12-1-2001

Discrete Maximum Principle for Nonsmooth Optimal Control Problems with Delays

Boris S. Mordukhovich
Wayne State University, boris@math.wayne.edu

Ilya Shvartsman
Wayne State University

DISCRETE MAXIMUM PRINCIPLE FOR NONSMOOTH OPTIMAL CONTROL PROBLEMS WITH DELAYS

BORIS S. MORDUKHOVICH and ILYA SHVARTSMAN

WAYNE STATE UNIVERSITY

Detroit, MI 48202

Department of Mathematics

Research Report

2001 Series
#12

This research was partly supported by the National Science Foundation.

To the memory of B. N. Pshenichnyi

DISCRETE MAXIMUM PRINCIPLE FOR NONSMOOTH OPTIMAL CONTROL PROBLEMS WITH DELAYS

BORIS S. MORDUKHOVICH [1] and ILYA SHVARTSMAN

Department of Mathematics, Wayne State University, Detroit, MI 48202, U.S.A.

boris@math.wayne.edu, ilya@math.wayne.edu

Abstract. We consider optimal control problems for discrete-time systems with delays. The main goal is to derive necessary optimality conditions of the discrete maximum principle type in the case of nonsmooth minimizing functions. We obtain two independent forms of the discrete maximum principle with transversality conditions described in terms of subdifferentials and superdifferentials, respectively. The superdifferential form is new even for non-delayed systems and may be essentially stronger than a more conventional subdifferential form in some situations.

Key words and phrases. Optimal control, discrete-time systems, time delays, maximum principle, nonsmooth variational analysis, subdifferentials and superdifferentials.

1 Introduction

This paper is devoted to the study of nonsmooth optimal control problems governed by discrete-time systems with time delays in state variables. As the basic model, we consider the following problem (P) of the Mayer type:

$$\text{minimize } \; J(x,u) := \varphi(x(t_1)) \tag{1.1}$$

over discrete control processes $\{x(\cdot), u(\cdot)\}$ satisfying

$$x(t+h) = x(t) + hf(t, x(t), x(t-\tau), u(t)), \quad x(t_0) = x_0 \in I\!R^n, \tag{1.2}$$

[1]Research was partly supported by the National Science Foundation under grant DMS-0072179 and by the Distinguished Faculty Fellowship at Wayne State University.

$$u(t) \in U, \quad t \in T := \{t_0, t_0 + h, \ldots, t_1 - h\}, \tag{1.3}$$

$$x(t) = c(t), \quad t \in T_0 := \{t_0 - \tau, t_0 - \tau + h, \ldots, t_0 - h\}, \tag{1.4}$$

where $h > 0$ is a discrete stepsize, $\tau = Nh$ is a time delay with some $N \in I\!N := \{1, 2, \ldots\}$, U is a compact set describing constraints on control values in (1.3), and $c(\cdot)$ is a given function describing the initial "tail" condition (1.4) for the delayed system (1.2). Problems of this type arise in variational analysis of delay-differential systems via discrete approximations; cf. [8, 9] and their predecessors for non-delayed systems in [14] and [6, 7]. They are important for many applications, especially to economic modelling, to qualitative and numerical aspects of optimization and control of various hereditary processes, to numerical solutions of control systems with distributed parameters, etc.; see, e.g., [1, 2, 8, 12, 16] and the references therein. Note that delayed discrete systems may be reduced to non-delayed ones of a bigger dimension by a multi-step procedure and that they both can be reduced to finite-dimensional mathematical programming. Nevertheless, optimal control problems of type (P) deserve a special attention in order to obtain results that take into account their particular dynamic structure and the influence of delays on the process of dynamic optimization.

It is well known that, while for continuous-time systems optimal controls satisfy the Pontryagin maximum principle without restrictive assumptions [11], its discrete analogue (the discrete maximum principle) does not generally hold unless a certain convexity is imposed a priori on the control system; see, e.g., [1, 4, 5, 12] and their references. A clear explanation of this phenomenon is given in Section 5.9 of Pshenichnyi's book [13] (the first edition), where it is shown why discrete systems require a convexity assumption for the validity of the maximum principle while continuous-time systems enjoy it automatically due to the so-called "hidden convexity". Relationships between convexity and the maximum principle are transparent from the viewpoint of nonsmooth analysis due to the special nature of the normal cone to convex sets; cf. [14] and [6].

The goal of this paper is to derive necessary optimality conditions in the form of the discrete maximum principle for problem (P) and some of its generalizations. Our *standing assumption* is that $f = f(t, x, y, u)$ is continuous with respect to all variables but t and continuously differentiable with respect to the state variables (x, y) for all $t \in T$ and $u \in U$ near the optimal solution under consideration. We do *not* assume *any smoothness* of the cost function φ and derive new versions of the discrete maximum principle with transversality conditions taking into account the nonsmoothness of φ. A striking result obtained in this paper, new for both delayed and non-delayed systems, is the *superdifferential form* of the discrete maximum principle, where the transversality condition is expressed in terms of the so-called Fréchet superdifferential. This is a rather surprising result, since it applies to *minimization* problems for which subdifferential forms of necessary optimality conditions are more conventional. We also obtain the discrete maximum principle for nonsmooth problems with transversality conditions of subdifferential type that extend known results to the case of delayed systems. We'll discuss relationships between the superdifferential and subdifferential forms of the discrete maximum principle: they are generally independent while the superdifferential one may be essentially stronger in some situations when it applies.

The rest of the paper is organized as follows. Section 2 presents basic definitions and preliminaries from nonsmooth analysis used in the sequel. In Section 2 we prove the superdifferential form of the discrete maximum principle and formulate some of its corollaries. Section 4 contains versions of the subdifferential discrete maximum principle for delayed systems and their comparison with the superdifferential version of Section 3.

