THE APPROXIMATE MAXIMUM PRINCIPLE IN CONSTRAINED OPTIMAL CONTROL [1]

BORIS S. MORDUKHOVICH and ILYA SHVARTSMAN

Department of Mathematics, Wayne State University, Detroit, MI 48202

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

Abstract. The paper concerns optimal control problems for dynamic systems governed by a parametric family of discrete approximations of control systems with continuous time. Discrete approximations play an important role in both qualitative and numerical aspects of optimal control and occupy an intermediate position between discrete-time and continuous-time control systems. The central result in optimal control of discrete approximations is the Approximate Maximum Principle (AMP), which is justified for smooth control problems with endpoint constraints under certain assumptions without imposing any convexity, in contrast to discrete systems with a fixed step. We show that these assumptions are essential for the validity of the AMP, and that the AMP does not hold, in its expected (lower) subdifferential form, for nonsmooth problems. Moreover, a new upper subdifferential form of the AMP is established in this paper for both ordinary and time-delay control systems. This solves a long-standing question about the possibility to extend the AMP to nonsmooth control problems.

Key words. optimal control, discrete approximations, approximate maximum principle, stability under perturbations, nonsmooth and variational analysis, lower and upper subgradients, time delays

AMS subject classification. 49K15, 93C55, 49M25, 49J52, 49J53.

1 Introduction and Preliminaries

This paper is devoted to *discrete approximations* of continuous-time control systems that, viewed as a *parametric process* with a decreasing discretization step, occupy an *intermediate* position between control systems with discrete and continuous times. As the basic model for our study, we consider discrete approximations of the following Mayer-type optimal control problem governed by ordinary differential equations with endpoint constraints:

$$(P)\qquad \begin{cases} \text{minimize } J(x,u) := \varphi_0(x(t_1)) \\ \text{subject to} \\ \dot{x}(t) = f(t,x(t),u(t)) \;\; \text{a.e.} \;\; t \in [t_0,t_1], \quad x(t_0) = x_0 \in I\!R^n, \\ u(t) \in U \;\; \text{a.e.} \;\; t \in [t_0,t_1], \\ \varphi_i(x(t_1)) \le 0, \quad i = 1,\ldots,m, \\ \varphi_i(x(t_1)) = 0, \quad i = m+1,\ldots,m+r, \end{cases}$$

[1]Research was partly supported by the National Science Foundation under grants DMS-0072179 and DMS-0304989.

over measurable controls $u(\cdot)$ and absolutely continuous trajectories $x(\cdot)$ on the fixed time interval $T := [t_0, t_1]$. It is well known that many other control problems (of Lagrange and Bolza types, with integral constraints, on variable time intervals, etc.) reduce to the form of (P). Observe also that the results of this paper can be easily extended to control problems with non-fixed initial vector (i.e., when φ_i in (P) depend on both endpoints $x(t_0)$ and $x(t_1)$ for all $i = 0, \ldots, m + r$) as well as to problems with continuously time-dependent control sets $U = U(t)$.

Problem (P) may be treated as an infinite-dimensional optimization problem with special equality-type dynamic constraints governed by differential operators as well with geometric constraints given by arbitrary control sets; this makes it to be *nonsmooth* even under all smooth functional data f and φ_i. On the other hand, it is natural to explore a different way to study continuous-time problems (P), which goes back to Leibnitz and Euler and which consists of approximating (P) by a family of discrete-time systems arising when the time-derivative $\dot{x}(t)$ is replaced with the *finite differences*

$$\dot{x}(t) \approx \frac{x(t+h) - x(t)}{h} \quad \text{as} \quad h \to 0.$$

Allowing also *perturbations* of the *endpoint constraints* (which is very essential for variational stability), problem (P) is replaced in this way by the following family of discrete-time problems (P_N) depending on the natural parameter $N = 1, 2, \ldots$:

$$(P_N) \qquad \begin{cases} \text{minimize } J(x_N, u_N) := \varphi_0(x_N(t_1)) \\ \text{subject to} \\ x_N(t + h_N) = x_N(t) + h_N f(t, x_N(t), u_N(t)), \quad x_N(t_0) = x_0 \in I\!\!R^n, \\ u_N(t) \in U, \quad t \in T_N := \{t_0, t_0 + h_N, \ldots, t_1 - h_N\}, \\ \varphi_i(x_N(t_1)) \le \gamma_{iN}, \quad i = 1, \ldots, m, \\ |\varphi_i(x_N(t_1))| \le \delta_{iN}, \quad i = m+1, \ldots, m+r, \\ h_N := \dfrac{t_1 - t_0}{N}, \quad N \in I\!\!N := \{1, 2, \ldots\}, \end{cases}$$

where $\gamma_{iN} \to 0$ and $\delta_{iN} \downarrow 0$ as $N \to \infty$ for all i. For each fixed $N \in I\!\!N$ problem (P_N) is *finite-dimensional* and seems to be simpler than the continuous-time problem (P). Indeed, applying well-developed methods of finite-dimensional variational analysis, it is possible to derive necessary optimality conditions in problems (P_N) even with nonsmooth data and general dynamic constraints governed by discrete inclusions and then obtain the corresponding results for optimal control of differential inclusions by passing to the limit from discrete approximations; see [4, 6, 13] for detailed proofs and discussions. However, this approach has some limitation regarding necessary optimality conditions of the *maximum principle* type.

As well known, the central result of the optimal control theory for continuous-time problems (P), the Pontryagin Maximum Principle (PMP) [11], holds with *no convexity* assumptions on the admissible velocity sets $f(t,x,U)$. This specific result, from the general viewpoint of optimization theory, is strongly due to continuous-time dynamic constraints in (P) governed by differential operators. It happens that continuous-type control systems enjoy a certain *hidden convexity*, which is deeply related to the classical Lyapounov theorem on the range convexity of nonatomic vector measures and eventually leads the to maximum principle form. It is not surprising therefore that an analogue of the maximum principle for discrete-time control systems *does not generally hold* without a priori convexity assumptions. This may create troubles for applications of the PMP in numerical calculations of nonconvex continuous-time control systems, which inevitably involve finite-difference approximations via time discretization. To avoid such troubles, it is sufficient to justify not a full analogue of the PMP, with the exact maximum condition, but its *approximate* counterpart, where an error in the maximum condition depends on the discretization stepsize *tending to zero* when the latter is decreasing.

The first result of this type in the absence of convexity assumptions was given by Gabasov and Kirillova [2, 3], under the name of "quasi-maximum principle," for parametric discrete systems with smooth cost and dynamics and with no endpoint constraints. The proof of this result, purely analytic, essentially exploited the unconstrained nature of the problem.

The following *Approximate Maximum Principle* (AMP) for the nonconvex constrained problems (P_N) was established by Mordukhovich [4, 5]. The proof in [4, 5] is geometric based on the discovered finite-difference counterpart of the hidden convexity property and the separation theorem. Denote

$$H(t,x,p,u) := \langle p, f(t,x,u)\rangle, \quad p \in I\!\!R^n, \tag{1.1}$$

the *Hamilton-Pontryagin function* for the dynamic constraints under consideration.

APPROXIMATE MAXIMUM PRINCIPLE. *Let the pairs* $(\bar{x}_N, \bar{u}_N)$ *be optimal to* (P_N) *for all* $N \in I\!\!N$*, where* U *is a compact subset of a metric space with the metric* $d(\cdot,\cdot)$*, where* f *is continuous with respect to its variables and continuously differentiable with respect to* x *in a tube containing the optimal trajectories* $\bar{x}_N(t)$ *for large* N*, and where each* φ_i *is continuously differentiable around the limiting point(s) of* $\{\bar{x}_N(t_1)\}$*. Impose the following assumptions:*

(a) *The* CONSISTENCY CONDITION *on the perturbation of the equality constraints meaning that*

$$\lim_{N\to\infty} \frac{h_N}{\delta_{iN}} = 0 \quad \textit{for all} \quad i = m+1, \ldots, m+r. \tag{1.2}$$

(b) *The* PROPERNESS *of the sequences of optimal controls* $\{\bar{u}_N\}$*, which means that for every increasing subsequence* $\{N\}$ *of natural numbers and every sequence of mesh points* $\tau_{\theta(N)} \in T_N$

satisfying $\tau_{\theta(N)} = t_0 + \theta(N)h_N$, $\theta(N) = 0, 1, \ldots, N-1$, *and* $\tau_{\theta(N)} \to t \in [t_0, t_1]$ *one has*

$$\text{either } d(u_N(\tau_{\theta(N)}), u_N(\tau_{\theta(N)+q})) \to 0 \quad \text{or } d(u_N(\tau_{\theta(N)}), u_N(\tau_{\theta(N)-q})) \to 0$$

as $N \to \infty$ *with any natural constant* q.

Then there are numbers $\{\lambda_{iN} | \, i = 0, \ldots, m+r\}$ *and a function* $\varepsilon(t, h_N) \to 0$ *as* $N \to \infty$ *uniformly in* $t \in T_N$ *such that*

$$H(t, \bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t)) = \max_{u \in U} H(t, \bar{x}_N(t), p_N(t+h_N), u) + \varepsilon(t, h_N) \tag{1.3}$$

for all $t \in T_N$ *and that*

$$\lambda_{iN}(\varphi_i(\bar{x}_N(t_1)) - \gamma_{iN}) = O(h_N), \quad i = 1, \ldots, m, \tag{1.4}$$

$$\lambda_{iN} \geq 0, \;\; i = 0, \ldots, m, \quad \text{and} \quad \sum_{i=0}^{m+r} \lambda_{iN}^2 = 1 \tag{1.5}$$

for all $N \in I\!N$, *where* $p_N(t)$, $t \in T_N \cup \{t_1\}$, *is the corresponding trajectory of the adjoint system*

$$p_N(t) = p_N(t+h_N) + h_N \frac{\partial H}{\partial x}(t, \bar{x}_N(t), p_N(t), \bar{u}_N(t)), \quad t \in T_N, \tag{1.6}$$

with the transversality condition

$$p_N(t_1) = -\sum_{i=0}^{m+r} \lambda_{iN} \nabla \varphi_i(\bar{x}_N(t_1)). \tag{1.7}$$

Observe that the closer h_N is to zero, the more precise the approximate maximum condition (1.3) and the approximate complementary slackness condition (1.4) are. This means that the AMP in (P_N) tends to the PMP in (P) as $N \to \infty$, which actually justifies the *stability* of the Pontryagin Maximum Principle with respect to discrete approximations under the assumptions made.

It has been shown in [4, 5] that the *consistency* condition in (a) is *essential* for the validity of the AMP in problems with equality constraints. The first goal of the paper is to examine the other two significant assumptions made in the above theorem: the *properness* condition in (b) and the *smoothness* of the initial data. We show in Section 2 that *both of these assumptions are essential for the validity of the AMP*.

Note that the properness of the sequence of optimal controls in (b) is a *finite-difference counterpart* of the piecewise continuity (or, more generally, of *Lebesgue regular points* having full measure) for optimal controls in continuous-time systems. It turns out that the situation when sequences of optimal controls are not proper in discrete approximations is not unusual for systems with nonconvex velocities, and it leads to the violation of the AMP already in the case of smooth problems with inequality constraints.

