

NECESSARY CONDITIONS FOR LIPSCHITZ CONTINUOUS DISCRETE CONTROL PROBLEMS

JAROSLAV DOLEŽAL

*Department of Electrical and Computer Engineering, Oregon State University,
Corvallis, Oreg., U.S.A.*

and

*Institute of Information Theory and Automation, Czechoslovak Academy of Sciences,
Prague, Czechoslovakia*

1. Introduction

At present time there exist a number of results dealing with the theory of discrete optimal control systems. For all let us recall at least the references [1], [2], [6], [10], where various classes of discrete systems are studied in detail and the respective necessary optimality conditions are presented. As a rule, the original discrete control problem is interpreted as a mathematical programming one, and the existing results in this area are then applied to derive necessary conditions for the original control problem. Application of the classical results of the mathematical programming theory usually requires to assume the continuous differentiability of problem describing functions. Moreover, if the aim is to obtain necessary conditions in the form of a discrete maximum principle, then one cannot avoid some additional assumptions concerning the convexity of the problem in question.

In recent years a considerable progress was achieved in the direction of releasing the rather stringent differentiability assumption. Most of the respective theory deals with the so-called locally Lipschitz functions. The basic theory including the definition of generalized gradients can be found in [3]–[5], [11]–[15]. In this way also the classical results of the mathematical programming theory were generalized to the case of locally Lipschitz functions.

Our aim is to apply some of these results to discrete optimal control problems and to derive necessary conditions for intrinsically “nonsmooth”

discrete control problems described by locally Lipschitz functions. To do this, first some basic definitions and results are summarized for convenience. It is shown that in the general case only rather formal necessary conditions can be deduced which, however, are still of practical and theoretical interest. More familiar form of necessary conditions is obtained under some additional assumptions which, on the other hand, considerably narrow the class of studied discrete optimal control problems. The main reason is a somewhat unsatisfactory definition of partial generalized gradients, and this fact turns out to be crucial for this application. Therefore, an alternative approach is suggested, based on a modification of the definition of partial generalized gradients. Then it is possible to deal also with the general case in a straightforward way and without additional assumptions. Some details in this respect were given by the author in [7]–[9].

2. Preliminary results

Throughout the paper we shall work with the class of locally Lipschitz real functions. This means that for every bounded set $Q \subset R^n$, there exists a constant c such that for all $x, y \in Q$

$$(1) \quad |f(x) - f(y)| \leq c \|x - y\|.$$

It is known that functions of this type have almost everywhere the derivative ∇f (gradient). It will be assumed that all functions appearing further are locally Lipschitz (in the vector case component-wise).

DEFINITION 1. The *generalized gradient of a function* $f: R^n \rightarrow R^1$ at x , denoted by $\partial f(x)$, is the set (co denotes the convex hull)

$$\partial f(x) = \text{co} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) \mid x_i \rightarrow x \right\},$$

with f differentiable at x_i for each i .

It can be easily shown that $\partial f(x)$ is a nonempty compact convex set in R^n . Many interesting properties, equivalent definitions and generalization of this concept to Banach spaces are listed in the mentioned references [3]–[5], [11]–[15], where also locally Lipschitz mathematical programming problems are studied from the point of view of necessary conditions. For convenience, let us include some basic results here, which will be used in the sequel. All vectors are treated as column-vectors for the sake of simplicity and, as usual, T denotes the transposition.

Let $c \in R^1$ and let $f, g: R^n \rightarrow R^1$ be locally Lipschitz. Then

$$(2) \quad \partial(cf)(x) = c \partial f(x),$$

$$(3) \quad \partial(f+g)(x) \subset \partial f(x) + \partial g(x),$$

$$(4) \quad f \text{ has a local extremum at } x \Rightarrow 0 \in \partial f(x).$$

If now $Q \subset R^n$ is nonempty and closed, denote by $d_Q(x)$ the real function giving the distance of x to Q . It is not difficult to prove [3] that the function d_Q is Lipschitz. Then as the *tangent cone to Q at the point $x \in Q$* let us define the set

$$(5) \quad T(Q; x) = \{v \in R^n \mid v^T q \leq 0 \text{ for all } q \in \partial d_Q(x)\},$$

i.e., $T(Q; x)$ is a polar cone to $\partial d_Q(x)$. The *normal cone $N(Q; x)$ at x* can be then defined as a polar cone to $T(Q; x)$, or directly as

$$(6) \quad N(Q; x) = \text{cl}\{\gamma q \mid \gamma > 0, q \in \partial d_Q(x)\},$$

i.e., $N(Q; x)$ is the closure of the conical hull of $\partial d_Q(x)$.