Our notation is basically standard; see, e.g., [15]. Let us mention that A^* stands for the adjoint (transposed) matrix to A and that

$$\operatorname*{Lim\,sup}_{x\to\bar{x}} F(x) := \Big\{ y \in I\!R^m \ \Big|\ \exists \text{ sequences } x_k \to \bar{x} \text{ and } y_k \to y \text{ with } y_k \in F(x_k) \text{ for all } k \in I\!N \Big\}$$

denotes the *Painlevé-Kuratowski upper (outer) limit* for a set-valued mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ as $x \to \bar{x}$. The expressions

$$\operatorname{cl}\Omega, \quad \operatorname{co}\Omega, \quad \text{and} \quad \operatorname{cone}\Omega := \Big\{ \alpha x \Big| \ \alpha > 0,\ x \in \Omega \Big\}$$

stand for the closure, convex hull, and conic hull of a set Ω, respectively. The notation $x \xrightarrow{\varphi} \bar{x}$ means that $x \to \bar{x}$ with $\varphi(x) \to \varphi(\bar{x})$.

2 Tools of Nonsmooth Analysis

In this section we review several constructions of nonsmooth analysis and their properties needed in what follows. For more information we refer the reader to [3, 6, 15].

Let Ω be a nonempty set in $I\!R^n$, and let

$$\Pi(x;\Omega) := \Big\{ w \in \operatorname{cl}\Omega \ \text{ with } \ |x - w| = \operatorname{dist}(x;\Omega) \Big\}$$

be the Euclidean projector of x to the closure of Ω. The *basic normal cone* [6] to Ω at $\bar{x} \in \text{cl}\,\Omega$ is defined by

$$N(\bar{x};\Omega) := \operatorname*{Lim\,sup}_{x\to\bar{x}} \Big[\text{cone}\,(x - \Pi(x;\Omega))\Big]. \tag{2.1}$$

This cone if often nonconvex, and its convex closure agrees with the Clarke normal cone [3].

Given an extended-real-valued function $\varphi: I\!R^n \to \overline{I\!R} := [-\infty,\infty]$ finite at $\bar{x}$, we define its *basic subdifferential* [6] by

$$\partial\varphi(\bar{x}) := \Big\{x^* \in I\!R^n \Big|\ (x^*,-1) \in N((\bar{x},\varphi(\bar{x}));\text{epi}\,\varphi)\Big\}, \tag{2.2}$$

where $\text{epi}\,\varphi := \{(x,\mu) \in I\!R^{n+1}|\ \mu \geq \varphi(x)\}$ stands for the epigraph of φ. If φ is locally Lipschitzian around $\bar{x}$, then $\partial\varphi(\bar{x})$ is a nonempty compact satisfying

$$(x^*,-\lambda) \in N((\bar{x},\varphi(\bar{x}));\text{epi}\,\varphi) \Longleftrightarrow \lambda \geq 0,\ x^* \in \lambda\partial\varphi(\bar{x}). \tag{2.3}$$

One always has $\overline{\partial}\varphi(\bar{x}) = \text{co}\,\partial\varphi(\bar{x})$ for the Clarke *generalized gradient* of locally Lipschitzian functions [3]. Note the the latter construction, in contrast to (2.2), possesses the classical plus-minus symmetry $\overline{\partial}(-\varphi)(\bar{x}) = -\overline{\partial}\varphi(\bar{x})$. If φ is lower semicontinuous around $\bar{x}$, then the basic subdifferential (2.2) admits the representation

$$\partial\varphi(\bar{x}) = \operatorname*{Lim\,sup}_{x\xrightarrow{\varphi}\bar{x}} \widehat{\partial}\varphi(x)$$

in terms of the so-called *Fréchet subdifferential* of φ at x defined by

$$\widehat{\partial}\varphi(x) := \Big\{x^* \in I\!R^n \Big|\ \liminf_{u\to x} \frac{\varphi(u)-\varphi(x)-\langle x^*,u-x\rangle}{|u-x|} \geq 0\Big\}. \tag{2.4}$$

The symmetric constructions

$$\partial^+\varphi(\bar{x}) := -\partial(-\varphi)(\bar{x}),\quad \widehat{\partial}^+\varphi(\bar{x}) := -\widehat{\partial}(-\varphi)(\bar{x}) \tag{2.5}$$

to (3.2) and (2.4) are called, respectively, the *basic superdifferential* and the *Fréchet superdifferential* of φ at $\bar{x}$. Note that

$$\widehat{\partial}^+\varphi(\bar{x}) := \Big\{x^* \in I\!R^n \Big|\ \limsup_{x\to\bar{x}} \frac{\varphi(x)-\varphi(\bar{x})-\langle x^*,x-\bar{x}\rangle}{|x-\bar{x}|} \leq 0\Big\} \tag{2.6}$$

and that both $\widehat{\partial}\varphi(\bar{x})$ and $\widehat{\partial}^+\varphi(\bar{x})$ are nonempty simultaneously if and only if φ is Fréchet differentiable at $\bar{x}$, in which case they both reduce to the classical (Fréchet) derivative of φ at this point:

$$\widehat{\partial}\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x}) = \Big\{\nabla\varphi(\bar{x})\Big\}. \tag{2.7}$$

