



**HAL**  
open science

# Tutorial on Nonsmooth Analysis and Optimal Control

Richard Vinter

► **To cite this version:**

Richard Vinter. Tutorial on Nonsmooth Analysis and Optimal Control. SADCO Summer School 2011 - Optimal Control, Sep 2011, London, United Kingdom. inria-00629428

**HAL Id: inria-00629428**

**<https://inria.hal.science/inria-00629428>**

Submitted on 5 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Nonsmooth Analysis and Optimal Control

Richard B. Vinter

Imperial College London

SADCO Summer School,

**Imperial College London, Sept. 5 – 9 2011**

# Outline of course

- Motivating examples
- Review of analytic techniques
- nonlinear programming
- Optimal control (Maximum Principle)

**The handouts provide a more detailed coverage of the material of this course.**

'How to control a dynamical system to optimize a cost. e.g. energy consumption, time, distance to target'

'System' can have many different interpretation

- mechanism
- social behaviour
- economic phenomena

**Early applications:** aerospace (selection of optimal flight paths)

**Modern-day applications:** process control, resource economics, robotics, engine management . .

# Example 1: Godard Problem

‘A rocket travels vertically. What should be the time-varying force (from burning fuel) to maximize final height?’

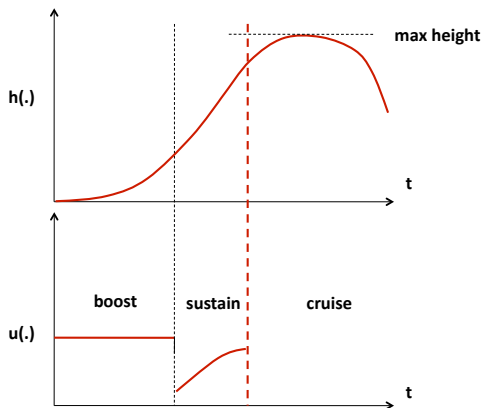
$h$  height  
 $v$  velocity  
 $m$  mass

$$\begin{aligned}\dot{v}(t) &= \frac{1}{m} [F(t) - D(v, h)] \\ \dot{h}(t) &= v(t) \\ \dot{m}(t) &= \frac{-F(t)}{c}\end{aligned}$$

where  $g$  is a constant, and drag model is  $D(v, h) = \frac{1}{2}\rho v^2 C_d e^{-\beta h}$ .

**Optimal Control Problem:** Find  $0 \leq F(t) \leq F_{\max}$  to maximize  $h(t_f)$  with  $v(0) = h(0) = 0$  and  $m(0), m(t_f)$  given.

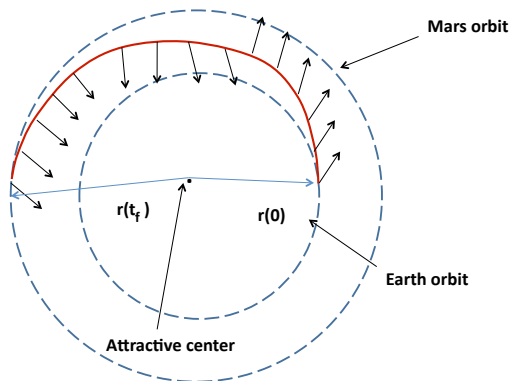
# Godard Problem, cont.



Godard Rocket Problem

# Orbit Transfer Problem

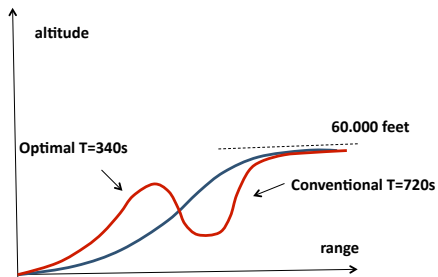
orbit.pdf



Minimum Time Path Orbit (Earth to Mars) – Thrust Direction every 10 days

# Example 3: Minimum Time Ascent of an F-4 aircraft

aircraft.pdf



Optimal climb manoeuvre for supersonic F4 (Phantom) aircraft



# Further Examples

- **Sustainable harvesting:** 'Optimal exploitation of fish resources under sustainability constraint'
- **Process control:** 'Optimize purity of distillate from reflux ratio control'
- **Sports Science:** 'Golf swing - maximize impact velocity of club on ball'

# Nonsmooth Analysis

**Nonsmooth Analysis:** local approximation of non-differentiable functions and of sets with non-differentiable boundaries.

Relevant to: optimization, dynamical systems, control . . .

**Key question:**

How should classical concepts of 'gradients' be adapted, to give provide useful local information about non-differentiable functions?

# Origins

Take a closed set  $C \in R^n$ , a function  $f : R^n \rightarrow R$  and  $\bar{x} \in R^n$ .

Assume

- boundary  $\{C\}$  is an  $n - 1$  dimensional  $C^1$  manifold
- $f(\cdot)$  is continuously differentiable

The normal vector  $\eta$  to boundary  $\{C\}$  at  $\bar{x}$  is the (unit) normal to the tangent space of the manifold at  $\bar{x}$ , oriented to 'point out of  $C$ '.

If  $C = \{x \in R^n \mid g(x) \leq 0\}$ , then  $\eta = \nabla g(\bar{x})$ .

(normal vector provides a dual description of tangent space approximation to the boundary of  $C$  near  $\bar{x}$ ).

The gradient  $\nabla f(\bar{x})$  provides a linear approximation to  $f(\cdot)$  near  $\bar{x}$ .

$$\nabla f(\bar{x})(x - \bar{x}) \simeq f(x) - f(\bar{x})$$

# Origins, continued

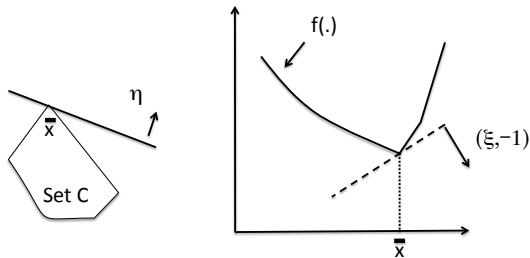
Suppose  $C$  and  $f(\cdot)$  are **smooth AND convex**.

$\eta$  and  $\xi = \nabla f(\bar{x})$  can be equivalently defined to satisfy the properties

$$\eta \cdot (x - \bar{x}) \leq 0 \text{ and } \xi \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}).$$

Now assume  $C$  and  $f(\cdot)$  are **merely convex**.

**set of normal vectors**  $N_C(\bar{x}) := \{\nu \mid \nu \cdot (x - \bar{x}) \leq 0, \forall x \in C\}$   
**set of subgradients**  $\partial f(\bar{x}) := \{\xi \mid \xi \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}), \forall x \in \mathbb{R}^n\}.$



# Proximal Normal Cone

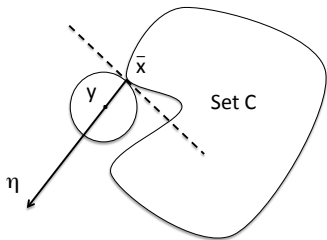
First step: define the **outward normal vector** to closed set in  $R^k$ .