Consider further a problem of minimization of a function $f: R^n \rightarrow R^1$ subject to the constraints $x \in Q \subset R^n$, and $h_i(x) = 0$, $i = 1, \dots, p$, and $g_j(x) \leq 0$, $j = 1, \dots, q$, where h_i and g_j are real functions on R^n , and all the functions are locally Lipschitz. According to [14], if \hat{x} is the minimizing point, then there exist a number $\mu \leq 0$, multipliers ψ_1, \dots, ψ_p , and ν_1, \dots, ν_q , with $\nu_j \leq 0$, $j = 1, \dots, q$, not all zero, such that

$$(7) \quad \partial\left(\mu f + \sum_{i=1}^p \psi_i h_i + \sum_{j=1}^q \nu_j g_j\right)(\hat{x}) \cap N(Q; \hat{x}) \neq \emptyset,$$

with $\nu_j g_j(\hat{x}) = 0$, $j = 1, \dots, q$.

For $Q \subset R^n$, denote by $\text{int}Q$ the interior of the set Q in R^n . Further, let Q_1 and Q_2 be nonempty and closed sets in R^n and let $x \in Q_1 \cap Q_2$ be a point for which $T(Q_1; x) \cap \text{int}T(Q_2; x) \neq \emptyset$. Then, according to [15], one has

$$(8) \quad N(Q_1 \cap Q_2; x) \subset N(Q_1; x) + N(Q_2; x).$$

Further details and consequences are given in the mentioned references, e.g. also the conditions that guarantee $\mu \neq 0$ (constraint qualification). Most of the existing results are moreover valid in Banach spaces, however certain precautions are to be observed. Here we deal exclusively with R^n context.

Certain drawback of the existing theory, having in mind the later application, is the definition of partial generalized gradients [5]. To do this, let $f: R^n \times R^m \rightarrow R^1$ be locally Lipschitz. For each $x \in R^n$ denote the generalized gradient of the function $f(x, \cdot)$ by $\partial_y f(x, y)$, and in a similar way also for $\partial_x f(x, y)$. Such a definition of partial generalized gradients seems to be quite obvious and reasonable. However, there is no relationship between the sets $\partial f(x, y)$ and $\partial_x f(x, y) \times \partial_y f(x, y)$, as there are simple examples showing that neither of these sets is contained in the other.

This is also the main reason that after giving formal necessary conditions for the general case, a more special class of discrete control problems is studied in detail. Namely, the system in question will be assumed

"additive" in state and control variables, which overcomes the indicated difficulty and the resulting necessary optimality conditions assumes a more familiar form of [1], [2], [6], [10].

3. Discrete optimal control problem

Let us consider a discrete dynamical system described by the relations (K denotes the prescribed number of stages, $x \in R^n$ —the state, and $u \in R^m$ —the control)

$$(9) \quad x_{k+1} = f_k(x_k, u_k), \quad k = 0, 1, \dots, K-1,$$

$$(10) \quad \begin{bmatrix} x_k \\ u_k \end{bmatrix} \in M_k \subset R^n \times R^m, \quad k = 0, 1, \dots, K-1,$$

$$(11) \quad x_K \in A_K \subset R^n.$$

The aim is to minimize the functional

$$(12) \quad J = g(x_K) + \sum_{k=0}^{K-1} h_k(x_k, u_k).$$

As above, it is assumed that all functions are locally Lipschitz and all sets are nonempty and closed. Here $f_k: R^n \times R^m \rightarrow R^n$, $h_k: R^n \times R^m \rightarrow R^1$ and $g: R^n \rightarrow R^1$.

It is easy to see that the above problem (9)–(12) represents a mathematical programming problem in the space of the dimension $mK + n(K+1)$, i.e., we have to work with the variable $z = (x_0, x_1, \dots, x_K, u_0, u_1, \dots, u_{K-1})^T$. The special structure of this mathematical programming problem makes it possible to decompose it with respect to the discrete time variable k . To derive the theorem stated below one has to realize that if a function f does not depend on a certain variable, the corresponding component of all vectors belonging to ∂f is zero. Moreover, if $Q \subset R^n$ and $x \in Q$, the normal cone $N(Q; x)$ is also the normal cone of the set $Q \times R^m$ in $R^n \times R^m$ at (x, y) , $y \in R^m$. Then relations (7)–(8) are applied together with properties (2)–(3) to the overall mathematical programming problem indicated above. The obvious details of this procedure are omitted.