In contrast, the basic subdifferential and superdifferential are simultaneously nonempty for every locally Lipschitzian function; they may be essentially different, e.g., for $\varphi(x) = |x|$ on $I\!R$ when $\partial\varphi(0) = [-1, 1]$ while $\partial^+\varphi(0) = \{-1, 1\}$. Note also that if φ is Lipschitz continuous around $\bar{x}$, then

$$\partial\varphi(\bar{x}) = \partial^+\varphi(\bar{x}) = \left\{\nabla\varphi(\bar{x})\right\} \tag{2.8}$$

if and only if φ is *strictly* differentiable at $\bar{x}$, i.e.,

$$\lim_{\substack{x \to \bar{x} \\ x' \to \bar{x}}} \frac{\varphi(x) - \varphi(x') - \langle\nabla\varphi(\bar{x}), x - x'\rangle}{|x - x'|} = 0,$$

which happens, in particular, when φ is continuously differentiable around $\bar{x}$. The singleton relations (2.8) may be violated if φ is just differentiable but not strictly differentiable at $\bar{x}$. For example, if $\varphi(x) = x^2 \sin(1/x)$ for $x \neq 0$ with $\varphi(0) = 0$, then

$$\partial\varphi(0) = \partial^+\varphi(0) = [-1, 1] \;\text{ while }\; \widehat{\partial}\varphi(0) = \widehat{\partial}^+\varphi(0) = \{0\}.$$

Recall [6] that φ is *lower regular* at $\bar{x}$ if $\partial\varphi(\bar{x}) = \widehat{\partial}\varphi(\bar{x})$. It happens, in particular, when φ is either strictly differentiable at $\bar{x}$ or convex. Moreover, lower regularity holds for the class of *weakly convex* functions [10], which includes both smooth and convex functions and is closed with respect to taking the maximum over compact sets. Note that the latter class is a subclass of quasidifferentiable functions in the sense of Pshenichnyi [13].

A large class of lower regular functions (in somewhat stronger sense) has been studied in [15] under the name of *amenability*. It was shown there that the class of amenable functions enjoys a fairly rich calculus and includes a large core of functions frequently encountered in finite-dimensional minimization.

Symmetrically, φ is *upper regular* at $\bar{x}$ if $\partial^+\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x})$. It follows from (2.5) that this property is equivalent to the lower regularity of $-\varphi$ at $\bar{x}$. Thus all the facts about subdifferentials and lower regularity relative to minimization can be symmetrically transferred to superdifferentials and upper regularity relative to maximization. The point is that in the next section we are going to apply *superdifferentials and upper regularity to minimization problems.* The following proposition is useful in this respect.

Proposition 2.1 *Let $\varphi\colon I\!R^n \to \overline{I\!R}$ be Lipschitz continuous around $\bar{x}$ and upper regular at this point. Then $\emptyset \neq \widehat{\partial}^+\varphi(\bar{x}) = \overline{\partial}\varphi(\bar{x})$.*

Proof. The nonemptiness of $\widehat{\partial}^+\varphi(\bar{x})$ follows directly from $\partial\varphi(\bar{x}) \neq \emptyset$ for locally Lipschitzian functions and the definition of upper regularity. Due to $\overline{\partial}\varphi(\bar{x}) = \text{co}\,\partial\varphi(\bar{x})$, any local Lipschitzian function is lower regular at $\bar{x}$ if and only if $\widehat{\partial}\varphi(\bar{x}) = \overline{\partial}\varphi(\bar{x})$. Hence the upper regularity of φ at $\bar{x}$ and the plus-minus symmetry of the generalized gradient imply that

$$\widehat{\partial}^+\varphi(\bar{x}) = -\widehat{\partial}(-\varphi)(\bar{x}) = -\overline{\partial}(-\varphi)(\bar{x}) = \overline{\partial}\varphi(\bar{x}),$$

which ends the proof of the proposition. □

Note that all the assumptions of Proposition 2.1 hold for concave functions continuous around $\bar{x}$.

3 Superdifferential Form of the Discrete Maximum Principle

In this section we first study the discrete optimal control problem (P) defined in (1.1)–(1.4) and then consider its multiple delay generalization. Let $\{x(\cdot), u(\cdot)\}$ be a feasible process to (P), and let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to this problem. For convenience we introduce the following notation:

$$\xi(t) := (x(t), x(t-\tau)), \quad \bar{\xi}(t) := (\bar{x}(t), \bar{x}(t-\tau)),$$

$$f(t, \xi, u) := f(t, x(t), x(t-\tau), u(t)), \quad f(t, \bar{\xi}, u) := f(t, \bar{x}(t), \bar{x}(t-\tau), u(t)),$$

$$f(t+\tau, \xi, u) := f(t+\tau, x(t+\tau), x(t), u(t+\tau)),$$

$$\Delta x(t) := x(t) - \bar{x}(t), \quad \Delta f(t) := f(t, \xi, u) - f(t, \bar{\xi}, \bar{u}), \quad \Delta_u f(t) := f(t, \bar{\xi}, u) - f(t, \bar{\xi}, \bar{u}).$$

Using this notation, we define the *adjoint system*

$$p(t) = p(t+h) + h\frac{\partial f^*}{\partial x}(t, \bar{\xi}, \bar{u})p(t+h) + h\frac{\partial f^*}{\partial y}(t+\tau, \bar{\xi}, \bar{u})p(t+\tau+h), \quad t \in T, \tag{3.1}$$

to (2.2) along the optimal process $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$. Consider the *Hamilton-Pontryagin function*

$$H(t, p(t+h), \xi(t), u(t)) := \Big\langle p(t+h), f(t, \xi(t), u(t)) \Big\rangle, \tag{3.2}$$

which allows us to rewrite the adjoint system (3.1) in the simplified form

$$p(t) = p(t+h) + h\Big[\frac{\partial H}{\partial x}(t) + \frac{\partial H}{\partial y}(t+\tau)\Big]$$

with $H(t) := H(t, p(t+h), \bar{\xi}(t), \bar{u}(t))$. Form the set

$$\Lambda(\bar{u}(t)) := \Big\{u \in U \Big| \; f(t,\bar{\xi},u) \in \sigma(f(t,\bar{\xi},\bar{u}); f(t,\bar{\xi},U))\Big\}, \tag{3.3}$$

where $\sigma(q;Q)$ denotes the *star-neighborhood* of $q \in Q$ relative to Q defined by

$$\sigma(q;Q) := \Big\{a \in Q \Big| \; \exists\, \varepsilon_k \downarrow 0 \quad \text{such that} \quad q + \varepsilon_k(a-q) \in Q \quad \text{for all} \;\; k \in I\!N\Big\}. \tag{3.4}$$

