal{U}^f$ are decomposed as in (10) and/or $\mathcal{P}$ and $\mathcal{P}^f$ are similarly decomposed, we may consider relaxed versions of Algorithm 4.1 with respect to these decompositions. Notice that this is not the relaxation scheme considered in the following algorithm. To introduce it, assume first that Ψ takes the form $\Psi_1(u) + \Psi_2(p)$. Then, the auxiliary problem above splits into two *parallel problems*: one minimization in u and one maximization in p. We now consider *sequential problems* of these kinds.

Algorithm 4.2. In Algorithm 4.1, replace Step (ii) by the following two steps.

(ii-a) Solve the auxiliary problem

$$(\mathrm{AP}_u^k) \qquad \min_{u \in \mathcal{U}^f} \Psi(u, p^k) + \langle \epsilon \Phi'_u(u^k, p^k) - \Psi'_u(u^k, p^k), u \rangle.$$

Let u^{k+1} be a solution.

(ii-b) Solve the auxiliary problem

$$(\mathrm{AP}_p^k) \quad \max_{p \in \mathcal{P}^f} \Psi(u^{k+1}, p) + \langle \rho \Phi'_p(u^{k+1}, p^k) - \Psi'_p(u^{k+1}, p^k), p \rangle.$$

Let p^{k+1} be a solution.

Notice that Problem (AP_p^k) cannot be solved before Problem (AP_u^k) has been solved. This is what we mean by *sequential problems*. If u^{k+1} is replaced everywhere by u^k in Problem (AP_p^k), then we obtain an algorithm of the type of Algorithm 4.1, with the step-dependent core

$$\Psi^k : (u, p) \mapsto \Psi(u^k, p) + \Psi(u, p^k).$$

We shall not study Algorithms 4.1 and 4.2 in the general form given above. Instead, we shall apply these algorithms to the particular case when Φ is the Lagrangian functional coming from a constrained optimization problem. Moreover, we shall retain particular choices of cores Ψ in order to ultimately recover known algorithms and extend them. Corresponding respectively to Algorithm 4.1 and Algorithm 4.2, so-called *one-level algorithms* and *two-level algorithms* are examined in the following sections.

5. One-Level Algorithms in Minimization Problems with Explicit Constraints

5.1. Preliminaries. As in Section 2.1, we gather here a few facts to make the paper sufficiently self contained and to introduce some notation. So far, we have considered implicitly constrained problems by introducing feasible sets (such as $\mathscr{U}^f$). However, when applying our results to decomposition, we have restricted these implicit constraints to the form (10) in order to avoid any coupling through them. The motivation behind considering explicit constraints is to deal now with possibly coupling constraints. We thus consider the master problem.

$$\text{(MP)} \qquad \min_{u \in \mathscr{U}^f} J(u), \tag{21-1}$$

subject to

$$\Theta(u) = 0 \qquad \text{or} \qquad \Theta(u) \in -C, \tag{21-2}$$

where J is the same as previously and Θ is a mapping from $\mathscr{U}$ to a Hilbert space $\mathscr{C}$ supplied with a closed convex cone C (the positive cone) in case we need to consider inequality constraints. The interior $\mathring{C}$ of C is assumed to be nonempty. Notice that, at least formally, we can imbed the case of equality constraints by taking $C = \{0\}$.[3] In this case, applying the definition of the conjugate cone $C^* \subset \mathscr{C}^*$ (see Ref. 10) yields $C^* = \mathscr{C}^*$. This is well suited for the duality theory, that is, when a multiplier $p \in C^*$ is associated with the constraints to define the Lagrangian functional

$$L(u, p) = J(u) + \langle p, \Theta(u) \rangle, \tag{22}$$

since $p \in \mathscr{C}^*$ in the equality case. The mapping Θ is assumed to be convex, in the following sense:

$$\begin{gathered} \text{for all } u, w \in \mathscr{U}, \qquad \text{for all } \alpha \in [0, 1], \\ \alpha\Theta(u) + (1-\alpha)\Theta(w) - \Theta(\alpha u + (1-\alpha)w) \in C. \end{gathered} \tag{23}$$

This implies that, for all $p \in C^*$, the functional $u \mapsto \langle p, \Theta(u) \rangle$ is convex in the classical sense. Moreover, we assume it to be lower semicontinuous. Notice that, in the case $C = \{0\}$, (23) means that Θ is affine.

It is well known (Ref. 10) that, if L has a saddle point (u^*, p^*) on $\mathscr{U}^f \times C^*$, then u^* is a solution of (21). Conversely, in the convex case considered here, and provided that some so-called constraint qualification condition is met, if u^* is a solution to (21), then there exists $p^* \in C^*$ such that (u^*, p^*) is a saddle point of L. In our case, a possible constraint qualification

[3] Of course, $\mathring{C}$ is now empty, but this does not matter in the equality-constrained case.

condition is the following (see Ref. 10):

$$\exists u^0 \in \mathcal{U}^f, \text{ such that } \quad \Theta(u^0) \in -\mathring{C}.$$

We shall assume this to hold and shall search for the saddle point of L. Notice that we could also have considered augmented Lagrangian functionals (see Ref. 11).

5.2. One-Level Algorithms. We apply Algorithm 4.1 with $\Phi = L$, $\mathcal{P}^f = C^*$, and

$$\Psi(u, p) \triangleq K(u) + \langle p, \Omega(u) \rangle, \tag{24}$$

where K and Ω are mappings of the same kind as respectively J and Θ. One can check that the solution (u^{k+1}, p^{k+1}) of the auxiliary problem (if any) is also the solution of a constrained optimization problem given hereafter. As previously, to ensure that the latter problem is equivalent to the former, it is required that some constraint qualification condition hold. This is something difficult to guarantee at such a level of generality all along the algorithm. For the time being, we assume that this equivalence holds without going into details. More precise assumptions are considered in the next theorem.

Algorithm 5.1. In Algorithm 4.1, set $\Phi = L$, $\mathcal{P}^f = C^*$, and replace Step (ii) by the following step:

(ii) Solve the auxiliary problem

$$\min_{u \in \mathcal{U}^f} K(u) + \langle \epsilon J'(u^k) - K'(u^k), u \rangle + \langle p^k, (\epsilon \Theta'(u^k) - \Omega'(u^k)) \cdot u \rangle, \tag{25-1}$$

subject to

$$\Omega(u) + \rho \Theta(u^k) - \Omega(u^k) \in -C. \tag{25-2}$$

Let u^{k+1} be a solution, and let p^{k+1} be an optimal multiplier for the constraint (25-2).