**The Proximal Normal Cone.** Take a closed set  $C \subset R^k$  and a point  $\bar{x} \in C$ . A vector  $\eta \in R^k$  is said to be a **proximal normal vector** to  $C$  at  $\bar{x}$  if there exists  $M \geq 0$  such that

$$\eta \cdot (x - \bar{x}) \leq M|x - \bar{x}|^2 \quad \text{for all } x \in C. \quad (1)$$

The cone of all proximal normal vectors to  $C$  at  $\bar{x}$  is called the **proximal normal cone** to  $C$  at  $\bar{x}$  and is denoted by  $N_C^P(\bar{x})$ :

$$N_C^P(\bar{x}) := \{ \eta \in R^k : \exists M \geq 0 \text{ s.t. (1) is satisfied: } \}.$$



Defining property of proximal normal vectors  $\eta$ :

$$\eta \cdot (x - \bar{x}) \leq M|x - \bar{x}|^2 \text{ for all } x \in C.$$

Equivalently:

$$\exists y \text{ and } \alpha \geq 0 \text{ s.t. } \bar{x} = \text{Proj}_C(y) \text{ and } \eta = \alpha(y - \bar{x})$$

# Normal Vectors

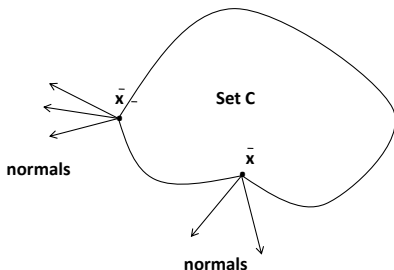
Take a closed set  $C \subset R^k$  and a point  $\bar{x} \in C$ . A vector  $\eta$  is said to be a **normal vector** to  $C$  at  $\bar{x} \in C$  if there exist sequences  $x_i \xrightarrow{C} \bar{x}$  and  $\eta_i \rightarrow \eta$  such that

$$\eta_i \in N_C^P(x_i) \text{ for all } i.$$

The set of normal vectors is called the **normal cone** to  $C$  at  $\bar{x}$  and is written  $N_C(\bar{x})$ :

$$N_C(\bar{x}) := \{\eta \in R^k : \exists x_i \xrightarrow{C} \bar{x} \text{ and } \eta_i \rightarrow \eta \text{ such that } \eta_i \in N_C^P(x_i) \text{ for all } i\}.$$

(The notation  $x_i \xrightarrow{C} x$  indicates that  $x_i \rightarrow x$  and  $x_i \in C$  for all  $i$ .)



Normal vectors at different base points  $\bar{x}$ .

- $N_C(\bar{x})$  is a **closed cone**.
- $N_C(\bar{x})$  **may not be convex** (see diagram)
- $N_C(\bar{x})$  **contains non-zero points**, if  $\bar{x}$  is a boundary point.



# Proximal Subgradients

Take an extended valued, lower semicontinuous function  $f : \mathbb{R}^k \rightarrow \mathbb{R} \cup \{+\infty\}$  and a point  $\bar{x} \in \text{dom}\{f\}$ .

A vector  $\eta \in \mathbb{R}^k$  is said to be a **proximal subgradient** of  $f$  at  $\bar{x}$  if there exists  $\epsilon > 0$  and  $M \geq 0$  such that

$$\eta \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}) + M|x - \bar{x}|^2$$

for all points  $x$  which satisfy  $|x - \bar{x}| \leq \epsilon$ .

(The notation  $\text{dom}\{f\}$  denotes the set  $\{y : f(y) < +\infty\}$ .)

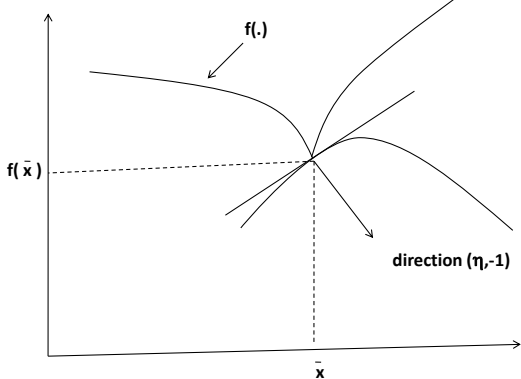
# Proximal Subdifferentials (continued)

The set of all proximal subgradients of  $f$  at  $\bar{x}$  is called the *proximal subdifferential* of  $f$  at  $\bar{x}$  and is denoted by  $\partial^P f(\bar{x})$ :

$$\partial^P f(\bar{x}) := \{ \eta \in \mathbb{R}^k \mid \exists \epsilon > 0 \text{ and } M \geq 0 \text{ such that (2) is satisfied} \}.$$

$$(2) \quad \eta \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}) + M|x - \bar{x}|^2$$

for all points  $x$  which satisfy  $|x - \bar{x}| \leq \epsilon$ .



**Geometric interpretation:** a proximal subgradient to  $f$  at  $\bar{x}$  is the slope at  $x = \bar{x}$  of a paraboloid,

$$y = \eta \cdot (x - \bar{x}) + f(\bar{x}) - M|x - \bar{x}|^2,$$

which coincides with  $f$  at  $x = \bar{x}$  and which lies below the graph of  $f$  on a neighbourhood of  $\bar{x}$ .

# Subgradients

## The Subdifferential.

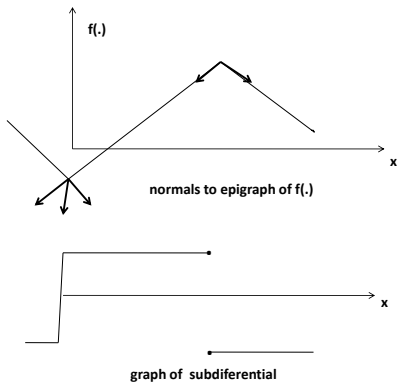
Take an extended valued, lower semicontinuous function  $f : \mathbb{R}^k \rightarrow \mathbb{R} \cup \{+\infty\}$  and a point  $\bar{x} \in \text{dom}\{f\}$ . A vector  $\eta \in \mathbb{R}^k$  is said to be a **subgradient** of  $f$  at  $\bar{x}$  if there exist sequences  $x_i \xrightarrow{f} \bar{x}$  and  $\eta_i \rightarrow \eta$  such that

$$\eta \in \partial^P f(x_i) \text{ for all } i.$$

The set of all subgradients of  $f$  at  $\bar{x}$  is called the **subdifferential** of  $f$  at  $\bar{x}$  and is denoted by  $\partial f(\bar{x})$ .

$$\partial f(\bar{x}) := \{\eta : \exists x_i \xrightarrow{f} \bar{x} \text{ and } \eta_i \rightarrow \eta \text{ such that } \eta_i \in \partial^P f(x_i) \text{ for all } i\}.$$

(The notation  $x_i \xrightarrow{f} \bar{x}$  means  $x_i \rightarrow \bar{x}$  and  $f(x_i) \rightarrow f(\bar{x})$  as  $i \rightarrow \infty$ .)



**Figure:** The Graph of the Subdifferential

# Properties

Properties of the subdifferential  $\partial f(\bar{x})$ .

- $\partial f(\bar{x})$  is a **closed (but not always convex)** set.
- For  $f(\cdot)$  convex,  $\partial f(\bar{x})$  is the subdifferential of convex analysis.
- It is possible that

$$\partial f(\bar{x}) \neq -\partial(-f)(\bar{x}).$$

(There are, in fact, a number of equivalent ways of defining subgradients, limits of strict subgradients, by means of normals to epigraph sets, etc.)

