KYBERNETIKA — VOLUME 25 (1989), NUMBER 3

EVALUATION OF CLARKE'S GENERALIZED GRADIENT IN OPTIMIZATION OF VARIATIONAL INEQUALITIES

TOMÁŠ ROUBÍČEK

Optimization of systems governed by variational inequalities with linear constraints in $\mathbb{R}^n$ yields nonsmooth cost functions to be minimized. After discussion about various numerical methods to solve such optimization problems, we propose usage of a bundle or a subgradient algorithm and deal with the only problem how to evaluate in such case the generalized gradient of Clarke, required by it. As the problem in general is very complicated, an effective procedure is proposed only for certain special data. However, using the transversality theory, it is shown that "almost all" (possibly in the generic sense) sufficiently smooth data fulfil the conditions that guarantee the validity of the procedure proposed.

1. FORMULATION OF THE PROBLEM AND CLASSICAL METHODS

This paper contains a new approach to such optimization problems where the state is governed by a variational inequality on a finite-dimensional space. For simplicity we confine ourselves to the elliptic case with control-independent linear constraints. Yet, our method could be extended to evolution variational inequalities (after a discretization both in space and in time) or to linear constraints depending on a control parameter as well. On the other hand, general convex constraints or nonlinear monotone operators in the variational inequality would cause probably considerable complications.

First we formulate our problem. Let $b_i \in \mathbb{R}^n$, $c_i \in \mathbb{R}$, $i \in I_K$, I_K be a finite index set, $\langle \cdot, \cdot \rangle$ denote the usual scalar product in $\mathbb{R}^n$, $n \geqq 1$. We consider the convex polyhedral set:

$$K = \{x \in \mathbb{R}^n;\ \forall i \in I_K: \langle b_i, x\rangle \geqq c_i\}\ .$$

Let $A: \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^n$ be a matrix-valued function, $m \geqq 1$, and $f: \mathbb{R}^m \to \mathbb{R}^n$. For every control parameter $u \in \mathbb{R}^m$ we consider the variational inequality:

(P_u) find $x = x(u) \in K$ such that $\langle A(u)\, x, y - x\rangle \geqq \langle f(u), y - x\rangle \quad \forall y \in K$

Furthermore, let $j\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a cost function. We will suppose:

(1) $A\colon \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^n$, $f\colon \mathbb{R}^m \to \mathbb{R}^n$, and $j\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ are C^1-mappings, $A(u)$ is a symmetric positive definite matrix for all $u \in \mathbb{R}^m$

(C^1 means continuously differentiable). As for the convex set K, we will suppose:

(2) K is nonempty, and

$$\forall x \in R^n\colon \{b_i;\ \langle b_i, x\rangle = c_i\} \quad \text{is a linearly independent set,}$$

(in other words, the binding constraints are always linearly independent). It is well known that (P_u) has exactly one solution $x(u)$ which is, at the same time, the only minimizer of the quadratic function $x \mapsto \langle A(u)\, x - 2 f(u), x\rangle$ over the convex set K. Also, $x(u)$ is the projection of $A^{-1}(u) f(u)$ onto K with respect to the metric induced by the matrix $A(u)$. We will denote the respective projector by $\mathrm{Pr}_{K,A(u)}\colon \mathbb{R}^n \to \mathbb{R}^n$, and write $x(u) = \mathrm{Pr}_{K,A(u)}\, A^{-1}(u) f(u)$. Thus we can see that the family of the variational inequalities $\{(\mathrm{P}_u);\ u \in \mathbb{R}^m\}$ forms by $u \mapsto x(u)$ a state operator $\mathbb{R}^m \to \mathbb{R}^n$, and the following optimization problem may be introduced:

$(\mathscr{P})$ minimize $J(u) = j(u, x(u))$ where $x(u)$ solves (P_u), $u \in \mathbb{R}^m$.

Problems of such type appear, after a discretization, in optimization (e.g. optimal control, optimal shape design, identification of coefficients, etc.) of an elliptic partial differential equation with, say, a unilateral boundary condition (or another unilateral elliptic problem; see [4]). We suppose that $(\mathscr{P})$ has got a solution, which can be ensured, e.g., by some coercivity of j. The question analysed in this paper is how to solve the optimization problem $(\mathscr{P})$ numerically by existing optimization algorithms. First we briefly touch some more or less standard methods. Of course, the main difficulty lies in the fact that the state operator is nonsmooth, thus also $J\colon \mathbb{R}^m \to \mathbb{R}$ is nonsmooth.

In principle we have two possibilities: either minimize J given by the implicit state operator without any constraint (i.e. the problem $(\mathscr{P})$), or minimize j over the product of the spaces of controls, states, and here also the Lagrange multipliers, but with constraints characterizing the state operator. For the latter possibility we can use the Kuhn-Tucker necessary (and here also sufficient) conditions for x to solve (P_u):

(3)
$$A(u)\, x - f(u) = \sum_{i \in I_K} \lambda_i b_i$$

$$\lambda_i \geqq 0\,, \quad \lambda_i r_i = 0\,, \quad r_i = \langle b_i, x\rangle - c_i \geqq 0\,,$$

where $\lambda_i = \lambda_i(u)$ are the uniquely determined (thanks to (2)) Lagrange multipliers regarding to the constraints forming the polyhedral set K. However such approach has the following disadvantage: in applications, n is typically far larger than m. E.g. in optimal shape design of an elastic 2D-body we have got, say, $m = 10$ design parameters, while the elliptic partial differential equation (i.e. the Lamé system)

is discretized by the finite element method using, say, 500 mesh point (hence $n = 1000$), and the boundary with a unilateral boundary condition (e.g. the Signorini problem) has got, say, $\mathrm{card}(I_K) = 20$ mesh points. Thus we would replace the unconstrained problem with 10 variables by a constrained problem with 1030 variables, which is certainly not much effective. This disadvantage hardly can be compensated by the fact that we need not to solve the variational inequality (P_u) (recall that we have enough effective numerical methods to solve (P_u); see [4]).