The impact of nonsmoothness to the validity of the AMP happens to be even more striking: the AMP does *not hold* in the expected conventional subdifferential form already for minimizing *convex* cost functions in discrete approximations of linear systems with no endpoint constraints, as well as for problems with nonsmooth dynamics. It seems that the *AMP is one of very few results on necessary optimality conditions that do not have expected counterparts in nonsmooth settings.*

On the other hand, we derive the AMP in problems (P_N) with nonsmooth functions describing the objective and inequality constraints in a new *upper subdifferential* (or superdifferential) form, which is also new for necessary optimality conditions in continuous-time control systems. The main difference between the conventional subdifferential form, which does not hold for the AMP but holds for the PMP, and the new one is that the latter involves upper (not lower) subgradients of nonsmooth functions in transversality conditions. This form applies to a class of *uniformly upper subdifferentiable* functions described in this paper, which particularly contains smooth and concave continuous functions being closed with respect to taking the minimum over compact sets. The results obtained solve a long-standing question about the possibility to establish the AMP in nonsmooth control problems. We also derive the upper subdifferential form of the AMP in discrete approximations of control systems with *time delays*, for which no results of this type have been known before. The main results of this paper have been announced in [9]

The rest of the paper is organized as follows. Section 2 contains examples on the *violation of the AMP* in smooth problems (P_N) without the properness condition as well as in problems with nonsmooth cost functions and/or nonsmooth dynamics. In Section 3 we discuss appropriate tools of nonsmooth analysis paying the main attentions to the concepts of *upper regularity* and uniform *upper subdifferentiability*, which are new in the study of *minimization* problems. The main Section 4 is devoted to the derivation of the AMP for the discrete approximation problems (P_N) in the *upper subdifferential form*; it contains three slightly different modifications of this results in somewhat distinct settings. In the final Section 5 we extend the AMP to discrete approximations of constrained *time-delay* systems, where the results obtained are new in both smooth and nonsmooth frameworks. We also present an example on the *violation* of the AMP in discrete approximations of functional-differential control systems of *neutral type*, even under smoothness assumptions in the absence of endpoint constraints.

Throughout the paper we use standard notation with some special symbols defined in the text where they are introduced.

2 Counterexamples

Let us start with an example on the violation of the AMP in discrete approximations of linear control systems with linear cost functions and linear endpoint inequality constraints but with *no properness condition.*

Example 2.1 (AMP does not hold in smooth control problems with no properness condition). *There is a two-dimensional linear control problem with an inequality constraint such that optimal controls in the sequence of its discrete approximations do not satisfy the Approximate Maximum Principle.*

Proof. Let us consider a linear continuous-time optimal control problem (P) with a two-dimensional state $x = (x_1, x_2) \in I\!R^2$ in the following form:

$$(2.1)\qquad \begin{cases} \text{minimize } \varphi(x(1)) := -x_1(1) \\ \text{subject to} \\ \dot{x}_1 = u, \quad \dot{x}_2 = x_1 - at, \quad x_1(0) = x_2(0) = 0, \\ u(t) \in U := \{0,1\}, \quad 0 \le t \le 1, \\ x_2(1) \le -\dfrac{a-1}{2}, \end{cases}$$

where $a > 1$ is a given constant. Observe that the only "unpleasant" feature of this problem is that the control set $U = \{0,1\}$ is *nonconvex*, and hence the feasible velocity sets $f(t,x,U)$ are nonconvex as well. It is clear that $\bar{u}(t) \equiv 1$ is the unique optimal solution to problem (2.1), and that the corresponding optimal trajectory is $\bar{x}_1(t) = t$, $\bar{x}_2(t) = -\dfrac{a-1}{2}t^2$. Moreover, the inequality constraint is active, since $\bar{x}_2(1) = -\dfrac{a-1}{2}$.

Let us now discretize this problem with the stepsize $h_N := \frac{1}{2N}$, $N \in I\!N$. For the notation convenience we omit the index N in what follows. Thus the discrete approximation problems (P_N) corresponding to (2.1) are written as:

$$(2.2)\qquad \begin{cases} \text{minimize } \varphi(x(1)) = -x_1(1) \\ \text{subject to} \\ x_1(t+h) = x_1(t) + hu(t), \quad x_1(0) = 0, \\ x_2(t+h) = x_2(t) + h(x_1(t) - at), \quad x_2(0) = 0, \\ u(t) \in \{0,1\}, \quad t \in \{0, h, \ldots, 1-h\}, \\ x_2(1) \le -\dfrac{a-1}{2} + h^2, \end{cases}$$

i.e., we put $\gamma_N := h_N^2$ in the constraint perturbation for (P_N).

To proceed, we compute the trajectories of (2.2) corresponding to $u(t) \equiv 1$. It is easy to see that $x_1(t) = t$ for this u. To compute $x_2(t)$, observe that

$$[y(t+h) = y(t) + ht,\ y(0) = 0] \Longrightarrow y(t) = \frac{t^2}{2} - \frac{th}{2}.$$

Indeed, one has by the direct calculation that

$$y(t) = h\sum_{\tau=0}^{t-h} = [\text{put } \tau = kh] = h^2 \sum_{k=0}^{\frac{t}{h}-1} k = h^2 \frac{\frac{t}{h}(\frac{t}{h}-1)}{2} = \frac{t^2}{2} - \frac{th}{2}.$$

Therefore for $x_2(t)$ corresponding to $u(t) \equiv 1$ in (2.2) we have

$$x_2(t) = h\sum_{\tau=0}^{t-h}(\tau - a\tau) = -\frac{a-1}{2}t^2 + \frac{a-1}{2}ht.$$

By this calculation we see that, for h sufficiently small, $x_2(t_1)$ no longer satisfies the endpoint constraint, and thus $u(t) \equiv 1$ is not a feasible control to problem (2.2) for all h close to zero. This implies that an optimal control to (2.2) for small h, which obviously exists, must have at least one *switching point* s such that $u(s) = 0$, and hence the maximum value of the corresponding endpoint $x_1(1)$ will be less than or equal to $1 - h$. Put

$$u(t) := \begin{cases} 1 & t \neq s, \\ 0 & t = s \end{cases}$$

and show that

$$x_2(t) = \begin{cases} -\dfrac{a-1}{2}t^2 + \dfrac{a-1}{2}ht, & t \leq s, \\ -\dfrac{a-1}{2}t^2 + \dfrac{a-1}{2}ht - h(t-s) + h^2, & t \geq s+h, \end{cases} \tag{2.3}$$

for the corresponding trajectories in (2.2) depending on h and s. We only need to justify the second part of this formula. To compute $x_2(t)$ for $t \geq s+h$, substitute $x_1(t) = t - h$ into (2.2). It is easy to see that the increment $\Delta x_2(t)$ compared to the case when $u(t) \equiv 1$ is

$$h\sum_{\tau=s+h}^{t-h}(-h) = -h(t-h-s) = -h(t-s) + h^2,$$

which justifies (2.3). Now let us specify the parameters of the above control putting $a = 2$ and $s = 0.5$ for all N, i.e., considering the discrete-time function

$$\bar{u}(t) := \begin{cases} 1 & t \neq 0.5, \\ 0 & t = 0.5; \end{cases}$$

note that the point $t = 0.5$ belongs to the grid T_N for all N due to $h_N := \frac{1}{2N}$. Observe that the sequence of these controls does *not satisfy the properness property* in the assumption (b) of the

AMP formulated in Section 1. It follows from (2.3) that the corresponding trajectories satisfy the endpoint constraint in (2.2) for all $N \in I\!N$, since $\bar{x}_2(1) = -\frac{1}{2}t^2 + h^2$. Moreover, it is clear from the above calculations that the control $\bar{u}(t)$ is optimal to problem (P_N) in (2.2) for any N. Let us show that the sequence of optimal controls $\bar{u}(t)$ does not satisfy the approximate maximum condition (1.3) at the point of switch.

The adjoint system (1.6) for problem (2.2) with any h is

$$p(t) = p(t+h) + h\frac{\partial f^*}{\partial x}(t, \bar{x}_1, \bar{x}_2, \bar{u})p(t+h),$$

where the Jacobian matrix $\partial f / \partial x$ and its adjoint/transposed one equal

$$\frac{\partial f}{\partial x} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \frac{\partial f^*}{\partial x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Thus we have the adjoint trajectories

$$p_1(t) = p_1(t+h) + hp_2(t+h) \quad \text{and} \quad p_2(t) \equiv \text{const},$$

where (p_1, p_2) satisfy the transversality condition (1.7) with the corresponding sign/notriviality conditions (1.5) written as

$$p_1(1) = \lambda_0, \; p_2(1) = -\lambda_1; \quad \lambda_0 \geq 0, \; \lambda_1 \geq 0, \; \lambda_0^2 + \lambda_1^2 = 1.$$

This implies that $p_1(t)$ is a linear nondecreasing function. The corresponding Hamilton-Pontryagin function (1.1) equals

$$H(t, x(t), p(t+h), u(t)) = p_1(t+h)u(t) + \text{terms not depending on } u.$$

Examining the latter expression and taking into account that the optimal controls are equal to one for all t but $t = 0.5$, we conclude that the approximate maximum condition (1.3) holds only if $p_1(t)$ is either nonnegative or tends to zero everywhere except $t = 0.5$. Observe that $p_1(t) \equiv 0$ yields $\lambda_1 = \lambda_2 = 0$, which contradicts the nontriviality condition. Hence $p_1(t)$ must be positive away from $t = 0$. Therefore a sequence of controls having a point of switch not tending to zero as $h \downarrow 0$ *cannot* satisfy the approximate maximum condition at this point. This shows that the AMP does not hold for the sequence of optimal controls to (2.2) built above. □

Many examples of this type can be constructed based on the above idea, which essentially means the following. Take a continuous-time problem with active inequality constraints and *nonconvex* admissible velocity sets $f(t, x, U)$. It often happens that after the discretization the "former" optimal control becomes not feasible in discrete approximations, and the "new" optimal control in the sequence of discrete-time problems has a singular point of switch (thus making the sequence of optimal controls not proper), where the approximate maximum condition does not hold.

The next example demonstrates that the AMP may be violated in problems of minimizing *nonsmooth cost* functions in linear systems with no endpoint constraint.

Example 2.2 (AMP does not hold for linear systems with nonsmooth and convex cost functions). *There is a one-dimensional control problem of minimizing a nonsmooth and convex cost function over a linear system with no endpoint constraints such that a proper sequence of optimal controls to discrete approximations does not satisfy the Approximate Maximum Principle.*

Proof. Consider the following sequence of one-dimensional optimal control problem (P_N), $N \in I\!N$, for discrete-time systems:

$$(2.4) \qquad \begin{cases} \text{minimize } \ \varphi(x_N(1)) := |x_N(1) - r| \\ \text{subject to} \\ x_N(t + h_N) = x_N(t) + h_N u_N(t), \quad x_N(0) = 0, \\ u_N(t) \in U := \{0, 1\}, \quad t \in T_N := \{0, h_N, \ldots, 1 - h_N\}, \end{cases}$$

where r is a positive *irrational* number less than 1 whose choice will be specified below. The dynamics in (2.4) is a discretization of the simplest ODE control system $\dot{x} = u$. Observe that, since r is irrational and h_N is rational, we have $\bar{x}_N(1) \neq r$ for the endpoint of an optimal trajectory to (2.4) as $N \in I\!N$, while obviously $\bar{x}(1) = r$ for optimal solutions to the continuous-time counterpart. It is also clear that for sufficiently small h_N an optimal control to (2.4) will be neither $u_N(t) \equiv 0$ nor $u_N(t) \equiv 1$, but it will have at least one point of switch.

