2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Necessary Optimality Conditions in Discrete Nonsmooth Optimal Control

Ilya Shvartsman

Abstract— **In this paper we outline a simple proof of the maximum principle for a nonsmooth discrete-time optimal control problem. The methodology is general and encompasses all subdifferentials for which the Lagrange Multiplier rule and the Chain Rule hold. This includes, but is not limited to, Mordukhovich (limiting), Clarke and Michel-Penot subdifferentials.**

Keywords. Nonsmooth optimal control, necessary optimality conditions, subdifferential.

I. Introduction and Preliminaries

In this paper we consider the following discrete-time problem:

$$\begin{cases} \text{Minimize } \varphi_0(x(a+1),\ldots,x(b)) \\ x(t+1)=f(t,x(t),u(t)),\ t\in\{a,a+1,\ldots,b-1\}, \\ x(a)=x_0, \\ \varphi_j(x(a+1),\ldots,x(b))\le 0,\ j=1,\ldots,r_1 \\ \varphi_j(x(a+1),\ldots,x(b))=0,\ j=r_1+1,\ldots,r_1+r_2, \\ u(t)\in U(t),\ t\in\{a,a+1,\ldots,b-1\} \end{cases} \tag{1}$$

Denote

$$T:=\{a,a+1,\ldots,b-1\}$$

In problem (1) $f: T\times I\!R^n\times I\!R^m\to I\!R^n$, $\varphi_j: I\!R^{n|T|}\to I\!R$, $j=0,\ldots,r_1+r_2$, and $U(\cdot)$ is a map from T to $I\!R^m$.

It is well known that to ensure the validity of the Pontryagin Maximum Principle in discrete time in a general setting, convexity of the sets $f(t,\bar x(t),U(t))$, $t\in T$ needs to be assumed (see, for e.g. [3] for the proof of the "smooth" discrete Maximum Principle). There is a lot of literature on generalizations of the Pontryagin Maximum Principle to nonsmooth problems in continuous time, with generalized differentiation mostly based on the so-called *limiting subdifferential* ∂^L introduced by Mordukhovich ([5], [6], [7], [2], [10] to name a few). At the same time, to the best of our knowledge, necessary optimality conditions in the context of problem (1) in terms of the limiting subdifferential are derived only in [5], but the obtained results are weaker and the assumptions are stronger than those in Theorem 1 below. We refer the reader to [6], Section 4 for the definition and main properties of the limiting subdifferential. Another possible approach to nonsmooth discrete-time problems, via the so-called *Fréchet upper subdifferential*, is developed in [8].

I. Shvartsman is with Dept. of Mathematics and Computer Science, Penn State Harrisburg, Middletown, PA 17057, USA ius13@psu.edu

In what follows, along with the limiting subdifferential, we will be using the construction of *Michel-Penot subdifferential* described below.

Let $f: I\!R^n\to I\!R$ be a continuous function. Michel-Penot directional derivative at the point $\bar x$ is defined as

$$d^M f(\bar x,h):=\sup_{e\in I\!R^n}\limsup_{t\downarrow 0}\frac{f(\bar x+t(h+e))-f(\bar x+te)}{t}$$

and the subdifferential is the dual construction:

$$\partial^M f(\bar x)=\{x^*|\langle x^*,h\rangle\le d^M f(\bar x,h)\text{ for all } h\in I\!R\}.$$

Michel-Penot subdifferential is always convex and, consequently, can be larger than the nonconvex limiting subdifferential (for e.g., for $\varphi(x)=-|x|$, $\partial^L\varphi(0)=\{-1,1\}$, while $\partial^M\varphi(0)=[-1,1]$). At the same time, Michel-Penot subdifferential shrinks to $\{\nabla f(\bar x)\}$ if f is merely differentiable *at* $\bar x$, while the limiting subdifferential shrinks to the gradient only under a stronger condition of strict differentiability. Therefore, both strict inclusions $\partial^L\varphi(\bar x)\subset\partial^M\varphi(\bar x)$ and $\partial^L\varphi(\bar x)\supset\partial^M\varphi(\bar x)$ are possible. We refer the reader to [4], [9] and references therein for more information on Michel-Penot subdifferential.