The theorem of convergence will be restricted to a rather particular case, namely, $\mathcal{U}^f = \mathcal{U}$, $C = \{0\}$ (equality constraint), J, K quadratic functionals, and Θ, Ω affine. Let

$$J(u) = \tfrac{1}{2}\langle u - f, A(u - f) \rangle,$$

$$K(u) = \tfrac{1}{2}\langle u - g, B(u - g) \rangle,$$

$$\Theta(u) = D(u - d), \qquad \Omega(u) = E(u - e).$$

Here, A, B are self-adjoint, continuous, linear operators on the Hilbert space $\mathcal{U}$; D, E are continuous linear mappings from $\mathcal{U}$ to $\mathcal{C}$; and f, g, d, $e \in \mathcal{U}$.

Theorem 5.1. Assume that K and J are strongly monotone and that D and E are surjective. Under the assumption that[4]

$$DA^{-1}E^* + EA^{-1}D^* - DA^{-1}BA^{-1}D^* \text{ is strongly monotone on } \mathscr{C}^*, \quad (26\text{-}1)$$

the parameters ϵ and ρ in Algorithm 5.1 can be chosen such that $\rho = \epsilon$ and

$$B - \epsilon A/2 \text{ is strongly monotone on } \mathscr{U}, \quad (27\text{-}1)$$

$$DA^{-1}E^* + EA^{-1}D^* - DA^{-1}(B + \epsilon A/2)A^{-1}D^* \quad (27\text{-}2)$$

is strongly monotone on $\mathscr{C}^*$.

Then, for any given (u^0, p^0), the algorithm generates a unique sequence $\{(u^k, p^k)\}$ which converges to the unique solution (u^*, p^*) of the master problem (21).

Remark 5.1. Assumption (26-1) implies that

$$DA^{-1}E^* + EA^{-1}D^* \text{ is strongly monotone.} \quad (26\text{-}2)$$

One can say that this latter condition is also sufficient, because, if it is met, one can make use of λE instead of E, with λ positive and large enough for (26-1) to be met. Essentially the same effect could have been obtained by taking $\rho \neq \epsilon$, provided that ϵ is replaced by ρ in the last term of (25-1) (this can be justified easily). We shall deal with condition (26-2) later.

The proof given in the Appendix follows closely that given in a previous paper by Cohen and Joalland (Ref. 12). Let us comment on assumption (26-2) in the case when we attempt to use this algorithm for decomposition-coordination purposes. We consider a decomposition of $\mathscr{U}$ as in (10), but we also consider a decomposition of the space $\mathscr{P} = \mathscr{C}^*$ (i.e., eventually $\mathscr{C}$ into $\mathscr{C}_1 \times \cdots \times \mathscr{C}_N$), and we must choose a core Ψ in (24) assuming an additive form with respect to

$$\{(u_i, p_i) \in \mathscr{U}_i \times \mathscr{C}_i^*, i = 1, \ldots, N\}.$$

By doing so, Problem (25) splits into N independent problems [provided that $\mathscr{U}^f$ is also as in (10) and, similarly, the positive cone C is equal to $C_1 \times \cdots \times C_N$, where C_i is a closed convex cone in $\mathscr{C}_i$, which we assume]. Notice that the ith subproblem involves the variable u_i and a constraint space $\mathscr{C}_i$ only (that is, the decomposition of $\mathscr{C}$ is in fact a distribution of the constraints among the subproblems[5]), thus involving a multiplier $p_i \in C_i^*$.

[4] The asterisk denotes the adjoint operator.

[5] In Ref. 6, we further discuss the connection between the decompositions of $\mathscr{U}$ and $\mathscr{C}$ and the fact that the latter can have less than N component subspaces.

Let us consider more specifically the case in Theorem 5.1. Then, for the decomposition to occur, B must be block diagonal with respect to the decomposition of $\mathcal{U}$, and E must be block diagonal[6] with respect to the decompositions of $\mathcal{U}$ and $\mathcal{C}$. This structural constraint of choosing a block diagonal E interacts with condition (26-2) in a way which is not fully studied at the present time. If D itself is *nearly block diagonal* (that is, the off-diagonal terms are *small*), one expects that (26-2) can be met under the structural constraint on E, since, if D were exactly block diagonal, one could choose $E = D$ and meet (26-2). This case is the case of *weak coupling* through the constraints. However, this is not the only situation, and we give hereafter an example of an arbitrarily large coupling through D, which nevertheless allows us to meet (26-2). For this purpose, we use a trick borrowed from Sundareshan (Ref. 13). Assume that D can be written as $E + ESA$, where E is a self-adjoint, surjective operator from $\mathcal{U}$ to $\mathcal{C}$, S is skew symmetric in $\mathcal{U}(S^* = -S)$, and A is the same as in (26-2). Then, by choosing E in (26-2) precisely as E above, and by calculating the expression in (26-2), one finds $2EA^{-1}E^*$, which is coercive, since E is surjective. Notice that S (which generally introduces a coupling through the constraints) can be multiplied by a positive, arbitrarily large number without altering the conclusion.

In Ref. 6, we show how quasilinearization algorithms in optimal control can be imbedded in Algorithm 5.1, as well as the generalized Takahara algorithms (Refs. 2 and 12) in decomposition. Also, a coordination algorithm, previously proposed in Refs. 14 and 15, resting upon the principle of resource allocation (Geoffrion, Ref. 16) can be derived from Algorithm 5.1 (see Ref. 6).

6. Two-Level Algorithms and Price Coordination in Nonseparable Cases

We now make use of Algorithm 4.2 to find the saddle point of (22), with the following choice of the core Ψ:

$$\Psi(u, p) \triangleq K(u) - \tfrac{1}{2}\|p\|^2. \tag{28}$$

However, we introduce the following feature: the second term in the right-hand side of (22) will be considered as nondifferentiable in u and will be dealt with as we did with the part J_1 in Algorithm 2.1. For this purpose, we write a necessary and sufficient condition for the right-hand inequality of

[6] In case $\mathcal{C}$ has less than N component subspaces (see Footnote 5), the concept of a block diagonal E needs some appropriate arrangement, which, however, is straightforward.

(20), with L replacing Φ, in the following form:

$$\text{for all } u \in \mathscr{U}^f, \qquad \langle J'(u^*), u-u^*\rangle + \langle p^*, \Theta(u) - \Theta(u^*)\rangle \geq 0. \tag{29}$$

Eventually, we get the following algorithm.

Algorithm 6.1. In Algorithm 4.2, set $\Phi = L$, $\mathscr{P}^f = C^*$, and replace Steps (ii) by the following steps.

(ii-a) Solve the following auxiliary problem in u:

$$\min_{u \in \mathscr{U}^f} K(u) + \langle \epsilon J'(u^k) - K'(u^k), u\rangle + \epsilon\langle p^k, \Theta(u)\rangle. \tag{30}$$

Let u^{k+1} be a solution.