It easily follows from (3.3) and (3.4) that $\Lambda(\bar{u}(t)) = U$ if the set $f(t,\bar{\xi},U)$ is convex. The following theorem establishes a new *superdifferential form* of the discrete maximum principle for both delayed and non-delayed systems.

Theorem 3.1 *Let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to (P). Assume that $\varphi\colon I\!R^n \to \overline{I\!R}$ is finite at $\bar{x}(t_1)$ and that $\widehat{\partial}^+\varphi(\bar{x}(t_1)) \neq \emptyset$. Then for any $x^* \in \widehat{\partial}^+\varphi(\bar{x}(t_1))$ one has the discrete maximum principle*

$$H(t,p(t+h),\bar{x}(t),\bar{x}(t-\tau),\bar{u}(t)) = \max_{u \in \Lambda(\bar{u}(t))} H(t,p(t+h),\bar{x}(t),\bar{x}(t-\tau),u), \quad t \in T. \tag{3.5}$$

where $p(\cdot)$ is an adjoint trajectory satisfying (3.1) *and the transversality conditions*

$$p(t_1) = -x^*, \qquad p(t) = 0 \;\; \textit{for} \;\; t > t_1. \tag{3.6}$$

The maximum condition (3.5) *is global over all $u \in U$ if the set $f(t,\xi,U)$ is convex.*

Proof. Take an arbitrary $x^* \in \widehat{\partial}^+\varphi(\bar{x}(t_1))$. It follows from (2.6) that

$$\varphi(x) - \varphi(\bar{x}(t_1)) \le \langle x^*, x - \bar{x}(t_1)\rangle + o(|x - \bar{x}(t_1)|) \tag{3.7}$$

for all x sufficiently close to $\bar{x}(t_1)$. Put $p(t_1) := -x^*$ and derive from (3.7) and (1.1) that

$$J(x,u) - J(\bar{x},\bar{u}) = -\langle p(t_1), \Delta x(t_1)\rangle + o(|\Delta x(t_1)|) \ge 0 \tag{3.8}$$

for all feasible processes $\{x(\cdot), u(\cdot)\}$ to (P) such that $x(t_1)$ is sufficiently close to $\bar{x}(t_1)$. One always has the identity

$$\langle p(t_1), \Delta x(t_1)\rangle = \sum_{t=t_0}^{t_1-h} \Big\langle p(t+h) - p(t), \Delta x(t)\Big\rangle + \sum_{t=t_0}^{t_1-h} \Big\langle p(t+h), \Delta x(t+h) - \Delta x(t)\Big\rangle. \tag{3.9}$$

Due to (1.2) we get the representation

$$\Delta x(t+h) - \Delta x(t) = h\Delta f(t) = h\Big[\Delta_u f(t) + \frac{\partial f}{\partial x}(t,\bar{\xi},\bar{u})\Delta x(t) + \frac{\partial f}{\partial y}(t,\bar{\xi},\bar{u})\Delta x(t-\tau) + \eta(t)\Big],$$

where the remainder $\eta(t)$ is computed by

$$\begin{aligned}\eta(t) &= \Big(\frac{\partial f}{\partial x}(t,\bar{\xi},u) - \frac{\partial f}{\partial x}(t,\bar{\xi},\bar{u})\Big)\Delta x(t) + \Big(\frac{\partial f}{\partial y}(t,\bar{\xi},u) - \frac{\partial f}{\partial y}(t,\bar{\xi},\bar{u})\Big)\Delta x(t-\tau)\\ &+ o(|\Delta x(t)|) + o(|\Delta x(t-\tau)|).\end{aligned}$$

This allows us to represent the second sum in (3.9) as

$$\begin{aligned}\sum_{t=t_0}^{t_1-h}\Big\langle p(t+h), \Delta x(t+h) - \Delta x(t)\Big\rangle &= h\sum_{t=t_0}^{t_1-h}\Big\langle p(t+h), \Delta_u f(t) + \frac{\partial f}{\partial x}(t,\bar{\xi},\bar{u})\Delta x(t)\\ &+ \frac{\partial f}{\partial y}(t,\bar{\xi},\bar{u})\Delta x(t-\tau) + \eta(t)\Big\rangle.\end{aligned}$$

Using the equalities

$$\Delta x(t) = 0 \ \text{ for } \ t \le t_0, \qquad p(t+h) = 0 \ \text{ for } \ t \ge t_1$$

and shifting the summation above, one has

$$\sum_{t=t_0}^{t_1-h}\Big\langle p(t+h), \frac{\partial f}{\partial y}(t,\bar{\xi},\bar{u})\Delta x(t-\tau)\Big\rangle = \sum_{t=t_0}^{t_1-h}\Big\langle p(t+\tau+h), \frac{\partial f}{\partial y}(t+\tau,\bar{\xi},\bar{u})\Delta x(t)\Big\rangle. \tag{3.10}$$

Finally, substituting (3.1), (3.9), and (3.10) into (3.8), we obtain

$$J(x,u) - J(\bar{x},\bar{u}) = -h\sum_{t=t_0}^{t_1-h}\Delta_u H(t) - h\sum_{t=t_0}^{t_1-h}\Big\langle p(t+h),\eta(t)\Big\rangle + o(|\Delta x(t_1)|) \ge 0 \tag{3.11}$$

with $\Delta_u H(t) := H(t,p(t+h),\bar{\xi}(t),u(t)) - H(t,p(t+h),\bar{\xi}(t),\bar{u}(t))$ whenever $\Delta x(t_1)$ is sufficiently small.