THEOREM 1. *If $\hat{x}_0, \hat{x}_1, \dots, \hat{x}_K$ and $\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{K-1}$ is a solution of the discrete optimal control problem (9)–(12) and if $\text{int} T(M_k; (\hat{x}_k, \hat{u}_k)) \neq \emptyset$, $k = 0, 1, \dots, K-1$, then there exists a number $\mu \leq 0$ and vectors $\lambda_k \in R^n$, $k = 1, \dots, K$, not all zero, such that*

$$\partial_{x,u} H_{k+1}(\hat{x}_k, \hat{u}_k) \cap \begin{bmatrix} \lambda_k \\ 0 \end{bmatrix} + N(M_k; (\hat{x}_k, \hat{u}_k)) \neq \emptyset, \quad k = 0, 1, \dots, K-1,$$

with $\lambda_0 = 0$, and

$$\mu \partial_{x,u} g(\hat{x}_K) \cap \lambda_K + N(A_K; \hat{x}_K) \neq \emptyset,$$

where, as usual,

$$H_{k+1}(x, u) := \mu h_k(x, u) + \lambda_{k+1}^T f_k(x, u), \quad k = 0, 1, \dots, K-1.$$

One can observe an evident analogy of the obtained necessary conditions with those of [1], [2], [6], [10]. However, because of the indicated property of partial generalized gradients ∂_x and ∂_u , one cannot simply decompose the "adjoint" condition in the theorem as desirable to obtain separate relations for \hat{x} and \hat{u} . So the given composed form of the generalized gradient $\partial_{x,u}$ must be maintained in this general case. Therefore also the general implicit constraints of the "mixed" type (10) were assumed to allow more general formulation. The fairly general form of the stated necessary conditions, although interesting from a theoretical point of view, seems not to possess much practical impact and a more concrete form is desirable.

4. Special class of problems

One way to overcome the encountered difficulty is to impose some additional assumptions on the studied control problem. This will evidently narrow the class of covered problems, but, on the other hand, it will enable to refine the above results in the intended way. First, one can simply assume the so-called subdifferential regularity [15] of functions f_k , h_k , and g . Then one has, e.g., that $\partial_{x,u} f_k(x, u) \subset \partial_x f_k(x, u) \times \partial_u f_k(x, u)$, and this fact makes the required decomposition possible. Still under this rather stringent assumption the class of treated problems is of practical importance including nonsmooth and nonconvex problems having the max type objective function.

Other possibility is to assume the "additive" structure of f_k and h_k . The dimensions of functions introduced further are the same as in problem (9)–(12). In such case the necessary gradient inclusion indicated above is preserved. Thus

$$(13) \quad x_{k+1} = f_k^1(x_k) + f_k^2(u_k), \quad k = 0, 1, \dots, K-1,$$

$$(14) \quad x_k \in A_k \subset R^n, \quad k = 0, 1, \dots, K,$$

$$(15) \quad u_k \in U_k \subset R^m, \quad k = 0, 1, \dots, K-1.$$

The aim is to minimize the functional

$$(16) \quad J = g(x_K) + \sum_{k=0}^{K-1} (h_k^1(x_k) + h_k^2(u_k)).$$

Again it is assumed that all functions are locally Lipschitz and all sets nonempty and closed. In this case Theorem 1 takes the following form.

THEOREM 2. *If $\hat{x}_0, \hat{x}_1, \dots, \hat{x}_K$ and $\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{K-1}$ is a solution of the discrete optimal control problem (13)–(16) with $\text{int}T(A_k; \hat{x}_k) \neq \emptyset$ and $\text{int}T(U_k; \hat{u}_k) \neq \emptyset$, $k = 0, 1, \dots, K-1$, then there exists a number $\mu \leq 0$ and vectors $\lambda_k \in R^n$, $k = 1, \dots, K$, not all zero, such that*