(ii-b) Update p according to

$$p^{k+1} = P(p^k + \rho\Theta(u^{k+1})), \tag{31}$$

where P is the projection onto C^* [$P = I$, if $C = \{0\}$] and $\rho > 0$.

Notice that (31) gives the explicit solution of the auxiliary problem in p.

Theorem 6.1. We still assume that J (respectively, K) is convex, with a G-derivative strongly monotone with constant a (respectively, b) and Lipschitz with constant A (respectively, B). The mapping Θ is convex [that is, it meets (23)] and Lipschitz with constant . We assume that a saddle point (u^*, p^*) of L exists; then, u^* is unique. Then, for $k > 0$, u^{k+1} exists and is unique; hence, p^{k+1} is well defined. And, if ϵ and ρ meet the following conditions:

$$0 < \epsilon \leq b/A \qquad \text{and} \qquad 0 < \rho < 2a/\tau^2, \tag{32}$$

the sequence $\{u^k\}$ generated by Algorithm 6.1 converges toward u^*. The sequence $\{p^k\}$ remains bounded, and any weak cluster point $\bar{p}$ (at least one exists) is such that $(u^*, \bar{p})$ is a saddle-point of L.

By choosing $K = J$, $\epsilon = 1$ in (30), Algorithm 6.1 turns out to be the Uzawa algorithm. Now, if one chooses

$$K(u) = \tfrac{1}{2}\|u\|^2,$$

if $\mathscr{U}^f = \mathscr{U}$, and if Θ is affine, then the solution of (30) is given by

$$u^{k+1} = u^k - \epsilon L'_u(u^k, p^k);$$

that is, we recover the Arrow–Hurwicz algorithm (see Ref. 7).

Notice that Theorem 6.1 then gives a proof of convergence for this algorithm when J is convex (as far as the author knows, J is generally assumed to be quadratic when dealing with such a proof in the literature, see

Ref. 7). Even more, we have generalized the algorithm to the case $\mathcal{U}^f \neq \mathcal{U}$ and to a large class of cores K.

From the point of view of decomposition, assuming (10), it is well known that the Uzawa algorithm is the basis of the so-called *price coordination principle* (Ref. 3) or *interaction balance principle* (Ref. 5). However, this requires that J and Θ assume additive forms with respect to the decomposition (10). Algorithm 6.1 provides an extension of this method to the case of a nonadditive J. One then chooses an additive K, and (30) splits into N independent problems. However, when Θ is also nonadditive, one must use the following algorithm, assuming Θ to be differentiable.

For the purpose of being able to give a proof of convergence, we shall also modify (31) using the following definitions. Let

$$\mathcal{B}_R = \{p \mid \|p\| \leq R\},$$

where $R > 0$, and let P_R be the projection onto $C^* \cap \mathcal{B}_R$.

Algorithm 6.2. In Algorithm 6.1, replace (30) and (31) by

$$\min_{u \in \mathcal{U}^f} K(u) + \langle \epsilon J'(u^k) - K'(u^k), u \rangle + \epsilon \langle p^k, \Theta'(u^k) \cdot u \rangle, \tag{33}$$

$$p^{k+1} = P_R(p^k + \rho\Theta(u^{k+1})). \tag{34}$$

Remark 6.1. The computation of (34) can be made in two steps. First,

$$q^{k+1} = P(p^k + \rho\Theta(u^{k+1}));$$

and second,

$$p^{k+1} = [\min(1, R/\|q^{k+1}\|)]q^{k+1}.$$

Remark 6.2. If Θ is affine (and thus also additive with respect to a possible decomposition of u), problems (30) and (33) are equivalent. In this case, Algorithm 6.1, which is simpler, will be preferably used.

Theorem 6.2. We consider the assumptions of Theorem 6.1 and these additional ones: Θ', the G-derivative of Θ, exists and meets the following condition: $\exists T \in C$, such that,

$$\text{for all } u, w \in \mathcal{U}^f, \qquad (\Theta'(u) - \Theta'(w)) \cdot (u - w) - \|u - w\|^2 T \in -C. \tag{35}$$

If R is large enough for some optimal p^* to belong to $\mathcal{B}_R$ and if

$$0 < \epsilon \leq b/(A + R\|T\|) \qquad \text{and} \qquad 0 < \rho < 2a/\tau^2, \tag{36}$$

then Algorithm 6.2 generates a well-defined sequence $\{u^k, p^k\}$, for which the conclusions of Theorem 6.1 hold true.

Remark 6.3. The property (35) can be derived from a more classical one when $\mathscr{C}$ is finite dimensional (say, $\mathbb{R}^r$) and when C is the classical positive cone of this space. Then, let Θ_i map $\mathscr{U}$ into the ith component of $\mathbb{R}^r$, if Θ_i' is Lipschitz with constant T_i; then, $T^t = (T_1, \ldots, T_r)$ meets (35).

Remark 6.4. There is some difficulty in using Theorem 6.2, since one would attempt to take a very large R in order to have a reasonable chance that some p^* (unknown) does belong to $\mathscr{B}_R$; but this forces one to choose a very small ϵ, according to (36), and this is intuitively not favorable for the convergence speed. However, one wonders to what extent these sufficient conditions are also necessary and, moreover, whether the sequence $\{p^k\}$, if generated by (31), should not remain bounded without using the additional projection on $\mathscr{B}_R$ as in (34).

As already discussed, Algorithm 6.2 and Theorem 6.2 provide an extension of the Arrow–Hurwicz algorithm when $\mathscr{U}^f$ is not $\mathscr{U}$, J is not quadratic, Θ is not affine, and $K(u)$ is not $\frac{1}{2}\|u\|^2$. From the decomposition viewpoint, it provides an extension of price coordination to the case when both J and Θ are not separable.

7. Discussion and Conclusions

Motivated by the decomposition-coordination algorithms, we have been led to the auxiliary problem principle, which provides a unified approach to the plentiful literature devoted to the topic in the last fifteen years (and especially after 1970). This has been more extensively discussed in Ref. 6. However, it is important to notice that the relevance of the auxiliary problem principle to this field stems from the introduction of an additional feature, namely, the choice of additive cores (with respect to some decompositions of the spaces involved). Without this feature, the auxiliary problem principle nevertheless has its own interest, since it has allowed us to recover and to extend some well-known algorithms, such as projected gradient, quasilinearization, Uzawa, Arrow–Hurwicz, and so on. Generally speaking, it allows one to replace the master problem by a sequence of auxiliary problems and, taking this opportunity, to give particular desirable features to the latter problems [well conditioning, see Gabay–Mercier (Ref. 17); decomposability, etc.]. It may also provide a framework for studying discretization or finite-element methods, although this task has not yet been tackled.