# Properties of Normal Cones and Sub-Differentials

## (A): (Upper Semicontinuity)

- (i) Take a closed set  $C \subset R^k$  and a point  $x \in C$ . Then for any convergent sequences  $x_i \xrightarrow{C} x$  and  $\eta_i \rightarrow \eta$  such that

$$\eta_i \in N_C(x_i) \quad \text{for all } i,$$

we have  $\eta \in N_C(x)$ .

- (ii) Take a lower semicontinuous function  $f : R^k \rightarrow R \cup \{+\infty\}$  and a point  $x \in \text{dom}\{f\}$ . Then for any convergent sequences  $x_i \xrightarrow{f} x$  and  $\xi_i \rightarrow \xi$  such that

$$\xi_i \in \partial f(x_i) \quad \text{for all } i,$$

we have  $\xi \in \partial f(x)$ .

# Further Properties

## (B): (Links between Normal Cones and Subdifferentials)

Take a closed set  $C \subset \mathbb{R}^k$  and a point  $\bar{x} \in C$ . Then

$$N_C(\bar{x}) = \partial \Psi_C(\bar{x}).$$

and

$$N_C(\bar{x}) \cap B = \partial d_C(\bar{x}).$$

Here,  $\Psi_C : \mathbb{R}^k \rightarrow \mathbb{R} \cup \{+\infty\}$  is the indicator function of the set  $C$

$$\Psi_C(x) := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

and  $d_C : \mathbb{R}^k \rightarrow \mathbb{R}$  is the distance function of the set  $C$

$$d_C(x) := \inf_{y \in C} |x - y|.$$



# Limits of Derivatives for Lipschitz Functions

C:(Limits of Derivatives).

Take a Lipschitz continuous function  $f : R^k \rightarrow R$ , a point  $\bar{x} \in R^k$  and any set  $S$  of  $R^k$  having zero Lebesgue measure. Then

$$\text{co } \partial f(\bar{x}) = \text{co } \{ \eta : \exists x_i \rightarrow \bar{x} \text{ such that } \nabla f(x_i) \text{ exists and } x_i \notin S \text{ for all } i \text{ and } \nabla f(x_i) \rightarrow \eta \}.$$

**A Lipschitz continuous function  $f : R^k \rightarrow R$  on  $R^k$  is differentiable almost everywhere with respect to  $k$ -dimensional Lebesgue measure. (Rademacher's Theorem).**

So the right side always defines a non-empty set.

# Properties of Lipschitz Functions

Take a lower semicontinuous function  $f(\cdot) : R^k \rightarrow R$  and a point  $\bar{x} \in R^k$ .

We deduce from Property (C):

$$\text{'}f(\cdot) \text{ is Lipschitz continuous' } \implies \text{co } \partial f(\bar{x}) = -\text{co } \partial(-f)(\bar{x}).$$

Note also the useful properties:

- $'|\xi| \leq K \text{ on } \bar{x} + \epsilon B' \implies 'f(\cdot) \text{ is } K\text{-Lipschitz on } \bar{x} + \epsilon B'$
- $'f(\cdot) \text{ is } C^1 \text{ on } \bar{x} + \epsilon B' \implies '\partial f(\bar{x}) = \{\nabla f(\bar{x})\}'.$

**A fundamental property of the gradient of a differentiable function is that it vanishes at a local minimizer of the function. The proximal subdifferential has a similar role in identifying possible minimizers:**

**(F): (Subdifferentials at Minima)**

**Conditions on Local Minimizers**

Assume that, for some  $\epsilon > 0$ , the function  $f : R^k \rightarrow R \cup \{+\infty\}$  satisfies

$$f(\bar{x}) \leq f(x) \text{ for all } x \in \bar{x} + \epsilon B.$$

for some  $\bar{x} \in R^k$ . Then

$$0 \in \partial^P f(\bar{x}).$$

It follows that  $0 \in \partial f(\bar{x})$ .

# Calculus of Differentials

A calculus is available to estimate subdifferentials of composite functions. Some the most important are as follows:

**(G): (Positive Homogeneity).** Take a lower semicontinuous function  $f : R^k \rightarrow R \cup \{+\infty\}$ , a point  $\bar{x} \in \text{dom } f$  and  $\alpha \geq 0$ . Then

$$\partial(\alpha f)(\bar{x}) = \alpha \partial f(\bar{x}).$$

**(H): (Sum Rule).** Consider a collection  $f_i : R^k \rightarrow R \cup \{+\infty\}$ ,  $i = 1, \dots, n$  of lower semicontinuous extended valued functions and a point  $\bar{x} \in \bigcap_{i=1, \dots, n} \text{dom } f_i$ . Assume that all the  $f_i$ 's except possibly one are Lipschitz continuous on a neighbourhood of  $\bar{x}$ .

**Then** the subdifferential of  $(f_1 + \dots + f_n)(x) = f_1(x) + \dots + f_n(x)$  satisfies

$$\partial(f_1 + \dots + f_n)(\bar{x}) \subset \partial f_1(\bar{x}) + \dots + \partial f_n(\bar{x}).$$

## (I): (Max Rule)

Consider a collection  $f_i : R^k \rightarrow R \cup \{+\infty\}$ ,  $i = 1, \dots, n$  of lower semicontinuous extended valued functions and a point  $\bar{x} \in \bigcap_{i=1, \dots, n} \text{dom } f_i$ . Assume that all the  $f_i$ 's, except possibly one, are Lipschitz continuous on a neighbourhood of  $\bar{x}$ .

Then the subdifferential of

$$(\max\{f_1, \dots, f_n\})(x) = \max\{f_1(x), \dots, f_n(x)\}$$

satisfies

$$\partial(\max_{i=1, \dots, n} f_i)(\bar{x}) \subset \left\{ \sum_{i \in J} \alpha_i \partial f_i(\bar{x}) : \alpha_i \geq 0 \text{ for all } i \in J \text{ and } \sum_{i \in J} \alpha_i = 1 \right\}$$

in which

$$J := \{i \in \{1, \dots, n\} : f_i(x) = \max_j f_j(\bar{x})\}.$$

## (J): (Chain Rule)

Take Lipschitz continuous functions  $G : R^k \rightarrow R^m$  and  $g : R^m \rightarrow R$  and a point  $\bar{x} \in R^k$ . Then the subgradient  $\partial(g \circ G)(\bar{x})$  of the composite function  $x \rightarrow g(G(x))$  at  $\bar{x}$  satisfies:

$$\partial(g \circ G)(\bar{x}) \subset \{\eta : \eta \in \partial(\gamma \cdot G)(\bar{x}) \text{ for some } \gamma \in \partial g(G(\bar{x}))\}.$$

Many of these rules are extensions of familiar principles of differential calculus. (But equality has been replaced by set inclusion).

An exception is the Max Rule above. This does not have a classical analogue.

(Note: pointwise infimum of smooth functions creates corners . . .)

