Nested Monotony for Variational Inequalities over Product of Spaces and Convergence of Iterative Algorithms[1]

G. COHEN[2] AND F. CHAPLAIS[3]

Communicated by D. Q. Mayne

Abstract. The auxiliary problem principle has been proposed by the first author as a framework to describe and analyze iterative algorithms such as gradient as well as decomposition/coordination algorithms for optimization problems (Refs. 1–3) and variational inequalities (Ref. 4). The key assumption to prove the global and strong convergence of such algorithms, as well as of most of the other algorithms proposed in the literature, is the strong monotony of the operator involved in the variational inequalities. In this paper, we consider variational inequalities defined over a product of spaces and we introduce a new property of strong nested monotony, which is weaker than the ordinary overall strong monotony generally assumed. In some sense, this new concept seems to be a minimal requirement to insure convergence of the algorithms alluded to above. A convergence theorem based on this weaker assumption is given. Application of this result to the computation of Nash equilibria can be found in another paper (Ref. 5).

Key Words. Monotony, convergence of algorithms, variational inequalities, optimization problems, decomposition/coordination algorithms.

1. Introduction

For classical optimization problems, Cohen (Refs. 1 and 2) and Cohen and Zhu (Ref. 3) introduced the so-called auxiliary problem principle as a general framework to describe and analyze computational algorithms ranging from ordinary gradient to decomposition/coordination algorithms. This principle has been further extended to deal with more general (nonsymmetric) variational inequalities (Ref. 4). Generally speaking, to prove the global convergence of such algorithms, one needs to assume that the operator involved in the variational inequality (be it considered directly or derived from the optimization problem under consideration) is monotone (strong monotony is even needed to prove strong convergence in infinite-dimensional Hilbert spaces).

One noticeable exception is the case of saddle-point problems. In this case, the classical assumption is that the corresponding saddle function, which is defined over the product of two spaces, is convex-concave. This is equivalent to the fact that its partial derivative with respect to the first argument (or more generally its subdifferential) is monotone (or monotone positive) when the first argument only is allowed to vary, and the derivative with respect to the second argument is monotone negative (that is, its opposite is monotone) when this latter argument varies. For twice-differentiable saddle functions, the two blocks on the diagonal of the Hessian are respectively positive and negative. Of course, the off-diagonal blocks are adjoint to each other, which will not be the case for nonsymmetric variational inequalities. As we shall see later on, this case of saddle-point problems appears to be a very specific case where the magnitude of the off-diagonal terms can be, in some sense, arbitrarily large and where the (positive or negative) monotony of the block-diagonal terms is sufficient to prove convergence of algorithms such as those of Uzawa, Arrow-Hurwicz or other more general algorithms (Ref. 2).

For the other situations, but still for operators defined over a product of spaces, it is easy to convince oneself that the monotony of the block operators on the diagonal is not a sufficient assumption to insure stability of the commonly used computational algorithms. On the other hand, assuming the monotony of the global operator, as generally done, may prove to be a too strong and not very natural assumption in some situations. This is, for example, the case of Nash problems as discussed in Ref. 5. The main contribution of this paper is to propose a new notion of (strong) nested monotony that will allow us to prove a general convergence theorem, and that is weaker than the usual monotony assumption. Some simple examples suggest that, in some sense, this requirement is minimal to insure convergence.

2. Basic Results on Variational Inequalities

2.1. Monotone Operators. We recall some results on existence of solutions to variational inequalities based on monotony assumptions. They are drawn from Ekeland and Temam (Ref. 6). We notice that other types of existence results are based on continuity (rather than monotony) assumptions on compact sets [see Aubin and Ekeland (Ref. 7)], but this latter approach is not generally the way in which computational algorithms are obtained. In Ref. 7, the case of multivalued (point-to-set) operators is also discussed. We shall not consider this case in this paper, and we shall limit ourselves to ordinary (single-valued) operators.

Let Ψ be a mapping from a reflexive Banach space U into its dual U^*, and let φ be a proper convex l.s.c. function from U into $\mathbf{R}$. Let U^f be a closed convex subset of U. One looks for $u^* \in U^f$ such that

$$\langle \Psi(u^*), u - u^* \rangle + \varphi(u) - \varphi(u^*) \geq 0, \qquad \forall u \in U^f. \tag{1}$$

In Ref. 6, one can find the following existence theorem.

Theorem 2.1. Assume the following:
(A1) Ψ is weakly continuous over every finite-dimensional subspace of U;
(A2) Ψ is monotone over U^f, that is,

$$\langle \Psi(u) - \Psi(v), u - v \rangle \geq 0, \qquad \forall u, v \in U^f; \tag{2}$$

(A3) there exists $w \in \operatorname{dom} \varphi$ such that

$$\lim_{\substack{\|u\| \to +\infty \\ u \in U^f}} \frac{\langle \Psi(u), u - w \rangle + \varphi(u) - \varphi(w)}{\|u\|} = +\infty. \tag{3}$$

Then, there exists a solution u^* to (1).

Assumption (A3) is of course useless if U^f is bounded. Moreover, it is also met if we strengthen the monotony assumption (A2) by requiring strong monotony of Ψ over U^f (with modulus a), which means that the following assumption is satisfied:
(A4) $\exists a > 0$: $\langle \Psi(u) - \Psi(v), u - v \rangle \geq a \|u - v\|^2, \forall u, v \in U^f.$ \tag{4}
Under (A4), u^* is unique. When Ψ is the derivative of a function J [which is convex, from (A2)], then u^* minimizes $(J + \varphi)(u)$ over U^f. When J is not differentiable, Ψ must be identified with the subdifferential ∂J that is now a point-to-set mapping. When Ψ is not a derivative or a subdifferential, problem (1) cannot generally receive an interpretation in terms of a minimization problem. Pang (Ref. 8) gives examples of such problems.