Hence we will consider only the former approach: minimize J over the space of controls. We have still two possibilities: either approximate $(\mathscr{P})$ by the usual smoothing technique, or treat directly the nonsmooth problem $(\mathscr{P})$. In our special problem the former, quite efficient method is based on the usual smooth penalty-function technique: replace the original variational inequality (P_u) by a smooth variational equality $(\mathrm{P}_u^\varepsilon)$:

$$(\mathrm{P}_u^\varepsilon) \qquad \text{find} \quad x = x(u) \in \mathbb{R}^n \quad \text{such that}$$

$$\langle A(u)\, x, y\rangle + \varepsilon^{-1}\langle \beta(x), y\rangle = \langle f(u), y\rangle \quad \forall y \in \mathbb{R}^n ,$$

or we may also write briefly: $A(u)\, x + \beta(x)/\varepsilon = f(u)$, where $\varepsilon > 0$ and $\beta\colon \mathbb{R}^n \to \mathbb{R}^n$ is an appropriate monotone C^1-mapping with $\mathrm{Ker}\,\beta = K$; cf. [4; Chap. 1, § 3.2]. However this approximation method has got all unpleasant properties of the penalty technique: for ε large we get smooth problems but $(\mathrm{P}_u^\varepsilon)$ approximates (P_u) badly, while for small $\varepsilon > 0$ we get a good approximation but with "rapidly changing" gradients, hence the approximation is "numerically nonsmooth" like the original one. Moreover, the approximate equation, being nonlinear, can be solved only iteratively with a certain prescribed accuracy. Yet, even small errors in evaluating of the cost function and its derivatives are very unpleasant for optimization algorithms and often cause their failure.

During last decade, efficient optimization algorithms for nonsmooth cost functions have been developed, thus we may solve directly the original nonsmooth problem. We have again two possibilities: either use an optimization algorithm for quasi-differentiable functions (see V. F. Demyanov [3]) requiring the directional derivatives of J, or use a bundle algorithm (see C. Lemaréchal et al. [8]) or a subgradient algorithm (B. T. Polak [10]) requiring some element of Clarke's generalized gradient of J. In any case, we need an efficient procedure yielding gradient information about J. However, it is not a trivial matter at all, and classical approaches do not seem much effective as explained below.

For every $u \in \mathbb{R}^m$ we classify the indices from I_K:

$$I_n(u) = \{i \in I_K;\ r_i(u) = \langle b_i, x(u)\rangle - c_i > 0\} ,$$
$$I_a(u) = \{i \in I_K;\ \lambda_i(u) > 0\} ,$$
$$I_0(u) = I_K \setminus (I_n(u) \cup I_a(u)) ,$$

where $x(u) = \mathrm{Pr}_{K,A(u)}\, A^{-1}(u) f(u)$ is the solution of (P_u), and $\lambda_i(u)$ are the Lagrange multipliers from the Kuhn-Tucker conditions (3) for the projection $\mathrm{Pr}_{K,A(u)}$ of

$A^{-1}(u)f(u)$. The indices from $I_n(u)$, $I_a(u)$, and $I_0(u)$ will be referred to as non-active, strongly active, and semi-active, respectively. As a consequence of (1) and (2), both the mappings $u \mapsto x(u)$ and $u \mapsto \lambda_i(u)$ are locally Lipschiz continuous (cf. W. W. Hager [5; Theorem 3.1]). If $I_0(u)$ is empty at some $u \in \mathbb{R}^m$, then J is differentiable at u. Indeed, by the mentioned continuity of $x(u)$ and $\lambda_i(u)$, we have $I_n(v) = I_n(u)$ and $I_a(v) = I_a(u)$ for every v belonging to a sufficiently small neighbourhood B of u. Then we have also $J(v) = J_{I_a(u)}(v)$ for all $v \in B$, where $J_I\colon \mathbb{R}^m \to \mathbb{R}$ is defined by:

$$J_I(u) = j(u, \mathrm{Pr}_{K(I),A(u)}\, A^{-1}(u)f(u)), \quad \text{with} \tag{4}$$

$$K(I) = \{x \in \mathbb{R}^n;\ \forall i \in I\colon \langle b_i, x\rangle = c_i\}\,.$$

Obviously, for $I = I_a(u)$, $K(I)$ is a nonempty linear variety in $\mathbb{R}^n$, and J_I is a C^1-function. Therefore J is differentiable at u because it coincides with J_I on B.