Suppose that for some subsequence $N_k \to \infty$ one has $\bar{x}_{N_k}(1) > r$; put $\{N_k\} = I\!N$ without loss of generality. Let us show that in this case the approximate maximum condition does *not* hold at points $t \in T_N$ for which $\bar{u}_N(t) = 1$. Indeed, we have

$$H(\bar{x}_N(t), p_N(t), u) = p_N(t + h_N)u \quad \text{and} \quad p_N(t) \equiv -1$$

for the Hamilton-Pontryagin function and the adjoint trajectory in (1.6) and (1.7), since $\bar{x}_N(1) > r$ along the optimal solution to (2.4). Thus

$$\max_{u \in U} H(\bar{x}_N(t), p_N(t + h_N), u) = 0 \ \text{ for all } \ t \in T_N, \quad \text{while} \ \ H(\bar{x}_N(s), p_N(s + h_N), \bar{u}_N(s)) = -1$$

at the points $s \in T_N$ of control switch, where $\bar{u}_N(s) = 1$ regardless of h_N.

Let us specify the choice of r in (2.4) ensuring that $\bar{x}_N(1) > r$ along some subsequence of natural numbers. We claim that $\bar{x}_N(1) > r$ if $r \in (0, 1)$ is an irrational number whose decimal representation contains infinitely many digits from the set $\{5,6,7,8,9\}$; e.g., $r = 0.676676667....$ Indeed, put $h_N := 10^{-N}$, which is a subsequence of $h_N = N^{-1}$ as required in (2.4). It is easy to see that in this case the set of all reachable points at $t = 1$ is the set of rational numbers between 0 and 1 with exactly N digits in the fractional part of their decimal representations. In particular, for $N = 3$ this set is $\{0, 0.001, 0.002, ..., 0.999, 1\}$. Therefore, by the construction of r, the closest point to r from the reachable set is greater than r, and such point must be the endpoint of the optimal trajectory $\bar{x}_N(1)$.

It remains to show that one always can choose a sequence of optimal control to (2.4) satisfying the properness condition. Taking r as above, we denote by $s(N) \in T_N$ the point of the grid closest to r. It is easy to see that the control

$$\bar{u}_N(t) := \begin{cases} 1, & t \le s(N), \\ 0, & t \ge s(N) + h_N, \end{cases}$$

is optimal to (2.4) for each $N \in I\!N$, and the sequence $\{\bar{u}_N\}$ satisfies the properness condition. □

Example 2.2 contradicts the AMP with the transversality condition in the *conventional subdifferential form*, which is

$$-p_N(t_1) \in \partial\varphi(\bar{x}_N(t_1))$$

for problems with no endpoint constraints. In our example the function $\varphi(x) = |x - r|$ is *convex*, and hence the subdifferential ∂ is understood in the sense of convex analysis. Note that we actually showed that the subdifferential agrees with the gradient

$$\partial\varphi(\bar{x}_N(1)) = \{\nabla\varphi(\bar{x}_N(1))\} = \{1\} \text{ for all } N \in I\!N$$

along the optimal trajectories in (2.4) due to the choice of r. Since any reasonable (lower) subdifferential for nonsmooth functions must reduce to the convex subdifferential for convex ones, Example 2.2 proves that there is *no hope for an extension of the AMP in the conventional subdifferential form to problems with nonsmooth costs.*

The next example, complemented to Example 2.2, shows that the AMP fails even for problems with *differentiable* but *not continuously differentiable* cost functions.

Example 2.3 (AMP does not hold for linear systems with differentiable but not C^1 cost functions). *There is a one-dimensional control problem of minimizing a Fréchet differentiable but not continuously differentiable cost function over a linear system with no endpoint constraints such that a proper sequence of optimal controls to discrete approximations do not satisfy the Approximate Maximum Principle.*

Proof. Consider the same control system as in (2.4) and construct a minimizing function $\varphi(x)$ satisfying the above requirements. Let $\psi(x)$ be a C^1 function with the properties:

$$\psi(x) \geq 0, \quad \psi(x) = \psi(-x), \quad \psi(x) \equiv 0 \text{ if } |x| > 2,$$
$$|\nabla\psi(x)| \leq 1 \text{ for all } x, \quad \text{and } \nabla\psi(-1) = a > 0.$$

Define the cost function $\varphi(x)$ by

$$\varphi(x) := \Big(x - \frac{1}{9}\Big)^2 + \sum_{n=1}^{\infty} 10^{-2n-3}\psi\Big(10^{2n+3}(x - \sum_{k=1}^{n} 10^{-k}) - 1\Big),$$

which is continuously differentiable around every point but $x = \frac{1}{9}$, where it is differentiable and attains its absolute minimum at. As in Example 2.2, we put $h_N := 10^{-N}$, and then the point $x = \frac{1}{9}$ *cannot be reached* by discretization. It is not hard to check that the endpoint of the optimal trajectory $\bar{x}_N$ for each N is

$$\bar{x}_N(1) = \sum_{k=1}^{N} 10^{-k} \quad \text{with} \quad \nabla\varphi(\bar{x}_N(1)) = a + \varepsilon_N,$$

where $\varepsilon_N \downarrow 0$ as $N \to \infty$. Proceeding as in Example 2.2, with the same sequence of optimal controls, we have $H(\bar{x}_N(t), p_N(t + h_N), u) \equiv -au$, and the approximate maximum condition (1.3) does not hold at those points where $\bar{u}_N(t) = 1$. □

The last example in this section concerns systems with *nonsmooth dynamics*. We actually consider a finite-difference analogue of minimizing an integral functional over a one-dimensional control system, which is equivalent to a two-dimensional optimal control problem of the Mayer type. The discrete "integrand" in this problem is nonsmooth with respect to the state variable x; it happens to be continuously differentiable with respect to x *along* the optimal process $\{\bar{x}_N(\cdot), \bar{u}_N(\cdot)\}$ under consideration but *not uniformly* in N. Thus the example below demonstrates that the *uniform smoothness* assumption on f in a tube containing optimal trajectories is essential for the validity of the AMP formulated in Section 1.

Example 2.4 (violation of AMP for control problems with nonsmooth dynamics). *The AMP does not hold in discrete approximations of a minimization problem for an integral functional over a one-dimensional linear control system with no endpoint constraints such that the integrand is linear with respect to the control variable while convex and nonsmooth with respect to the state one. Moreover, the integrand in this problem happens to be C^1 with respect to the state variable along the sequence of optimal solutions to the discrete approximations (P_N) for all $N \in I\!N$ but not uniformly in N.*

Proof. First we consider the following continuous-time optimal control problem:

$$(2.5)\qquad \begin{cases} \text{minimize } \; J(x,u) := \displaystyle\int_0^{t_1} (u(t) + |x(t) - t^2/2|)dt \\ \text{subject to} \\ \dot{x} = tu, \quad x(0) = 0, \\ u(t) \in U := \{1, c\}, \quad 0 \le t \le t_1, \end{cases}$$

where the terminal time t_1 and the number $c > 1$ will be specified below. It is obvious that the optimal control to (2.5) is $\bar{u}(t) \equiv 1$ and the corresponding optimal trajectory is $\bar{x}(t) = t^2/2$.

Discretizing (2.5), we get the sequence of finite-difference control problems:

$$(2.6)\qquad \begin{cases} \text{minimize } \; J(x_N, u_N) := h_N \displaystyle\sum_{t \in T_N} (u_N(t) + |x_N(t) - t^2/2|) \\ \text{subject to} \\ x_N(t + h_N) = x_N(t) + h_N t u_N(t), \quad x_N(0) = 0, \\ u_N(t) \in U = \{1, c\}, \quad t \in T_N := \{kh_N\}_{k=0}^{N-1}. \end{cases}$$

Let us first show that $\bar{u}_N(t) \equiv 1$ remains to be the (unique) optimal control to (2.6) if the stepsize h_N is sufficiently small and (t_1, c) are chosen appropriately. Indeed, similarly to Example 2.1 we compute the trajectory to (2.6) corresponding to the control $u_N(t) \equiv 1$:

$$x_N(t) = \frac{t^2}{2} - \frac{th_N}{2} \quad \text{for all} \quad N \in I\!N.$$

The value $J_N(1)$ of the cost functional at $u_N(t) \equiv 1$ equals

$$(2.7)\qquad J_N(1) = t_1 + h_N^2 \sum_{t \in T_N} \frac{t}{2} = t_1 + \frac{t_1^2 h_N}{4} + o(h_N).$$

If we replace $u_N(t) = 1$ by $u_N(t) = c$ at some point $t \in T_N$, then the increment of the summation $h_N \sum_{t \in T_N} u_N(t)$ equals $(c-1)h_N$. Hence

$$J(x_N, u_N) = h_N \sum_{t \in T_N} u_N(t) + h_N \sum_{t \in T_N} |x_N(t) - t^2/2| > h_N \sum_{t \in T_N} u_N(t) \ge t_1 + (c-1)h_N$$

for any feasible control $u_N(t)$ to (2.6), which is not $u_N(t) \equiv 1$. Comparing the latter with (2.7), we conclude that the control $u_N(t) \equiv 1$ is *optimal* to (2.6) if $t_1^2/4 < c - 1$ and N is sufficiently large.

Let us finally show that for $t_1 > 2$ and $c > t^2/4 + 1$ (e.g., for $t_1 = 3$ and $c = 4$) the sequence of optimal controls $\bar{u}_N \equiv 1$ does *not* satisfy the approximate maximum condition (1.3) at points $t \in T_N$ sufficiently close to $t = t_1/2$. Compute the Hamilton-Pontryagin function (1.1) as a function of $t \in T_N$ and $u \in U$ at the optimal trajectory $\bar{x}_N(t)$ corresponding to the optimal control under consideration with the adjoint trajectory $p_N(t)$ to (1.6). Reducing (2.6) to the standard Mayer form and taking into account that $\bar{x}_N(t) < t^2/2$ for all $t \in T_N$ due to above formula for the trajectory of (2.6) corresponding to $u_N(t) \equiv 1$, we get

$$\begin{aligned} H(t, \bar{x}_N(t), p_N(t), u) &= tp_N(t+h_N)u - u - |\bar{x}_N(t) - t^2/2| \\ &= (tp_N(t+h_N) - 1)u + (\bar{x}_N(t) - t^2/2), \end{aligned}$$

where $p_N(t)$ satisfies the equation

$$p_N(t) = p_N(t+h_N) + h_N, \quad p_N(t_1) = 0,$$

whose solution is $p_N(t) = t_1 - t$. Therefore

$$H(t, \bar{x}_N(t), p_N(t), u) = (t(t_1 - t + h_N) - 1)u + O(h_N) = (-t^2 + t_1 t - 1)u + O(h_N).$$

The multiplier $-t^2 + t_1 t - 1$ is positive in the neighborhood of $t = t_1/2$ if its discriminant $t_1^2 - 4$ is positive. Thus $u = c$, but not $u = 1$, provides the maximum to the Hamilton-Pontryagin function around $t = t_1/2$ if h_N is sufficiently small. □

Observe that the constructions in Example 2.2 and 2.4 are actually based on the same idea. The crucial point in Example 2.2 (and similarly in Example 2.3) is that, due to the *incommensurability* of the reachable set and the ideal point of minimum $x_N(1) = r$, the endpoint of the optimal trajectory $\bar{x}_N(1)$ turns out to be in the zone, where the discontinuous derivative of the cost function has the "wrong sign". A similar situation is in Example 2.4, but in this case the function $\frac{\partial H}{\partial x}$ is discontinuous with respect to x, and the optimal trajectory in the discrete problem deviates to the "wrong side" of the ideal (continuous-time) optimal trajectory.