Let us prove that (3.11) implies that $\Delta_u H(t) \le 0$ for any $t \in T$ and $u \in \Lambda(\bar{u}(t))$, which is equivalent to the discrete maximum principle (3.5). Assuming the contrary, we find

$$\theta \in T \quad \text{and} \quad u \in \Lambda(\bar{u}(\theta)) \quad \text{with} \quad \Delta_u H(\theta) := a > 0. \tag{3.12}$$

By definitions (3.3) and (3.4) there are sequences $\varepsilon_k \downarrow 0$ and $u_k \in U$ such that

$$f(\theta,\bar{\xi},\bar{u}) + \varepsilon_k\Big(f(\theta,\bar{\xi},u) - f(\theta,\bar{\xi},\bar{u})\Big) := f(\theta,\bar{\xi},u_k) \in f(\theta,\bar{\xi},U),$$

which is equivalent to

$$\Delta_{u_k} f(\theta) := f(\theta, \bar{\xi}, u_k) - f(\theta, \bar{\xi}, \bar{u}) = \varepsilon_k \Big(f(\theta, \bar{\xi}, u) - f(\theta, \bar{\xi}, \bar{u}) \Big) := \varepsilon_k \Delta_u f(\theta).$$

Now let us consider *needle variations* of the optimal control defined by

$$v_k(t) = \begin{cases} u_k & \text{if } t = \theta, \\ \bar{u}(t) & \text{if } t \in T \setminus \{\theta\}, \end{cases}$$

which are feasible to (P) for all $k \in I\!N$, and let $\Delta_k x(t)$ be the corresponding perturbations of the optimal trajectory generated by $v_k(t)$. One can see that

$$\Delta_k x(t) = 0 \ \text{ for } \ t = t_0, \dots, \theta \quad \text{ and } \quad |\Delta_k x(t)| = O(\varepsilon_k) \ \text{ for } \ t = \theta + h, \dots, t_1.$$

This implies that

$$\Big(\frac{\partial f}{\partial x}(t, \bar{\xi}, v_k) - \frac{\partial f}{\partial x}(t, \bar{\xi}, \bar{u})\Big) \Delta_k x(t) + \Big(\frac{\partial f}{\partial y}(t, \bar{\xi}, v_k) - \frac{\partial f}{\partial y}(t, \bar{\xi}, \bar{u})\Big) \Delta_k x(t - \tau) = 0, \quad t \in T,$$

and that $\eta_k(t) = o(\varepsilon_k)$, $k \in I\!N$, for the corresponding remainders $\eta_k(\cdot)$ defined above. Hence

$$J(x_k, v_k) - J(\bar{x}, \bar{u}) = -h \Delta_{u_k} H(\theta) - h \sum_{t=t_0}^{t_1 - h} \Big\langle p(t+h), \eta_k(t) \Big\rangle = -\varepsilon_k h a + o(\varepsilon_k) < 0$$

for all large $k \in I\!N$ due to (3.12). Since $x_k(t_1) \to \bar{x}(t_1)$ as $k \to \infty$, this contradicts (3.11) and completes the proof of the theorem. □

Let us present two important corollaries of Theorem 3.1. The first one assumes that φ is (Fréchet) differentiable *at* the point $\bar{x}(t_1)$. Note that it may not be strictly differentiable (and hence not upper regular) at this point as for the function $\varphi(x) = x^2 \sin(1/x)$ for $x \neq 0$ with $\varphi(0) = 0$; see Section 2. If φ is continuously differentiable around $\bar{x}(t_1)$ and $f = f(t, x, u)$ in (1.2), then this result and its proof go back to the discrete maximum principle for non-delayed systems established in [4, Chapter IX].

Corollary 3.2 *Let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to (P), where φ is assumed to be differentiable at $\bar{x}(t_1)$. Then one has the discrere maximum principle* (3.5) *with $p(\cdot)$ satisfying* (3.1) *and*

$$p(t_1) = -\nabla \varphi(\bar{x}(t_1)), \qquad p(t) = 0 \ \textit{for} \ \ t > t_1. \tag{3.13}$$

Proof. Follows from Theorem 3.1 due to the second relation in (2.7), which ensures that (3.6) reduces to (3.13). □

The next corollary provides a striking result for upper regular and Lipschitz continuous cost functions φ. In this case the discrete maximum principle holds with the transversality condition $p(t_1) = -x^*$ given by *any* vector x^* from the generalized gradient $\overline{\partial}\varphi(\bar{x}(t_1))$ while conventional results ensure such conditions only for *some* subgradient; see Section 4 for more discussions.

Corollary 3.3 *Let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to (P), where φ is assumed to be Lipschitz continuous around $\bar{x}(t_1)$ and upper regular at this point. Then for any vector $x^* \in \overline{\partial}\varphi(\bar{x}(t_1)) \neq \emptyset$ one has the maximum principle* (3.5) *with $p(\cdot)$ satisfying* (3.1) *and* (3.6).