Hence, classical and decomposition algorithms are imbedded in the same theory, which thus provides a link between the former and the latter types of algorithms. Also, the approach by basic algorithms not only allows one to give general proofs of convergence for a large class of algorithms dealt with simultaneously, but it gives systematic guidelines for designing new algorithms. Notice, for example, that we have used a very narrow part of the potentiality of Algorithms 4.1 and 4.2 when applied to (22), since we have limited ourselves to the rather particular cores, respectively, (24) and (28). Even in this restricted class, we outlined in Ref. 6 how to mix Algorithms 5.1 and 6.1 (or 6.2) in one single algorithm to obtain a new algorithm dealing with two sets of constraints differently. Since the possibilities of choices for cores and of combinations of the above basic algorithms are almost unlimited, one must wait for motivations arising from particular applications before going further in these directions.

As far as decomposition-coordination is concerned, we shall not discuss their advantages in detail in this paper (see Ref. 6). Let us just say that *sequential decomposition* (that is, relaxed algorithms) seems to be useful for solving problems of a size larger than that one can deal with using the largest computer facilities available at the present time [especially in dynamic programming methods; see Joalland–Cohen (Ref. 18), for example]. Notice that we could also have discussed relaxed versions of Algorithms 5.1, 6.1, 6.2, as we did for Algorithm 2.1. *Parallel decomposition* may be a way of achieving fast computation in online situations (arising, for example, in process control), provided that ad hoc tools be used (for example, multi-microprocessors). We lack experience in this field.

However, the main advantage of the auxiliary problem principle seems to be that it makes clear what makes decomposition possible in nonseparable situations. As a matter of fact, when looking at (8), we noticed that, if K is chosen as an additive functional with respect to (10), then the functional G^v is also additive. This is because the coupling introduced by J is removed by taking a linear approximation of this functional. But this is possible, because we have assumed that J is Gateaux differentiable. Even subdifferentiability (and so existence of directional derivatives, but no linearity of these derivatives with respect to the directions) would not have sufficed. One can see this by several ways. Recall that, for a subdifferentiable J, the derivative of J at v in the direction u can be obtained by

$$DJ(v; u) = \max_{g \in \partial J(v)} \langle g, u \rangle.$$

If this expression were to replace $\langle J'(v), u \rangle$ in (8), it would not be separable with respect to a decomposition of u as in (10), and the auxiliary problem would no longer split into independent subproblems.

Another way for looking at the same phenomenon is to recall that for a non-G-differentiable functional being stationary at a point in the directions of component subspaces [as given in (10)] does not imply that the functional is indeed stationary in all directions. For example, let us consider the following functional on $\mathbb{R}^2$:

$$J:(u_1, u_2)\mapsto \max(u_1^2+(u_2-1)^2, u_2^2+(u_1-1)^2). \tag{37}$$

This functional is convex and G-differentiable, except on the subspace $u_1 = u_2$. If one considers the restricted mappings

$$u_1 \mapsto J(u_1, 0) \qquad \text{and} \qquad u_2 \mapsto J(0, u_2),$$

they are minimal at $u_1 = 0$ and $u_2 = 0$, respectively. Hence, with an algorithm which proceeds by minimizations in u_1 and u_2 separately, starting at (0, 0) one would stop immediately. To be able to go down further, one must go in a direction $(u_1 > 0, u_2 > 0)$, but this requires one to move u_1 and u_2 in a coordinated way, while preserving some decentralization of computations. It would perhaps require some elaboration of the auxiliary problem principle, which is not clear at the present time.

Without this extension, one cannot decompose problems involving non-G-differentiable functionals, except if they are already separable, as supposed for J_1 in Section 2. However, in some situations, the difficulty can be bypassed, if not overcome. Consider for example J_1 being the indicator function of a feasible set, which would be nondecomposable as in (10); hence, J_1 would be nonadditive. By taking an explicit representation of this constraint, and using duality theory, we have gotten around this obstacle in Sections 5 and 6. We are currently investigating the same kind of ideas in order to apply our approach to minimization of functionals defined by a maximization operation as in (37), which inevitably gives rise to nondifferentiability.

8. Appendix: Proofs

8.1. Proof of Theorem 2.1. Assumptions (iii) and (iv) ensure that Assumption (A) holds for Problem (AP^k), giving the existence of u^{k+1}. The uniqueness is a consequence of K^k being coercive.

We consider the difference

$$\begin{aligned}\Delta_{k+1}^k &\triangleq (J+J_1)(u^k)-(J+J_1)(u^{k+1})\\ &\geq \langle J'(u^k), u^k-u^{k+1}\rangle - \tfrac{1}{2}A\|u^k-u^{k+1}\|^2 + J_1(u^k)-J_1(u^{k+1}),\end{aligned} \tag{38}$$

from (4). The solution u^{k+1} of Problem (AP^k) meets the following condition:

$$\begin{gathered} u^{k+1}\in \mathcal{U}^f \qquad \text{and, for all } u\in\mathcal{U}^f,\\ \langle (K^k)'(u^{k+1})-(K^k)'(u^k)+\epsilon^k J'(u^k), u-u^{k+1}\rangle\\ +\epsilon^k(J_1(u)-J_1(u^{k+1}))\geq 0. \end{gathered} \tag{39}$$

Using (39) with $u = u^k$, which belongs to $\mathcal{U}^f$, and placing it in (38) yield

$$\Delta_{k+1}^k \geq 1/\epsilon^k \langle (K^k)'(u^{k+1}) - (K^k)'(u^k), u^{k+1} - u^k \rangle - \tfrac{1}{2} A \|u^k - u^{k+1}\|^2.$$

By the strong monotony of $(K^k)'$, we have

$$\Delta_{k+1}^k \geq (b^k/\epsilon^k - A/2)\|u^{k+1} - u^k\|^2 \geq (\beta/2)\|u^{k+1} - u^k\|^2,$$

the latter from (12). This proves that the sequence $\{(J + J_1)(u^k)\}$ is strictly decreasing unless $u^k = u^{k+1}$, but then $u^k = u^*$ from Lemma 2.2. Moreover, that sequence is bounded from below by $\{(J + J_1)(u^*)\}$, so that it converges to some μ. Hence, $\Delta_{k+1}^k \to 0$, when $k \to +\infty$; and, from the last inequality,

$$\|u^k - u^{k+1}\| \to 0.$$

We have that

$$(J + J_1)(u^{k+1}) \geq \mu \geq (J + J_1)(u^*). \tag{40}$$

We prove that indeed equality holds in the second inequality. As a matter of fact, we have

$$\begin{aligned} \Delta_{k+1}^* &\triangleq (J + J_1)(u^*) - (J + J_1)(u^{k+1}) \\ &\geq \langle J'(u^{k+1}), u^* - u^{k+1} \rangle + J_1(u^*) - J_1(u^{k+1}), \end{aligned}$$

from the convexity of J. Then, from (39), with $u = u^*$, we get

$$\begin{aligned} \Delta_{k+1}^* \geq {} & \langle J'(u^{k+1}) - J'(u^k), u^* - u^{k+1} \rangle \\ & + (1/\epsilon^k)\langle (K^k)'(u^k) - (K^k)'(u^{k+1}), u^* - u^{k+1} \rangle. \end{aligned}$$