3 Uniformly Upper Subdifferentiable Functions

In this section we present some tools of nonsmooth analysis needed for the formulation and proofs of the main *positive* results of the paper: the Approximate Maximum Principle for ordinary and time-delay systems in the new *upper subdifferential* form. Results in this form are definitely non-traditional in optimization, since they concern *minimization* problems for which *lower* subdifferential constructions are usually employed. However, we saw in the preceding section that results of the conventional lower type simply do not hold for the AMP. In Sections 4 and 5 we are going to employ *upper* subdifferential constructions for nonsmooth minimization problems of optimal control, which happen to work for a special class of *uniformly upper subdifferentiable* functions we describe and discuss in this section.

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

$$\widehat{\partial}^+\varphi(\bar{x}) := \Big\{x^* \in I\!\!R^n \Big|\ \limsup_{x\to\bar{x}} \frac{\varphi(x) - \varphi(\bar{x})\langle x^*, x - \bar{x}\rangle}{\|x - \bar{x}\|} \le 0\Big\}. \tag{3.1}$$

This construction is known also as the "Fréchet superdifferential" or the "viscosity superdifferential"; it is extensively used in the theory of viscosity solutions. The set (3.1) is symmetric to the (lower) Fréchet subdifferential

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

which is widely used in variational analysis under the name of "regular" or "strict" subdifferential; see, e.g., [12] and [14]. The upper subdifferential (3.1) is our *primary* generalized differential construction in this paper. This set is closed and convex but may be empty for many functions useful in minimization. In fact, 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

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

Following [5], we define the *basic upper subdifferential* of φ at $\bar{x}$ by

$$\partial^+\varphi(\bar{x}) := \Big\{x^* \in I\!\!R^n |\ \exists\, x_k \to \bar{x}\ \text{ with }\ \varphi(x_k) \to \varphi(\bar{x})\ \text{ and }\ \exists\, x_k^* \in \widehat{\partial}^+\varphi(x_k)\ \text{ with }\ x_k^* \to x^*\Big\}$$

and call φ to be *upper regular* at $\bar{x}$ if $\partial^+\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x})$. This class includes, in particular, all strictly differentiable functions as well as proper concave functions. In the concave case $\widehat{\partial}^+\varphi(\bar{x})$ reduces to the upper subdifferential of convex analysis, which is nonempty whenever $\bar{x} \in \mathrm{ri}(\mathrm{dom}\,\varphi)$. Moreover, $\widehat{\partial}^+\varphi(\bar{x}) \ne \emptyset$ if φ is upper regular at $\bar{x}$ and Lipschitz continuous around this point. In the latter case the upper regularity of φ agrees with the subdifferential regularity of $-\varphi$ at the same point in the sense of [12]. It is interesting to observe that, for Lipschitzian upper regular functions, the Fréchet upper subdifferential (3.1) agrees with Clarke's generalized gradient $\overline{\partial}\varphi(\bar{x})$ of [1]. Indeed, one has

$$\overline{\partial}\varphi(\bar{x}) = \mathrm{co}\,\partial^+\varphi(\bar{x})$$

if φ is Lipschitz continuous around $\bar{x}$; see, e.g., [5, Theorem 2.1]. Since $\partial^+\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x})$ for upper regular functions and since $\widehat{\partial}^+\varphi(\bar{x})$ is always convex, we arrive at $\overline{\partial}\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x})$.

Let us now define a class of functions for which we obtain an extension of the AMP to nonsmooth control problems in the next section.

Definition 3.1 (uniform upper subdifferentiability). *A function* $\varphi\colon I\!\!R^n \to \overline{I\!\!R}$ *is* UNIFORMLY UPPER SUBDIFFERENTIABLE *around a point* $\bar{x}$ *where it is finite, if there is a neighborhood* V *of* $\bar{x}$ *such that for every* $x \in V$ *there exists* $x^* \in I\!\!R^n$ *with the following property: given any* $\varepsilon > 0$, *there is* $\eta > 0$ *for which*

$$\varphi(v) - \varphi(x) - \langle x^*, v - x\rangle \le \varepsilon\|v - x\| \tag{3.2}$$

whenever $v \in V$ *with* $\|v - x\| \le \eta$.

It is easy to check that the class of uniformly upper subdifferentiable functions includes all continuously differentiable functions, concave continuous functions, and also it is closed with respect to taking the minimum over compact sets. The uniform upper subdifferentiability property of φ around $\bar{x}$ is actually a localization of the so-called weak convexity property for $-\varphi$ in the sense of [10], which has been broadly used in numerical optimization. Note that if φ is Lipschitz continuous and differentiable at some point, it may not be uniformly upper subdifferentiable around it. Example: $\varphi(x) = x^2\sin(1/x)$ for $x \ne 0$ with $\varphi(0) = 0$. The following result shows, in particular, that uniformly upper subdifferential functions enjoy upper regularity and fully describes the set of x^* satisfying (3.2).

Proposition 3.2 (upper regularity of uniformly upper subdifferentiable functions). *Let* φ *be uniformly upper regular around* $\bar{x}$. *Then it is upper regular at* $\bar{x}$, *Lipschitz continuous around this point, and property* (3.2) *holds for all* $x^* \in \widehat{\partial}^+\varphi(x)$ *with* x *around* $\bar{x}$,

Proof. Denote by $G(x)$ the set of x^* for which (3.2) holds. This set is nonempty and, as directly follows from (3.2), it is closed, convex, and bounded for all $x \in V$. One can immediately observe from the comparison of (3.1) and (3.2) that $G(x) \subset \widehat{\partial}^+\varphi(x)$. Let us show that in fact $G(x) = \widehat{\partial}^+\varphi(x)$ whenever $x \in V$.

It follows from the results of [10, Section 1.1] that φ is locally Lipschitzian around $\bar{x}$, directionally differentiable on V in any direction, and its directional derivative admits the representation

$$\varphi'(x; w) = \min\{\langle x^*, w\rangle \mid x^* \in G(x)\} \quad \text{for all} \quad x \in V,\ w \in I\!\!R^n. \tag{3.3}$$

As well known, the Fréchet upper subdifferential (3.1) of a locally Lipschitzian function φ at x is representable as

$$\widehat{\partial}^+\varphi(x) = \Big\{x^* \in I\!\!R^n \Big|\ \langle x^*, w\rangle \ge \limsup_{\tau\downarrow 0} \frac{\varphi(x + \tau w) - \varphi(x)}{\tau} \ \text{ for all }\ w \in I\!\!R^n\Big\}.$$

Comparing the latter with (3.3), we get $G(x) = \widehat{\partial}^+\varphi(x)$ for all $x \in V$. Furthermore, it is not hard to show directly from the definition that the mapping $G\colon V \rightrightarrows I\!\!R^n$ is closed-graph on any compact subset of V. Taking finally into account the construction of the basic subdifferential, we conclude that $\partial^+\varphi(x) = \widehat{\partial}^+\varphi(x)$ for all $x \in V$. $\square$

Note that Proposition 3.2 is in accordance with [12, Theorem 9.16], which gives a characterization of the simultaneous Lipschitz continuity and subdifferential regularity of a function on a open set via the existence of the classical directional derivative and its upper semicontinuity with respect to directions. Note also that we need an extra requirement on the *uniform* upper subdifferentiability in Definition 3.1, which essentially restricts the class of functions suitable for applications to the AMP in the upper subdifferential form, due to the *parametric* nature of finite-difference systems viewed as a process as $N \to \infty$. In particular, the Lipschitz continuity and upper regularity are *not* needed for upper subdifferential results related to necessary optimality condition for *fixed* solutions in various problems of constrained optimization and optimal control; cf. [7, 8].

4 AMP in Upper Subdifferential Form

This is a central section of the paper, which collects the main positive results on the fulfillment of the AMP in the upper subdifferential form for the discrete approximation problems (P_N). We derive three closely related versions of the AMP in somewhat different settings of (P_N). The first version applies to problems with *no endpoint constraints* and establishes the upper subdifferential form of the AMP with *no properness* requirement on the sequence of optimal controls and with an *error estimate* as $\varepsilon(t, h_N) = O(h_N)$ in the approximate maximum condition. The second result, with a different proof, is the major version of the AMP for the *constrained nonsmooth* problems (P_N), which extends the one formulated in Section 1. The last version of the AMP concerns discrete approximation problems in the case of *incommensurability* of the time interval $t_1 - t_0$ and the discretization step h_N. This version is basic for the extension of the AMP to time-delay systems obtained in the next section.

Let us start with the upper subdifferential form of the AMP for problems with no endpoint constraints. Throughout this section impose the following *standing assumptions* on the mapping f and the control set U:

(H1) $f = f(t, x, u)$ is continuous with respect to all its variables and continuously differentiable with respect to the state variable x in some tube containing optimal trajectories for all u from the compact set U in a metric space and for all $t \in T_N$ uniformly in $N \in I\!N$.

Theorem 4.1 (AMP for problems with no endpoint constraints). *Let the pairs $(\bar{x}_N, \bar{u}_N)$ be optimal to problems (P_N) with $\varphi_i = 0$ for all $i = 1, \ldots, m + r$. Assume in addition to* (H1) *that φ_0 is uniformly upper subdifferentiable around the limiting point(s) of the sequence $\{\bar{x}_N(t_1)\}$,*

$N \in I\!N$. Then for every sequence of upper subgradients $x_N^* \in \widehat{\partial}^+\varphi_0(\bar{x}_N(t_1))$ there is $\varepsilon(t,h_N) \to 0$ as $N \to \infty$ uniformly in $t \in T_N$ such that the approximate maximum condition (1.3) holds for all $t \in T_N$, where each $p_N(t)$ satisfies the adjoint system (1.6) with the transversality condition

$$p_N(t_1) = -x_N^* \ \text{ for all } \ N \in I\!N. \tag{4.1}$$

Moreover, $\varepsilon(t,h_N) = O(h_N)$ in (1.3) if φ_0 is locally concave around $\bar{x}_N(t_1)$ uniformly in N while $\partial f(\cdot,u,t)/\partial x$ is locally Lipschitz around $\bar{x}_N(t)$ with a constant uniform in $u \in U$, $t \in T_N$, $N \in N$.