2.2. Operators over a Product of Spaces. When the space U is the product of N subspaces, we introduce the following notations. U, without subscript, denotes the whole space; U_i denotes the ith component subspace; and

$$u = (u_1, \ldots, u_N) \in U, \qquad u_i \in U_i,$$

$$U = \prod_{i=1}^{N} U_i, \qquad U^f = \prod_{i=1}^{N} U_i^f, \qquad U_i^f \subset U_i, \tag{5}$$

where U_i^f is a closed convex subset. To shorten notations, we shall write u_{-i} instead of $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_N)$, and we shall identify (u_i, u_{-i}) with u. Similarly, $u_{<i}$ will stand for $(u_1, \ldots, u_{i-1})$, $u_{\geqslant i}$ for $(u_i, \ldots, u_N)$, $u_{>i}$ for $(u_{i+1}, \ldots, u_N)$, etc. Since Ψ maps U into its dual space U^* which is the product of $\{U_i^*; i = 1, \ldots, N\}$, Ψ_i will denote the composition of Ψ and of the canonical projection onto U_i^*. Hence,

$$\Psi(u) = (\Psi_1(u), \ldots, \Psi_N(u)).$$

Finally, we shall always assume that φ is additive, that is,

$$\varphi(u) \equiv \sum_{i=1}^{N} \varphi_i(u_i). \tag{6}$$

Lemma 2.1. Under (5)-(6), the variational inequality (1) is equivalent to the following system of N coupled variational inequalities: find $u^* \in U^f$ such that, for $j = 1, \ldots, N$,

$$\langle \Psi_j(u^*), u_j - u_j^* \rangle + \varphi_j(u_j) - \varphi_j(u_j^*) \geqslant 0, \qquad \forall u_j \in U_j^f. \tag{7}$$

The proof is straightforward.

Assume that Ψ is differentiable, and let $\Psi'_{ij}(u)$ denote the partial derivative of Ψ_i with respect to u_j evaluated at point u. This is a linear operator from U_j to U_i^*. Then, we say that Ψ is symmetric iff $\Psi'_{ij}(u)$ is the adjoint of $\Psi'_{ji}(u)$ for all u and all pair (i, j).

Consider now the case $N = 2$ and $\varphi = 0$, for simplicity. We consider the following four particular situations.

Situation (a). Ψ is symmetric and monotone. Again, this is the case when u^* is a global minimum over U^f of a cost function of which Ψ is the derivative.

Situation (b). Ψ is symmetric but not monotone. However, for all i, Ψ_i is monotone as a function of u_i alone, u_{-i} being considered as a parameter. Then, u^* can be interpreted as what is called a person-by-person minimum in team theory, but this may not be a global or even a local minimum of the above cost function.

Situation (c). We keep the same assumptions about monotony as above, but Ψ is no longer assumed to be symmetric. However, we still assume that $\Psi'_{ii}(u)$ is self-adjoint for all u and i. Then, u^* can be interpreted as a Nash equilibrium (the cost function of player i is such that Ψ_i is equal to the partial derivative of this cost function with respect to u_i).

Situation (d). Ψ is symmetric but not monotone. Index j in (7) assumes two values only. For $j = 1$, Ψ_1 is monotone as a function of u_1. For $j = 2$, the corresponding inequality (7) keeps the same left-hand side but it is written with ≤ 0 instead of ≥ 0. Moreover, $-\Psi_2$ is assumed monotone as a function of u_2 (we say that Ψ_2 is negative monotone). Then, u^* can be interpreted as a saddle point.

Example 2.1. Consider the case when $U_1 = U_2 = \mathbf{R}$ and Ψ is linear and thus representable by a $(2, 2)$-dimensional matrix. We give four matrices corresponding to the four particular situations described above:

(a) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$;

(b) $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$;

(c) $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$;

(d) $\begin{bmatrix} 1 & 9 \\ 9 & -1 \end{bmatrix}$.

Recall that, for case (d), the sign of the second line should be reversed to fit into (7).

3. A General Class of Computational Algorithms: Nested Monotony and Convergence

In Ref. 4, the following class of algorithms, based on the so-called auxiliary problem principle, have been proposed to compute the solution of (1). Let us consider a differentiable auxiliary functional $K : U \rightarrow \mathbf{R}$ and a positive number ϵ.

Algorithm 3.1. (i) At $k = 0$, start with some initial u^0.
(ii) At step k, solve the auxiliary problem

$$\min_{u \in U^f} K(u) + \langle \epsilon \Psi(u^k) - K'(u^k), u \rangle + \epsilon \varphi(u). \tag{8}$$

Let u^{k+1} denote the solution of this problem (assume that it does exist and that it is unique).

(iii)   Stop if $\|u^{k+1} - u^k\|$ is below some threshold. Otherwise, go back to Step (ii) with $k \leftarrow k+1$.

The following theorem is proved in Ref. 4.

Theorem 3.1.   (i)   Assume (A1), (A4) and that $\varphi: U \to \mathbf{R}$ is a proper convex l.s.c. functional. Then, there exists a unique solution u^* to (1).

(ii)   Assume that $K: U \to \mathbf{R}$ is a proper convex and differentiable functional and that its derivative K' is strongly monotone with modulus b over U^f. Then, there exists a unique solution u^{k+1} to (8).

(iii)   Assume that Ψ is Lipschitz with modulus L over U^f, that is,

$$\exists L > 0: \|\Psi(u) - \Psi(v)\| \leq L\|u - v\|, \forall u, v \in U^f, \tag{9}$$

and we take

$$0 < \epsilon < 2ab/L^2. \tag{10}$$

Then, the sequence $\{u^k\}$ strongly converges toward u^*.

If we are in the situation (5)-(6), it is natural to choose a functional K that is also additive with respect to the decomposition of $U[K(u) = \sum K_i(u_i)]$, so that the auxiliary problem (8) splits up into N independent subproblems.

Theorem 3.1 essentially covers situation (a) considered in Section 2.2 (strong monotony should replace monotony, the latter weaker assumption being considered in Ref. 2) and the situation when Ψ is not symmetric but still strongly monotone as a whole. It also covers situation (d) as shown by the following lemma. But it does not cover situations (b) and (c), nor does it cover more general situations without symmetry assumptions. The purpose of the next sections is to give a new and less restrictive convergence theorem. But let us come back to the saddle point case with the following lemma, which is well known but for which we provide a proof for the sake of completeness.

Lemma 3.1.   Let $J: (u, v) \to J(u, v)$ be an $\mathbf{R}$-valued smooth functional, convex in u [respectively, strongly convex in u with modulus a uniformly in v] and concave in v [respectively, strongly concave in v with modulus b uniformly in u]. Then, the operator $(u, v) \to (J'_u(u, v), -J'_v(u, v))$ is monotone [respectively, strongly monotone with modulus $\min(a, b)$].