For $p\in I\!R^n$ denote

$$H(t,p,x,u)=p^T f(t,x,u)$$

and let $(\bar x,\bar u)$ be an optimal process in (1). The main result of this paper is the following theorem.

Theorem 1. *Let $f(t,\cdot,\cdot)$ be continuous around the optimal process, $f(t,\cdot,\bar u(t))$ be differentiable at $\bar x(t)$ for all $t\in T$, functions φ_j, $j=0,\ldots,r_1+r_2$ be Lipschitz continuous at $(\bar x(a+1),\ldots,\bar x(b))$ and the sets $f(t,\bar x(t),U(t))$ be convex for all $t\in T$. Then there exist*

(a) *numbers λ_j, $j=0,\ldots,r_1+r_2$, not all zero, such that $\lambda_j\ge 0$ and $\lambda_j\varphi_j(\bar x(a+1),\ldots,\bar x(b))=0$, $j=0,\ldots,r_1$,*

(b) *elements*

$$z_j^*=(z_{a+1j}^*,\ldots,z_{bj}^*)\in\partial^M(\lambda_j\varphi_j)(\bar x(a+1),\ldots,\bar x(b)),\quad j=0,\ldots,r_1+r_2, \tag{2}$$

(c) *function $p(\cdot)$ satisfying the adjoint equation*

$$p(t)=H_x(t,p(t+1),\bar x(t),\bar u(t))-\sum_{j=0}^{r_1+r_2}z_{tj}^*,\quad t\in T$$

$$p(b)=-\sum_{j=0}^{r_1+r_2}z_{bj}^* \tag{3}$$

such that there holds the maximum condition

$$H(t,p(t{+}1),\bar{x}(t),\bar{u}(t)) = \max_{u\in U(t)} H(t,p(t{+}1),\bar{x}(t),u),\ t\in T \quad (4)$$

Furthermore, if $f(t,\cdot,\bar{u}(t))$ *is* strictly *differentiable at* $\bar{x}(t)$ *for all* $t \in T$ *(rather than merely differentiable), then the assertion of the theorem is valid in terms of the limiting subdifferential with* (2) *replaced by*

$$z_j^* = (z_{a+1j}^*,\dots,z_{bj}^*) \in \partial^L(\lambda_j\varphi_j)(\bar{x}(a+1),\dots,\bar{x}(b)),\quad j=0,\dots,r_1+r_2 \quad (5)$$

In the special case when φ_j, $j = 0,\dots,r_1+r_2$ are functions of $x(b)$ only, assertions (b) and (c) of the previous theorem take a form analogous to the conditions of the Nonsmooth Pontryagin Maximum Principle in continuous time (e.g., [10], Theorem 6.2.3):

(b') $z_j^* \in \partial(\lambda_j\varphi_j)(\bar{x}(b))$, $j = 0,\dots,r_1+r_2$, (here '∂' stands for either the limiting or Michel-Penot subdifferential)

(c') $p(\cdot)$ satisfies the adjoint equation

$$p(t) = H_x(t,p(t+1),\bar{x}(t),\bar{u}(t)),\quad t\in T$$
$$p(b) = -\sum_{j=0}^{r_1+r_2} z_j^*$$

Remark 1. *The property* $\partial(\lambda\varphi)(x) = \lambda\partial\varphi(x)$ *is satisfied by Michel-Penot subdifferential for any* λ *and by the limiting subdifferential for* $\lambda \ge 0$*. Therefore, conditions* (2) *and* (5) *can be restated accordingly.*

Consider an optimization problem

$$\begin{cases}\text{Minimize } f_0(y)\\ \text{subject to constraints}\\ f_j(y)\le 0,\ j=1,\dots,r_1\\ f_j(y)=0,\ j=r_1+1,\dots,r_1+r_2\\ y\in\Omega.\end{cases}$$

Here $f_i : I\!R^n \to I\!R$, $i = 0,\dots,r_1+r_2$. The following Lagrange Multiplier Rule holds in terms of Michel-Penot subdifferential ([4]) and in terms of the limiting subdifferential ([7], Theorem 5.21):