Hence we might arrange our computation as follows: to solve (P_u) we use an algorithm that gives (after a finite number of steps) the solution $x(u)$ together with the corresponding Lagrange multipliers $\lambda_i(u)$, hence we can determine the partition $I_K = I_n(u) \cup I_0(u) \cup I_a(u)$. If it happens that $I_0(u) = \emptyset$, we compute the derivative $\mathrm{D}J(u)$ ("D" will denote the derivative) as $\mathrm{D}J_I(u)$ via the adjoint-equation technique considering only the constraints with the indices from $I = I_a(u)$. This technique, though quite difficult to programme, is very efficient. Being standard in optimal control, we will not explain it here; for the case of the so-called simple constraints see also [6]. It should be emphasized that the case $I_0(u) = \emptyset$ may be expected as very frequent: by the well-known Rademacher theorem (i.e. the locally Lipschitzian function J is differentiable a.e. in $\mathbb{R}^m$) and by the transversality theory used below we could demonstrate that, roughly speaking, for "randomly" chosen smooth data A and f, and for almost all $u \in \mathbb{R}^n$, the set $I_0(u)$ will be empty.

Nevertheless, a minimizer of J cannot be considered as chosen randomly, and we may expect that $I_0(u)$ will be often nonempty when u is an optimal solution of $(\mathscr{P})$. Thus the analysis we will perform below is particularly useful at an optimal solution of $(\mathscr{P})$, or near it if round-off or other computational errors are taken into consideration; cf. also the numerical experiments made in [6; § 6].

Hence we have to analyse the case $I_0(u) \neq \emptyset$, i.e. the case when the so-called strict complementarity conditions for P_u are not satisfied. First we discuss the method based on the directional derivatives. K. Jittorntrum [7] proved that there exist directional derivatives of $x(u)$ and $\lambda_i(u)$, and proposed a procedure how to evaluate them: take some $I \in \mathscr{I}(u)$, where

$$\mathscr{I}(u) = \{I \subset I_K;\ I_a(u) \subset I \subset I_a(u) \cup I_0(u)\}\,, \tag{5}$$

then evaluate the directional derivatives of the function J_I and of the corresponding multipliers, check whether these derivatives fulfils certain system of inequalities derived by differentiation (with respect to u) of the Kuhn-Tucker conditions (3)

for (P_u), and possibly repeat all computation with another $I \in \mathcal{I}(u)$; cf. [7; Remark (iii)]. Since I_K is finite, such algorithm actually find the directional derivative after a finite number of trials and error choices of the index set I. However, if there is, say, 10 semi-active indices, then there is 1024 possibilities how to choose I, and it may be expected that we make in average 512 error choices per one evaluation of the directional derivative, which is certainly too much computational effort wasted. K. Malanowski [9], using the results of [7], evaluated the directional derivative (somewhat more effectively) via solving a quadratic programming problem. However, in any case we obtain only the derivative of J in some direction, and thus we need additionally a procedure solving the so-called direction-finding subproblem (see e.g. [3]) which will require to evaluate the directional derivative in many directions at each current control.

Now we come to our new approach based on employing a bundle or a subgradient algorithm to minimize J. Due to (1) and (2) the state operator $u \mapsto x(u)$ is locally Lipschitzian ([5; Theorem 3.1]), hence also J is locally Lipschitzian. Thus the Clarke generalized gradient $\partial J(u)$ of J at u is well defined by the formula [2; Theorem 2.5.1]:

$$\partial J(u) = \operatorname{co}\{\lim_{u_k \to u} \mathrm{D}J(u_k)\ ;\quad J \text{ is differentiable at every } u_k \text{ and the limit exists}\}\ ,$$

where "co" denotes the convex hull. Using the bundle [8] or the subgradient [10] algorithm to minimize J, we need an effective procedure that yields only one element of $\partial J(u)$. We will compute it again as $\mathrm{D}J_I(u)$. As we have already seen above, there is the fundamental problem how to choose I. In the case of the directional derivative we may even expect that there is only one correct choice of I from $\mathcal{I}(u)$. Since the generalized gradient in the nondifferentiable points contains more than one element, we may expect that now there is a large amount of correct choices of I from $\mathcal{I}(u)$. It is highly advantageous, compensating thus the fact that (from the optimization point of view) Clarke's generalized gradient describes local behaviour of J worse than the directional derivatives.

However, it must be pointed out that, in some special situations, there may exist incorrect choice of I from $\mathcal{I}(u)$, i.e. $\mathrm{D}J_I(u) \notin \partial J(u)$ for some $I \in \mathcal{I}(u)$; cf. the example in [6]. On the other hand, we demonstrate in what follows that, roughly speaking, for a "random" choice of a sufficiently smooth mapping f and for every $u \in \mathbb{R}^m$ (thus for a minimizer of J too), every choice of I from $\mathcal{I}(u)$ yields an element of the generalized gradient and, moreover, taking all I from $\mathcal{I}(u)$, we obtain even the whole generalized gradient (of course, after making the convex hull). Supposing that we dispose of a package of, e.g., a bundle algorithm and of a solver for (P_u), our strategy is now evident:

i) for u stated by the bundle algorithm, solve (P_u) by a method that yields also the Lagrange multipliers,

ii) then determine $\mathcal{I}(u)$ from (5) and take some $I \in \mathcal{I}(u)$ (in other words, I contains

obligatorily all strongly active indices, and optionally some semi-active indices),

iii) evaluate $DJ_I(u)$ by the standard adjoint-equation technique,

iv) answer the bundle algorithm by the value of the cost function $J(u)$ and by the gradient information $DJ_I(u)$.

This very simple strategy is a straightforward extension of the standard technique used in smooth optimal control. For its simplicity and efficiency we must pay the only price: in some special situations the procedure may produce element $DJ_I(u)$ that does not belong to $\partial J(u)$, which might cause a failure of the bundle algorithm. On the other hand, such situations will be shown to be rare in a certain sense, and thus we may rely on the algorithm to work without any failure (cf. also Remark 3.2).