Proof.   We only need to consider the strong version of the lemma; the weak version follows by making $a = b = 0$. Strong convexity in u with modulus a uniform in v means that there exists $a > 0$ such that, for all v, u, u' and $\alpha \in [0, 1]$,

$$J(\alpha u + (1 - \alpha)u', v) \leq \alpha J(u, v) + (1 - \alpha)J(u', v)$$
$$- a\alpha(1 - \alpha)\|u - u'\|^2/2.$$

This is also equivalent to the strong monotony of $J'_u(\cdot, v)$. Similar definitions stand for strong concavity. One has

$$J(u, v) - J(u', v') = [J(u, v) - J(u', v)] + [J(u', v) - J(u', v')]$$

$$\leq \langle J'_u(u, v), u - u' \rangle - a\|u - u'\|^2/2$$

$$+ \langle J'_v(u', v'), v - v' \rangle - b\|v - v'\|^2/2,$$

the inequalities being a consequence of the strong convexity and of the strong concavity (see Ref. 2). Similarly,

$$J(u, v) - J(u', v') = [J(u, v) - J(u, v')] + [J(u, v') - J(u', v')]$$

$$\geq \langle J'_v(u, v), v - v' \rangle + b\|v - v'\|^2/2$$

$$+ \langle J'_u(u', v'), u - u' \rangle + a\|u - u'\|^2/2.$$

Combining the inequalities above yields the result. $\qquad\square$

4. Strong Nested Monotony

For an operator Ψ defined over a product of sets as described in Section 2.2, we introduce the following definition.

Definition 4.1. We say that an operator Ψ is strongly nested monotone relatively to the variational inequality (1) and to the decomposition (5) iff the following holds true.

(i)   Initialization. The operator

$$u_N \to \Omega_N(u_N; u_{<N}) \triangleq \Psi_N(u_{<N}, u_N)$$

is strongly monotone in u_N over U^f_N uniformly in $u_{<N}$ ranging in $U^f_{<N}$. According to Theorem 2.1, and modulo a continuity assumption of the type (A1), the variational inequality in u_N (parametrized by $u_{<N}$)

$$u_N \in U^f_N: \langle \Omega_N(u_N; u_{<N}), v_N - u_N \rangle$$

$$+ \varphi(v_N) - \varphi(u_N) \geq 0, \forall v_N \in U^f_N,$$

has a unique solution denoted by $u^*_N(u_{<N})$. We also set $W_N(u_{<N})$ equal to this solution.

(ii)   Stage j (repeat for $j = N, \ldots, 1$; for $j = 1$, $u_{<j}$ must be omitted). The operator

$$u_j \to \Omega_j(u_j; u_{<j}) \triangleq \Psi_j(u_{<j}, u_j, W_{j+1}(u_{<j}, u_j)) \tag{11}$$

is strongly monotone in u_j over U_j^f uniformly in $u_{<j}$ ranging in $U_{<j}^f$. Again, modulo a continuity assumption, the variational inequality in u_j (parametrized by $u_{<j}$) analogous to the above has a unique solution, denoted by $u_j^*(u_{<j})$. We further set

$$W_j(u_{<j}) \triangleq \{u_j^*(u_{<j}), W_{j+1}(u_{<j}, u_j^*(u_{<j}))\} \in \prod_{k=j}^{N} U_k^f.$$

There are several possible sets of assumptions that would ensure property (A1) for the successive Ω_j. We shall not discuss further this point, because, for algorithmic purposes, we shall consider even stronger assumptions later on.

Definition 4.1 introduces recursively a sequence of nested operators Ω_j and of associated variational inequalities, the definition of rank j operator incorporating the solutions of downstream inequalities (through W_{j+1}), whereas the upstream $u_{<j}$ appears as a parameter. Alternatively, and for a given $i, 1 \le i < N$, one may consider the set of parallel and coupled variational inequalities (7), based on Ψ_j but for j ranging from this i to N, this set being parametrized by $u_{<i}$. It is easy to realize that $W_i(u_{<i})$ is indeed the solution of this set of inequalities. This is stated as a proposition hereafter. Therefore, anticipating the fact that strong nested monotony is a weaker assumption than strong monotony, and modulo the continuity assumptions alluded to above, we see a way of ensuring the existence and uniqueness of the solution to a variational inequality defined over a product of sets without dwelling upon the usual assumption of strong monotony.

Proposition 4.1. Let us consider the following (system of) variational inequalities (with obvious notations)

$$u_{\ge i} \in U_{\ge i}^f : \langle \Psi_{\ge i}(u_{<i}, u_{\ge i}), v_{\ge i} - u_{\ge i} \rangle$$

$$+ \sum_{j=i}^{N} [\varphi_j(v_j) - \varphi_j(u_j)] \ge 0, \forall v_{\ge i} \in U_{\ge i}^f. \tag{12}$$

Then, $W_i(u_{<i})$ as introduced in Definition 4.1 is a solution of (12).

Coming back to Definition 4.1, we notice first that u_{j+1}^* might be nonunique if we assume Ω_{j+1} only simply monotone instead of strongly monotone and then Ω_j might be a point-to-set operator even if Ψ is an ordinary operator. This is one reason why we have given directly a definition of strong nested monotony (for the solution uniqueness, strict monotony would have been enough, but strong nested monotony will be required for algorithmic purposes). Definition of simple nested monotony is possible at the price of dealing with multivalued operators.

Provided that the burden caused by the complexity of notations be overcome, this new notion of nested monotony is simple conceptually. However, it might be difficult to check in some practical situations, since from the definition it involves the solutions u_j^* of partial variational problems parametrized by $u_{<j}$. Nevertheless, looking for it may serve as a guideline in circumstances when global monotony (or saddle point interpretation) cannot be expected.

It should be realized that many factors may affect this property. A first factor is not only the decomposition (5) itself, but even the order in which the set $\{1, \ldots, N\}$ is considered in the definition (we have considered the order from N to 1, but any other permutation of this set might have been considered as well). A second factor is the presence of constraints (represented here by the feasible sets U_1^f) and of the functionals φ_i, since these elements affect obviously the value of u_i^* entering the definition of Ω_{i-1}. Finally, a third factor, which is even more subtle, will be explained directly through the second example below. These observations show that nested monotony is not an intrinsic property of Ψ alone, but rather of Ψ considered within the framework of the variational inequality (1), the decomposition (5), etc.

Example 4.1. We consider the case

$$N = 2, \qquad U_1 = U_1^f = \mathbf{R}, \qquad U_2 = U_2^f = \mathbf{R}^2, \qquad \varphi \equiv 0,$$

and Ψ linear and represented by the following matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ -3 & 1 & 0 \\ 4 & 0 & 2 \end{bmatrix}.$$

A straightforward application of Definition 4.1 leads to checking for the strong monotony (or the positive definiteness) of

$$\Omega_1 = A - BD^{-1}C,$$

which is indeed equal to 1 in this case. If we consider the reversed order ($j = 1$ then 2), we are led to check for the strong monotony of

$$\Omega_2 = D - CA^{-1}B,$$

or equivalently for the positive definiteness of $(\Omega_2 + \Omega_2^*)/2$, which is given by

$$(\Omega_2 + \Omega_2^*)/2 = \begin{bmatrix} 7 & 0.5 \\ 0.5 & -10 \end{bmatrix}.$$

Thus, the property does not hold with this reversed order, whereas it does with the former order.