The right-hand side of this inequality is proved to converge to zero by taking its absolute value, by making use of the Schwarz inequality and the Lipschitz continuity of J' and $(K^k)'$, and by remembering that $\|u^k - u^{k+1}\| \to 0$ and that $\{u^k\}$ is a bounded sequence. This last fact comes from Assumption (A), together with the fact that $(J + J_1)(u^k)$ converges to a finite limit. Hence, we have that $\lim \Delta_{k+1}^* \geq 0$; but the reverse inequality also holds true from (40), which proves that

$$\mu = (J + J_1)(u^*). \tag{41}$$

Since the sequence $\{u^k\}$ is bounded, it has at least one cluster point in the weak topology. Let $\bar{u}$ denote such a point, and let $\{u^{k'}\}$ denote a subsequence weakly converging to $\bar{u}$. Notice that $\bar{u} \in \mathcal{U}^f$, since $\mathcal{U}^f$ is convex and thus weakly closed. Since $J + J_1$ is convex and lower semicontinuous in the strong topology, it is also lower semicontinuous in the weak topology. Therefore, we have that

$$\lim_{k' \to \infty} (J + J_1)(u^{k'}) = \mu \geq (J + J_1)(\bar{u});$$

but equality must hold from (41). This proves that $\bar{u}$ is a solution u^* of Problem (MP).

We now prove the last statements of the theorem by making use of Assumption (v), which implies that u^* is unique. Adding (39) with $u = u^*$ and (7) multiplied by ϵ^k with $u = u^{k+1}$ yields

$$\langle (K^k)'(u^{k+1}) - (K^k)'(u^k) + \epsilon^k (J'(u^k) - J'(u^*)), u^* - u^{k+1} \rangle \geq 0.$$

We rewrite it as follows:

$$\langle (K^k)'(u^{k+1}) - (K^k)'(u^k) + \epsilon^k (J'(u^k) - J'(u^{k+1})), u^* - u^{k+1} \rangle$$
$$\geq \epsilon^k \langle J'(u^{k+1}) - J'(u^*), u^{k+1} - u^* \rangle \geq \epsilon^k a \|u^{k+1} - u^*\|^2.$$

the last inequality following from (1). Making use of the Schwarz inequality in the first member and then the Lipschitz inequalities on $(K^k)'$ and J', we obtain Ineq. (13) after division by $\|u^{k+1} - u^*\|$, which we assume nonnull (otherwise, the result is trivial). This inequality proves that

$$\|u^{k+1} - u^*\| \to 0,$$

since $\|u^{k+1} - u^k\|$ does so, and B^k/ϵ^k is bounded from above by B/α. □

8.2. Proof of Theorem 2.2. We only give the beginning of the proof, since the rest follows as previously. We again consider Ineq. (38) and substitute the right-hand side of (14) for u^{k+1}. This yields

$$\Delta_{k+1}^k \geq \rho^k \langle J'(u^k), u^k - \hat{u}^{k+1} \rangle - (\rho^k)^2 (A/2) \|u^k - \hat{u}^{k+1}\|^2$$
$$+ J_1(u^k) - J_1(\rho^k \hat{u}^{k+1} + (1-\rho^k) u^k).$$

Inequality (39) is replaced by

$$\hat{u}^{k+1} \in \mathcal{U}^f \quad \text{and} \quad \forall u \in \mathcal{U}^f,$$
$$\langle (K^k)'(\hat{u}^{k+1}) - (K^k)'(u^k) + J'(u^k), u - \hat{u}^{k+1} \rangle + J_1(u) - J_1(\hat{u}^{k+1}) \geq 0.$$

We use it with $u = u^k$ in the previous one. This yields

$$\Delta_{k+1}^k \geq \rho^k \langle (K^k)'(u^k) - (K^k)'(\hat{u}^{k+1}), u^k - \hat{u}^{k+1} \rangle - (\rho^k)^2 (A/2) \|u^k - \hat{u}^{k+1}\|^2$$
$$+ (1-\rho^k) J_1(u^k) + \rho^k J_1(\hat{u}^{k+1}) - J_1(\rho^k \hat{u}^{k+1} + (1-\rho^k) u^k).$$

If $J_1 \not\equiv 0$, by taking $\rho^k \leq 1$, we use the convexity of J_1 and the strong monotony of $(K^k)'$ to obtain

$$\Delta_{k+1}^k \geq \rho^k (b^k - \rho^k (A/2)) \|u^k - \hat{u}^{k+1}\|^2 \geq \alpha^2 (\beta/2) \|u^k - \hat{u}^{k+1}\|^2.$$

The latter inequality follows from (12) applied to ρ^k. Notice that we must not a priori restrict ρ^k to be not larger than 1 if $J_1 \equiv 0$. However, we must ensure that $u^k \in \mathcal{U}^f$ for all k, since we have implicitly used this assumption. □

8.3. Proof of Theorem 3.1. The solutions (u^*, v^*), u^{k+1}, v^{k+1} of, respectively, Problems (MP'), (AP_u^k), (AP_v^k) are feasible points characterized by the following relations: for all $u \in \mathcal{U}^f$, $v \in \mathcal{V}^f$,

$$\langle J'_u(u^*, v^*), u - u^*\rangle + \langle J'_v(u^*, v^*), v - v^*\rangle \geq 0, \tag{42}$$

$$\langle K'_u(u^{k+1}, v^k) - K'_u(u^k, v^k) + \epsilon_1 J'_u(u^k, v^k), u - u^{k+1}\rangle \geq 0, \tag{43}$$

$$\langle K'_v(u^{k+1}, v^{k+1}) - K'_v(u^{k+1}, v^k) + \epsilon_2 J'_v(u^{k+1}, v^k), v - v^{k+1}\rangle \geq 0. \tag{44}$$

We now consider

$$\Delta_{k+1}^k \triangleq [J(u^k, v^k) - J(u^{k+1}, v^k)] + [J(u^{k+1}, v^k) - J(u^{k+1}, v^{k+1})].$$

We now proceed as in (38) for the two brackets above (respective constants A_1 and A_2) and then we use (43) with $u = u^k$ and (44) with $v = v^k$. At last, we make use of the strong monotony assumptions on the restricted mappings derived from K (constants b_1 and b_2). We obtain

$$\Delta_{k+1}^k \geq (b_1/\epsilon_1 - A_1/2)\|u^{k+1} - u^k\|^2 + (b_2/\epsilon_2 - A_2/2)\|v^{k+1} - v^k\|^2,$$

which proves that $J(u^k, v^k)$ decreases and that

$$\|u^{k+1} - u^k\| \to 0 \qquad \text{and} \qquad \|v^{k+1} - v^k\| \to 0,$$

as in the proof of Theorem 2.1 and with the help of (18).