Proof. Considering a sequence of optimal solutions $(\bar{x}_N, \bar{u}_N)$ to (P_N), we suppose that $x_N(t)$ belong to the uniform neighborhoods in the assumptions made for all $N \in I\!N$. It follows from the uniform upper subdifferentiability of φ_0 by Proposition 3.2 that $\widehat{\partial}^+\varphi_0(\bar{x}_N(t_1)) \neq \emptyset$ and that inequality (3.2) holds for any $x^* \in \widehat{\partial}^+\varphi_0(\bar{x}_N(t_1))$ as $N \to \infty$. Now taking an arbitrary sequence of $x_N^* \in \widehat{\partial}^+\varphi_0(\bar{x}_N(t_1))$, we get

$$\varphi_0(x) - \varphi_0(\bar{x}_N(t_1)) \le \langle x_N^*, x - \bar{x}_N(t_1)\rangle + o(\|x - \bar{x}_N(t_1)\|), \tag{4.2}$$

where $\dfrac{o(\|x - \bar{x}_N(t_1)\|)}{\|x - \bar{x}_N(t_1)\|} \to 0$ as $x \to x_N(t_1)$ *uniformly* in N. Moreover, one can clearly eliminate o, i.e., put $o(\|x - \bar{x}_N(t_1)\|) \equiv 0$ if φ_0 is assumed to be uniformly locally concave. Letting $p_N(t_1) := -x_N^*$ as in (4.1), we derive from (4.2) that

$$J(x_N,u_N) - J(\bar{x}_N,\bar{u}_N) \le -\langle p_N(t_1), \Delta x_N(t_1)\rangle + o(\|\Delta x_N(t_1)\|), \tag{4.3}$$

with $\Delta x_N(t) := x_N(t) - \bar{x}_N(t)$, for all feasible processes (x_N, u_N) to (P_N) whenever $x_N(t_1)$ is sufficiently close to $\bar{x}_N(t_1)$. From the identity

$$\langle p_N(t_1), \Delta x_N(t_1)\rangle = \sum_{t \in T_N} \Big\langle p_N(t+h_N) - p_N(t), \Delta x_N(t)\Big\rangle + \sum_{t \in T_N} \Big\langle p_N(t+h_N), \Delta x_N(t+h_N) - \Delta x_N(t)\Big\rangle$$

and (4.3) we get

$$\begin{aligned} &0 \le J(x_N,u_N) - J(\bar{x}_N,\bar{u}_N) \le -\langle p_N(t_1), \Delta x_N(t_1)\rangle + o(\|\Delta x_N(t_1)\|) = \\ &= -\sum_{t \in T_N} \Big\langle p_N(t+h_N) - p_N(t), \Delta x_N(t)\Big\rangle - h_N \sum_{t \in T_N} \Big\langle p_N(t+h_N), \frac{\partial f}{\partial x}(t,\bar{x}_N,\bar{u}_N)\Delta x_N(t)\Big\rangle \\ &-h_N \sum_{t \in T_N} \Delta_u H(t,\bar{x}_N(t),p_N(t+h_N),\bar{u}_N(t)) + h_N \sum_{t \in T_N} \eta_N(t) + o(\|\Delta x_N(t_1)\|), \end{aligned} \tag{4.4}$$

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

$$\begin{aligned} \eta_N(t) = \ &\Big\langle \frac{\partial H}{\partial x}(t,\bar{x}_N(t),p_N(t+h_N),u_N(t)) - \frac{\partial H}{\partial x}(t,\bar{x}_N(t),p_N(t+h_N),\bar{u}_N(t)), \\ &\Delta x_N(t)\Big\rangle + o(\|\Delta x_N(t)\|), \end{aligned} \tag{4.5}$$

with $o(\|\Delta x_N(t)\|)$ uniform in N due to (H1), and where

$$\Delta_u H(t, \bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t)) := H(t, \bar{x}_N(t), p_N(t+h_N), u_N(t)) - H(t, \bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t)).$$

One can easily see that $o(\|\Delta x_N(t)\|) = O(\|\Delta x_N(t)\|^2)$ in (4.5), uniformly in N, under the additional Lipschitzian assumption on $\partial f(\cdot, u, t)/\partial x$ in the theorem.

Now let us consider *needle variations* of the optimal controls $\bar{u}_N$ in the following form:

$$u_N(t) = \begin{cases} v & \text{if } t = \tau, \\ \bar{u}_N(t) & \text{if } t \in T_N \setminus \{\tau\}\ , \end{cases} \tag{4.6}$$

where $v \in U$ and $\tau = \tau(N) \in T_N$ as $N \in I\!N$. All the controls (4.6) are feasible to the discrete problems with no endpoint constraints. The trajectory increments corresponding to the needle variations (4.6) satisfy

$$\Delta x_N(t) = 0 \text{ for } t = t_0, \ldots, \tau \quad \text{and} \quad |\Delta x_N(t)| = O(h_N) \text{ for } t = \tau + h_N, \ldots, t_1.$$

Taking this into account and substituting (4.6) into (4.4), we get

$$0 \le J(x_N, u_N) - J(\bar{x}_N, \bar{u}_N) \le -h_N \Delta_u H(\tau, \bar{x}_N(\tau, p_N(\tau + h_N), \bar{u}_N(\tau)) + o(h_N), \tag{4.7}$$

where one obviously has $o(h_N) = O(h_N^2)$ under the additional concavity and Lipschitzian assumptions made in the theorem. Arguing by contradiction, we derive from (4.7) the approximative maximum condition (1.3), with the specification of $\varepsilon(t, h_N)$ under the additional assumptions, and complete the proof of the theorem. □

Remark 4.2 (upper versus lower subdifferential forms of transversality conditions). The main difference between the conventional (lower) subdifferential form, which is *not* actually fulfilled in the case of AMP, and the upper subdifferential form of Theorem 4.1 is that the transversality condition (4.1) holds for *every* upper subgradient $x_N^* \in \widehat{\partial}^+\varphi_0(\bar{x}_N(t_1))$ instead of just for *some* lower subgradient in the conventional transversality conditions for continuous-time and discrete-time (with a fixed step) systems. In particular, for discrete-time systems with convex velocity sets both lower and upper subdifferential forms of the (exact) discrete maximum principle hold; see [8], where the upper subdifferential/ superdifferential form of the discrete maximum principle has been established under milder assumptions on φ_0 in comparison with Theorem 4.1. If φ_0 is Lipschitz continuous and upper regular and hence $\widehat{\partial}^+\varphi_0(\bar{x}) = \overline{\partial}\varphi_0(\bar{x})$, which is the case for uniformly upper subdifferential functions by Proposition 3.2, there is indeed a *dramatic* difference between the upper subdifferential form of transversality conditions and a well-recognized form in terms of the Clarke subdifferential: instead of the fulfillment transversality just for some element of $\overline{\partial}\varphi_0(\bar{x}(t_1))$ we establish its fulfillment for the *whole set*! Similar situation takes place for continuous-time systems, where the upper subdifferential form of transversality in the maximum principle can be proved, in problems with no endpoint constraints, in the line of arguments of Theorem 4.1. Observe, however, that there is a more subtle lower subdifferential form of transversality conditions for continuous-time and discrete-time (of a fixed step) systems that involves basic/limiting subgradients but not Clarke ones; see [5, 14]. This form is generally independent on the upper subdifferential form of transversality conditions. Note that the major drawback of the upper subdifferential form is that it applies to a restrictive class of functions. But, as we saw in Section 2, there is *no alternative* to this form for the Approximate Maximum Principle.

Next let us consider a sequence of the discrete approximation problems (P_N) with *endpoint constraints* of the inequality and equality types. We are going to derive an extension of the AMP formulated in Section 1 to these problems involving *nonsmooth* functions that describe the cost and inequality constraints. The following upper subdifferential version of the AMP for constrained problems impose the *uniform upper subdifferentiability* property on the cost and inequality constraint functions, the *properness* assumption on the sequence of optimal controls, and the *consistency* condition on perturbations of the equality constraints. As we saw in Section 2, all the three requirements are essential.

The proof of the AMP for constrained problems is substantially different from the one in Theorem 4.1 and much more involved, although it employs the same approach to handle nonsmoothness. The major part of the proof goes back to the smooth setting and is based on a finite-difference counterpart of the *hidden convexity* properties for sequences of discrete approximations.

Before formulating and proving the theorem, we need an auxiliary result that actually reflects the hidden convexity property in the nonsmooth setting under consideration. Let us first recall some definitions from [4, 5]. Given a sequence of feasible solutions (x_N, u_N) to (P_N), we say that the inequality constraint

$$\varphi_i(x_N(t_1)) \le \gamma_{iN} \ \text{ with } \ i \in \{1, \ldots, m\}$$

is *essential* for x_N along a subsequence $\mathcal{M}$ of natural numbers if $\varphi_i(x_N(t_1)) - \gamma_{iN} = O(h_N)$ as $h_N \to 0$, i.e., there is $K_i \ge 0$ such that

$$-K_i h_N \le \varphi_i(x_N(t_1)) - \gamma_{iN} \le 0 \ \text{ as } \ N \to \infty, \ \ N \in \mathcal{M}.$$

This constraint is *unessential* for x_N along $\mathcal{M}$ if for any $K > 0$ there is $N_0 \in I\!N$ such that

$$\varphi_i(x_N(t_1)) - \gamma_{iN} \le -K h_N \ \text{ for all } \ N \ge N_0, \ \ N \in \mathcal{M}.$$

Note that the notion of essential constraints in sequences of discrete approximation corresponds to the notion of *active* constraints in nonparametric optimization problems. Without loss of generality we suppose that for the sequence of optimal trajectories $\bar{x}_N$ to (P_N) under consideration the first $l \in \{1,\ldots,m\}$ inequality constraints are essential while the other $m-l$ constraints are unessential along all natural numbers, i.e., with $\mathcal{M} = \mathcal{N}$.

Assume now that $\widehat{\partial}^+\varphi_i(\bar{x}_N(t_1)) \neq \emptyset$ for all $i = 0,\ldots,l$ and $N \in I\!\!N$ sufficiently large and fix some sequence of upper subgradients $x^*_{iN} \in \widehat{\partial}^+\varphi_i(\bar{x}_N(t_1))$ for such i and N. Denote by $\Delta_{\tau,v}\bar{x}_N(t_1)$ the *endpoint increment* generated by the *needle variation* (4.6) of the optimal control $\bar{u}_N$ with some $\tau \in T_N$ and $v \in U$. Form the set

$$S_N := \{(s_0,\ldots,s_l) \in I\!\!R^n \mid s_i = \langle x^*_{iN}, \Delta_{\tau,v}\bar{x}_N(t_1)\rangle,\ \tau \in T_N,\ \ v \in U\} \tag{4.8}$$

along the fixed sequences of the above upper subgradients x^*_{iN} and consider the *negative orthant* in R^{l+1} given by

$$I\!\!R^{l+1}_< := \{(x_0,\ldots,x_l) \in I\!\!R^{l+1} \mid x_i < 0 \ \text{ for all } \ i = 0,\ldots,l\}.$$

The following result is due to the hidden convexity property of finite-difference systems established in [4, 5], with the adjustment to nonsmoothness via uniform upper subdifferentiability.

Lemma 4.3 (hidden convexity). *Let $(\bar{x}_N,\bar{u}_N)$ be a sequence of optimal solutions to problems (P_N) with no equality constraints and with the inequality constraints such that the first $l \in \{1,\ldots,m\}$ of them are essential for the sequence of $\bar{x}_N$ while the other are unessential for this sequence. In addition to* (H1) *assume that each φ_i, $i = 0,\ldots,l$, is uniformly upper subdifferentiable around the limiting point(s) of $\{\bar{x}_N(t_1)\}$, $N \in I\!\!N$, and that*

(H2) *the sequence of optimal controls $\{\bar{u}_N\}$ is proper.*

Then there is a sequence of $(l+1)$-dimensional quantities of order $o(h_N)$ as $h_N \downarrow 0$ such that

$$(\operatorname{co} S_N + o(h_N)) \cap I\!\!R^{l+1}_< = \emptyset \ \ \textit{for large} \ \ N \in I\!\!N, \tag{4.9}$$

where $\operatorname{co} S_N$ stands for the convex hull of the set S_N in (4.8) *built upon the given sequences of upper subgradients $x^*_{iN} \in \widehat{\partial}^+\varphi_i(\bar{x}_N(t_1))$, $i = 0,\ldots,l$.*

Proof. It follows the proof of Lemma 3 in [4] based on the hidden convexity property of Theorem 1 therein (respectively, Lemma 16.3 and Theorem 15.1 in [5]). The only essential difference is that the equalities

$$\varphi_i(\bar{x}_N(t_1)) - \varphi_i(\bar{x}_N(t_1)) - \langle \nabla\varphi_i(\bar{x}_N(t_1)), \Delta x_N(t_1)\rangle + o(\|\Delta x_N(t_1)\|) = 0, \quad i = 0,\ldots,l,$$

in the smooth case of [4, 5] are replaced with the inequalities

$$\varphi_i(\bar{x}_N(t_1)) - \varphi_i(\bar{x}_N(t_1)) - \langle x^*_{iN}, \Delta x_N(t_1)\rangle + o(\|\Delta x_N(t_1)\|) \le 0, \quad i = 0,\dots,l,$$

due to the uniform upper subdifferentiability of φ_i; cf. the proof of Theorem 4.1. □

Based on Lemma 4.3, we get the following extension of the AMP to finite-difference problems with nonsmooth inequality and smooth equality constraints.