Example 4.2. Consider the same situation as above, but with a one-dimensional u_2 and with Ψ given by matrix (b) of Example 2.1. It is easily checked that Ψ is not nested monotone, whichever order is considered. However, we can allow the following renumbering that will not change the system of linear equations to which (1) or (7) reduces. By simply writing Eq. (7) for $j = 2$ above (and not below) the same equation for $j = 1$, matrix (b) of Example 2.1 is converted into matrix (a). The latter is positive definite; thus, as we shall see hereafter, it is nested monotone for all decomposition and ordering of the variables. What we have done actually is to renumber the subspaces in the range space of Ψ, namely U^*, without performing the same permutation among the component subspaces of the domain space, namely, U. This is conceivable only when the U_i's are of the same nature and thus interchangeable, and also because we do not pay attention to any symmetry consideration, a property that is not preserved by such manipulations.

Let us now show that nested monotony is weaker than classical monotony.

Proposition 4.2. If Ψ is strongly monotone over U^f with modulus a, and if we assume that, at every stage of the successive variational inequalities introduced in Definition 4.1, Assumption (A1) can be ensured (existence of u_i^*), then strong nested monotony holds true with uniform modulus a.

Proof. For all i and for all u_i and v_i in U_i^f, we must prove that

$$\langle \Omega_i(u_i; u_{<i}) - \Omega_i(v_i; u_{<i}), u_i - v_i \rangle \geq a \| u_i - v_i \|^2, \tag{13}$$

uniformly in $u_{<i}$ ranging in $U_{<i}^f$. To shorten notations, let us set

$$S = W_{i+1}(u_{<i}, u_i) \quad \text{and} \quad T = W_{i+1}(u_{<i}, v_i).$$

Then, from the strong monotony of Ψ over U^f, we have that

$$\langle \Psi_i(u_{<i}, u_i, S) - \Psi_i(u_{<i}, v_i, T), u_i - v_i \rangle$$
$$+ \langle \Psi_{>i}(u_{<i}, u_i, S) - \Psi_{>i}(u_{<i}, v_i, T), S - T \rangle$$
$$\geq a(\| u - v \|^2 + \| S - T \|^2) \geq a \| u - v \|^2.$$

From the definition of S and from Proposition 4.1 [see Eq. (12)], we have

$$\langle \Psi_{>i}(u_{<i}, u_i, S), T - S \rangle + \varphi_{>i}(T) - \varphi_{>i}(S) \geq 0.$$

A similar inequality holds true by substituting v_i for u_i and by exchanging the roles of S and T. Adding these two inequalities to the previous one and remembering the definition (11) of Ω_i yield the result. $\square$

To show that the converse statement of Proposition 4.2 is not true, consider again the situation of Example 4.2, but with Ψ given by the matrix (c) of Example 2.1. It is easily seen that strong nested monotony does hold, but Ψ is not even monotone (its symmetric part is not definite). It turns out, however, that, in the situation when there is no constraint $(U^f = U)$, when $\varphi \equiv 0$ and when Ψ is both linear and symmetric, nested monotony is equivalent to monotony. But this result is no longer true if we relax either the assumption of symmetry (as just shown) or that of linearity as shown by the following example.

Example 4.3. We consider again the situation of Example 4.2, but with a nonlinear Ψ given by

$$\Psi_1(u_1, u_2) = 3(u_1)^{1/3} + 2u_1 - 2u_2,$$

$$\Psi_2(u_1, u_2) = -2u_1 + (u_2)^3 + u_2.$$

It can be checked that Ψ is symmetric, in the sense that its Jacobian matrix is symmetric. Moreover, this Jacobian is not everywhere definite [check it at $(u_1, u_2) = (1, 0)$, for example], so that Ψ cannot be monotone. On the other hand, if we solve the equation $\Psi_2(u_1, u_2) = 0$ for u_2 by using Cardan's formula, we get

$$u_2^\#(u_1) = [u_1 + (u_1^2 + 1/27)^{1/2}]^{1/3} + [u_1 - (u_1^2 + 1/27)^{1/2}]^{1/3}.$$

Then, it can be checked that $\Psi_1(u_1, u_2^\#(u_1))$ is strongly monotone in u_1 (that is, here, that it is strictly increasing).

Since strong nested monotony does not imply in general strong monotony, it cannot be expected that, if strong nested monotony holds true for some decomposition (5), it also does so in general for a coarser decomposition. The converse statement is true: if the property holds true for some decomposition, it does so for all finer decompositions. Since the order in which component subspaces are considered is important as shown by Example 4.1, we must be specific in what we mean by a finer decomposition than (5).

Definition 4.2. Given a decomposition as (5), all finer decompositions are obtained by decomposing some components U_i into two or more subcomponents $U_{ij}, j = 1$ to N_i (for each i, the order in which the U_{ij}'s are numbered is arbitrary), and then by renumbering the new components sequentially (from 1 to $N' > N$) following the lexicographic order. Of course, U^f [respectively, φ] must be decomposable with respect to this finer decomposition [see (5) and (6) respectively].

Proposition 4.3. Strong nested monotony related to some decomposi-
tion (5) implies the same property for all finer decompositions.

Proof. Let $I \subset \{1, \dots, N\}$ be the set of indices j of the initial decompo-
sition for which U_j has indeed been subdivided. Let k be the largest index
in I. To each index i of the initial decomposition, there corresponds an
index (if $i \notin I$) or a set of indices (if $i \in I$) in the new numbering, that we
denote by $\nu(i)$. Let m be the larger index in $\nu(k)$. While checking for
Definition 4.1, starting with the larger index N' of the new decomposition,
there is no difference with the initial decomposition before the index m is
reached by moving backward (the correspondence is one to one between
the indices of the two decompositions until that stage). Then, since
$\Omega_k(u_k; u_{<k})$ was assumed to be strongly monotone in the old decomposition
by Definition 4.1, this is true for all Ω_t, $t \in \nu(k)$, in the new decomposition,
as implied by Proposition 4.2 applied to operator Ω_k and to the overall
space U_k and its decomposition, whatever this decomposition may be (here,
$u_{<k}$ is just a parameter). Once we finished with the old index k and the
new indices in $\nu(k)$, the situation is again the same in both cases (down-
stream u_i's have been replaced by functions $u_i^\#$ of upstream variables, these
functions being intrinsically defined as shown by Proposition 4.1). Hence,
the same reasoning as above can be repeated for the next index in I (moving
backward) and the corresponding indices of the new decomposition and
so on.   □

5. Convergence Theorem Based on Strong Nested Monotony

In this section, our purpose is to give a new convergence theorem for
a special case of Algorithm 3.1 when the auxiliary cost function K is additive
with respect to the decomposition (5); this is of course a desirable feature
in this situation, since then problem (8) splits up into N-independent
subproblems. Unlike Theorem 3.1, this new theorem will not require strong
global monotony of Ψ, but only strong nested monotony.