Proof. This follows from Theorem 3.1 and Proposition 2.1. □

Now let us consider an extension (P_1) of problem (P) to the case of *multiple delays*: minimize (1.1) over discrete control processes $\{x(\cdot), u(\cdot)\}$ satisfying the system

$$x(t+h) = x(t) + hf(t, x(t), x(t-\tau_1), \ldots, x(t-\tau_m), u(t)), \quad x(t_0) = x_0 \in I\!R^n, \tag{3.14}$$

with many delays $\tau_i = N_i h$ for $N_i \in I\!N$ and $i = 1, \ldots, m$ subject to constraints (1.3) and (1.4), where $f = f(t, x, x_1, \ldots, x_m, u)$ satisfies our standing assumption and where the initial interval T_0 is correspondingly modified. Denote $\bar{\xi}(t) := (\bar{x}(t), \bar{x}(t-\tau_1), \ldots, \bar{x}(t-\tau_m))$ and define $p(\cdot)$ satisfying (3.6) and the adjoint system

$$p(t) = p(t+h) + h\frac{\partial f}{\partial x}^*(t, \bar{\xi}, \bar{u})p(t+h) + h\sum_{i=1}^{m}\frac{\partial f}{\partial x_i}^*(t+\tau_i, \bar{\xi}, \bar{u})p(t+\tau_i+h) \tag{3.15}$$

for $t \in T$, which can be rewritten in the Hamiltonian form

$$p(t) = p(t+h) + h\frac{\partial H}{\partial x}(t) + h\sum_{i=1}^{m}\frac{\partial H}{\partial x_i}(t+\tau_i)$$

in terms of (3.2) with $H(t) := H(t, p(t+h), \bar{\xi}(t), \bar{u}(t))$. The proof of the following theorem is similar to the basic case of Theorem 3.1 and can be omitted.

Theorem 3.4 *Let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to (P_1) with $\widehat{\partial}^+\varphi(\bar{x}(t_1)) \neq \emptyset$. Then for any $x^* \in \widehat{\partial}^+\varphi(\bar{x}(t_1))$ one has the discrete maximum principle*

$$H(t, p(t+h), \bar{\xi}(t), \bar{u}(t)) = \max_{u \in \Lambda(\bar{u}(t))} H(t, p(t+h), \bar{\xi}(t), u) \quad \text{for all } t \in T, \tag{3.16}$$

where $p(\cdot)$ is an adjoint trajectory satisfying (3.6) *and* (3.15).

Of course, we have the corollaries of Theorem 3.4 similar to the above ones for Theorem 3.1. Let us obtain another corollary of Theorem 3.4 for a counterpart (P_2) of the optimal control problem (P) involving discrete systems of *neutral type*

$$x(t+h) = x(t) + hf\Big(t, x(t), x(t-\tau), \frac{x(t-\tau+h) - x(t-\tau)}{h}, u(t)\Big), \quad t \in T, \tag{3.17}$$

where $\dfrac{x(t-\tau+h)-x(t-\tau)}{h}$ can be treated as an analogue of the delayed derivative $\dot{x}(t-\tau)$ under the time discretization and where $f = f(t,x,y,z,u)$ satisfies our standing assumption.

Given an optimal process $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ to (P_2), we put

$$\bar{\xi}(t) := \Big(\bar{x}(t), \bar{x}(t-\tau), \frac{\bar{x}(t-\tau+h) - \bar{x}(t-\tau)}{h}\Big), \quad t \in T, \tag{3.18}$$

and define the adjoint discrete neutral type system

$$\begin{aligned} p(t) &= p(t+h) + h\frac{\partial f^*}{\partial x}(t, \bar{\xi}, \bar{u})p(t+h) + h\frac{\partial f^*}{\partial y}(t+\tau, \bar{\xi}, \bar{u})p(t+\tau+h) \\ &+ \frac{\partial f^*}{\partial z}(t+\tau-h, \bar{\xi}, \bar{u})p(t+\tau) - \frac{\partial f^*}{\partial z}(t+\tau, \bar{\xi}, \bar{u})p(t+\tau+h), \quad t \in T. \end{aligned} \tag{3.19}$$

Corollary 3.5 *Let $\{\bar{x}(\cdot), \bar{u}(\cdot)\}$ be an optimal process to (P_2) with $\widehat{\partial}^+\varphi(\bar{x}(t_1)) \neq \emptyset$. Then for any $x^* \in \widehat{\partial}^+\varphi(\bar{x}(t_1))$ one has the discrete maximum principle* (3.16), *where $\bar{\xi}(\cdot)$ is defined in* (3.18) *and where $p(\cdot)$ is an adjoint trajectory satisfying* (3.6) *and* (3.19).

Proof. Observe that the neutral system (3.17) can be easily reduced to (3.14) with two delays. Thus this corollary follows from Theorem 3.4 via simple calculations. □

A drawback of the the superdifferential form of the discrete maximum principle established above is that the Fréchet superdifferential may be *empty* for nice functions important in nonsmooth minimization, e.g., for convex functions that are not differentiable at minimum points. In the next section we derive results on the discrete maximum principle that cover delayed problems of type (P) with general nonsmooth cost functions φ. Results of the latter subdifferential type are applicable to a broad class of nonsmooth problems, but they may not be that sharp as the superdifferential form of Theorem 3.1 when it applies.

4 The Discrete Maximum Principle via Basic Normals and Subgradients

In this final section of the paper we present nonsmooth versions of the discrete maximum principle for the delayed problem (P) in (1.1)–(1.4) with transversality conditions expressed in terms of basic normals and subgradients defined in Section 2. The corresponding modifications for problems (P_1) and (P_2) can be made similarly to Section 3.