(a) *the vectors λ_k satisfy the relations*

$$\partial_x(\mu h_k^1 + \lambda_{k+1}^T f_k^1)(\hat{x}_k) \cap \lambda_k + N(A_k; \hat{x}_k) \neq \emptyset, \quad k = 0, 1, \dots, K-1,$$

with $\lambda_0 = 0$ and

$$\mu \partial_x g(x_K) \cap \lambda_K + N(A_K; \hat{x}_K) \neq \emptyset,$$

(b) *the optimal control sequence satisfies the relations*

$$\partial_u(\mu h_k^2 + \lambda_{k+1}^T f_k^2)(\hat{u}_k) \cap N(U_k; \hat{u}_k) \neq \emptyset, \quad k = 0, 1, \dots, K-1.$$

One can see that the promised analogy with classical problems studied in [1], [2], [6], [10] is more apparent, because the adjoint system (a) and the optimality conditions (b) are separated.

5. An alternative approach

In this section let us briefly explore the possible alternative definition of partial generalized gradients, which is especially suitable for our decomposition purpose. This approach is described in more detail in [8] and the subsequent theory will appear in [9]. To overcome the indicated difficulty connected with the earlier definition of partial generalized gradients, several attempts can be found in the existing results. Let us mention the idea of [14] which inspired the results given in [8] and is briefly described in the sequel. Let us only recall the basic definition.

DEFINITION 2. Let $f: R^n \times R^m \rightarrow R^1$ be locally Lipschitz. As *partial generalized gradients of f with respect to the first, resp. the second, variable at (x, y)* we define the sets

$$\hat{\partial}_x f(x, y) = \text{co} \left\{ \lim_{i \rightarrow \infty} \nabla_x f(x_i, y_i) \mid (x_i, y_i) \rightarrow (x, y) \right\},$$

resp.

$$\hat{\partial}_y f(x, y) = \text{co} \left\{ \lim_{i \rightarrow \infty} \nabla_y f(x_i, y_i) \mid (x_i, y_i) \rightarrow (x, y) \right\},$$

with f differentiable at (x_i, y_i) for each i .

It follows directly from this definition that

$$(17) \quad \partial f(x, y) \subset \hat{\partial}_x f(x, y) \times \hat{\partial}_y f(x, y).$$

In fact it can be shown [9] that:

$$(18) \quad \hat{\partial}_x f(x, y) = \text{pr}_x[\partial f(x, y)],$$

where pr_x denotes the projection on the x -axis, and analogously,

$$(19) \quad \hat{\partial}_y f(x, y) = \text{pr}_y[\partial f(x, y)].$$

It would be possible to derive the analogical theorems to those resulting from the original definition of the generalized gradient [5]. Here let us only recall that there is a direct analogy with (2)–(4). Namely, for $c \in \mathbb{R}^1$ and $f, g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^1$ locally Lipschitz, one obtains

$$(20) \quad \hat{\partial}_i(cf)(x, y) = c\hat{\partial}_i f(x, y), \quad i = x, y,$$

$$(21) \quad \hat{\partial}_i(f+g)(x, y) \subset \hat{\partial}_i f(x, y) + \hat{\partial}_i g(x, y), \quad i = x, y,$$

$$(22) \quad f \text{ has a local extremum at } (x, y) \Rightarrow 0 \in \hat{\partial}_i f(x, y), \quad i = x, y.$$

Moreover, as it was to be expected, in the general case,

$$(23) \quad \partial_i f(x, y) \subset \hat{\partial}_i f(x, y), \quad i = x, y.$$

Now we are in a position to formulate necessary optimality conditions for the general case of a discrete control problem described as follows:

$$(24) \quad x_{k+1} = f_k(x_k, u_k), \quad k = 0, 1, \dots, K-1,$$

$$(25) \quad x_k \in A_k \subset \mathbb{R}^n, \quad k = 0, 1, \dots, K,$$

$$(26) \quad u_k \in U_k \subset \mathbb{R}^m, \quad k = 0, 1, \dots, K-1,$$

$$(27) \quad J = g(x_K) + \sum_{k=0}^{K-1} h_k(x_k, u_k).$$

Let us immediately formulate the final result. Again, $H_{k+1}(x, u)$ will have the same meaning as in Theorem 1. All functions are defined as in (9)–(12) and all sets are assumed to be nonempty and closed.