As we shall see, the proof of this new theorem is rather involved.
Therefore, we shall restrict ourselves to the case of $N = 2$ in (5) and we
shall redefine completely our notations in this section in order to avoid
subscripts as far as possible.

5.1. New Notations. Table 1 gives previous notations on the left-hand
side and their present counterparts on the right-hand side. Moreover, about
operator Ω_1 introduced in Definition 4.1, we now set

$$\Omega(u) \triangleq A(u, v^*(u)) \equiv \Omega_1(u_1) \equiv \Psi_1(u_1, u_2^\#(u_1)). \tag{14}$$

Table 1. Notations.

$\Psi_1(u_1, u_2): A(u, v)$	$\Psi_2(u_1, u_2): B(u, v)$	$\varphi_1(u_1): f(u)$
$\varphi_2(u_2): g(v)$	$K_1(u_1): K(u)$	$K_2(u_2): L(v)$

5.2. Algorithm. For the sake of completeness, we rewrite Algorithm 3.1 with these new notations. We use two different steplengths ϵ and ρ (standing for ϵ_1 and ϵ_2) in the two auxiliary subproblems. At stage k, knowing u^k and v^k, one has to solve

$$\min_{u \in U^f} K(u) + \langle \epsilon A(u^k, v^k) - K'(u^k), u \rangle + \epsilon f(u), \tag{15}$$

$$\min_{v \in V^f} L(v) + \langle \rho B(u^k, v^k) - L'(v^k), v \rangle + \rho g(v). \tag{16}$$

The corresponding solutions are denoted by u^{k+1} and v^{k+1}, respectively. Notice that this scheme corresponds to parallel decomposition. Sequential decomposition would correspond to solving (16) first and then using its solution v^{k+1} instead of v^k in (15); we consider here the subproblems in this order to stick to the order introduced in Definition 4.1. Essentially, we expect the following results to apply to sequential schemes too.

5.3. Assumptions and Convergence Theorem. We require the following Lipschitz and strong monotony assumptions. There exist positive constants such that, for all u, u' in U^f and all v, v' in V^f,

$$\|A(u, v) - A(u', v)\| \leq X \|u - u'\|, \tag{17}$$

$$\|A(u, v) - A(u, v')\| \leq Y \|v - v'\|, \tag{18}$$

$$\|B(u, v) - B(u', v)\| \leq Z \|u - u'\|, \tag{19}$$

$$\|B(u, v) - B(u, v')\| \leq T \|v - v'\|, \tag{20}$$

$$\|K'(u) - K'(u')\| \leq C \|u - u'\|, \tag{21}$$

$$\|L'(v) - L'(v')\| \leq D \|u - u'\|, \tag{22}$$

$$\langle K'(u) - K'(u'), u - u' \rangle \geq c \|u - u'\|^2, \tag{23}$$

$$\langle L'(v) - L'(v'), v - v' \rangle \geq d \|v - v'\|^2, \tag{24}$$

$$\langle B(u, v) - B(u, v'), v - v' \rangle \geq t \|v - v'\|^2, \tag{25}$$

$$\langle \Omega(u) - \Omega(u'), u - u' \rangle \geq r \|u - u'\|^2. \tag{26}$$

Notice that (25)-(26) are nothing else than the strong nested monotony assumption.

Remark 5.1.   Consider the following variational inequality in v parametrized by u:

$$w \in V^f : \langle B(u, w), v - w \rangle + g(v) - g(w) \geq 0, \forall v \in V^f. \tag{27}$$

We see that all of the assumptions of Theorem 2.1, including (A1), are satisfied by the assumptions above. Therefore, the solution w or $v^*(u)$ does exist and it is unique.

We also state the following lemma about the Lipschitz property of Ω.

Lemma 5.1.   Under the above assumptions, for all u, u' in U^f, one has

$$\|v^*(u) - v^*(u')\| \leq S\|u - u'\|, \tag{28}$$

$$\|\Omega(u) - \Omega(u')\| \leq R\|u - u'\|, \tag{29}$$

where

$$S = Z/t \text{ and } R^2 = (X^2 + Y^2)(1 + S^2).$$

Proof.   Let us again set $w = v^*(u)$ and likewise for w'. Adding Inequality (27) with v set equal to w' with the same inequality for w' with v set equal to w, one gets

$$\langle B(u', w) - B(u', w'), w - w' \rangle \leq \langle B(u, w) - B(u', w), w' - w \rangle.$$

Using (25) to get a lower bound of the left-hand side and (19) to get an upper bound of the right-hand side, one proves (28). Then,

$$\|\Omega(u) - \Omega(u')\|^2 \leq [\|A(u, w) - A(u', w)\| + \|A(u', w) - A(u', w')\|]^2$$
$$\leq (X^2 + Y^2) \cdot (\|u - u'\|^2 + \|w - w'\|^2),$$

using the definition (14) of Ω, and then successively the triangular inequality, (17)–(18) and the Hölder inequality. The proof is completed by using (28). $\square$

Remark 5.2.   Considering the variational inequality in u,

$$u^* \in U^f : \langle \Omega(u^*), u - u^* \rangle + f(u) - f(u^*) \geq 0, \forall u \in U^f, \tag{30}$$

we see that all the assumptions of Theorem 2.1 are now satisfied. Therefore, there exists a unique solution u^*. Let $v^* = v^*(u^*)$. From Proposition 4.1, we know that (u^*, v^*) is the unique solution of the original variational inequality in (u, v) involving (A, B, f, g).

Theorem 5.1. Under the previous assumptions, starting from any (u^0, v^0) in $U^f \times V^f$, algorithm (15)–(16) generates a well-defined sequence (u^k, v^k). Moreover, if

$$0 < \rho < 2dt/T^2 \quad \text{and} \quad 0 < \epsilon < h(\rho), \tag{31}$$

where h is a certain positive function of ρ for positive values of the argument, then the sequence (u^k, v^k) converges (in norm) toward (u^*, v^*).