Lagrange Multiplier Rule. *Let* $\bar{y}$ *be a solution to* (I)*,* Ω *be a closed convex set and the functions* f_i*,* $i = 0,\dots,r_1+r_2$ *be Lipschitz around* $\bar{y}$*. Then there exist numbers* λ_j*,* $j = 0,\dots,r_1+r_2$*, not all zero, such that, for* $j = 0,\dots,r_1$*,* $\lambda_j \ge 0$ *and* $\lambda_j f_j(\bar{y}) = 0$*, and the following inclusion holds:*

$$0 \in \sum_{j=0}^{r_1+r_2} \partial(\lambda_j f_j)(\bar{y}) + N(\bar{y},\Omega), \quad (6)$$

where $N(\bar{y},\Omega)$ *denotes the normal cone of convex analysis.*

Remark 2. *Due to the property in Remark 1, inclusion* (6) *can be written as*

$$0 \in \sum_{j=0}^{r_1+r_2} \lambda_j\partial^M f_j(\bar{y}) + N(\bar{y},\Omega)$$
$$0 \in \sum_{j=0}^{r_1} \lambda_j\partial^L f_j(\bar{y}) + \sum_{j=r_1+1}^{r_1+r_2} \partial^L(\lambda_j f_j)(\bar{y}) + N(\bar{y},\Omega)$$

The following Chain Rule holds for both Michel-Penot and the limiting subdifferentials (([9], Proposition 2.7) and ([7], Corollary 3.43), respectively):

Chain Rule. *Let* $g : I\!R^n \to I\!R^k$ *be differentiable at* $\bar{y}$ *and* $f : I\!R^k \to I\!R$ *be Lipschitz around* $\bar{z} := g(\bar{y})$*. Then*

$$\partial^M(f\circ g)(\bar{y}) \subset \bigcup_{z^*\in\partial^M f(\bar{z})} \langle z^*, \nabla g(\bar{y})\rangle$$

If, moreover, g *is strictly differentiable at* $\bar{y}$*, then*

$$\partial^L(f\circ g)(\bar{y}) \subset \bigcup_{z^*\in\partial^L f(\bar{z})} \langle z^*, \nabla g(\bar{y})\rangle$$

The idea of the proof of Theorem 1 is as follows: the optimal control problem (1) is reduced to an optimization problem (6), to which the Lagrange Multiplier Rule, in conjunction with the Chain Rule, is applied. This methodology is used in [9] for a continuous-time problem, but the proof in [9] is significantly more difficult, since it involves limiting procedure based on measure theory. However, despite relative simplicity, the present paper still produces a new result. We also remark that this methodology is general and can be applied, for example, to the Clarke's generalized gradient ([1]).

In the following section we present an outline of the proof of Theorem 1. Full proof will be presented in a paper currently under review.

II. Outline of the Proof of Theorem 1

Take $\theta \in T$, $v \in U(\theta)$ and consider the variation of the optimal control

$$u(t) = \begin{cases}\bar{u}(t), & t\ne\theta\\ v, & t=\theta\end{cases} \quad (7)$$

Denote

$$y = f(\theta,\bar{x}(\theta),v),\quad \bar{y} = f(\theta,\bar{x}(\theta),\bar{u}(\theta)) \quad (8)$$

and let $x^y(\cdot)$ be the trajectory of (1) corresponding to control (7). The fact that $v = \bar{u}(\theta)$ minimizes $\varphi_0(x^y(a+1),\dots,x^y(b))$ on the trajectories of (1) can be expressed as: $y = \bar{y}$ is a solution of the problem

$$\begin{cases}\text{Minimize } \varphi_0(x^y(a+1),\dots,x^y(b))\\ \varphi_j(x^y(a+1),\dots,x^y(b))\le 0,\ j=0,\dots,r_1\\ \varphi_j(x^y(a+1),\dots,x^y(b))=0,\ j=r_1+1,\dots,r_1+r_2\\ \text{over } y\in f(\theta,\bar{x}(\theta),U(\theta))\end{cases}$$

Due to Lagrange Multiplier Rule (6) there exist numbers $\lambda_j,\, j = 0, \ldots, r_1 + r_2$ as in the assertion of Theorem 1 such that

$$0 \in \sum_{j=0}^{r_1+r_2} \partial_y(\lambda_j \varphi_j)(x^{\bar{y}}(a+1), \ldots, x^{\bar{y}}(b)) + N(\bar{y}, f(\theta, \bar{x}(\theta), U(\theta))), \tag{9}$$

where $x^{\bar{y}}(\cdot) \equiv \bar{x}(\cdot)$. (Here and below '$\partial$' stands for either subdifferential).