Theorem 4.4 (AMP for problems with endpoint constraints). *Let the pairs $(\bar{x}_N, \bar{u}_N)$ be optimal to problems (P_N), where the first $l \in \{1,\dots,m\}$ inequality constraints are essential for $\bar{x}_N$, $N \in I\!N$. In addition to* (H1) *and* (H2) *assume that φ_i are uniformly upper subdifferentiable around the limiting point(s) of $\{\bar{x}_N(t_1)\}$ for $i = 0,\dots,l$ and continuously differentiable around them for $i = m+1,\dots,m+r$, and that*

(H3) *the consistency condition* (1.2) *holds for the perturbations δ_{iN} of the equality constraints.*

Then for any sequences of upper subgradients $x^*_{iN} \in \widehat{\partial}^+\varphi_i(\bar{x}_N(t_1))$, $i = 0,\dots,l$, there are numbers $\{\lambda_{iN} |\ i = 0,\dots,l, m+1,\dots,m+r\}$ and a function $\varepsilon(t,h_N) \to 0$ as $N \to \infty$ uniformly in $t \in T_N$ such that the approximate maximum condition (1.3) *is fulfilled with the adjoint trajectory $p_N(t)$ to* (1.6) *satisfying the transversality condition*

$$p_N(t_1) = -\sum_{i=0}^{l} \lambda_{iN} x^*_{iN} - \sum_{i=m+1}^{m+r} \lambda_{iN} \nabla\varphi_i(\bar{x}_N(t_1)) \tag{4.10}$$

along with

$$\varphi_i(\bar{x}_N(t_1)) - \gamma_{iN} = O(h_N) \quad \textit{for} \quad i = 1,\dots,l, \tag{4.11}$$

$$\lambda_{iN} \ge 0 \ \textit{ for } \ i = 0,\dots,l, \quad \textit{and} \quad \sum_{i=0}^{l} \lambda^2_{iN} + \sum_{i=m+1}^{m+r} \lambda^2_{iN} = 1. \tag{4.12}$$

Proof. Let us first consider the case of inequality constraints, i.e., when $\varphi_i = 0$ for the indices $i = m+1,\dots,m+r$ in (P_N). Take arbitrary sequences of $x^*_{iN} \in \widehat{\partial}^+\varphi_i(\bar{x}_N(t_1))$ as $i = 0,\dots,l$. By Lemma 4.3 we apply the separation theorem to the convex sets in (4.9). It follows from the structures of these sets that there are $\lambda_{iN} \ge 0$ for $i = 0,\dots,l$ and all N sufficiently large satisfying $\lambda^2_{0N} + \ldots + \lambda^2_{lN} = 1$ and

$$\sum_{i=0}^{l} \langle x^*_{iN}, \Delta_{\tau,v}\bar{x}_N(t_1)\rangle + o(h_N) \ge 0$$

for any $\tau \in T_N$ and $v \in U$. Then considering the trajectory $p_N(t)$ of the adjoint system (1.6) with the transversality condition

$$p(t_1) = -\sum_{i=0}^{l} \lambda_{iN} x^*_{iN}$$

and arguing as in the proof of Theorem 4.1, we get the inequality

$$h_N[H(\tau,\bar{x}_N(\tau),p_N(\tau+h_N),v)-H(\tau,\bar{x}_N(\tau),p_N(\tau+h_N),\bar{u}_N(\tau))]+o(h_N)\le 0$$

held for all $\tau\in T_N$ and $v\in U$. This easily implies the approximate maximum condition (1.3). Since (4.11) just means that the inequality constraints are essential for $\bar{x}_N$ as $i=1,\ldots,l$, we arrive at all the conclusions of the theorem for the case of inequality constraints. Observe that the result obtained ensures the fulfillment of the AMP with *zero multipliers* corresponding to unessential inequality constraints.

Next let us consider the general case of (P_N) when the equality constraints are present as well. Each equality constraint $\varphi_{iN}(x)=0$ can be obviously represented as the two inequality constraints

$$\varphi^+_{iN}(x):=\varphi_i(x)-\delta_{iN}\le 0,\quad \varphi^-_{iN}(x):=-\varphi_i(x)-\delta_{iN}\le 0 \tag{4.13}$$

for $i=m+1,\ldots,m+r$. Let us show that if one of the constraints (4.13) is essential for $\bar{x}_N$ along some subsequence $\mathcal{M}\subset I\!N$, then the other is unessential along the same subsequence under the consistency condition (1.2). Indeed, suppose for definiteness that the constraint $\varphi^+_{iN}(\bar{x}_N(t_1))\le 0$ is essential for some $i\in\{m+1,\ldots,m+r\}$ along $\mathcal{M}$. Then by (1.2) we have

$$\varphi^-_{iN}(\bar{x}_N(t_1))=-\varphi_i(\bar{x}_N(t_1))+\delta_{iN}-2\delta_{iN}=-\varphi^+_{iN}(\bar{x}_N(t_1))-2\delta_{iN}\le Kh_N,\ \ N\in\mathcal{M},$$

for any $K>0$ as $N\to\infty$, which means that the constraint $\varphi^-_{iN}(\bar{x}_N(t_1))\le 0$ is unessential. Since φ_i is assumed to be C^1 for $i=m+1,\ldots,m+r$, both φ^+_{iN} and φ^-_{iN} are uniformly upper subdifferentiable around the reference points. Applying now the inequality case of the theorem that has been already proved, we find either λ^+_{iN} or λ^-_{iN} corresponding to one of the essential constraints $\varphi^+_{iN}(\bar{x}_N(t_1))\le 0$ and $\varphi^-_{iN}(\bar{x}_N(t_1))\le 0$, respectively. Finally putting

$$\lambda_{iN}:=\lambda^+_{iN}\quad\text{or}\quad\lambda_{iN}:=-\lambda^-_{iN},\quad i=m+1,\ldots,m+r,$$

depending on which constraint is essential, we complete the proof of the theorem. □

Note that both Theorems 4.1 and 4.4 concern the discrete approximation problems (P_N) with $t_1-t_0=Nh_N$, i.e., when the time interval and the discretization step are commensurable. Of course, it is not always the case in applications. Moreover, to extend the AMP to time-delay systems in the next section, we reduce them to systems with no delays but with *incommensurable* t_1-t_0 and h_N. To proceed in this way, one needs to use modifications of the above results in the case of incommensurability. Let us present the corresponding modification of Theorem 4.1 for problems with no endpoint constraints. For simplicity we use the notation $f(t,x_N,u_N):=f(t,x_N(t),u_N(t))$ and consider the following sequences of discrete approximations with the grid

$$T_N:=\{t_0,t_0+h_N,\ldots,t_1-\widetilde{h}_N-h_N\},\quad h_N:=\frac{t_1-t_0}{N},\quad \widetilde{h}_N:=t_1-t_0-h_N\Big[\frac{t_1-t_0}{h_N}\Big],$$

where $[a]$ stands as usual for the greatest integer less than or equal to the real number a. The modified problems are written as

$$
(\widetilde{P}_N) \qquad \begin{cases} \text{minimize } J(x_N,u_N) := \varphi(x_N(t_1)) \\ \text{subject to} \\ x_N(t+h_N) = x_N(t) + h_N f(t,x_N,u_N), \quad t \in T_N, \quad x_N(t_0) = x_0 \in I\!R^n, \\ x_N(t_1) = x_N(t_1 - \widetilde{h}_N) + \widetilde{h}_N f(t_1 - \widetilde{h}_N, x_N, u_N), \\ u_N(t) \in U, \quad t \in T_N. \end{cases}
$$

Theorem 4.5 (AMP for problems with incommensurability). *Let the pairs* $(\bar{x}_N, \bar{u}_N)$ *be optimal to problems* $(\widetilde{P}_N)$. *Assume in addition to* (H1) *that* φ *is uniformly upper subdifferentiable around the limiting point(s) of the sequence* $\{\bar{x}_N(t_1)\}$, $N \in I\!N$. *Then for every sequence of upper subgradients* $x_N^* \in \widehat{\partial}^+\varphi(\bar{x}_N(t_1))$ *there is* $\varepsilon(t,h_N) \to 0$ *as* $N \to \infty$ *uniformly in* $t \in T_N$ *such that the approximate maximum condition*

$$
H(t,\bar{x}_N,p_N,u_N) = \max_{u \in U} H(t,\bar{x}_N,p_N,u) + \varepsilon(t,h_N)
$$

holds for all $t \in \widetilde{T}_N := T_N \cup \{t_1 - \widetilde{h}_N\}$, *where the Hamilton-Pontryagin function is defined by*

$$
H(t,x_N,p_N,u) := \begin{cases} \langle p_N(t+h_N), f(t,x_N,u)\rangle & \text{if } t \in T_N, \\ \langle p_N(t), f(t-\widetilde{h}_N,x_N,u)\rangle & \text{if } t = t_1 - \widetilde{h}_N, \end{cases}
$$

and where each $p_N(t)$ *satisfies the adjoint system*

$$
\begin{cases} p_N(t) = p_N(t+h_N) + h_N \dfrac{\partial f}{\partial x}^*(t,\bar{x}_N,\bar{u}_N)p_N(t+h_N), \quad t \in T_N, \\ p_N(t_1 - \widetilde{h}_N) = p_N(t_1) + \widetilde{h}_N \dfrac{\partial f}{\partial x}^*(t_1 - \widetilde{h}_N,\bar{x}_N,\bar{u}_N)p_N(t_1) \end{cases}
$$

with the transversality condition $p_N(t_1) = -x_N^*$.

Proof. It is similar to the proof of Theorem 4.1 with the following modification of the increment formula for the minimizing functional:

$$
\begin{aligned}
0 \ &\le J(x_N,u_N) - J(\bar{x}_N,\bar{u}_N) \le -\langle p_N(t_1), \Delta x_N(t_1)\rangle + o(\|\Delta x_N(t_1)\|) \\
&= -\sum_{t \in T_N} \Big\langle p_N(t+h_N) - p_N(t), \Delta x_N(t)\Big\rangle - \Big\langle p_N(t_1) - p_N(t_1 - \widetilde{h}_N), \Delta x_N(t_1 - \widetilde{h}_N)\Big\rangle \\
&-h_N \sum_{t \in T_N} \Big\langle p_N(t+h_N), \frac{\partial f}{\partial x}(t,\bar{x}_N,\bar{u}_N)\Delta x_N(t)\Big\rangle - \widetilde{h}_N \Big\langle p_N(t_1), \frac{\partial f}{\partial x}(t_1 - \widetilde{h}_N,\bar{x}_N,\bar{u}_N)\Delta x_N(t_1 - \widetilde{h}_N)\Big\rangle \\
&-h_N \sum_{t \in \widetilde{T}_N} \Delta_u H(t,\bar{x}_N,p_N,\bar{u}_N) + h_N \sum_{t \in \widetilde{T}_N} \eta_N(t) + o(\|\Delta x_N(t_1)\|),
\end{aligned}
$$

where $\Delta_u H$ and $\eta_N(t)$ are defined as above. Substituting now the adjoint trajectory into this formula and using the needle variation (4.6), we arrive at the conclusions of the theorem. □

Similarly to the proof of Theorem 4.4 we can get its modification to the case of incommensurability with the transversality and related conditions (4.10)–(4.12).