Remark 5.3. Notice that the upper bound of ρ is the exact analog of the bound of ϵ in Theorem 3.1 [see (10)]. More will be told about $h(\rho)$ and the respective choices of ϵ and ρ at the end of the proof.

Proof. For the sake of simplicity, we set $w^k = v^*(u^k)$ throughout this proof. The auxiliary subproblems (15)–(16) are equivalent to the following variational inequalities (see Ref. 6), which have unique solutions u^{k+1} and v^{k+1}, respectively, from Theorem 2.1 and the assumptions

$$\langle K'(u^{k+1}) - K'(u^k) + \epsilon A(u^k, v^k), u - u^{k+1} \rangle$$
$$+ \epsilon(f(u) - f(u^{k+1})) \geq 0, \qquad \forall u \in U^f, \tag{32}$$

$$\langle L'(v^{k+1}) - L'(v^k) + \rho B(u^k, v^k), v - v^{k+1} \rangle$$
$$+ \rho(g(v) - g(v^{k+1})) \geq 0, \qquad \forall v \in V^f. \tag{33}$$

Set u equal to u^* in (32), and combine it with (30) multiplied by ε, in which u has been set equal to u^{k+1}. This yields

$$\langle K'(u^k) - K'(u^{k+1}), u^* - u^{k+1} \rangle$$
$$\leq \epsilon \langle A(u^k, v^k) - \Omega(u^*), u^* - u^{k+1} \rangle$$
$$= \epsilon \langle A(u^k, v^k) - A(u^k, w^k), (u^* - u^k) + (u^k - u^{k+1}) \rangle$$
$$+ \epsilon \langle \Omega(u^k) - \Omega(u^*), u^* - u^k \rangle + \epsilon \langle \Omega(u^k) - \Omega(u^*), u^k - u^{k+1} \rangle$$
$$\leq \epsilon Y \| v^k - w^k \| (\| u^* - u^k \| + \| u^k - u^{k+1} \|)$$
$$- \epsilon r \| u^k - u^* \|^2 + \epsilon R \| u^k - u^* \| \| u^k - u^{k+1} \|, \tag{34}$$

where we have used (26) for the second term above, the Schwarz and the triangular inequalities, and then (18) for the first term, the Schwarz inequality and (29) for the last term. Similarly, set v equal to w^k in (33), and combine it with (27) multiplied by ρ, and where u, w, v are set equal to u^k, w^k, v^{k+1}, respectively. This yields

$$\langle L'(v^k) - L'(v^{k+1}), w^k - v^{k+1} \rangle$$
$$\leq \rho \langle B(u^k, v^k) - B(u^k, w^k), w^k - v^{k+1} \rangle$$
$$= \rho \langle B(u^k, v^k) - B(u^k, w^k), w^k - v^k \rangle$$
$$+ \rho \langle B(u^k, v^k) - B(u^k, w^k), v^k - v^{k+1} \rangle$$
$$\leq -\rho t \| v^k - w^k \|^2 + \rho T \| v^k - w^k \| \| v^k - v^{k+1} \|, \tag{35}$$

where we have used (25) for the former term above, and the Schwarz inequality, and then (20) for the latter.

On the other hand, let us consider the following functional:

$$\Lambda(u, v) \triangleq K(u^*) - K(u) - \langle K'(u), u^* - u \rangle$$
$$+ \alpha [L(v^*(u)) - L(v) - \langle L'(v), v^*(u) - v \rangle],$$

where α is a positive constant to be chosen later on. From the convexity of K and L, Λ takes only nonnegative values. Let us study the difference $\Lambda(u^{k+1}) - \Lambda(u^k)$ that we write symbolically $\kappa + \alpha\mu$, where κ includes all terms from K and μ all terms from L. After some manipulations, one gets

$$\kappa = [K(u^k) - K(u^{k+1}) - \langle K'(u^k), u^k - u^{k+1} \rangle]$$
$$+ [\langle K'(u^k) - K'(u^{k+1}), u^* - u^{k+1} \rangle].$$

The latter expression in brackets already appeared in (34), whereas the former is bounded from above by $-c \|u^{k+1} - u^k\|^2/2$ [this is a consequence of (23), see Lemma 2.1 in Ref. 2]. Similarly,

$$\mu = [L(v^k) - L(v^{k+1}) - \langle L'(v^k), v^k - v^{k+1} \rangle] + [L(w^{k+1}) - L(w^k)]$$
$$+ [\langle L'(v^k) - L'(v^{k+1}), w^k - v^{k+1} \rangle] + [\langle L'(v^{k+1}), w^k - w^{k+1} \rangle].$$

The third expression in brackets already appeared in (35), the first expression is bounded from above by $-d \|v^{k+1} - v^k\|^2/2$ as a consequence of (24), the second expression is also bounded from above by $\langle L'(w^k), w^{k+1} - w^k \rangle + D \|w^{k+1} - w^k\|^2/2$ [this is a consequence of (22), see Lemma 2.1 in Ref. 2]. Therefore,

$$\mu \leqslant -d \|v^{k+1} - v^k\|^2/2 + D \|w^{k+1} - w^k\|^2/2 - \rho t \|v^k - w^k\|^2$$
$$+ \rho T \|v^k - w^k\| \|v^k - v^{k+1}\|$$
$$+ \langle (L'(w^k) - L'(v^k)) + (L'(v^k) - L'(v^{k+1})), w^{k+1} - w^k \rangle.$$

In the last expression, we use the Schwarz and the triangular inequalities and also (22). We then collect all these inequalities for $\kappa + \alpha\mu$, and we get

$$\kappa + \alpha\mu \leqslant \epsilon Y \|v^k - w^k\| (\|u^k - u^*\| + \|u^k - u^{k+1}\|)$$
$$- \epsilon r \|u^k - u^*\|^2 - c \|u^{k+1} - u^k\|^2/2$$
$$+ \epsilon R \|u^k - u^*\| \|u^k - u^{k+1}\| - \alpha d \|v^{k+1} - v^k\|^2/2$$
$$+ \alpha D \|w^{k+1} - w^k\|^2/2 - \alpha \rho t \|v^k - w^k\|^2$$
$$+ \alpha \rho T \|v^k - w^k\| \|v^k - v^{k+1}\|$$
$$+ \alpha D \|w^{k+1} - w^k\| (\|w^k - v^k\| + \|v^{k+1} - v^k\|).$$