Next, we add (42) with

$$(u, v) = (u^{k+1}, v^{k+1}),$$

(43) with $u = u^*$ divided by ϵ_1, and (44) with $v = v^*$ divided by ϵ_2. This yields

$$\begin{aligned}
&(1/\epsilon_1)\langle K'_u(u^{k+1}, v^k) - K'_u(u^k, v^k), u^* - u^{k+1}\rangle \\
&+ (1/\epsilon_2)\langle K'_v(u^{k+1}, v^{k+1}) - K'_v(u^{k+1}, v^k), v^* - v^{k+1}\rangle \\
&+ \langle [J'_u(u^k, v^k) - J'_u(u^{k+1}, v^k)] \\
&+ [J'_u(u^{k+1}, v^k) - J'_u(u^{k+1}, v^{k+1})], u^* - u^{k+1}\rangle \\
&+ \langle J'_u(u^{k+1}, v^{k+1}) - J'_u(u^*, v^*), u^* - u^{k+1}\rangle \\
&+ \langle [J'_v(u^{k+1}, v^k) - J'_v(u^{k+1}, v^{k+1})] \\
&+ [J'_v(u^{k+1}, v^{k+1}) - J'_v(u^*, v^*)], v^* - v^{k+1}\rangle \geq 0.
\end{aligned}$$

We make repeated uses of the Schwarz inequality, the Lipschitz assumptions, and the global strong monotony assumption on the derivatives of J, and we get

$$\begin{aligned} a(\|u^{k+1}-u^*\|^2+\|v^{k+1}-v^*\|^2) &\le [(B_1/\epsilon_1+A_1)\|u^k-u^{k+1}\|+A_{12}\|v^k-v^{k+1}\|]\,\|u^*-u^{k+1}\| \\ &\quad +(B_2/\epsilon_2+A_2)\|v^k-v^{k+1}\|\,\|v^*-v^{k+1}\| \\ &\le (\|u^{k+1}-u^*\|^2+\|v^{k+1}-v^*\|^2)^{1/2} \\ &\quad \times([(B_1/\epsilon_1+A_1)\|u^k-u^{k+1}\|+A_{12}\|v^k-v^{k+1}\|]^2 \\ &\quad +[(B_2/\epsilon_2+A_2)\|v^k-v^{k+1}\|]^2)^{1/2}. \end{aligned}$$

The latter inequality results from the Hölder inequality. We divide by

$$(\|u^{k+1}-u^*\|^2+\|v^{k+1}-v^*\|^2)^{1/2},$$

which we assume nonnull (otherwise, the result is trivial), and take the square. This yields

$$\begin{aligned} a^2(\|u^{k+1}-u^*\|^2+\|v^{k+1}-v^*\|^2) &\le [(B_1/\epsilon_1+A_1)\|u^k-u^{k+1}\|+A_{12}\|v^k-v^{k+1}\|]^2 \\ &\quad +(B_2/\epsilon_2+A_2)^2\|v^k-v^{k+1}\|^2. \end{aligned}$$

We now easily complete the proof. □

8.4. Proof of Theorem 5.1. Necessary and sufficient conditions for problem (25) are

$$B(u^{k+1}-u^k)+\epsilon A(u^k-f)+E^*(p^{k+1}-p^k)+\epsilon D^*p^k=0, \quad (45\text{-}1)$$

$$E(u^{k+1}-u^k)+\epsilon D(u^k-d)=0; \quad (45\text{-}2)$$

and, in the same way, for the equality-constrained problem (21),

$$A(u^*-f)+D^*p^*=0, \quad (46\text{-}1)$$

$$D(u^*-d)=0. \quad (46\text{-}2)$$

Since A and B are strongly monotone, they have continuous inverse operators. Moreover, since D and E are surjective, $DA^{-1}D^*$ and $EB^{-1}E^*$ are also strongly monotone. From these facts, one can write explicit expressions for the inverse operators of the composite operators

$$\begin{bmatrix} B & E^* \\ E & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & D^* \\ D & 0 \end{bmatrix}, \quad (47)$$

which are involved, respectively, in (45) and (46). This proves that these equations have unique solutions, respectively, (u^{k+1}, p^{k+1}) and (u^*, p^*).

Moreover, it is easy to check that, if (26) is met, then one can choose ϵ small enough but positive, so that (27) is also met. We consider the following positive functional:

$$F(u,p) \triangleq \tfrac{1}{2}\langle p-p^*, DA^{-1}D^*(p-p^*)\rangle$$
$$+\tfrac{1}{2}\langle A(u-u^*)+D^*(p-p^*), A^{-1}[A(u-u^*)+D^*(p-p^*)]\rangle$$

and the difference

$$\Delta_{k+1}^k \triangleq F(u^k,p^k)-F(u^{k+1},p^{k+1}).$$

After some calculations, which make use of (45), one gets

$$\epsilon\Delta_{k+1}^k = \tfrac{1}{2}\langle u^k-u^{k+1}, (2B-\epsilon A)(u^k-u^{k+1})\rangle$$
$$+\langle u^k-u^{k+1}, (2BA^{-1}D^*-\epsilon D^*)(p^k-p^{k+1})\rangle$$
$$+\langle p^k-p^{k+1}, (DA^{-1}E^*+EA^{-1}D^*-\epsilon DA^{-1}D^*)(p^k-p^{k+1})\rangle,$$

which we may also write as

$$\epsilon\Delta_{k+1}^k = G(u^k-u^{k+1}, p^k-p^{k+1}), \tag{48}$$

with an obvious definition of the quadratic functional G, with $G(0,0)=0$. We notice that conditions (27) ensure that G is strongly monotone. Hence, (48) proves that the sequence $\{F(u^k,p^k)\}$ is decreasing and bounded from below by zero. Thus, it converges and $\Delta_{k+1}^k \to 0$. From the coercivity of G, we have that

$$(u^k-u^{k+1}, p^k-p^{k+1}) \to (0,0).$$

We now subtract (46) multiplied by ϵ from (45), and we pass to the limit, which proves that

$$(u^k-u^*, p^k-p^*) \to (0,0);$$

here, we use the continuity of the inverse of the second operator in (47). □

8.5. Proof of Theorem 6.1. From the assumptions on K and Θ, Assumption (A) holds for the functional in (30). Hence, u^{k+1} exists and is unique from the strong monotony of K. The left-hand side inequality in (20) (for $\Phi = L$) is equivalent to

$$\text{for all } \rho \geq 0, \qquad p^* = P(p^* + \rho\Theta(u^*)). \tag{49}$$

Since the projection is a nonexpansive mapping, from (49) and (31) we get

$$\|p^{k+1}-p^*\| \leq \|p^k - p^* + \rho(\Theta(u^{k+1})-\Theta(u^*))\|.$$

By taking the square in both sides and using the Lipschitz assumption on Θ

we get

$$2\langle p^k - p^*, \Theta(u^*) - \Theta(u^{k+1})\rangle \leq (1/\rho)(\|p^k - p^*\|^2 - \|p^{k+1} - p^*\|^2) + \rho\tau^2\|u^{k+1} - u^*\|^2. \tag{50}$$