5 AMP for Discrete Approximations of Delay Systems

This section is devoted to the extension of the AMP in the upper subdifferential form to finite-difference approximations of *time-delay* control systems. Actually we are not familiar with any previous results on the AMP for optimal control problems with delays, so the results obtained below seem to be new even for smooth delay problems.

We pay the main attention to discrete approximations of the following time-delay problem with no endpoint constraints:

$$(D) \qquad \begin{cases} \text{minimize } J(x,u) := \varphi(x(t_1)) \\ \text{subject to} \\ \dot{x}(t) = f(t,x(t),x(t-\theta),u(t)) \;\; \text{a.e. } \; t \in [t_0,t_1], \\ x(t) = c(t), \quad t \in [t_0-\theta,t_0], \\ u(t) \in U \;\; \text{a.e. } \; t \in [t_0,t_1] \end{cases}$$

over measurable controls and absolute continuous trajectories, where $\theta > 0$ is the constant time-delay, and where $c\colon [t_0-\theta,t_0] \to I\!R^n$ is a given function defining the initial "tail" condition that is necessary to start the delay system. Based on the above constructions for non-delayed systems, one can derive similar results for delay systems with endpoint constraints. We may also extend the results obtained to more complicated delay systems involving variable delays, set-valued tail conditions, etc. On the other hand, we show in the end of this section that the AMP does *not hold* for discrete approximations of functional-differential systems of *neutral type* that contain time-delays not only in state variables but in velocity variables as well.

Let us build discrete approximations of the time-delay problem (D) based on the Euler finite-difference replacement of the derivative. In the case of time-delay systems we need to ensure that the point $t-\theta$ belongs to the discrete grid when t does. It can be achieved by defining the discretization step as $h_N := \dfrac{\theta}{N}$ in contrast to $h_N = \dfrac{t_1-t_0}{N}$ for the non-delayed problems (P_N). In such a scheme the length of the time interval t_1-t_0 is generally *no longer commensurable* with the discretization step h_N.

To this end we consider the following sequences of discrete approximations of the delay problem (D) with the grid on the main interval $[t_0,t_1]$ given by

$$T_N := \{t_0, t_0+h_N, \ldots, t_1-\widetilde{h}_N-h_N\}, \quad h_N := \frac{\theta}{N}, \quad \widetilde{h}_N := t_1-t_0-h_N\Big[\frac{t_1-t_0}{h_N}\Big],$$

but also involving the grid T_{0N} on the initial interval $[t_0-\theta,t_0]$ as below:

$$(D_N)\qquad \begin{cases} \text{minimize } J(x_N,u_N) := \varphi(x_N(t_1)) \\ \text{subject to} \\ x_N(t+h_N) = x_N(t) + h_N f(t,x_N(t),x_N(t-Nh_N),u_N(t)),\ t\in T_N, \\ x_N(t_1) = x_N(t_1-\widetilde{h}_N) + \widetilde{h}_N f(t_1-\widetilde{h}_N, x_N(t_1-\widetilde{h}_N), u_N(t_1-\widetilde{h}_N)), \\ x_N(t) = c(t), \quad t\in T_{0N} := \{t_0-\theta, t_0-\theta+h_N,\ldots,t_0\}, \\ u_N(t)\in U, \quad t\in T_N. \end{cases}$$

To derive the AMP for the sequence of problems (D_N), we reduce these problems to the ones *without delays* and employ the results of Section 4. This follows the "method of steps" developed by Warga [15] in the case of delay problems for continuous-time systems. Our assumptions on the initial data of (P) are similar to those in Section 4 for non-delay systems. A counterpart of (H1) is formulated as:

(H) $f=f(t,x,y,u)$ is continuous with respect to all its variables and continuously differentiable with respect to (x,y) in some tube containing optimal trajectories for all u from the compact set U in a metric space and for all $t\in\widetilde{T}_N := T_N\cup\{t_1-\widetilde{h}_N\}$ uniformly in $N\in I\!N$.

For convenience we introduce the following notation:

$$\xi_N(t) := (x_N(t), x_N(t-\theta)), \quad \bar{\xi}_N(t) := (\bar{x}_N(t),\bar{x}_N(t-\theta)),$$
$$f(t,\xi_N,u_N) := f(t,x_N(t),x_N(t-\theta),u_N(t)), \quad f(t,\bar{\xi}_N,u_N) := f(t,\bar{x}_N(t),\bar{x}_N(t-\theta),u_N(t)),$$

in which terms the *adjoint system* to (D_N) is written as

$$\begin{aligned} p_N(t) &= p_N(t+h_N) + h_N\frac{\partial f^*}{\partial x}(t,\bar{\xi}_N,\bar{u}_N)p_N(t+h_N) \\ &+h_N\frac{\partial f^*}{\partial y}(t+\theta,\bar{\xi}_N,\bar{u}_N)p_N(t+\theta+h_N) \quad \text{for } t\in T_N, \end{aligned}$$

$$p_N(t_1-\widetilde{h}_N) = p_N(t_1) + \widetilde{h}_N\frac{\partial f^*}{\partial x}(t_1-\widetilde{h}_N,\bar{\xi}_N,\bar{u}_N)p_N(t_1)$$

along the optimal processes $(\bar{x}_N,\bar{u}_N)$ to the delay problems for each $N\in I\!N$. Introducing the corresponding *Hamilton-Pontryagin function*

$$(5.1)\qquad H(t,x_N,y_N,p_N,u) := \begin{cases} \langle p_N(t+h_N), f(t,x_N,y_N,u)\rangle & \text{if } t\in T_N, \\ \langle p_N(t), f(t-\widetilde{h}_N,x_N,y_N,u)\rangle & \text{if } t=t_1-\widetilde{h}_N, \end{cases}$$

with $\bar{y}_N(t)=\bar{x}_N(t-\theta)$, we rewrite the adjoint system as

$$(5.2)\ \begin{cases} p_N(t) = p_N(t+h_N) + h_N\Big[\dfrac{\partial H}{\partial x}(t,\bar{\xi}_N,p_N,\bar{u}_N) + \dfrac{\partial H}{\partial y}(t+\theta,\bar{\xi}_N,p_N,\bar{u}_N)\Big],\ t\in T_N, \\ p_N(t_1-\widetilde{h}_N) = p_N(t_1) + \widetilde{h}_N\dfrac{\partial H}{\partial x}(t_1-\widetilde{h}_N,\bar{\xi}_N,p_N,\bar{u}_N). \end{cases}$$

Theorem 5.1 (AMP for delay systems). *Let the pairs $(\bar{x}_N, \bar{u}_N)$ be optimal to problems (D_N). Assume in addition to* (H) *that φ is uniformly upper subdifferentiable around the limiting point(s) of the sequence $\{\bar{x}_N(t_1)\}$, $N \in \mathbb{N}$. Then for every sequence of upper subgradients $x_N^* \in \widehat{\partial}^+\varphi(\bar{x}_N(t_1))$ the approximate maximum condition*

$$H(t, \bar{\xi}_N, p_N, \bar{u}_N) = \max_{u \in U} H(t, \bar{\xi}_N, p_N, u) + \varepsilon(t, h_N), \quad t \in \widetilde{T}_N, \tag{5.3}$$

holds with the Hamilton-Pontryagin function (5.1) *and with some $\varepsilon(t, h_N) \to 0$ as $h_N \to 0$ uniformly in $t \in \widetilde{T}_N$, where the adjoint trajectory p_N satisfies* (5.2) *and the transversality relations*

$$p_N(t_1) = -x_N^*, \qquad p_N(t) = 0 \quad \text{as} \quad t > t_1. \tag{5.4}$$

Proof. Let us reduce the delay discrete approximation problems to the ones with no delay by the following multistep procedure. Denote

$$\begin{aligned}
&y_{1N}(t) := x_N(t - h_N), \quad t \in \{t_0 + 2h_N, \ldots, t_1 - \widetilde{h}_N\},\\
&y_{1N}(t) := c_N(t - h_N), \quad t \in \{t_0 - \theta + h_N, \ldots, t_0 + h_N\},\\
&y_{2N}(t) := y_{1N}(t - h_N), \quad t \in \{t_0 - \theta + 2h_N, \ldots, t_1 - \widetilde{h}_N\},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&y_{NN}(t) := y_{N-1,N}(t - h_N), \quad t \in \{t_0, \ldots, t_1 - \widetilde{h}_N\},
\end{aligned}$$

and observe that the values of $y_{1N}(t_1), \ldots, y_{NN}(t_1)$ can be defined arbitrarily, since they do not enter either the adjoint system or the cost function. To match the setup of Theorem 4.1, define

$$y_{1N}(t_1) := x_N(t_1 - \widetilde{h}_N), \;\; y_{2N}(t_1) := y_{1N}(t_1 - \widetilde{h}_N), \ldots, y_{NN}(t_1) := y_{N-1,N}(t_1 - \widetilde{h}_N).$$

After the change of variables one has

$$y_{NN}(t) = \begin{cases} x_N(t - \theta), & t \in \{t_0 + \theta + h_N, \ldots, t_1 - \widetilde{h}_N\},\\ c(t - \theta), & t \in \{t_0, \ldots, t_0 + \theta\}. \end{cases}$$

The original system in (D_N) is thereby reduced, for each $N \in \mathbb{N}$, to the following *non-delayed* system of dimension $\mathbb{R}^{(N+1)n}$:

$$\begin{cases} z_N(t + h_N) = z_N(t) + h_N g(t, z_N, u_N), \;\; t \in T_N,\\ z_N(t_1) = z_N(t_1 - \widetilde{h}_N) + \widetilde{h}_N g(t_1 - \widetilde{h}_N, z_N, u_N), \end{cases} \tag{5.5}$$

with the state vector $z_N(t) := (x_N(t), y_{1N}(t), \ldots, y_{NN}(t))^*$ (the star stands for transposing) and with the mapping $g(t, z_N, u_N)$ given by

$$g(t, z_N(t), u_N(t)) := \begin{pmatrix} f(t, x_N(t), y_{NN}(t), u_N(t))\\ \dfrac{x_N(t) - y_{1N}(t)}{h_N}\\ \ldots\ldots\\ \dfrac{y_{N-1,N}(t) - y_{NN}(t)}{h_N} \end{pmatrix},$$

where h_N should be replaced by $\widetilde{h}_N$ for $t = t_1 - \widetilde{h}_N$ in the last formula.