It may be also said that our algorithm is based on the fact, that, roughly speaking, for randomly chosen sufficiently smooth f and for every $u \in \mathbb{R}^m$, the set $I_0(u)$ of the semi-active indices contains at most m elements, each of them can become strongly active or non-active (independently of the others) when u moves a little.

The plan of the paper is the following: in Section 2 we deal with the evaluation of $\partial J(u)$ by the outlined way for the case that the mapping f is in a "good" position with respect to the other data of the problem, and then in Section 3 we investigate, by using the deep results of the transversality theory, the question when f is in such position.

2. Ψ-TRANSVERSALITY AND EVALUATION OF $\partial J(u)$

First, in a simple case when A does not depend on u, we explain "geometrically" the principle we will use. Let M_k denote the set of all $w \in \mathbb{R}^n$ such that the projection $\mathrm{Pr}_{K,A}$ of $A^{-1}w$ has at least k semi-active indices. In view of (2) we have obviously $\emptyset = M_{n+1} \subset M_n \subset \ldots M_1 \subset M_0 = \mathbb{R}^n$, and each M_k can be compound with some polyhedral parts of $(n-k)$-dimensional hyper-planes (each M_k looks like the union of several "broken" $(n-k)$-dimensional hyper-planes). This would not enable to use the results of smooth analysis (i.e. the transversality theory). Yet we can enlarge each M_k to $\overline{M}_k$ taking for $\overline{M}_k$ the union of not only the parts, but of all respective hyper-planes. It is important that the dimension of the hyper-planes forming $\overline{M}_k$ is again $n-k$. Now we consider the m-dimensional variety $F = f(\mathbb{R}^m)$. We intuitively feel that, taking f "randomly", there is hardly a chance for $F \cap \overline{M}_k$ to be nonempty for $k > m$. If $k \leqq m$, we may suppose that F intersects $\overline{M}_k$, but again there is hardly a chance that, if $w \in F \cap \overline{M}_k$, the dimension of the space generated by the tangent hyper-planes to F and $\overline{M}_k$ at w is less than $m + n - k$. These intuitive assertions actually hold provided f is smooth enough; cf. the Sard-Brown theorem below. The first assertion says that there is hardly a chance that for any control $u \in \mathbb{R}^m$ the set of the semi-active indices will contain more than m elements. The second assertion will enable to prove simply that these indices become actually either strongly active or non-active in a neighbourhood of u. Thus we can avoid

also somewhat complicated situations with an oscillating nature as in [7; the proof of Theorem 3, the case $R_2 \neq \emptyset$], but the main aim is to avoid the situation when $I_0(u)$ contains more than m indices (recall an example in [6] showing that (6) actually need not be valid in this case). If f is in such a "good" position with respect to the other data A and K, we shall say that f is Ψ-transversal to A and K.

Now we define the Ψ-transversality precisely. We treat straight the case when A depends on u, which requires making considerations not in the space $\mathbb{R}^n$ like in the above heuristical explanation, but in $\mathbb{R}^m \times \mathbb{R}^n$.

Definition 2.1. A triple (I_1, I_2, I_3) will be called admissible if $I_i \subset I_K$ for $i = 1, 2, 3$, $I_i \cap I_j = \emptyset$ for $i \neq j$, and the vectors $\{b_i;\ i \in I_1 \cup I_2 \cup I_3\}$ are linearly independent.

Definition 2.2. A C^1-mapping $f\colon \mathbb{R}^m \to \mathbb{R}^n$ will be called Ψ-transversal with respect to the other data A and K (briefly Ψ-transversal) if for every admissible (I_1, I_2, I_3) and every $u \in \mathbb{R}^m$ such that $\Psi(I_1, I_2, I_3)(u, f(u)) = 0$ we have

$$\text{Range } D\Psi(I_1, I_2, I_3)(u, f(u)) = \mathbb{R}^p \quad \text{with} \quad p = \text{card}(I_2 \cup I_3),$$

where the function $\Psi(I_1, I_2, I_3)\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ is defined by

$$[\Psi(I_1, I_2, I_3)(u, w)]_i = \begin{cases} [(B_I A^{-1}(u) B_I^{\mathrm{T}})^{-1} (c_I + B_I A^{-1}(u) w)]_i \\ \qquad\qquad \text{for} \quad i \in I_2, \\ [(B_I A^{-1}(u) (w - B_I^{\mathrm{T}}(B_I A^{-1}(u) B_I^{\mathrm{T}})^{-1} (c_I + \\ \quad + B_I A^{-1}(u) w)) - c_I]_i \qquad \text{for} \quad i \in I_3, \end{cases}$$

where $I = I_1 \cup I_2$, $[\cdot]_i$ denotes the component with the index i, B_I is the matrix whose rows are just the vectors b_i with $i \in I$, B_I^{T} is its transpose, and c_I is the vector with the components c_i for $i \in I$.

Note that the components of $\Psi(I_1, I_2, I_3)(u, w)$ with $i \in I_2$ are just the corresponding Lagrange multipliers of the projection $\text{Pr}_{K(I), A(u)}$ of $A^{-1}(u) w$, while the components with $i \in I_3$ are equal to $\langle b_i, x \rangle - c_i$ with $x = \text{Pr}_{K(I), A(u)} A^{-1}(u) w$, representing thus the violation of the constraints with $i \in I_3$ that are not included to $K(I)$. Note also that (1) guarantees Ψ to be a C^1-mapping. For simplicity we have considered all admissible triples (I_1, I_2, I_3), not only the triples which actually appear within the projection onto K.