It can be shown that the map $y \mapsto x^y(t),\ t \in T$ is differentiable at $y = \bar{y}$ with respect to the set $f(t, \bar{x}(t), U(t))$, that is, there exists an $n \times m$ matrix $D_y x^{\bar{y}}(t)$ such that

$$x^y(t) - \bar{x}(t) = D_y x^{\bar{y}}(t)(y - \bar{y}) + o(|y - \bar{y}|) \quad \text{for } y \in f(t, \bar{x}(t), U(t))$$

and, moreover,

$$D_y x^{\bar{y}}(t) = \begin{cases} 0, & t \le \theta \\ \Phi(\theta + 1, t), & t \ge \theta + 1 \end{cases} \tag{10}$$

where Φ is the transition matrix of the system

$$r^T(\tau) = r^T(\tau + 1) f_x(\tau, \bar{x}(\tau), \bar{u}(\tau)),\ \tau \in T\,.$$

Due to (9), there exist $\xi^* \in N(\bar{y}, f(\theta, \bar{x}(\theta), U(\theta)))$ and $y_j^* \in \partial_y(\lambda_j \varphi_j)(x^{\bar{y}}(a+1), \ldots, x^{\bar{y}}(b))$, $j = 0, \ldots, r_1 + r_2$, such that

$$\sum_{j=0}^{r_1+r_2} y_j^* + \xi^* = 0$$

Therefore, for y defined in (8), we have

$$\sum_{j=0}^{r_1+r_2} \langle y_j^*, y - \bar{y} \rangle \ge 0 \tag{11}$$

Due to the Chain Rule and (10), for each j there exists $z_j^* = (z_{a+1j}^*, \ldots, z_{bj}^*) \in \partial(\lambda_j \varphi_j)(\bar{x}(a+1), \ldots, \bar{x}(b))$ (here '∂' is the subdifferential with respect to x) such that

$$\begin{aligned} \langle y_j^*, y - \bar{y} \rangle &= \langle z_j^*, \left(D_y x^{\bar{y}}(a+1), \ldots, D_y x^{\bar{y}}(b)\right)(y - \bar{y}) \rangle \\ &= \sum_{s=\theta+1}^{b} \langle z_{sj}^*, \Phi(\theta + 1, s)(y - \bar{y}) \rangle \end{aligned} \tag{12}$$

Define

$$p^T(t) = -\sum_{j=0}^{r_1+r_2} \sum_{s=t}^{b} z_{sj}^* \Phi(t, s) \tag{13}$$

It follows from (11) and (12) that

$$p^T(t+1)(y - \bar{y}) \equiv p^T(t+1)(f(\theta, \bar{x}(\theta), v) - f(\theta, \bar{x}(\theta), \bar{u}(\theta)) \le 0,$$

which is the assertion of the maximum principle (4) and it is straightforward to show that (13) is equivalent to the adjoint equation (3).

References

[1] F. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, 1983.

[2] F. Clarke, M.R. de Pinho, *Optimal Control Problems with Mixed Constraints*, SIAM J. Control and Optim., 48(7), 2010, pp. 4500-4524.

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

[4] A. D. Ioffe, *A Lagrange Multiplier Rule with Small Convex Valued Subdifferentials for Nonsmooth Problems of Mathematical Programming involving Equality and Nonfunctional Constraints*, Math. Programming, 58, 1993, pp. 137-145.

[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 and Optim., 33, 1995, pp. 882-915.

[7] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation* Springer, 2005.

[8] B.S. Mordukhovich, I. Shvartsman, *Discrete Maximum Principle for Nonsmooth Optimal Control Problems with Delays*, Cybernetics and System Analysis, No.2, pp. 245-254, 2002.

[9] I. Shvartsman, *New Approximation Method in the Proof of the Maximum Principle for Nonsmooth Optimal Control Problems with State Constraints*, J. of Mathematical Analysis and Applications, Vol. 326(2), pp. 974-1000, 2007.

[10] R. B. Vinter. *Optimal Control*. Birkhäuser, Boston, 2000.