Theorem 4.1 *Let $\{\bar{x}(\cdot),\bar{u}(\cdot)\}$ be an optimal process to problem (P), and let $\bar{x} := \bar{x}(t_1)$. Assume that the set $f(t,x,y,U)$ is convex around $(\bar{x}(t),\bar{x}(t-\tau))$ for all $t \in T$. Then one has the following assertions.*

(i) *Let φ be lower semicontinuous around $\bar{x}$. Then there is a nonzero vector $(x^*,\lambda) \in \mathbb{R}^{n+1}$ such that $\lambda \geq 0$, $(x^*,-\lambda) \in N((\bar{x},\varphi(\bar{x}));\operatorname{epi}\varphi)$, and the discrete maximum principle*

$$H(t,p(t+h),\bar{x}(t),\bar{x}(t-\tau),\bar{u}(t)) = \max_{u\in U} H(t,p(t+h),\bar{x}(t),\bar{x}(t-\tau),u), \quad t \in T. \tag{4.1}$$

holds with $p(\cdot)$ satisfying (3.1) *and* (3.6).

(ii) *Let φ be Lipschitz continuous around $\bar{x}$. Then there is $x^* \in \partial\varphi(\bar{x})$ such that* (4.1) *holds with $p(\cdot)$ satisfying* (3.1) *and* (3.6).

Proof. We'll proceed similarly to the non-delayed case using the *method of metric approximation*; cf. [6, Section 11]. This method allows us to approximate the original nonsmooth problem by a family of smooth discrete problems with delays and then arrive at the desired conclusions by a limiting procedure involving the corresponding results and constructions of Sections 2 and 3.

Let us first prove assertion (i). Taking a parameter $\gamma \in \mathbb{R}$, we consider a parametric family of the following optimal control problems (P_γ) for delayed discrete systems with the distance-type cost functional:

$$\text{minimize} \quad J_\gamma(x,u) := \operatorname{dist}((x(t_1),\gamma);\operatorname{epi}\varphi) + \sum_{t=t_0}^{t_1} |x(t)-\bar{x}(t)|^2$$

over control processes $\{x(\cdot),u(\cdot)\}$ subject to constraints (1.2)–(1.4).

Let $\bar{\gamma} := \varphi(\bar{x}(t_1))$, and let $\{\bar{x}_\gamma(\cdot),\bar{u}_\gamma(\cdot)\}$ be optimal processes to (P_γ) that obviously exist by the classical Weierstrass theorem due to the standing assumptions made in Section 1. It follows from the structure of (P_γ) and the optimality of $\{\bar{x}(\cdot),\bar{u}(\cdot)\}$ in the original problem (P) that $\bar{x}_\gamma(t) \to \bar{x}(t)$ as $\gamma \to \bar{\gamma}$ for all $t \in T \cup \{t_1\}$. Moreover,

$$m_\gamma := \operatorname{dist}((\bar{x}_\gamma(t_1),\gamma);\operatorname{epi}\varphi) > 0 \quad \text{whenever} \quad \gamma < \bar{\gamma}. \tag{4.2}$$

The latter allows us to conclude that, for any $\gamma < \bar{\gamma}$, the process $\{\bar{x}_\gamma(\cdot),\bar{u}_\gamma(\cdot)\}$ is optimal to the *smooth* problem $(\overline{P}_\gamma)$ of minimizing the functional

$$\bar{J}_\gamma(x,u) := \left(|x(t_1)-x_\gamma|^2 + |\gamma - w_\gamma|^2\right)^{1/2} + \sum_{t=t_0}^{t_1} |x(t)-\bar{x}(t)|^2$$

subject to (1.2)–(1.4), where (x_γ, w_γ) is an arbitrary element of the Euclidean projector $\Pi((\bar{x}_\gamma(t_1),\gamma); \mathrm{epi}\,\varphi)$ of $(\bar{x}_\gamma(t_1),\gamma)$ to the closed set $\mathrm{epi}\,\varphi$. Introducing an additional state variable $x_{n+1}(t)$ by

$$x_{n+1}(t+h) = x_{n+1}(t) + |x(t) - \bar{x}(t)|^2, \quad x_{n+1}(t_0) = 0, \tag{4.3}$$

we rewrite problem $(\overline{P}_\gamma)$ in the equivalent form of minimizing the Mayer-type functional

$$\bar{J}_\gamma(x, x_{n+1}, u) = \Big(|x(t_1) - x_\gamma|^2 + |\gamma - w_\gamma|^2\Big)^{1/2} + x_{n+1}(t_1) + |x(t_1) - \bar{x}(t_1)|^2 \tag{4.4}$$

over $\{x(\cdot), x_{n+1}(\cdot), u(\cdot)\}$ satisfying (1.2)–(1.4) and (4.3). Denote $\bar{\xi}_\gamma(t) := (\bar{x}_\gamma(t), \bar{x}_\gamma(t-\tau))$ and observe that the sets $f(t, \bar{\xi}_\gamma(t), U)$ are convex for all $t \in T$ while the cost function in (4.4) is differentiable at $(\bar{x}_\gamma(t_1), \bar{x}_{n+1}(t_1))$, where $\bar{x}_{n+1}(\cdot)$ is generated by $\bar{x}_\gamma(\cdot)$ in (4.3). Now applying Corollary 3.2 to problem (P_γ) as $\gamma < \bar{\gamma}$ and taking into account the structure of the cost function (4.4), we arrive at the discrete maximum principle

$$H(t, p_\gamma(t+h), \bar{\xi}_\gamma(t), \bar{u}_\gamma(t)) = \max_{u \in U} H(t, p_\gamma(t+h), \bar{\xi}_\gamma(t), u), \quad t \in T,$$

where $p_\gamma(\cdot)$ satisfies the adjoint system (3.1) along $\{\bar{x}_\gamma(\cdot), \bar{u}_\gamma(\cdot)\}$ with the transversality conditions

$$p_\gamma(t_1) = -\frac{\bar{x}_\gamma(t_1) - x_\gamma}{m_\gamma} - 2(\bar{x}_\gamma(t_1) - \bar{x}(t_1)), \qquad p_\gamma(t) = 0 \ \text{ for } \ t > t_1,$$

where $m_\gamma > 0$ is given in (4.2), and where

$$\left(\frac{|\bar{x}_\gamma(t_1) - x_\gamma|}{m_\gamma}\right)^2 + \left(\frac{|\gamma - w_\gamma|}{m_\gamma}\right)^2 = 1.$$

Passing to the limit as $\gamma \uparrow \bar{\gamma}$ in the above relations and using the construction of the basic normal cone (2.1), we arrive at all the conclusions of (i).