One point should be emphasized: we do not expect that, for a concrete data f, A, and K, the Ψ-transversality of f may be effectively verified, although under some quite restrictive assumptions it is possible (cf. Remark 2.1 below or a special case in [6; Theorem 5.2]). From our computational viewpoint the only important facts are the following:

i) if f is Ψ-transversal, we state an effective procedure to evaluate the Clarke generalized gradient of J (for the bundle algorithm it suffices even to find only one element of its),

ii) the cases when f is not Ψ-transversal are rare in some sense, which justify our algorithm for practical cases (i.e. the Ψ-transversality need not be verified).

The point i) is solved by Theorem 2.1 together with the algorithm outlined in Section 1; the point ii) will be treated in the next section by Theorems 3.1. and 3.2.

Theorem 2.1. Let (1) and (2) hold. If f is Ψ-transversal, then for every $u \in \mathbb{R}^m$:

$$\partial J(u) = \operatorname{co} \{\mathrm{D}J_I(u); I \in \mathscr{I}(u)\}, \tag{6}$$

where J_I and $\mathscr{I}(u)$ are defined by (4) and (5), respectively.

Proof. First, we will prove that $\mathrm{D}J_I(u) \in \partial J(u)$ for every $I \in \mathscr{I}(u)$. We employ the mapping $\Psi(I_1, I_2, I_3)$: $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ defined above, taking $I_1 = I_a(u)$, $I_2 = I \setminus I_a(u)$, and $I_3 = I_K \setminus (I_n(u) \cup I)$. For brevity we denote this mapping as $\psi_{I,u}$: $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$. Of course, $p = \operatorname{card}(I_2 \cup I_3) = \operatorname{card}(I_0(u))$, where $I_0(u) = I_K \setminus (I_a(u) \cup I_n(u))$ is clearly the set of the semi-active indices of the projection in question. Note that, in view of (2), the triple of the index sets used for the definition of $\psi_{I,u}$ is admissible for any $I \in \mathscr{I}(u)$. Obviously, $\operatorname{Pr}_{K,A(u)} A^{-1}(u) f(u) = \operatorname{Pr}_{K(I),A(u)} A^{-1}(u) f(u)$, the Lagrange multipliers with the indices from I of both of these projections being the same. Since the multipliers λ_i as well as the residua r_i of the former projection are equal to zero for every $i \in I_0(u)$, we see that

$$\psi_{I,u}(u, f(u)) = 0 .$$

Thanks to the Ψ-transversality, $v \mapsto \psi_{I,u}(v, f(v))$ is a C^1-mapping with a surjective derivative at u. Hence we may use the well-known inverse mapping theorem (see e.g. [1; Chap. 2, § 1]), and choose a sequence $\{u_k\}$ such that $u_k \to u$ and

$$\forall u_k: \psi_{I,u}(u_k, f(u_k)) > 0 ,$$

i.e. each component is positive (if $m = p$, then use the inverse mapping theorem directly, and if $m > p$, the first employ an arbitrary C^1-mapping φ: $\mathbb{R}^m \to \mathbb{R}^{m-p}$ such that the mapping $v \mapsto (\psi_{I,u}(v, f(v)), \varphi(v))$ has a surjective derivative at u, and afterwards apply the mentioned theorem). Thus we observe that, if u_k is sufficiently near to u, all the components of the vector $(B_I A^{-1}(u_k) B_I^{\mathrm{T}})^{-1} (c_I + B_I A^{-1}(u_k) . f(u_k))$ are positive (note that these components are just the multipliers of the projection $\operatorname{Pr}_{K(I),A(u_k)}$ of $A^{-1}(u_k) f(u_k)$), and also $\langle b_i, \operatorname{Pr}_{K(I),A(u_k)} A^{-1}(u_k) f(u_k) \rangle - c_i > 0$ for every $i \in I_K \setminus I$. From these facts we conclude that $\operatorname{Pr}_{K(I),A(u_k)} A^{-1}(u_k) f(u_k) = \operatorname{Pr}_{K,A(u_k)} A^{-1}(u_k) f(u_k)$ and, moreover, $I_a(u_k) = I$ and $I_n(u_k) = I_K \setminus I$. Particularly, $I_0(u_k) = \emptyset$ and therefore J is differentiable at u_k with $\mathrm{D}J(u_k) = \mathrm{D}J_I(u_k)$. Since J_I is continuously differentiable, the sequence $\mathrm{D}J_I(u_k)$ has a limit, and we have clearly

$$\lim_{u_k \to u} \mathrm{D}J(u_k) = \mathrm{D}J_I(u) .$$

From the definition of ∂J it immediately follows that $\mathrm{D}J_I(u) \in \partial J(u)$. Because of the convexity of $\partial J(u)$, we have proved the inclusion

$$\operatorname{co} \{\mathrm{D}J_I(u);\ I \in \mathscr{I}(u)\} \subset \partial J(u) .$$