Let us apply Theorem 4.1 to the minimizing the same functional as in (D_N) on the feasible pairs (z_N, u_N) of the non-delayed system (5.5). The adjoint system in this problem, with respect to the new adjoint variable $q \in I\!R^{(N+1)n}$, has the form

$$
(5.6) \qquad \begin{cases} q_N(t) = q_N(t+h_N) + h_N \dfrac{\partial g}{\partial z}^*(t, \bar{z}_N, \bar{u}_N) q(t+h_N), & t \in T_N, \\ q_N(t_1 - \widetilde{h}_N) = q_N(t_1) + \widetilde{h}_N \dfrac{\partial g}{\partial z}^*(t_1 - \widetilde{h}_N, \bar{z}_N, \bar{u}_N) q_N(t_1) \end{cases}
$$

with the transversality condition

$$
(5.7) \qquad q_N(t_1) = -\Big(x_N^*, 0, \ldots, 0\Big)^* \ \text{ for } \ x_N^* \in \widehat{\partial}^+ \varphi(\bar{x}_N(t_1)).
$$

Computing the matrix $\dfrac{\partial g}{\partial z}$, we get

$$
\frac{\partial g^*}{\partial z} = \frac{1}{h_N} \begin{pmatrix} h_N \dfrac{\partial f^*}{\partial x} & 1 & \ldots & 0 & 0 \\ 0 & -1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ h_N \dfrac{\partial f^*}{\partial y_{NN}} & 0 & \ldots & 0 & -1 \end{pmatrix}.
$$

Taking this into account and performing elementary calculations, we arrive at the adjoint system (5.2) and the transversality relations (5.4) for the first component $p_N(t)$ of the the adjoint trajectory $q_N(t)$ satisfying (5.6) and (5.7). Denoting by $\widetilde{H}(t, z_N, q_N, u)$ the Hamilton-Pontryagin function (5.1) to the non-delayed system (5.5), one has

$$
\begin{aligned} \widetilde{H}(t, \bar{z}_N, q_N, u) &= \langle q_N(t+h_N), g(t, \bar{z}_N, u)\rangle = \langle p_N(t+h_N), f(t, \bar{\xi}_N, u)\rangle + r(t, \bar{z}_N, q_N, h_N) \\ &= H(t, \bar{\xi}_N, p_N, u) + r(t, \bar{z}_N, q_N, h_N), \quad t \in T_N, \end{aligned}
$$

and similarly for $t = t_1 - \widetilde{h}_N$, where H is given in (5.1), and where the remainder $r(t, \bar{z}_N, q_N, h_N)$ does not depend on u. Finally applying the approximate maximum condition (1.3) from Theorem 4.1 to system (5.5), we arrive at (5.3) and complete the proof of the theorem. □

Note that in the case of continuously differentiable cost functions φ around $\bar{x}_N(t_1)$ uniformly in N, the transversality relations (5.4) reduce to

$$
p_N(t_1) = -\nabla \varphi(\bar{x}_N(t_1)), \qquad p_N(t) = 0 \ \text{ as } \ t > t_1.
$$

Similarly to the proof of Theorem 5.1 we can deduce from Theorem 4.4 its delay counterpart for discrete approximation problems with *endpoint constraints*. In this result we add assumptions

(H2) and (H3) to those in (H) and replace the transversality relations (5.4) in Theorem 5.1 by the conditions (4.10)–(4.12) with $p_N(t) = 0$ as $t > t_1$.

Finally in this section, we consider optimal control problems for finite-difference approximations of the so-called *functional-differential systems of neutral type*

$$\dot{x}(t) = f(t, x(t), x(t-\theta), \dot{x}(t-\theta), u(t)) \quad \text{a.e.} \quad t \in [t_0, t_1], \tag{5.8}$$

which contain time-delays not only in state but also in *velocity* variables. A finite-difference counterpart of (5.8) with the stepsize h and with the grid $T := \{t_0, t_0 + h, \ldots, t_1 - h\}$ is

$$x(t+h) = x(t) + hf\Big(t, x(t), x(t-\theta), \frac{x(t-\theta+h) - x(t-\theta)}{h}, u(t)\Big), \quad t \in T, \tag{5.9}$$

and the adjoint system is given by

$$\begin{aligned} p(t) \ &= p(t+h) + h\frac{\partial f^*}{\partial x}(t, \bar{\xi}, \bar{u})p(t+h) + h\frac{\partial f^*}{\partial y}(t+\theta, \bar{\xi}, \bar{u})p(t+\theta+h) \\ &+h\frac{\partial f^*}{\partial z}(t+\theta-h, \bar{\xi}, \bar{u})p(t+\theta) - h\frac{\partial f^*}{\partial z}(t+\theta, \bar{\xi}, \bar{u})p(t+\theta+h), \quad t \in T, \end{aligned} \tag{5.10}$$

where $(\bar{x}, \bar{u})$ is an optimal solution to the neutral analogue of problem (D_N), and where

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

It has been proved in [8] that optimal solutions to problems like (D_N) for discrete systems of the neutral type (5.9) satisfy the *exact discrete maximum principle* with transversality conditions in the *upper subdifferential form* provided that the velocity sets $f(t, x, y, z, U)$ are *convex* around $\bar{\xi}(t)$. What about an analogue of the approximate maximum principle with no convexity assumptions on the velocity sets? The following example shows that the AMP is *not fulfilled* for finite-difference neutral systems, in contrast to ordinary and delay ones, even in the case of *smooth* cost functions.

Example 5.2 (AMP does not hold for neutral systems). *There is a two-dimensional control problem of minimizing a linear function over a smooth neutral system with no endpoint constraints such that some sequence of optimal controls to discrete approximations does not satisfy the approximative maximum principle regardless of the stepsize and a mesh point.*

Proof. Consider the following parametric family of discrete optimal control problems with the parameter $h > 0$:

$$(5.11)\quad \begin{cases} \text{minimize } J(x_1, x_2, u) := x_2(2) \\ \text{subject to} \\ x_1(t+h) = x_1(t) + hu(t), \;\; t \in T := \{0, h, \ldots, 2-h\}, \\ x_2(t+h) = x_2(t) + h\Big(\dfrac{x_1(t-1+h) - x_1(t-1)}{h}\Big)^2 - hu^2(t), \;\; t \in T, \\ x_1(t) \equiv x_2(t) \equiv 0, \;\; t \in T_0 := \{-1, \ldots, 0\}, \\ |u(t)| \le 1, \;\; t \in T. \end{cases}$$

It is easy to see that

$$x_2(1) = -h \sum_{t=0}^{1-h} u^2(t) \quad \text{and}$$

$$\begin{aligned} J(x_1, x_2, u) := x_2(2) \; &= x_2(1) + h \sum_{t=1}^{2-h} \Big(\frac{x_1(t-1+h) - x_1(t-1)}{h}\Big)^2 - h \sum_{t=1}^{2-h} u^2(t) \\ &= -h \sum_{t=0}^{1-h} u^2(t) + h \sum_{t=0}^{1-h} u^2(t) - h \sum_{t=1}^{2-h} u^2(t) = -h \sum_{t=1}^{2-h} u^2(t). \end{aligned}$$

Thus the control

$$\bar{u}(t) = \begin{cases} 0, & t \in \{0, \ldots, 1-h\}, \\ 1, & t \in \{1, \ldots, 2-h\}, \end{cases}$$

is the only optimal control to (5.11) for any h. The corresponding trajectory is

$$\bar{x}_1(t) = \begin{cases} 0, & t \in \{0, \ldots, 1-h\}, \\ t-1, & t \in \{1, \ldots, 2-h\}; \end{cases} \qquad \bar{x}_2(t) = \begin{cases} 0, & t \in \{0, \ldots, 1-h\}, \\ -t+1, & t \in \{1, \ldots, 2-h\}. \end{cases}$$

Computing the partial derivatives of f in (5.11), we get

$$\frac{\partial f}{\partial x} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \frac{\partial f}{\partial y} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and}$$

$$\frac{\partial f}{\partial z}(t+1) = \frac{1}{h} \begin{pmatrix} 0 & 0 \\ 2(x_1(t+h) - x_1(t)) & 0 \end{pmatrix}.$$

Hence the adjoint system (5.10) reduces to

$$\begin{aligned} &p_1(t) = p_1(t+h) + 2(\bar{x}_1(t) - \bar{x}_1(t-h))p_2(t+1) - 2(\bar{x}_1(t+h) - \bar{x}_1(t))p_2(t+1+h) \\ &p_2(t) \equiv \text{const}, \quad t \in \{0, \ldots, 2-h\}, \end{aligned}$$

with the transversality conditions

$$p_1(2) = 0, \;\; p_2(2) = -1; \quad p_1(t) = p_2(t) = 0 \;\text{ for }\; t > 2.$$

The solution of this system is

$$p_1(t) \equiv 0, \ \ p_2(t) \equiv -1 \ \text{ for all } \ t \in \{0, \dots, 2-h\}.$$

Thus the Hamilton-Pontryagin function along the optimal solution is

$$\begin{aligned} H(t, \bar{x}_1, \bar{x}_2, p_1, p_2, u) \ &= p_1(t+h)u + p_2(t+h)\Big\{\Big(\frac{x_1(t-1+h) - x_1(t-1)}{h}\Big)^2 - u^2\Big\} \\ &= u^2 \ \text{ for all } \ t \in \{0, \dots, 1-h\}. \end{aligned}$$

This shows that the optimal control $\bar{u}(t) = 0$ does not provide the approximate maximum to the Hamilton-Pontryagin function regardless of h and the mesh point $t \in \{0, \dots, 1-h\}$. Note at the same time that another sequence of optimal controls with $\bar{u}(t) = 1$ for all $t \in \{0, \dots, 2-h\}$ satisfies the exact discrete maximum principle regardless of h. □

References

[1] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[2] R. GABASOV and F. M. KIRILLOVA, *On the extension of the maximum principle by L.S. Pontrygin to discrete systems,* Autom. Telem., 27 (1966), pp. 46-61.

[3] R. GABASOV and F. M. KIRILLOVA, *Qualitative Theory of Optimal Processes,* Marcel Dekker, New York, 1976.

[4] B. S. MORDUKHOVICH, *Approximate maximum principle for finite-difference control systems,* U.S.S.R. Comput. Maths. Math. Phys., 28 (1988), pp. 106–114.

[5] B. S. MORDUKHOVICH, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988.

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

[7] B. S. MORDUKHOVICH, *Necessary conditions in nonsmooth minimization via lower and upper subgradients*, to appear in Set-Valued Anal.

[8] B. S. MORDUKHOVICH and I. SHVARTSMAN, *Discrete maximum principle for nonsmooth optimal control problems with delays*, Cybern. Syst. Anal., 2002, No. 2, pp. 123–133.

[9] B. S. MORDUKHOVICH and I. SHVARTSMAN, *The approximate maximum principle for constrained control systems*, Proc. 41nd IEEE Conf. Dec. Cont., Las Vegas, NV, 2002, pp. 4345–4350.

[10] E. A. NURMINSKII, *Numerical Methods of Solutions to Deterministic and Stachastic Minimax Problems*, Naukova Dumka, Kiev, 1979.

[11] L. S. PONTRYAGIN, V. G. BOLTYANSKII, R. V. GAMKRELIDZE and E. F. MISHCHENKO, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.

[12] R. T. ROCKAFELLAR and R. J.-B. WETS, *Variational Analysis*, Springer, Berlin, 1998.

[13] G. V. SMIRNOV, *Introduction to the Theory of Differential Inclusions*, American Mathematical Society, Providence, RI, 2002.

[14] R. B. VINTER, *Optimal Control*, Birkhäuser, Boston, 2000.

[15] J. WARGA, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.