## Ekeland's Theorem.

Take a complete metric space  $(X, d(\cdot, \cdot))$ , a lower semicontinuous function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , a point  $x_0 \in \text{dom } f$  and a number  $\epsilon > 0$  such that

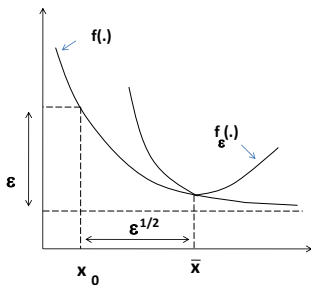
$$f(x_0) \leq \inf_{x \in X} f(x) + \epsilon.$$

Then there exists  $\bar{x} \in X$  such that

$$d(\bar{x}, x_0) \leq \epsilon^{\frac{1}{2}}$$

and

$$f(\bar{x}) (= f(\bar{x}) + \epsilon^{\frac{1}{2}} d(x, \bar{x})|_{x=\bar{x}}) = \inf_{x \in X} \{f(x) + \epsilon^{\frac{1}{2}} d(x, \bar{x})\}.$$



**Figure:** Illustration of Ekeland's Theorem



## Borwein-Preiss Variational Principle

Take a complete Hilbert space  $X$ , a lower semicontinuous function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , a point  $x_0 \in \text{dom } f$  and a number  $\epsilon > 0$  such that

$$f(x_0) < \inf_{x \in X} f(x) + \epsilon.$$

Then for any  $\lambda > 0$ , there exist  $y, x \in X$  such that

$$\|z - x_0\| \leq \lambda, \|y - x_0\| \leq \lambda, f(y) \leq f(x_0), \text{ and}$$

$$x \rightarrow f(x) + \frac{\epsilon}{\lambda^2} \|x - z\|^2$$

has a unique minimum at  $x = y$ .

# Inf Convolution

Assume  $g(\cdot) : R^k \rightarrow R$  is Lipschitz continuous with Lipschitz constant  $K$ . For each  $\epsilon > 0$ , take  $g_\epsilon(\cdot)$  the function

$$g_\epsilon(x) = \min_y \{g(y) + \epsilon^{-1}|y - x|^2\} \quad \text{for all } x \in R^k \quad (2)$$

Take any  $x \in R^k$  and  $\epsilon > 0$ . Let  $\bar{y} \in R^k$  be a vector achieving the infimum in (2) (one such vector exists), and set

$$\eta_\epsilon(x) := 2\epsilon^{-1}(x - \bar{y}).$$

Then, for each  $x \in R^k$ , we have

- (i):  $g_\epsilon(\cdot)$  is Lipschitz continuous with Lipschitz constant  $K$ .
- (ii):  $g(x) \geq g_\epsilon(x) \geq g(x) - K\epsilon$ .
- (iii):  $g_\epsilon(x') - g_\epsilon(x) \leq \eta \cdot (x' - x) + \epsilon^{-1}|x' - x|^2$  for all  $x' \in R^k$
- (iv):  $\eta \in \partial^P g(\bar{y})$
- (v):  $|x - \bar{y}| \leq K\epsilon$

We illustrate the use of the calculus rules by proving a multiplier rule for a mathematical programming problem with functional inequality constraints:

$$\left\{ \begin{array}{l} \text{Minimize } f(x) \\ \text{over } x \in R^k \text{ satisfying} \\ g_i(x) \leq 0, \text{ for } i = 1, \dots, n \end{array} \right. \quad (3)$$

Here  $f : R^k \rightarrow R$  and  $g_i : R^k \rightarrow R, i = 1, \dots, n$ , are given functions.

## Multiplier Rule I (Inequality Constraints)

Let  $\bar{x}$  be a minimizer for problem (3). Assume that  $f, g_i, i = 1, \dots, n$ , are Lipschitz continuous.

Then there exist non-negative numbers  $\lambda_0, \lambda_1, \dots, \lambda_n$  such that

$$\sum_{i=0}^n \lambda_i = 1,$$

$$\lambda_i = 0 \text{ if } g_i(\bar{x}) < 0 \text{ for } i = 1, \dots, n$$

and

$$0 \in \lambda_0 \partial(\bar{x}) + \sum_{i=1}^n \lambda_i \partial g_i(\bar{x}). \quad (4)$$

**Proof.** Consider the function

$$\phi(x) = \max\{f(x) - f(\bar{x}), g_1(x), \dots, g_n(x)\}.$$

We claim that  $\bar{x}$  minimizes  $\phi$  over  $x \in R^k$ . If this were not the case, it would be possible to find  $x'$  such that

$$\phi(x') < \phi(\bar{x}) = 0.$$

It would then follow that

$$f(x') - f(\bar{x}) < 0, g_1(x') < 0, \dots, g_n(x') < 0,$$

in contradiction of the minimality of  $\bar{x}$ . The claim has been confirmed.

According to Rule (F)

$$0 \in \partial\phi(\bar{x}).$$

We deduce from Rule (I) that there exist non-negative numbers  $\lambda_0, \dots, \lambda_n$  such that

$$\sum_{i=0}^n \lambda_i = 1,$$

and

$$0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i=1}^n \lambda_i \partial g_i(\bar{x}).$$

Notice that, if  $g_i(\bar{x}) < 0$ , then

$$g_i(\bar{x}) < \phi(\bar{x}).$$

So Rule (I) supplies the supplementary information that

$$\lambda_i = 0 \quad \text{if} \quad g_i(\bar{x}) < 0.$$

The multiplier rule is proved.

Ekeland's Theorem and the preceding nonsmooth calculus rules can be used, for example, to derive a multiplier rule for a nonlinear program problem with a functional equality constraint:

$$\left\{ \begin{array}{l} \text{Minimize } f(x) \\ \text{over } x \in R^k \text{ satisfying} \\ G(x) = 0. \end{array} \right.$$

The data for this problem comprise functions  $f : R^k \rightarrow R$  and  $G : R^k \rightarrow R^m$ .

## Theorem

### A Multiplier Rule (Equality Constraints)

Let  $x_0$  be a minimizer for the nonlinear program with an equality constraint. Assume that  $f$  and  $G$  are Lipschitz continuous.

Then there exist number  $\lambda_0 \geq 0$  and a vector  $\lambda_1 \in R^m$  such that

$$\lambda_0 + |\lambda_1| = 1$$

and

$$0 \in \lambda_0 \partial(x_0) + \partial(\lambda_1 \cdot G)(x_0). \quad (5)$$



Is there a very simple approach?

Try to use the fact that  $x_0$  is an unconstrained minimizer of the function

$$\gamma(x) := \max\{f(x) - f(x_0), |G(x)|\}.$$

By Rule (F),  $0 \in \partial\gamma(x_0)$ . We deduce from the Max Rule (I) that there exist multipliers  $\lambda_0 \geq 0$  and  $\nu_1 \geq 0$  such that  $\lambda_0 + \nu_1 = 1$  and

$$0 \in \lambda_0 \partial f(x_0) + \nu_1 \partial |G(x)| |_{x=x_0}.$$

Trivially satisfied with  $\lambda_0 = 0$ ,  $\lambda_1 = 1$ , because  $0 \in \partial |G(x)| |_{x=x_0}$ .

Gives no useful information!