To prove the converse inclusion, we consider a neighbourhood M of the set $\{\mathrm{D}J_I(u); I \in \mathscr{I}(u)\}$. Since J_I are C^1-functions, we can take a neighbourhood N_u of u such that $\mathrm{D}J_I(v) \in M$ whenever $v \in N_u$ and $I \in \mathscr{I}(u)$. Moreover, taking N_u sufficiently small, we may suppose that $I_a(u) \subset I_a(v)$ and $I_n(u) \subset I_n(v)$ for all $v \in N_u$ as a consequence of the continuity of the mappings $v \mapsto r_i(v)$ and $v \mapsto \lambda_i(v)$. Therefore, $I_0(v) \subset I_0(u)$ and also $\mathscr{I}(v) \subset \mathscr{I}(u)$. Now, we will investigate the set $S = \{v \in N_u; I_0(v) \neq \emptyset\}$. For every $v \in N_u$, $i \in I_0(v)$, and $I \in \mathscr{I}(v)$, we have $[\psi_{I,u}(v, f(v))]_i = 0$ because the values r_i as well as the multipliers λ_i with $i \in I_0(v)$ of the projection $\mathrm{Pr}_{K(I \cup I_0(v)),A(v)}$ of $A^{-1}(v) f(v)$ are equal to zero and $[\psi_{I,v}(v, f(v))]_i = [\psi_{I,u}(v, f(v))]_i$. Hence the following inclusion is obviously valid:

$$S \subset \bigcup_{\substack{I \in \mathscr{I}(u) \\ i \in I_0(u)}} S_{I,i}$$

where $S_{I,i} = \{v \in N_u; [\psi_{I,u}(v, f(v))]_i = 0\}$. Employing again the Ψ-transversality, which implies local surjectivity of the C^1-mapping $v \mapsto \psi_{I,u}(v, f(v))$ at the point u, and taking N_u sufficiently small, we conclude that $S_{I,i}$ has the Lebesgue measure zero in $\mathbb{R}^m$ (use the well-known implicit function theorem to construct a C^1-function from a subset of $\mathbb{R}^{m-1}$ to $\mathbb{R}$ whose graph is just $S_{I,i}$). Thus also S has zero measure. We have already explained that J is differentiable on $N_u \setminus S$ and, considering $v \in N_u \setminus S$, we have got $\mathrm{D}J(v) = \mathrm{D}J_I(v)$ with $I = I_a(v) = I_K \setminus I_n(v)$ (then obviously $\mathscr{I}(v) = \{I\}$). Therefore $\mathrm{D}J(v) = \mathrm{D}J_I(v)$ for some $I \in \mathscr{I}(u)$ and we observe that $\mathrm{D}J(v) \in M$ whenever $v \in N_u \setminus S$. Since M has been an arbitrary neighbourhood of $\{\mathrm{D}J_I(u); I \in \mathscr{I}(u)\}$, we obtain the estimate

$$\{\lim_{u_k \to u} \mathrm{D}J(u_k);\ u_k \in N_u \setminus S \text{ and the limit exists}\} \subset$$
$$\subset \{\mathrm{D}J_I(u); I \in \mathscr{I}(u)\}\ .$$

This inclusion is preserved for the convex hulls as well. Since S has zero measure, from Thm. 2.5.1 in [2] it immediately follows that

$$\mathrm{co}\,\{\lim_{u_k \to u} \mathrm{D}J(u_k);\ u_k \in N_u \setminus S \text{ and the limit exists}\} = \partial J(u)\,,$$

which completes the proof. □

Remark 2.1. There is one particular case in which the Ψ-transversality is simply ensured: the matrix A does not depend on $u \in \mathbb{R}^m$ and the mapping f has a surjective derivative, i.e. $\mathrm{Range}\,\mathrm{D}f(u) = \mathbb{R}^n$ for all $u \in \mathbb{R}^m$, This case has been investigated by a slightly different technique in [6; § 5]. For the special case that $f: \mathbb{R}^n \to \mathbb{R}^n$ is identity cf. also K. Malanowski [9; § 4]. The surjectivity condition is, however, very restrictive: e.g. it cannot be fulfilled when $m < n$.

Remark 2.2. The qualification hypothesis (2) excludes equality constraints (e.g. $\langle b_i, x\rangle \geqq c_i$ with $b_1 = -b_2$ and $c_1 = -c_2$), but such constraints are not important from the viewpoint of nonsmooth analysis. Also situations of the type $K = \{x \in \mathbb{R}^n; \langle b_i, x\rangle \geqq 0\ \forall i \in I_K\}$, where I_K has more than n indices, are not allowed

because, in such situations, the analysis of local behaviour of J must be performed finer than it has been by (6). Nevertheless, it seems that in tasks arising by discretization of unilateral problems for partial differential equations the latter situation will not appear.

3. GENERICITY OF THE Ψ-TRANSVERSALITY

In this section we apply Thom's and Sard-Brown's theorems to show that, under some additional assumptions, the cases when f is not Ψ-transversal are "rare". The theory we will use is usually referred to as transversality theory; for a survey see [1; Chap. 2, §§ 6, 7].

Here we briefly present some definitions and assertions from the transversality theory, generality being restricted to our special case for the sake of simplicity. Let $P(f)$ be a statement about the points f of a complete metric space. V. We say that $P(f)$ is a generic property on V if the set $\{f \in V;\ P(f) \text{ is true}\}$ contains a dense G_δ subset of V. A C^1-mapping $\varphi: X \to \mathbb{R}^p$, X is a Banach space, is said to be transversal to a linear subspace $M \subset \mathbb{R}^p$ at a point $\chi \in X$ if either $\varphi(\chi) \notin M$ or $\mathrm{D}\varphi(\chi)(X) + M = \mathbb{R}^p$. Furthermore, φ is called transversal to M if it is transversal to M at every $\chi \in X$; in other words, $\mathrm{D}\varphi(\chi)(X) + M = \mathbb{R}^p$ whenever $\chi \in X$, $\varphi(\chi) \in M$. If V is a Banach space, $\Phi: \mathbb{R}^m \times V \to \mathbb{R}^p$ is a C^k-mapping (i.e. k-times continuously differentiable) with $k \geqq \max(1, m + 1 - \operatorname{codim} M)$ and Φ is transversal to M, then the statement $P(f) = \{\Phi(\cdot, f) \text{ is transversal to } M\}$ is a generic property on V. This assertion (in somewhat more general form) is known as Thom's transversality theorem. The space of all C^k-mappings from X to Y will be denoted by $C^k(X, Y)$.