THEOREM 3. *If $\hat{x}_0, \hat{x}_1, \dots, \hat{x}_K$ and $\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{K-1}$ is a solution of the discrete optimal control problem (24)–(27) with $\text{int}T(A_k; \hat{x}_k) \neq \emptyset$ and $\text{int}T(U_k; \hat{u}_k) \neq \emptyset$, $k = 0, 1, \dots, K-1$, then there exists a number $\mu \leq 0$ and vectors $\lambda_k \in \mathbb{R}^n$, $k = 1, \dots, K$, not all zero, such that*

(a) *the vectors λ_k satisfy the relations*

$$\hat{\partial}_x H_{k+1}(\hat{x}_k, \hat{u}_k) \cap \lambda_k + N(A_k; \hat{x}_k) \neq \emptyset, \quad k = 0, 1, \dots, K-1,$$

with $\lambda_0 = 0$ and

$$\mu \hat{\partial}_x g(\hat{x}_K) \cap \lambda_K + N(A_K; \hat{x}_K) \neq \emptyset,$$

(b) *the optimal control sequence satisfies the relations*

$$\hat{\partial}_u H_{k+1}(\hat{x}_k, \hat{u}_k) \cap N(U_k; \hat{u}_k) \neq \emptyset, \quad k = 0, 1, \dots, K-1.$$

To avoid more complicated notation, no explicit constraints given as a system of equalities and/or inequalities were taken into account in the above formulation. Owing to (8), such constraints can be always

included in the same way as described in [1], [2], [6], [10] for continuously differentiable cases. Sometimes the aim is also to have maximum principle formulation of necessary conditions. This tool, powerful when dealing with continuous time systems, seems not to be of such importance in the discrete case. The main reason is the need of additional convexity assumption with only a relatively small gain for the solution of practical problems. Some indications in this respect in the connection with locally Lipschitz formulation of discrete control problems were given in [8]. As was recently pointed out to the author by Prof. Rolewicz, similar attempt was done independently in [16], where primarily the maximum principle formulation was investigated.

6. Conclusions

In this contribution the possibility of application of new results in the field of nonsmooth analysis to problems of discrete optimal control was studied in detail. Certain difficulties were pointed out when dealing with the general case together with some indications and suggestions how to overcome them. In this way a set of necessary conditions for general discrete control problems was derived assuming only the locally Lipschitz continuity of the studied problem. Some subsequent cases were treated separately to enlighten some aspects of this approach.

References

- [1] V. G. Boltjanskiĭ, *Optimal Control of Discrete Systems*, Nauka, Moscow 1973 [in Russian].
- [2] M. D. Canon, C. D. Cullum, E. Polak, *Theory of Optimal Control and Mathematical Programming*, McGraw-Hill, New York 1970.
- [3] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. 205 (1975), 247-262.
- [4] —, *A new approach to Lagrange multipliers*, Math. Oper. Res. 1 (1976), 165-174.
- [5] —, *Generalized gradients of Lipschitz functionals*, Technical Summary Report No. 1687, Univ. of Wisconsin, Madison, Wis. 1976.
- [6] J. Doležal, *Necessary optimality conditions for discrete systems with state-dependent control region*, Kybernetika 11 (1975), 423-450.
- [7] —, *Nonsmooth problems in mathematical programming and discrete optimal control*, in: Proc. 10th Symposium Slovak Cyber. Soc. Bratislava 1980, 109-114 [in Czech].
- [8] —, *Nonsmooth and nonconvex problems in discrete optimal control*, Internat. J. Systems Sci. 13 (1982), 969-978.
- [9] —, *A note on partial generalized gradients*, to appear.
- [10] A. J. Dubovickiĭ, *The discrete maximum principle*, Avtomat. i Telemekh. 39 (1978), 55-71 [in Russian].

- [11] J. B. Hiriart-Urruty, *New concepts in nondifferentiable programming*, Journées d'Analyse Non Convexe, Univ. de Pau, 1977.
 - [12] —, *On optimality conditions in nondifferentiable programming*, Math. Programming 14 (1978), 73–86.
 - [13] —, *Tangent cones, generalized gradients and mathematical programming in Banach spaces*, Math. Oper. Res. 4 (1979), 79–97.
 - [14] —, *Refinements of necessary optimality conditions in nondifferentiable programming I*, Appl. Math. Optim. 5 (1979), 63–82.
 - [15] R. T. Rockafellar, *The theory of subgradients and its applications to problems of optimization*, Lecture Notes, Univ. of Montreal, Montreal Press, Montreal 1978.
 - [16] M. Studniarski, *Generalized partial derivatives and discrete maximum principle*, Thesis, Univ. of Łódź, Łódź 1980 [in Polish].
-