Instead **perturb**  $\phi(\cdot)$ :

Take a sequence  $\epsilon_i \downarrow 0$ . For each  $i$  define the function

$$\phi_i(x) := \max\{f(x) - f(x_0) + \epsilon_i, |G(x)|\}.$$

Note that  $\phi_i(x_0) = \epsilon_i$  and  $\phi_i(x) \geq 0$  for all  $x$ . It follows that

$$\phi_i(x_0) \leq \inf_x \phi(x) + \epsilon_i.$$

Apply Ekeland's Thm. to  $\phi_i(\cdot)$ , with Euclidean metric:

there exists  $x_i \in R^k$  such that

$$|x_i - x_0| \leq \epsilon_i^{\frac{1}{2}}$$

and

$$\tilde{\phi}_i(x_i) \leq \inf_{x \in R^k} \tilde{\phi}_i(x),$$

where

$$\tilde{\phi}_i(x) := \max\{f(x) - f(x_0) + \epsilon_i, |G(x)|\} + \epsilon_i^{\frac{1}{2}}|x - x_i|.$$

By non-smooth calculus rules,

$$0 \in \partial \tilde{\phi}_i(x_i) + \epsilon_i^{\frac{1}{2}} B.$$

**Important observation:** for each  $i$ ,

$$|G(x_i)| = 0 \quad \text{implies} \quad f(x) - f(x_0) + \epsilon > |G(x_i)|.$$

**Apply Max Rule to  $\tilde{\phi}_i$ :**

There exist non-negative numbers  $\lambda_0^i$  and  $\nu_1^i$  such that

$$\lambda_0^i + \nu_1^i = 1,$$

$$0 \in \lambda_0^i \partial f(x_i) + \nu_1^i \partial |G(x_i)| + \epsilon_i^{\frac{1}{2}} B..$$

and

$$|G(x_i)| = 0 \quad \text{implies} \quad \nu_1^i = 0.$$

For simplicity assume:  $G(\cdot)$  is scalar valued. Then,

$$G(x_j) \neq 0 \text{ implies } \nu_j \partial |G(x_j)| = \partial(\lambda_1^i G(x_j))$$

where

$$\lambda_1^i = \text{sign}\{G(x_j)\} \nu_j = (+/-) \nu_j .$$

Hence

$$\lambda_0^i + |\lambda_1^i| = 1,$$

$$0 \in \lambda_0^i \partial f(x_j) + \partial(\lambda_1^i G(x_j)) + \epsilon_j^{\frac{1}{2}} B..$$

Extract subsequences so that

$$(\lambda_0^i, \lambda_1^i) \rightarrow (\lambda_0, \lambda_1)$$

Pass to the limit:

$$\lambda_0 + |\lambda_1| = 1 \text{ and } 0 \in \lambda_0 \partial f(\bar{x}) + \partial(\lambda_1 G(\bar{x})) .$$

# Application of Inf Convolution

Suppose that  $x_0 \in R^n$  solves

$$\begin{cases} \text{Minimize } f(x) \\ \text{over } x \in R^k \text{ satisfying} \\ G(x) = 0, \text{ for } i = 1, \dots, n \end{cases}$$

where  $f : R^n \rightarrow R$ ,  $G : R^n \rightarrow R^k$  and

$f(\cdot)$  is Lipschitz,  $G(\cdot)$  is continuously differentiable.

Idea: Use 'smooth multiplier rule' to get multiplier rule with  $f(\cdot)$  Lipschitz continuous.

Take  $\epsilon_j \downarrow 0$ . Let  $f_j(\cdot)$  be inf-convolution of  $f(\cdot)$ . Then  $x_0$  is an  $\epsilon_j$  minimizer.

Apply Borwein-Preiss Thm. : there exist

$$\alpha_j \downarrow 0, \{M_j\} \subset R, y_j \rightarrow x_0, z_j \rightarrow x_0, d_j \rightarrow x_0,$$

$$\xi_j \in \partial^P f(d_j) \text{ for all } j, \xi_j \rightarrow \xi.$$

and, for each  $j$ ,  $y_j$  solves

$$\left\{ \begin{array}{l} \text{Minimize } \xi_j \cdot (x - y_j) + M_j |x - y_j|^2 + \alpha_j |x - z_j|^2 \\ \text{over } x \in R^k \text{ satisfying} \\ G(x) = 0, \text{ for } i = 1, \dots, n \end{array} \right.$$

Apply 'smooth' Multiplier Rule: there exists  $\lambda_0^i \geq 0$  and  $\lambda_1^i$  such that

$$0 = \lambda_0^i \xi + \lambda_1^i \cdot \nabla_x G(y_i) + \text{'small terms'}$$

i.e.

$$0 \in \partial^P f(d_i) + \lambda_1^i \cdot \nabla_x G(y_i) + \text{'small terms'}.$$

In limit

$$0 \in \partial^P f(x_i) + \lambda_1^i \cdot \nabla_x G(x_i).$$



# The Control System

$$(S) \begin{cases} \dot{x}(t) = f(t, x(t), u(t)), & \text{a.e. } t \in [S, T] \\ u(t) \in U & \text{a.e. } t \in [S, T] \end{cases}$$

Data:  $f : [S, T] \times R^n \times R^m \rightarrow R^n$  and  $U \subset R^m$ .

A process is pair  $(x(\cdot), u(\cdot))$  s.t.

$u(\cdot) : [S, T] \rightarrow R^m$  is measurable,  $u(t) \in U$  a.e.

$x(\cdot) : [S, T] \rightarrow R^n$  is absolutely continuous function s.t.

$$\dot{x}(t) = f(t, x(t), u(t)), \quad \text{a.e. } t \in [S, T]$$

# Hypotheses

(H1) (a):  $f(t, \cdot, x)$  is  $C^1$  for fixed  $(t, x)$

(b):  $f(\cdot, x, \cdot)$  is measurable w.r.t. the product  $\sigma$ -field  $\mathcal{L} \times \mathcal{B}$ .

$\mathcal{L}$  and  $\mathcal{B}$  are the  $\sigma$ -fields of Lebesgue subsets of  $[S, T]$   
and of Borel subsets of  $R^m$ , resp.

(b) ensures that  $t \rightarrow f(t, x, u(t))$  is  $\mathcal{L}$  measurable for each  $\mathcal{L}$  measurable function  $u(\cdot)$  and  $x \in R^n$ .

(H1) is satisfied if  $f(\cdot, x, u)$  is  $\mathcal{L}$ -measurable and  $f(t, \cdot, \cdot)$  is continuous

# Hypotheses, continued

(H2) (a): There exist  $\epsilon > 0$ ,  $k(\cdot) \in L^1$  and  $c(\cdot) \in L^1$  such that

$$|f(t, x, u) - f(t, x', u)| \leq k(t)|x - x'|$$

for all  $x, x' \in \bar{x}(t) + \epsilon B$  and  $u \in U$ , a.e.,

$$(b): \quad |f(t, x, u)| \leq c(t)$$

for all  $x \in \bar{x}(t) + \epsilon B$  and  $u \in U$ , a.e.

These hypotheses ensure state trajectories  $x(\cdot)$  are unique (for given  $u(\cdot)$  and  $x(S)$ ) + uniformly bounded.

# The optimal control problem

$$(P) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \text{ and } d(x(T)) = 0, \end{cases}$$

Data: function  $g(\cdot) : R^n \times R^n \rightarrow R$  and set  $C \subset R^n \times R^n$ .