Theorem 3.1. Let (2) hold, $A \in C^k(\mathbb{R}^m, \mathbb{R}^n \times \mathbb{R}^n)$ for some $k \geqq m$, $A(u)$ be symmetric positive definite for all $u \in \mathbb{R}^m$. Then the statement $P(f) = \{f \text{ is } \Psi\text{-transversal}\}$ is a generic property on $C^k(R^m, R^n)$.

Proof. We take an admissible triple (I_1, I_2, I_3) and investigate the mapping $\Phi: \mathbb{R}^m \times C^k(\mathbb{R}^m, \mathbb{R}^n) \to \mathbb{R}^p$, defined by $\Phi(u, f) = \Psi(I_1, I_2, I_3)(u, f(u))$ with $p = \operatorname{card}(I_2 \cup I_3)$. Denote by $\mathrm{D}_f\Phi$ the partial derivative of Φ with respect to f. It is easy to see that, for any $u_0 \in \mathbb{R}^m$, the mapping $\Phi(u_0, \cdot): C^k(\mathbb{R}^m, R^n) \to \mathbb{R}^p$ is affine, its derivative at a point f_0 is equal to $\mathrm{D}_f\Phi(u_0, f_0)$, and $D_f\Phi(u_0, f_0)(f) = G(u_0) f(u_0)$, where $G(u_0)$ is the matrix of the rank $p \times n$, whose ith row is equal to ith row of the matrix $H = (B_I A^{-1}(u_0) B_I^{\mathrm{T}})^{-1} B_I A^{-1}(u_0)$ if $i \in I_2$, and to the vector $v_i = A^{-1}(u_0) b_i - A^{-1}(u_0) B_I^{\mathrm{T}} H b_i$ if $i \in I_3$ (where $I = I_1 \cup I_2$). The matrix $A^{-1}(u_0) B_I^{\mathrm{T}}(B_I A^{-1}(u_0) . B_I^{\mathrm{T}})^{-1} B_I$ represents the projector with respect to the metric induced by $A(u_0)$ onto the space generated by $\{A^{-1}(u_0) b_i;\ i \in I\}$, from which immediately follows that the matrix H has linearly independent rows (recall that b_i, $i \in I$, are linearly independent since the triple (I_1, I_2, I_3) is admissible). Now, suppose that, for some $i \in I_3$, v_i is a linear combination of the vectors v_j, $j \in I_3 \setminus \{i\}$, and the rows of H. In other words, for every $y \in \mathbb{R}^n$, $\langle v_j, y\rangle = 0, j \in I_3 \setminus \{i\}$, $Hy = 0$ implies $\langle v_i, y\rangle = 0$.

Obviously, $\langle v_j, y\rangle = \langle A^{-1}(u_0)\, b_j, y\rangle$ for every $j \in I_3$ and y such that $Hy = 0$. Moreover, since the matrix $B_I\, A^{-1}(u_0)\, B_I^{\mathrm{T}}$ is regular, the condition $Hy = 0$ is equivalent to $\langle A^{-1}(u_0)\, b_j, y\rangle = 0$ for each $j \in I$. Consequently, for every $y \in \mathbb{R}^n$, $\langle A^{-1}(u_0)\, b_j, y\rangle = 0$ for each $j \in I \cup I_3 \setminus \{i\}$ should imply $\langle A^{-1}(u_0)\, b_i, y\rangle = 0$, which is obviously impossible because the vectors b_j, $j \in I \cup I_3$, are supposed to be linearly independent. Consequently, $G(u_0)$ has linearly independent rows and Range $\mathrm{D}_f = \mathbb{R}^p$ at every point (u_0, f_0).

As A is a C^k-mapping, $\Phi(\cdot, f) \in C^k(\mathbb{R}^m, \mathbb{R}^p)$ for every $f \in C^k(\mathbb{R}^m, \mathbb{R}^n)$. Since $\Phi(u, \cdot)$ is an affine (and hence also C^k-) mapping, we see that Φ is globally a C^k-mapping on its domain. Since Range $\mathrm{D}_f\Phi = \mathbb{R}^p$, Φ is transversal to every linear subspace $M \subset \mathbb{R}^p$. We take $M = \{0\}$. Due to Thom's transversality theorem and Lemma 6 from [1; Chapt. 2, § 6], the statement $Q(I_1, I_2, I_3)(f) = \{\Phi(\cdot, f)$ is transversal to $\{0\}$, where Φ is constructed by means of (I_1, I_2, I_3) as described above$\}$ is a generic property on $C^k(\mathbb{R}^m, \mathbb{R}^n)$ provided $k \geqq \max(1, m - p + 1)$. This condition is fulfilled since $p \geqq 1$ (the case $p = 0$ is not interesting) and $k \geqq m$ (see the assumptions).