From (28), we know that

$$\|w^{k+1} - w^k\| \leq S\|u^{k+1} - u^k\|,$$

where S is given by Lemma 5.1. We make use of this inequality in the expressions above. For all cross products, we also use systematically the fact that, for all numbers x, y, ξ (ξ positive),

$$xy \leq (\xi x^2 + y^2/\xi)/2.$$

Constants such as ξ will be denoted by ξ_i, i from 1 to 6, since there are 6 cross products above. These positive constants, together with α, will be chosen later on (we have not found any strategy to fix them at this stage of the proof). After straightforward calculations, we finally get

$$\kappa + \alpha\mu \leq [\eta_1\epsilon\|u^k - u^*\|^2 + \eta_2\alpha\|v^k - w^k\|^2$$

$$+ \eta_3\alpha\|v^{k+1} - v^k\|^2 + \eta_4\|u^{k+1} - u^k\|^2]/2, \tag{36}$$

with the following expressions of the η_i's:

$$\eta_1 = Y/\xi_1 + R\xi_3 - 2r,$$

$$\eta_2 = \epsilon Y(\xi_1 + \xi_2)/\alpha + \rho T\xi_6 + D\xi_5 - 2\rho t,$$

$$\eta_3 = \rho T/\xi_6 + D\xi_4 - d,$$

$$\eta_4 = \epsilon(Y/\xi_2 + R/\xi_3) + \alpha DS^2(1 + 1/\xi_4 + 1/\xi_5) - c.$$

Recall that $\kappa + \alpha\mu$ represents the difference $\Lambda(u^{k+1}, v^{k+1}) - \Lambda(u^k, v^k)$ and that Λ assumes nonnegative values. Then, suppose that we are able to choose positive values of α, ξ_1 to ξ_6, ρ, ϵ such that η_1 and η_2 are negative and such that η_3 and η_4 are null. In view of (36), this would mean that the sequence $\{\Lambda(u^k, v^k)\}$ is nonincreasing and that, being bounded from below by zero, it converges. Therefore, the difference of two successive terms would tend to zero. From (36) and the assumptions made about the η_i's, we could conclude that both $\|u^k - u^*\|$ and $\|v^k - w^k\|$ tend to zero. Since w [that is, $v^*(u)$] is Lipschitz continuous in u (Lemma 5.1), we could conclude that w^k tends toward $v^*(u^*) = v^*$ and that the same happens to v^k, thus completing the proof.

Therefore, it remains to prove that the assumptions on the η_i's can be fulfilled by a proper choice of all the parameters left free until now. It is really difficult to choose them without having a guideline. The idea that we have followed is to minimize some linear combination of the η_i's with arbitrary positive weights λ_i under the constraints that η_1 and η_2 be strictly negative and that η_3 and η_4 be null (nonpositive would be enough). This idea of minimization can be justified to some extent by the fact that we want to make the increase of the Lyapounov function Λ over any stage of

the algorithm as small as possible (or its decrease as large as possible). That this idea is heuristic does not affect the validity of the proof, since the only objective that we have to meet is to fulfill the conditions on the η_i's. For the same reason, we are not going to give the detailed calculations leading to a solution, but we shall simply exhibit this solution.

These calculations, which are straightforward but cumbersome, consist in differentiating the expression $\Sigma \lambda_i \eta_i$ with respect to the variables α and ξ_1 to ξ_6, and then in writing that these derivatives are null (again, that these stationarity conditions yield really a true minimum does not matter for our final purpose). Because of the special structure of the objective function (notice that all variables appear once linearly and once as a denominator), it is possible to get relations between the variables which are independent of the weights λ_i. Using these relations and those provided by the conditions $\eta_3 = \eta_4 = 0$, one can express all the other variables as the following functions of ξ_2, ϵ, ρ:

$$\xi_1 = \xi_2(c\xi_2 - 2\epsilon Y)/\epsilon(R\xi_2 + Y),$$

$$\xi_3 = \epsilon(R\xi_2 + Y)/(c\xi_2 - 2\epsilon Y),$$

$$\xi_4 = dS\xi_2/(DS\xi_2 + \rho T),$$

$$\xi_5 = S\xi_2,$$

$$\xi_6 = (DS\xi_2 + \rho T)/d,$$

$$\alpha = Yd(c\xi_2 + \epsilon R\xi_2 - \epsilon Y)/DS(R\xi_2 + Y)(S(D + d)\xi_2 + \rho T + d).$$

We let the reader check that indeed $\eta_3 = \eta_4 = 0$. Notice that, for ξ_1 and ξ_3 to be positive, we must impose

$$\epsilon < c\xi_2/2Y. \tag{37}$$

Placing the above expressions into those of η_1 and η_2, we get the conditions

$$\eta_1 = \epsilon(R\xi_2 + Y)^2/\xi_2(c\xi_2 - 2\epsilon Y) - 2r < 0, \tag{38}$$

$$\eta_2 = [DS^2(D + d)\xi_2^2 + 2DS(d + \rho T)\xi_2 + \rho(\rho T^2 - 2dt)]/d < 0. \tag{39}$$

Condition (38) translates into the following condition on ϵ:

$$\epsilon < 2rc\xi_2^2/[(R\xi_2 + Y)^2 + 4rY\xi_2]. \tag{40}$$

This is stronger than (37), which will thus be a consequence of (40). The right-hand side of (40) is an increasing function $p(\xi_2)$ (calculate the derivative). Therefore, the larger ξ_2 is, the larger the upper bound of ϵ can be. But, as we shall see, there is a tradeoff with the condition imposed by (39). Observe first that this latter condition cannot be met if ρ does not satisfy the condition in (31). If it does so, then (39) can always be satisfied by choosing ξ_2 small enough, but this may mean a small ϵ too [to meet (40)].

To be more specific, observe that the left-hand side of (39) is a poly-nomial of degree 2 in ξ_2, the coefficients of which are functions of ρ. Under condition (31) on ρ, this polynomial has two real roots, one negative that we disregard (ξ_2 must be positive) and one positive that we denote by $q(\rho)$. Once ρ has been chosen to satisfy the condition in (31), in order to meet (39), ξ_2 must be strictly smaller than $q(\rho)$; and then, in order to meet (40), since p is an increasing function, ϵ must be smaller than $h(\rho) \triangleq p(q(\rho))$ [this defines h appearing in (31)].