A process  $(\bar{x}(\cdot), \bar{u}(\cdot))$  is a minimizer if  $x(S) = x_0$ ,  $d(x(T)) = 0$

and  $g(x(T)) \geq g(\bar{x}(T))$  for all processes  $(x(\cdot), u(\cdot))$  s.t

$$x(S) = x_0, d(x(T)) = 0.$$

(Accommodate cost including integral term

$$\int_S^T L(t, x(t), u(t)) dt + g(x(T))$$

by 'state augmentation' - see notes.)

# Maximum Principle

Assume (H1) and (H2),

$g(\cdot)$  and  $d(\cdot)$  are continuously differentiable.

Then there exists an absolutely continuous  $p(\cdot) : [S, T] \rightarrow R^n$ ,

$\lambda \geq 0$  and  $\mu \in R^k$

such that . . .

# Maximum Principle continued . .

$$(\lambda, \mu) \neq 0$$

(Non-trivial Lagrange Multipliers)

$$-\dot{p}(t) = p(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)), \text{ a.e.},$$

(The Costate Equation)

$$p(t) \cdot f(t, \bar{x}(t), \bar{u}(t)) = \max_{u \in U} p(t) \cdot f(t, \bar{x}(t), u) \text{ a.e.},$$

(The Generalized Weierstrass Condition)

$$-p(T) \in \lambda \nabla_x g(\bar{x}(T)) + \mu \nabla_x d(\bar{x}(T))$$

(The Transversality Condition)

# Structure of Conditions

Conditions are positively homogeneous in  $(p(\cdot), \lambda, \mu)$ , i.e. for any  $\gamma > 0$

' $(p(\cdot), \lambda, \mu)$ ' can be replaced by ' $(\gamma p(\cdot), \gamma \lambda, \gamma \mu)$ '

So (by scaling) we can take  $\lambda = 0$  or  $1$

**Normal case:  $\lambda = 1$ .**

If conditions are not normal, they make no reference to the cost (trivial information)

In normal case,  $-p(T) = \nabla_x g(\bar{x}(T)) + \mu \nabla_x d(\bar{x}(T))$

i.e. transversality conditions impose  $n - k$  constraints on boundary values of  $p(\cdot)$ .

# Structure, cont.

$(\bar{x}(\cdot), p(\cdot))$  solves Hamilton's system of equations

( $2n$  scalar differential equations)

$$d/dt \begin{bmatrix} p(t) \\ \bar{x}(t) \end{bmatrix} = \begin{bmatrix} p(t) \nabla_x f(t, \bar{x}(t), p(t), u^*(\bar{x}(t))) \\ f(t, \bar{x}(t), p(t), u^*(\bar{x}(t))) \end{bmatrix} \quad \text{a.e. ,}$$

$\bar{x}(S) = x_0, d(\bar{x}(T)) = 0$ , and

$$-p(T) = \nabla_x g(\bar{x}(T)) + \mu \nabla_x d(\bar{x}(T))$$

( $n + k$  conditions on end-points of  $x(\cdot)$  and  $n - k$  conditions on endpoints of  $p(\cdot)$ , given  $2n$  conditions on endpoint so  $(p(\cdot), x(\cdot))$ )

|

So 'two point boundary value problem' has the correct number of end-point conditions.



# Systems of First Order, Linear Ordinary Differential Equations

Consider the ODE

$$(ODE) \quad \dot{x}(t) = A(t)x(t)$$

where  $A(\cdot)$  is integrable  $n \times n$  matrix valued function.

Let  $\Phi(\cdot, \cdot) : [S, T] \times [S, T] \rightarrow R^n \times R^n$  be transition matrix of (ODE). Then

$$\begin{cases} d/dt \Phi(t, s) = \nabla_x A(t) \Phi(t, s) & \text{a.e. } t \\ \Phi(s, s) = I_{n \times n} . \end{cases}$$

# Properties of Transition Matrix

(i):  $\Phi(s, s) = I_{n \times n}$  for all  $s$

(ii):  $\Phi^{-1}(t, s) = \Phi(s, t)$  for all  $t, s$

(iii): For each  $s$ ,  $d/dt \Phi(t, s) = A(t)\Phi(t, s)$  a.e.  $s$

(iv): For each  $t$ ,  $d/ds \Phi(t, s) = -\Phi(t, s)A(s)$  a.e.  $t$

**2. Lebesgue points** Take an integrable function  $z(\cdot) : [S, T] \rightarrow \mathbb{R}^n$ .

The set of Lebesgue points  $\mathcal{T}$  of  $z(\cdot)$  is:

$$\mathcal{T} = \{t \in (S, T) \mid \lim_{\delta \downarrow 0} \delta^{-1} \int_t^{(t+\delta) \wedge T} z(s) ds = z(t)\}.$$

**Lusin's Theorem.**

$$\mathcal{L}\text{-meas } \mathcal{T} = |T - S|.$$

# Simple Case: Free Right Endpoint Problem

Special Case: Free right Endpoint Problem:

$$(P) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \end{cases}$$

Data:  $C^1$  function  $g(\cdot) : R^n \rightarrow R$  and vector  $x_0 \in R^n$ .

Absence of right endpoint constraint greatly simplifies derivation of Maximum Principle

# Maximum Principle in special case

Let  $(\bar{x}(\cdot), \bar{u}(\cdot))$  be a minimizer. Assume (H1) and (H2).

Then there exists  $p(\cdot) \in W^{1,1}$  such that

$$-\dot{p}(t) = p(t) \nabla_x f(t, \bar{x}(t), \bar{u}(t))$$

$$p(T) = -\nabla_x g(\bar{x}(T)),$$

$$p(t) \cdot [f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \geq 0 \quad \text{for all } u \in U, \text{ a.e.}$$

# Proof of Max. Prin. for Free Right Endpoint Problems

Take an arbitrary countable set  $U^0$ :

$$U^0 \subset U.$$

Let  $\mathcal{T} \subset (S, T)$  comprise the Lebesgue points of

$t \rightarrow f(t, \bar{x}(t), v) - f(t, \bar{x}(t), \bar{u}(t))$ , for all  $v \in U^0$ .

This means: for each  $t \in \mathcal{T}$  and  $u \in U^0$ ,

$$\begin{aligned} \lim_{\alpha \downarrow 0} \alpha^{-1} \int_t^{t+\alpha} (f(s, \bar{x}(s), u) - f(s, \bar{x}(s), \bar{u}(s))) ds \\ \rightarrow (f(t, \bar{x}(t), v(t)) - f(t, \bar{x}(t), \bar{u}(t))). \end{aligned}$$

The set  $\mathcal{T}$  has measure  $|T - S|$ .

# Proof (continued)

Take  $\alpha_j \downarrow 0$ . Select  $t \in \mathcal{T}$ ,  $u \in U^0$ .