A necessary and sufficient condition for $u^{k+1} \in \mathcal{U}^f$ to be a solution of problem (30) is that

$$\forall u \in \mathcal{U}^f,\ \langle K'(u^{k+1}) - K'(u^k) + \epsilon J'(u^k), u - u^{k+1}\rangle + \epsilon\langle p^k, \Theta(u) - \Theta(u^{k+1})\rangle \geq 0. \tag{51}$$

Let us add (51) with $u = u^*$ to (29) multiplied by ϵ with $u = u^{k+1}$. We get

$$\langle K'(u^{k+1}) - K'(u^k), u^* - u^{k+1}\rangle + \epsilon\langle J'(u^k) - J'(u^*), u^* - u^{k+1}\rangle + \epsilon\langle p^k - p^*, \Theta(u^*) - \Theta(u^{k+1})\rangle \geq 0. \tag{52}$$

The first term in (52) yields

$$\begin{aligned}&\langle K'(u^{k+1}), u^* - u^{k+1}\rangle - \langle K'(u^k), u^* - u^k\rangle - \langle K'(u^k), u^k - u^{k+1}\rangle \\ &\leq \langle K'(u^{k+1}), u^* - u^{k+1}\rangle - \langle K'(u^k), u^* - u^k\rangle \\ &\quad + K(u^{k+1}) - K(u^k) - (b/2)\|u^k - u^{k+1}\|^2,\end{aligned}$$

from Ineq. (3) used for K. The second term in (52) yields

$$\begin{aligned}&\epsilon\langle J'(u^k), u^k - u^{k+1}\rangle + \epsilon\langle J'(u^k), u^* - u^k\rangle + \epsilon\langle J'(u^*), u^{k+1} - u^*\rangle \\ &\leq \epsilon(A/2)\|u^k - u^{k+1}\|^2 - \epsilon(a/2)(\|u^k - u^*\|^2 + \|u^{k+1} - u^*\|^2),\end{aligned}$$

by repeated applications of Lemma 2.1. Finally, for the third term in (52), we make use of (50); eventually, we get from (52) the following:

$$\begin{aligned}&K(u^{k+1}) + \langle K'(u^{k+1}), u^* - u^{k+1}\rangle - K(u^k) - \langle K'(u^k), u^* - u^k\rangle \\ &\quad + \tfrac{1}{2}(\epsilon A - b)\|u^k - u^{k+1}\|^2 + (\epsilon/2)a(\|u^{k+1} - u^*\|^2 - \|u^k - u^*\|^2) \\ &\quad + \epsilon(\rho(\tau^2/2) - a)\|u^{k+1} - u^*\|^2 + (\epsilon/2\rho)(\|p^k - p^*\|^2 - \|p^{k+1} - p^*\|^2) \geq 0,\end{aligned} \tag{53}$$

which, by defining

$$F(u, p) \triangleq (\epsilon/2\rho)\|p - p^*\|^2 + K(u^*) - K(u) - \langle K'(u), u^* - u\rangle - \epsilon(a/2)\|u - u^*\|^2,$$

can be written as

$$F(u^k, p^k) - F(u^{k+1}, p^{k+1}) \geq \epsilon(a - \rho(\tau^2/2))\|u^{k+1} - u^*\|^2 + \tfrac{1}{2}(b - \epsilon A)\|u^k - u^{k+1}\|^2. \tag{54}$$

From that and (32), we obtain that the sequence $\{F(u^k, p^k)\}$ is decreasing. Moreover, it is bounded from below, since we have

$$F(u, p) \geq (\epsilon/2\rho)\|p - p^*\|^2 + \tfrac{1}{2}(b - \epsilon a)\|u - u^*\|^2 \geq 0,$$

the first inequality coming from (3) applied to K and the second because $\epsilon \leq b/A \leq b/a$. Thus, the sequence $\{F(u^k, p^k)\}$ converges, and the left-hand side of (54) goes to zero when $k \to +\infty$. Looking at the positive right-hand side, and recalling (32), we see that $\|u^{k+1} - u^*\| \to 0$. This proves the first assertion of the theorem.

Notice that, since $F(u^k, p^k)$ and u^k have a limit when $k \to +\infty$, so does $\|p^k - p^*\|^2$. With this remark, we can then proceed as Bensoussan, Lions, and Temam (Ref. 19) to prove the last assertion. We first prove that $\langle p^k, \Theta(u^k)\rangle \to 0$.

Since $\|p^k - p^*\|^2$ has a limit, the sequence $\{p^k\}$ is bounded and has weak cluster points. Let $\bar{p}$ be one of them, and let $\{p^{k'}\}$ be a subsequence converging toward $\bar{p}$. Since C^* is closed and convex, hence weakly closed, $\bar{p} \in C^*$. From the previous results,

$$\langle p^{k'}, \Theta(u^{k'})\rangle \to \langle \bar{p}, \Theta(u^*)\rangle. \tag{55}$$

Equality (49) is equivalent to

$$\langle p^*, \Theta(u^*)\rangle = 0, \qquad \text{for all } p \in C^*, \qquad \langle p, \Theta(u^*)\rangle \leq 0. \tag{56}$$

Hence,

$$\langle \bar{p}, \Theta(u^*)\rangle \leq 0.$$

The reverse inequality will now be established. Equality (31) is equivalent to

$$p^{k+1} \in C^*, \qquad \text{for all } p \in C^*, \quad \langle p - p^{k+1}, p^k + \rho\Theta(u^{k+1}) - p^{k+1}\rangle \leq 0,$$

which we use with $p = p^*$, yielding

$$\begin{aligned} \rho\langle p^{k+1} - p^*, \Theta(u^{k+1})\rangle &\geq \langle p^{k+1} - p^*, (p^{k+1} - p^*) - (p^k - p^*)\rangle \\ &\geq \tfrac{1}{2}(\|p^{k+1} - p^*\|^2 - \|p^k - p^*\|^2). \end{aligned}$$

Consider the above inequality for indices k', and recall that $\|p^{k'} - p^*\|^2$ has a limit. The right-hand side goes to zero as $k \to +\infty$; thus, recalling (55),

$$\langle \bar{p}, \Theta(u^*)\rangle - \langle p^*, \Theta(u^*)\rangle \geq 0.$$

Using the first equality (56), we have established the reverse inequality that we were looking for. Thus,

$$\lim \langle p^{k'}, \Theta(u^{k'})\rangle = 0$$

for any subsequence, and so this holds for the whole sequence.