Finally, to make the tradeoff between the value of ρ and the bound on ϵ more precise, we must worry about the range of values of ρ for which q (and therefore h) is increasing or decreasing. It may be proved that q is increasing for values of ρ close to zero and decreasing for values close to the upper bound of ρ, that is, $2dt/T^2$. Actually, it can be proved that q is always decreasing for values of ρ larger than dt/T^2, the exact value for which q reaches its maximum depending on the condition numbers c/C and d/D. Heuristically, it may be said that ρ should always be chosen smaller than dt/T^2, although the algorithm still converges for values of ρ close to twice this latter bound, but then with ϵ very small.   □

6. Final Comments and Conclusions

In this paper, we have introduced a new notion of strong nested monotony for variational inequalities defined over a product of spaces. Under this assumption, and other technical assumptions, existence of a unique solution can be guaranteed, and this solution can be computed by Algorithm 3.1, provided that the auxiliary function K in (8) be chosen additive with respect to decomposition (5), which is assumed to hold together with (6). This general algorithm includes several algorithms already proposed in the literature, and it generalizes them to some extent [Pang (Ref. 8), Pang and Chan (Ref. 9), Glowinski and associates (Ref. 10)]. The decomposition aspects have been further developed in other papers (see Refs. 1-5).

Strong nested monotony has the advantage over classical strong monotony in that it is a weaker requirement. Since the latter property is generally assumed in order to prove convergence of algorithms such as Algorithm 3.1, Theorem 5.1 seems to be an original and useful contribution to this field (see Ref. 5 for an application to the computation of Nash equilibria).

Admittedly, nested monotony may not be so easy to check in some practical situations. Moreover, we have shown in Section 4 several factors affecting this property [decomposition into subspaces, although this may be imposed by the decomposability of U^f and φ, see (5) and (6); order in

which subspaces are numbered in both the domain and the range spaces of Ψ, the latter numbering having some flexibility with respect to the former when some U_i and U_j are identical, so that U_i may be associated with U_j^* and conversely as discussed in Example 4.2, etc.]. More insight would be needed from this point of view.

Nevertheless, we believe that this property is in some sense not only sufficient but also necessary in order to prove convergence of iterative algorithms such as Algorithm 3.1. Of course this fuzzy statement cannot be made more precise in general, but we shall show hereafter some simple situation where it can be proved.

We consider again the linear two-dimensional situation discussed in Example 4.2, but now with the general 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

as operator Ψ (recall that φ is null and that there are no constraints, so that the problem consists in solving a system of 2 linear equations with null right-hand side; the solution is $(0, 0)$]. Using Algorithm 5.2 (we now use the notations of Section 5), with

$$K(u) = u^2/2 \text{ and } L(v) = v^2/2,$$

one gets

$$u^{k+1} = u^k - \epsilon(au^k + bv^k),$$

$$v^{k+1} = v^k - \rho(cu^k + dv^k).$$

It is easy to figure out that, for this algorithm to converge for ϵ and ρ positive and small enough, it is necessary that $ad - bc$ and $\epsilon a + \rho d$ are positive, which means, for the latter condition, that at least a or d is positive (then, if, for example, d is positive, it is sufficient to choose the ratio ϵ/ρ small enough). We recognize in these conditions those of strong nested monotony for the only possible decomposition here, and for one or the other out of the two possible orders.

To close this paper, let us give another insight into the meaning of Theorem 5.1. To fix ideas, say that d is positive together with $ad - bc$. For a given ratio $\tau = \epsilon/\rho$ and for ϵ close to zero, the discrete-time dynamic system above behaves asymptotically as the differential system

$$\dot{u} = -au - bv,$$

$$\tau\dot{v} = -cu - dv.$$

Thus, making τ small (that is, ϵ small compared to ρ) corresponds to

creating a two-time-scale system that is stable: v quickly moves toward the curve $v^*(u) = -cu/d$ (this move is stable, because d is positive), while u slowly moves toward the equilibrium point $u^* = 0$. We know that the reduced system (in the terminology of singular perturbation theory), obtained by making $\tau = 0$, namely,

$$\dot{u} = -(a - bc/d)u,$$

is stable by the assumption that $ad - bc$ is positive. But, coming back to the framework of an iterative algorithm, sticking to the curve $v^*(u)$ is computationally costly because, for a frozen u^k, a theoretically infinite number of iterations on v would have to be performed. It is of course easier to let u and v evolve simultaneously. The two-time-scale property [which insures stability because one remains close to the path $v^*(u)$, which safely leads to the solution] is simply induced by the respective magnitude of ϵ and ρ as expressed by Theorem 5.1 (it is not even necessary to perform several iterations on v for one iteration on u). This interpretation does not necessarily mean that, practically ϵ should be smaller than ρ by several orders of magnitude. This depends, among other factors, on the relative degrees of stability of B, as an operator on v alone, and of Ω, as measured by the moduli t and r [see (25)-(26)].

References

1. COHEN, G., *Optimization by Decomposition and Coordination: A Unified Approach*, IEEE Transactions on Automatic Control, Vol. AC-23, No. 2, pp. 222-232, 1978.
2. COHEN, G., *Auxiliary Problem Principle and Decomposition of Optimization Problems*, Journal of Optimization Theory and Applications, Vol. 32, No. 3, pp. 277-305, 1980.
3. COHEN, G., and ZHU, D. L., *Decomposition Coordination Methods in Large-Scale Optimization Problems: The Nondifferentiable Case and the Use of Augmented Lagrangians*, Advances in Large-Scale Systems, Theory and Applications, Edited by J. B. Cruz, JAI Press, Greenwich, Connecticut, Vol. 1, pp. 203-266, 1984.
4. COHEN, G., *Auxiliary Problem Principle Extended to Variational Inequalities*, Journal of Optimization Theory and Applications, Vol. 59, No. 2, pp. 325-333, 1988.
5. COHEN, G., *Nash Equilibria: Gradient and Decomposition Algorithms*, Large Scale Systems, Vol. 12, pp. 173-184, 1987.
6. EKELAND, I., and TEMAM R., *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, The Netherlands, 1976.
7. AUBIN, J. P., and EKELAND, I., *Applied Nonlinear Analysis*, John Wiley and Sons, New York, New York, 1984.

8. PANG, J. S., *Asymmetrical Variational Inequality Problems Over Product Sets: Applications and Iterative Methods,* Mathematical Programming, Vol. 31, No. 2, pp. 206–219, 1985.

9. PANG, J. S., and CHAN, D., *Iterative Methods for Variational and Complementarity Problems,* Mathematical Programming, Vol. 24, No. 3, pp. 284–313, 1982.

10. GLOWINSKI, R., LIONS, J. L., and TREMOLIERES, R., *Analyse Numérique des Inéquations Variationnelles, Tome 1: Théorie Générale, Premières Applications,* Dunod, Paris, France, 1976.