For each  $j$  define the process  $(x_j(\cdot), u_j(\cdot))$ :

$$u_j(s) = \begin{cases} u & \text{for } s \in [t, (t + \alpha_j) \wedge T] \\ \bar{u}(s) & \text{otherwise} \end{cases}$$

$$x_j(S) = \bar{x}(S).$$

Define also

$$y_j(s) = \alpha_j^{-1} [x_j(s) - \bar{x}(s)] \quad \text{for all } s \in [S, T].$$

# Needle Variation

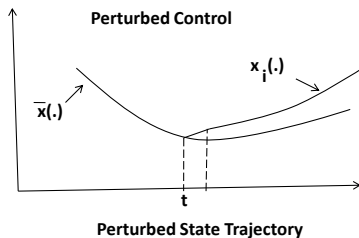
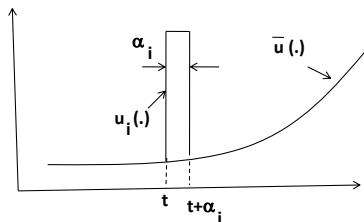


Figure: Needle Variation



# Limit Taking

Can show

$$(A): \|x_j(\cdot) - \bar{x}(\cdot)\|_{L^\infty} \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

(B):

$$y_j(T) - \Phi(T, t)(\alpha_j^{-1}) \int_t^{t+\alpha_j} [f(s, \bar{x}(s), u) - f(t, \bar{x}(s), \bar{u}(s))] \rightarrow 0$$

as  $j \rightarrow \infty$ .

(B) implies (since  $t$  is a Lebesgue point)

$$(B)': y_j(T) - \Phi(T, t)[f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \rightarrow 0$$

But

$$\alpha_j^{-1} [g(x_j(T)) - g(\bar{x}(T))] \geq 0 \text{ for each } i \text{ (by optimality)}$$

and

$$\alpha_j^{-1} [x_j(T) - \bar{x}(T)] \rightarrow y(T) \text{ as } i \rightarrow \infty.$$

It follows that:

$$\nabla_x g(\bar{x}(T)) \cdot y(T) \geq 0.$$

## Proof (continued)

Recall:  $\nabla_x g(\bar{x}(T)) \cdot y(T) \geq 0$ . It follows:

$$\nabla_x g(\bar{x}(T))\Phi(T, t)[f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \geq 0.$$

Write  $p(t) = -\nabla_x g(\bar{x}(T))\Phi(T, t)$ . Then

$$\begin{aligned} -\dot{p}(t) &= \nabla_x g(\bar{x}(T))\Phi(T, t)\nabla_x f(t, \bar{x}(t), \bar{u}(t)) \\ &= p(t)\nabla_x f(t, \bar{x}(t), \bar{u}(t)) \end{aligned}$$

(co-state equation)

$$p(T) = -\nabla_x g(\bar{x}(T)),$$

(transversality condition)

## Proof (continued)

We had:  $\nabla_x g(\bar{x}(T))\Phi(T, t)[f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \geq 0$ .

This implies

$$\rho(t) \cdot [f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \geq 0.$$

for all  $u \in U^0$ , a.e.  $t \in [S, T]$ .

Remains true for all  $u \in U$ , a.e.  $t \in [S, T]$

(Weierstrass condition)

This completes the proof.

# Nonsmooth Terminal Cost

Refinement: Nonsmooth Terminal Cost:

$$(P) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \end{cases}$$

Data: **Lipschitz continuous** function  $g(\cdot) : R^n \rightarrow R$  and vector  $x_0 \in R^n$ .

Transversality condition

$$-p(T) \stackrel{?}{=} \nabla_x g(\bar{x}(T))$$

no longer makes sense. Replace by

$$-p(T) \in \partial g(\bar{x}(T))$$

# Free Right Endpoint Problem with Nonsmooth Cost

Refinement: Allow **nonsmooth** terminal cost

Replace nonsmooth  $g(\cdot)$  by its **inf convolution**:

$$g^\epsilon(x) := \min\{g(y) + \epsilon^{-1}|x - y|^2\}$$

$$(\epsilon > 0)$$

$$(P_\epsilon) \begin{cases} \text{Minimize } g^\epsilon(x(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \end{cases}$$

# Properties of Inf Convolution

$$g^\epsilon(x) := \min\{g(y) + |x - y|^2\}$$

Take an  $K$ -Lipschitz fn  $g(\cdot) : R^k \rightarrow R$  and arbitrary  $x \in R^k$ .

Write  $\xi := 2(x - y)$

where  $y$  minimizes  $y' \rightarrow g(y) + |x - y|^2$ .

- $g^\epsilon(\cdot)$  is  $K$ -Lipschitz
- $g(x) \geq g^\epsilon(x) \geq g(x) - \epsilon$  for all  $x$
- $g^\epsilon(x') - g^\epsilon(x) \leq \xi \cdot (x' - x) + \epsilon^{-1}|x - x'|^2$
- $\xi \in \partial^P g(y)$  and  $|x - y| \leq K^{-1}\epsilon$

Reformulate problem as

$$\begin{cases} \text{Minimize } J^\epsilon(u(\cdot)) \\ \text{over } u(\cdot) \in \mathcal{M} \end{cases}$$

where

$$J^\epsilon(u(\cdot)) = g^\epsilon(x(T; x_0, u(\cdot)))$$

$(x(t; x_0, u(\cdot)))$  denotes sol. of  $\dot{x}(t) = f(t, x(t), u(t))$ ,  $x(S) = x_0$

$$\mathcal{M} = \{\text{measurable } u : [S, T] \rightarrow R^m \mid u(t) \in U \text{ a.e.}\}$$

**Fact.**  $(\mathcal{M}, d_{\mathcal{E}})(\cdot, \cdot)$  is a complete metric space with

$$d_{\mathcal{E}}(u(\cdot), u'(\cdot)) := \mathcal{L}\text{-meas}\{t \mid u(t) \neq u'(t)\}.$$

Also,  $u(\cdot) \rightarrow J^\epsilon(u(\cdot))$  is **continuous** w.r.t.  $d_{\mathcal{E}}(\cdot, \cdot)$  metric.



# Variational Analysis

By properties of inf convolution

$$g^\epsilon(x(T)) \geq g(x(T)) \Rightarrow J^\epsilon(\bar{u}(\cdot)) \leq \inf_{u(\cdot)} \{J^\epsilon(u(\cdot))\} + \epsilon$$

**Apply Ekeland's Theorem:** Take  $\epsilon_j \downarrow 0$ . Then, for each  $i$ ,

There exists a process  $x_i(\cdot), (u_i(\cdot))$  s.t.  $u_i(\cdot)$  is a minimizer for  $(P_i)'$ :

$$(P_i)' \left\{ \begin{array}{l} \text{Minimize } J^{\epsilon_i}(u(\cdot)) + \epsilon_i^1 d_{\mathcal{M}}(u(\cdot), u_i(\cdot)) \\ \text{over } u(\cdot) \in \mathcal{M} \end{array} \right.$$

and

$$d_{\mathcal{M}}(u(\cdot), u_i(\cdot)) \leq \epsilon_i^{1/2}.$$

# Analysis (continued)

Write as optimal control problem with integral cost:

$$(P_i)' \begin{cases} \text{Minimize } \epsilon_i^{1/2} \int_S^T m_i(t, u(t)) dt + g^{\epsilon_i}(x(T)) - g^{\epsilon_i}(x_i(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t. } x(S) = x_0 \end{cases}$$

in which

$$m_i(t, u) = \begin{cases} 1 & \text{if } u = u_i \\ 0 & \text{otherwise.} \end{cases}$$

To use smooth maximum principle, use properties of inf convolution: There exists  $\xi \in R^k$

$$g^{\epsilon_i}(x(T)) - g^{\epsilon_i}(x_i(T)) \leq \xi \cdot (x(T) - x_i(T)) + \epsilon^{-1} |x(T) - x_i(T)|^2$$

$$\xi \in \partial^P g(y_i) \text{ for some } y_i \in x_i(T) + \epsilon^{1/2} B$$

for all processes  $(x(\cdot), u(\cdot))$

# The Maximum Principle

inf convolution inequality  $\Rightarrow$

$(x_i(\cdot), u_i(\cdot))$  is a minimizer for

$$(P_i)'' \begin{cases} \text{Minimize } \epsilon_i^{1/2} \int_S^T m_i(t, u(t)) dt \\ \quad + \xi \cdot (x(T) - x_i(T)) + \epsilon^{-1} |x(T) - x_i(T)|^2 \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t. } x(S) = x_0 \end{cases}$$

in which  $\xi \in \partial^P g(y_i)$ , for some  $y_i \in x_i(T) + \epsilon^{1/2}B$ .