Since

$$\langle \bar{p}, \Theta(u^*)\rangle = 0,$$

from the inequality in (56) we have that

$$L(u^*, p) \leq L(u^*, \bar{p}), \qquad \text{for all } p \in C^*.$$

As far as the other inequality of the saddle point is concerned, we obtain an equivalent variational inequality by taking the limit in (51). Since this holds for any weak cluster point $\bar{p}$, the proof is now completed. □

8.6. Proof of Theorem 6.2. We just outline the differences with respect to the previous proof. Since we have assumed that there exists a solution of (49) which belongs to $\mathscr{B}_R$, then this p^* is also a solution of

$$p^* = P_R(p^* + \rho\Theta(u^*))$$

(see Remark 6.1). Using this and (34), as we did previously with (49) and (31), we can derive (50) once again. The other difference between Algorithms 6.1 and 6.2 lies in $\Theta'(u^k) \cdot u$ [see (33)] replacing $\Theta(u)$ [see (30)]. This causes the last term in (51) to be, with $u = u^*$,

$$\epsilon\langle p^k, \Theta'(u^k) \cdot (u^* - u^{k+1})\rangle.$$

Since the functional $u \mapsto \langle p^k, \Theta(u)\rangle$ is convex, we have

$$\langle p^k, \Theta'(u^k) \cdot (u^* - u^k)\rangle \leq \langle p^k, \Theta(u^*) - \Theta(u^k)\rangle.$$

On the other hand, using (35), one can prove an inequality similar to (4) for the functional above, with the constant A replaced by

$$\langle p^k, T\rangle \leq \|p^k\|\|T\| \leq R\|T\|.$$

Therefore,

$$\langle p^k, \Theta'(u^k) \cdot (u^k - u^{k+1})\rangle \leq \langle p^k, \Theta(u^k) - \Theta(u^{k+1})\rangle + (R/2)\|T\|\|u^k - u^{k+1}\|^2.$$

Adding the above two inequalities, we get

$$\langle p^k, \Theta'(u^k) \cdot (u^* - u^{k+1})\rangle \leq \langle p^k, \Theta(u^*) - \Theta(u^{k+1})\rangle + (R/2)\|T\|\|u^k - u^{k+1}\|^2.$$

This eventually shows that the previous proof can be carried on starting from (51), with $u = u^*$, with the only difference that we must add the new term

$$\epsilon R\|T\|\|u^k - u^{k+1}\|^2/2$$

in the left-hand side of this inequality. This modifies the last term in (54), which is now

$$(b - \epsilon(A + R\|T\|))\|u^k - u^{k+1}\|^2/2.$$

The first condition in (36) allows one to complete the proof as previously. □

References

1. ARROW, K. J., and HURWICZ, L., *Decentralization and Computation in Resource Allocation*, Essays in Economics and Econometrics, Edited by R. W. Pfouts, University of North Carolina Press, Rayleigh, North Carolina, 1960.
2. TAKAHARA, Y., *Multilevel Approach to Dynamic Optimization*, Systems Research Center, Case Western Reserve University, Report No. SRC-50-C-64-18, 1964.
3. LASDON, L. S., and SCHOEFFLER, J. D., *A Multilevel Technique for Optimization*, Proceedings of the Joint Automatic Control Conference, Troy, New York, 1965.
4. BROSILOW, C. B., LASDON, L. S., and PEARSON, J. D., *Feasible Optimization Methods for Interconnected Systems*, Proceedings of the Joint Automatic Control Conference, Troy, New York, 1965.
5. MESAROVIC, M. D., MACKO, D., and TAKAHARA, Y., *Theory of Hierarchical Multilevel Systems*, Academic Press, New York, New York, 1970.
6. COHEN, G., *Optimization by Decomposition and Coordination: a Unified Approach*, IEEE Transactions on Automatic Control, Special Issue on Large-Scale Systems, Vol. AC-23, No. 2, 1978.
7. EKELAND, I., and TEMAM, R., *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, Holland, 1970.
8. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.
9. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
10. LUENBERGER, D. G., *Optimization by Vector Space Methods*, John Wiley and Sons, New York, New York, 1969.
11. BERTSEKAS, D. P., *Multiplier Methods: A Survey*, Automatica, Vol. 12, pp. 133–146, 1976.
12. COHEN, G., and JOALLAND, G., *Coordination Methods by the Prediction Principle in Large Dynamic Constrained Optimization Problems*, Proceedings of the IFAC Symposium on Large-Scale Systems, Edited by G. Guardabassi and A. Locatelli, Udine, Italy, 1976.
13. SUNDARESHAN, M. K., *Generation of Multilevel Control and Estimation Schemes for Large-Scale Systems: A Perturbational Approach*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-7, No. 3, 1977.
14. COHEN, G., *Coordination of Constrained Optimization Problems by Resource Allocation*, Ecole des Mines, Centre d'Automatique-Informatique, Internal Report No. A/49, 1972.
15. BERNHARD, P., *Commande Optimale, Décentralisation, et Jeux Dynamiques*, Dunod, Paris, France, 1975.
16. GEOFFRION, A. M., *Primal Resource Directive Approach for Optimizing Nonlinear Decomposable Systems*, Operations Research, Vol. 18, No. 3, 1970.
17. GABAY, D., and MERCIER, B., *A Dual Algorithm for the Solution of a Nonlinear Variational Problem via Finite-Element Approximation*, Computers and Mathematics with Applications, Vol. 2, pp. 17–40, 1976.

18. JOALLAND, G., and COHEN, G., *Optimal Control of a Water Distribution Network by Two Multilevel Methods*, Proceedings of the 7th IFAC World Congress, Helsinki, Finland, Edited by A. Niemi, Pergamon Press, London, England, 1978.
19. BENSOUSSAN, A., LIONS, J. L., and TEMAM, R., *Sur les Méthodes de Décomposition, de Décentralisation, et de Coordination et Applications*, Méthodes Numériques d'Analyse de Systèmes, Vol. 2, Cahier IRIA No. 11, 1972.