To justify (ii) when φ is Lipschitz continuous around $\bar{x}(t_1)$, we observe that in this case one has $x^* \in \lambda\partial\varphi(\bar{x}(t_1))$ from (i) and (2.3). The latter implies that $\lambda \neq 0$, which yields (ii) by normalization. $\square$

Let us compare the superdifferential and subdifferential forms of the discrete maximum principle from Theorems 3.1 and 4.1, respectively. As mentioned above, Theorem 4.1 is applicable to a broad class of nonsmooth problems (P) while Theorem 3.1 requires that $\hat{\partial}^+\varphi(\bar{x}(t_1)) \neq \emptyset$, which excludes many nonsmooth functions. On the other hand, the superdifferential form has essential advantages for special classes of cost functions.

First we observe that Theorem 3.1 implies the gradient form (3.13) of transversality when φ is just differentiable at $\bar{x}(t_1)$ (it may even not be Lipschitz continuous around this point) while Theorem 4.1 ensures (3.13) only when φ is *strictly* differentiable at $\bar{x}(t_1)$; see (2.8) and the related discussion in Section 2. The most striking difference between subdifferential and superdifferential transversality conditions appears in the case of *upper regular* and locally Lipschitzian cost functions. In this case Theorem 4.1 provides the discrete maximum principle generated by *some* subgradient $x^* \in \partial\varphi(\bar{x}(t_1)) \subset \overline{\partial}\varphi(\bar{x}(t_1))$ in (3.6) while Corollary 3.3 ensures it for *every* $x^* \in \overline{\partial}\varphi(\bar{x}(t_1))$. This is a big difference!

To conclude, we present a simple illustrative *example* of a non-delayed problem, where the superdifferential form of the discrete maximum allows us to eliminate a non-optimal control but the subdifferential form fails to do it. Minimize the cost functional (1.1) with $\varphi(x) = -|x|$, $x \in I\!R$, and $t_1 = 1$ subject to the constraints

$$x(t+h) = x(t) + hu(t), \quad x(0) = 0,$$

$$u(t) \in U := [0,1], \quad t \in T := \{0, h, \ldots, 1-h\},$$

where $h = 1/N$ for some natural number $N \geq 2$. The control $\bar{u}(t) \equiv 0$ is obviously not optimal while Theorem 4.1 cannot eliminate it. Indeed, $\partial\varphi(0) = \{-1, 1\}$, and one may take $p(1) = -1 \in -\partial\varphi(0)$ due to this result. We see that the control $\bar{u}(t) \equiv 0$ satisfies the maximum condition (4.1) with $p(t) \equiv -1$. On the other hand, the discrete maximum principle does not hold for $\bar{u} \equiv 0$ if we select $p(1) = 1 \in \widehat{\partial}^+\varphi(0) = \overline{\partial}\varphi(0) = [-1, 1]$, i.e., this control can be eliminated by Corollary 3.3.

References

[1] V. G. Boltyanskii, *Optimal Control of Discrete Systems*, Nauka, Moscow, 1973.

[2] E. N. Chukwu, *Stability and Time-Optimal Control of Hereditary Systems*, Academic Press, Boston, 1992.

[3] F. H. Clarke, *Optmization and Nonsmooth Analysis*, Wiley, New York, 1983; Russian transl. in Nauka, Moscow, 1988.

[4] R. Gabasov and F. M. Kirillova, *Qualitative Theory of Optimal Process*, Nauka, Moscow, 1971; English transl. in Marcel Dekker, New York, 1976.

[5] H. Halkin, A maximum principle of the Pontryagin type for systems decsribed by nonlinear difference equations, *J. SIAM Control* **4** (1966), 90–112.

[6] B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988; expanded English edition to appear.

[7] B. S. Mordukhovich, Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions, *SIAM J. Control Optim.* **33** (1995), 882–915.

[8] B. S. Mordukhovich, Optimal control of difference, differential, and differential-difference inclusions, *J. Math. Sci.* **100** (2000), 2313–2632.

[9] B. S. Mordukhovich and R. Trubnik, Stability of discrete approximations and necessary optimality conditions for delay-differential inclusions, *Ann. Oper. Res.* **101** (2001), 149–170.

[10] E. A. Nurminskii, *Numerical Methods for Solutions of Deterministic and Stochastic Minimax Problems*, Naukova Dumka, Kiev, 1979.

[11] A. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Nauka, Moscow, 1961; English transl. in Wiley, New York, 1962.

[12] A. I. Propoi, *Elements of the Theory of Optimal Discrete Processes*, Nauka, Moscow, 1973.

[13] B. N. Pshenichnyi, *Necessary Conditions for an Extremum*, Nauka, Moscow, 1968; English translation in Marcel Dekker, New York, 1971.

[14] B. N. Pshenichnyi, *Convex Analysis and Extremal Problems*, Nauka, Moscow, 1980.

[15] R. T. Rockafelar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[16] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972; Russian transl. in Nauka, Moscow, 1977.