This problem has a  $C^1$  terminal cost, in which  $g(x(T))$  is replaced by

$$\tilde{g}(x) = \xi \cdot (x - x_i(T)) + \epsilon^{-1} |x - x_i(T)|^2$$

# Approximate Necessary Conditions

Apply smooth Maximum Principle  $\Rightarrow$

For each  $i$  there exists  $p_i(\cdot)$  such that

$$\begin{cases} -\dot{p}_i(t) = p_i(t) \cdot f_x(t, x_i(t), u_i(t)), & \text{a.e.}, \\ p_i(t) \cdot f(t, x_i(t), u_i(t)) = \max_{u \in U} p_i(t) \cdot f(t, u_i(t), u) - \epsilon_i^{1/2} & \text{a.e.}, \\ -p(T) = \xi \end{cases}$$

(  $\xi \in \partial^P g(y_i)$  for some  $y_i \in x_i(T) + \epsilon^{1/2} B$  )

But (along a subsequence)

$\text{meas}\{t \mid u_i(t) \neq \bar{u}(t)\} \rightarrow 0$ ,  $x_i(\cdot) \rightarrow \bar{x}(\cdot)$  uniformly,

$p_i(\cdot) \rightarrow p(\cdot)$  uniformly (for some  $p(\cdot) \in W^{1,1}$ ) such that

$$\begin{cases} -\dot{p}_i(t) = p_i(t) \cdot f_x(t, x_i(t), u_i(t)), & \text{a.e.}, \\ p_i(t) \cdot f(t, x_i(t), u_i(t)) = \max_{u \in U} p_i(t) \cdot f(t, u_i(t), u) - \epsilon_i^{1/2} & \text{a.e.}, \\ -p_i(T) = \xi \end{cases}$$

( $\xi \in \partial^P g(y_i)$  for some  $y_i \in x_i(T) + \epsilon^{1/2}B$ )

Passage to the limit as  $i \rightarrow \infty$  yields:

There exists  $p(\cdot) \in W^{1,1}$  and  $\eta \in \partial g(\bar{x})(T)$  such that

$$\begin{cases} -\dot{p}(t) = p(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)), & \mathbf{a.e.}, \\ p(t) \cdot f(t, \bar{x}(t), \bar{u}(t)) = \max_{u \in U} p(t) \cdot f(t, \bar{u}(t), u) & \mathbf{a.e.}, \\ -p(T) \in \partial g(\bar{x}(T)) \end{cases}$$

(limits of proximal normal vectors are normal vectors.)

**We have derived the Maximum Principle for nonsmooth terminal costs.**

# Fixed End-Point Problems

Consider now the problem, when an endpoint constraint is introduced:

Refinement: Allow **nonsmooth** terminal cost

$$(P) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \text{ and } d(x(T)) = 0 \end{cases}$$

in which  $g(\cdot) : R^n \rightarrow R$  and  $d(\cdot) : R^n \rightarrow R^k$  are given Lipschitz functions.

Inclusion of **endpoint constraints**

$$d(x(T)) = 0$$

makes derivation of necessary conditions of optimality much more difficult.

# Fixed end-Point Maximum Principle

Let  $(\bar{x}(\cdot), \bar{u}(\cdot))$  be a minimizer. Assume (H1) and (H2).

Then there exists  $p(\cdot) \in W^{1,1}$ ,  $\lambda \geq 0$ ,  $\mu \in R^k$  such that

$$\left\{ \begin{array}{l} \lambda + |\mu| = 1 \\ -\dot{p}(t) = p(t) \nabla_x f(t, \bar{x}(t), \bar{u}(t)), \quad \text{a.e.} \\ p(t) \cdot [f(t, \bar{x}(t), u) - f(t, \bar{x}(t), \bar{u}(t))] \geq 0 \quad \text{for all } u \in U, \text{ a.e.} \\ -p(T) \in \partial\{\lambda g(x) + \mu \cdot d(x)\}|_{x=\bar{x}(T)} \end{array} \right.$$



# Proof technique: endpoint constraint

Take  $\epsilon \downarrow 0$ . Use the fact that  $(\bar{x}(\cdot), \bar{u}(\cdot))$  is an  $\epsilon_j$  minimizer of **free right end-point problem**

$$\begin{cases} \text{Minimize } \max\{g(x(T)) + \epsilon_j, |d(x(T))|\} \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 \text{ and } d(x(T)) = 0, \end{cases}$$

because  $\max\{g(\bar{x}(T)) + \epsilon, |d(\bar{x}(T))|\} = \epsilon$

and

$\max\{g(x) + \epsilon_j, |d(x)|\} \geq 0$ , for all  $x$ .

(as in NLP,  $\epsilon_j > 0$  ensures **non-degeneracy**)

By Ekeland's Thm., there exists a minimizer  $(x_i(\cdot), u_i(\cdot))$  to

$$\left\{ \begin{array}{l} \text{Minimize } \epsilon_i^{1/2} \int_S^T m_i(t, u(t)) dt + \max\{g(x(T)) + \epsilon, |d(x(T))|\} \\ \text{over processes } (x(\cdot), u(\cdot)) \text{ s.t.} \\ x(S) = x_0 . \end{array} \right.$$

and

$$\text{meas}\{t \mid u_i(t) \neq \bar{u}(t)\} \leq \epsilon_i^{1/2}$$

Apply free end-point non-smooth Maximum Principle. Take limit.

For  $i$ 'th problem the co-state arc  $p_i(\cdot)$  satisfies

$$-p(T) \in \partial \max\{g(x) + \epsilon_i, |d(x)|\}|_{x=x_i(T)}$$

Key condition:  $|d(x_i(T))| = 0$  implies  $g(x_i(T)) + \epsilon_i > 0$

$$\begin{aligned} \text{So: } \partial \{ \max\{g(x) + \epsilon_i, |d(x)|\}|_{x=x_i(T)} \} \\ \subset \partial \{ \lambda_i g(x) + \mu_i \cdot d(x) \}|_{x=x_i(T)} \end{aligned}$$

for some  $\lambda \in [0, 1]$ ,  $\mu \in \mathbb{R}^k$  s.t.  $\lambda + |\mu| = 1$ .

(See Notes for details.)

The accompanying notes provide references, further discussion and extensions of the Maximum Principle described here, to allow for:

- Free end-times (the final time  $T$  is included in the choice variables)
- Pathwise state constraints.