However, saying that f is Ψ-transversal means precisely that $Q(I_1. I_2, I_3)(f)$ is true for every admissible triple (I_1, I_2, I_3). Since I_K has been supposed as finite, the collection of all admissible triples is finite too, and the genericity is preserved for the Ψ-transversality as well (because, in a complete metric space, the intersection of a finite number of dense G_δ subsets is again a dense G_δ subset). □

The genericity works on infinite-dimensional spaces like $C^k(\mathbb{R}^m, \mathbb{R}^n)$ where no analogue with the Lebesgue measure can be defined. However, on finite-dimensional spaces it may happen that a statement is generic, being false almost everywhere (a.e.). Nevertheless, if we confine ourselves to A independent of u and use the Sard-Brown theorem, we can show that the Ψ-transversality holds a.e. on certain finite-dimensional affine submanifolds of $C^k(\mathbb{R}^m, \mathbb{R}^n)$. The mentioned Sard-Brown theorem sounds as follows (cf. [1; Chapt. 2, § 7, Thm. 1]): If $\varphi \in C^k(\mathbb{R}^m, \mathbb{R}^p)$, $k \geqq \max(1, m - p + 1)$, then the set of critical values of φ has Lebesgue measure zero in $\mathbb{R}^p$. We recall that $q \in \mathbb{R}^p$ is a critical value of φ if there is $u \in \mathbb{R}^m$ such that Range $\mathrm{D}\varphi(u) \neq \mathbb{R}^p$ and $\varphi(u) = q$.

Theorem 3.2. Let (2) hold, $f \in C^m(\mathbb{R}^m, \mathbb{R}^n)$, and A symmetric positive definite independent of $u \in \mathbb{R}^m$. Then $f + a$ is Ψ-transversal for a.a. $a \in \mathbb{R}^n$.

Proof. Take an admissible triple (I_1, I_2, I_3). Since A does not depend on u, we may write $\Psi(I_1, I_2, I_3)(u, w) = Gw + g$, where the matrix G, now independent of u, has been already derived in the proof of Theorem 3.1, and g is a vector from $\mathbb{R}^p$, $p = \mathrm{card}\,(I_2 \cup I_3)$. Due to the Sard-Brown theorem, the set of the critical values of $Gf + g\colon \mathbb{R}^m \to \mathbb{R}^p$ has the Lebesgue measure zero in $\mathbb{R}^p$ (note that f is smooth enough because $p \geqq 1$; the case $p = 0$ is not interesting). Clearly, $q \in \mathbb{R}^p$ is a critical value of $Gf + g$ iff there is $a \in \mathbb{R}^n$ such that $Ga = q$ and zero is the critical value of $G(f + a) + g$. Since G has linearly independent rows (see the proof of Thm. 3.1), the set $\{a \in \mathbb{R}^n;\, 0$ is the critical value of $G(f + a) + g\}$ has the Lebesgue measure

zero in $\mathbb{R}^n$. However, saying that $f + a$ is not Ψ-transversal means precisely that zero is the critical value of $G(f + a) + g$ for some admissible triple (I_1, I_2, I_3) by means of which G and g has been constructed. Since there is only a finite number of possibilities how to choose (I_1, I_2, I_3) and the union of a finite number of sets with a zero measure has again a zero measure, the assertion is proved. □

Remark 3.1. Since the number m of parameters to be optimized is usually large, the assumptions of Theorems 3.1 and 3.2 practically force us to use C^∞-mappings for A and f (while j is to be only C^1-function).

Remark 3.2. Roughly speaking, Theorems 3.1 and 3.2 assert that for a randomly chosen f, we may expect f to be Ψ-transversal. On the other hand, we must be very cautious if there is a symmetry in the problem because then f cannot be considered as chosen randomly. In such case the index set I used for evaluation of an element of $\partial J(u)$ must be taken not only to fulfil the condition $I_a(u) \subset I \subset I_K \setminus I_n(u)$, but also to preserve the symmetry of the problem. Thus here it is particularly recommandable to exploit as much symmetry as possible to treat smaller problem without any symmetry.

ACKNOWLEDGEMENT

The author is indebted to Dr. J. V. Outrata for many valuable comments.

(Received December 17, 1987.)

REFERENCES

[1] J.-P. Aubin and I. Ekeland: Applied Nonlinear Analysis. J. Wiley, New York 1984.
[2] F. H. Clarke: Optimization and Nonsmooth Analysis. J. Wiley, New York 1983.
[3] V. F. Demyanov: Quasidifferentiable functions: Necessary conditions and descent directions. Math. Programming Study *29* (1986), 20—43.
[4] R. Glowinski, J. L. Lions and R. Trémolieres: Analyse Numérique des Inéquations Variationnelles. Dunod, Paris 1976.
[5] W. W. Hager: Lipschitz continuity for constrained processes. SIAM J. Control Optim. *17* (1979), 321—338.
[6] J. Haslinger and T. Roubíček: Optimal control of variational inequalities. Approximation theory and numerical realization. Appl. Math. Optim. *14* (1986), 187—201.
[7] K. Jittorntrum: Solution point differentiability without strict complementarity in nonlinear programming. Math. Programming Study *21* (1984), 127—138.
[8] C. Lemaréchal, J. J. Strodiot and A. Bihain: On a bundle algorithm for nonsmooth optimization. In: Nonlinear Programming 4 (O. L. Mangasarian et al., eds.). Academic Press, New York 1981.
[9] K. Malanowski: Differentiability with respect to parameters of solutions to convex programming problems. Math. Programming *33* (1985), 352—361.
[10] B. T. Polak: Introduction to Optimization. Nauka, Moscow 1983. In Russian.

Ing. Tomáš Roubíček, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.