State Constrained Dynamic Optimization

Piernicola Bettiol

University of Brest

44th International Summer School of Automatic Control

Grenoble, 28th August - 1st September 2023

Outline of this lecture

- Overview of classical methods for studying/solving Dynamic Optimization problems
- Necessary optimality conditions, Maximum Principle
- Dynamic programming
- State constraints free $\rightarrow$ enter the state constraints
- Nonsmooth Analysis: basic notions
- Exercises
- References

A standard Dynamic Optimization Problem

$$(P)\begin{cases}\textbf{Minimize } g(x(T)) + \int_S^T L(t,x(t),u(t))dt \\ \textbf{over meas. functions } u : [S,T] \to \mathbb{R}^m\ , \\ \textbf{and arcs } x \in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) \ = \ f(t,x(t),u(t)) \quad \text{a.e. } t \in [S,T] \\ u(t) \in U(t) \subset \mathbb{R}^m \quad \text{a.e. } t \in [S,T] \\ h(x(t)) \le 0 \quad \text{for all } t \in [S,T] \\ x(S) = x_0 \end{cases}$$

The data for this problem comprise:

$[S,T]$	**time interval**
$g : \mathbb{R}^n \to \mathbb{R}$	**endpoint cost function**
$L : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$	**running cost (Lagrangian)**
$f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$	**dynamics**
$U : [S,T] \rightsquigarrow \mathbb{R}^m$	**control set**
$h : \mathbb{R}^n \to \mathbb{R}$	**state constraint**
$x_0 \in \mathbb{R}^n$	**left-end point**

A standard Dynamic Optimization Problem

$$(P) \begin{cases} \textbf{Minimize } g(x(T)) \\ \textbf{over meas. functions } u : [S,T] \to \mathbb{R}^m \,, \\ \textbf{and arcs } x \in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) \;=\; f(t, x(t), u(t)) \quad \text{a.e. } t \in [S,T] \\ u(t) \in U(t) \subset \mathbb{R}^m \quad \text{a.e. } t \in [S,T] \\ h(x(t)) \le 0 \quad \text{for all } t \in [S,T] \\ x(S) = x_0 \end{cases}$$

Some Application Areas

1. *Aerospace:* flight trajectories
2. *Economics:* growth/consumption, optimal harvesting
3. *Chemical engineering, Biology:* optimize yield
4. *Medicine:* anti-cancer treatments, etc.

Example: A Growth/Consumption Model

A ‘growth versus consumption’ problem of neoclassical macro-economics, based on the Ramsey model of economic growth.

Question: what balance should be struck between investment and consumption to **maximize overall investment in social programmes** over a fixed period of time?

$$\begin{cases} \textbf{Maximize} \ \int_0^T (1-u(t))x^\alpha(t)dt \\ \text{subject to} \\ \dot{x}(t) = -ax(t) + bu(t)x^\alpha(t) \quad \text{for a.e. } t \in [0,T], \\ u(t) \in [0,1] \quad \text{for a.e. } t. \in [0,T]\,, \\ x(t) \geq 0 \text{ for all } t \in [0,T]\,, \\ x(0) = x_0 \ . \end{cases}$$

Here, $a > 0$, $b > 0$, $x_0 \geq 0$ and $\alpha \in (0,1)$ are given constants and $[0, T]$ is a given interval.

A Growth/Consumption Model...

$$\begin{cases} \textbf{Maximize} \ \int_0^T (1-u(t))x^\alpha(t)dt \\ \text{subject to} \\ \dot{x}(t) = -ax(t) + bu(t)x^\alpha(t) \quad \text{for a.e. } t \in [0,T], \\ u(t) \in [0,1] \quad \text{for a.e. } t. \in [0,T]\,, \\ x(t) \geq 0 \text{ for all } t \in [0,T]\,, \\ x(0) = x_0\ . \end{cases}$$

Data/model interpretation:

$x \rightarrow$ **global economic output**
$r(x) = bx^\alpha \rightarrow$ financial return from economic output x
$-ax \rightarrow$ fixed costs reducing growth
$u \rightarrow$ **the proportion to invest in industry**
$1-u \rightarrow$ **the proportion to invest in social programmes**

A Growth/Consumption Model...

$$
\begin{cases}
\text{Minimize } -\int_0^T (1-u(t))x^\alpha(t)dt \\
\text{subject to} \\
\dot{x}(t) = -ax(t) + bu(t)x^\alpha(t) \quad \text{for a.e. } t \in [0,T], \\
u(t) \in [0,1] \quad \text{for a.e. } t. \in [0,T]\,, \\
-x(t) \leq 0 \text{ for all } t \in [0,T]\,, \\
x(0) = x_0\ .
\end{cases}
$$

Data/model interpretation:

$x \to$ **global economic output**
$r(x) = bx^\alpha \to$ financial return from economic output x
$-ax \to$ fixed costs reducing growth
$u \to$ **the proportion to invest in industry**
$1-u \to$ **the proportion to invest in social programmes**

A 'simplified' Dynamic Optimization Problem

$$(P)\begin{cases} \textbf{Minimize } g(x(T)) \\ \textbf{over meas. functions } u:[S,T]\to\mathbb{R}^m, \\ \textbf{and arcs } x\in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) = f(t,x(t),u(t)) \quad \text{a.e. } t\in[S,T] \\ u(t)\in U(t)\subset\mathbb{R}^m \quad \text{a.e. } t\in[S,T] \\ x(S)=x_0 \end{cases}$$

Data: $g:\mathbb{R}^n\to\mathbb{R}$, $f:\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$, $U(t)\subset\mathbb{R}^m$, $x_0\in\mathbb{R}^n$

Rmk: $\int_S^T L$ can be '**removed**' by state augmentation technique

$\Rightarrow$ **no state constraints** at present

A **minimizer**: an **admissible process** (trajectory/control pair) $(\bar{x},\bar{u})$ s.t.

$$g(\bar{x}(T))\le g(x(T)) \quad \text{for all } \textit{admissible } (x,u)$$

Differential Inclusion Formulation

$$(DI)\begin{cases} \textbf{Minimize } g(x(T)) \\ \textbf{over arcs } x \in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) \ \in \ F(t,x(t)) \quad \text{a.e. } t \in [S,T] \\ x(S) = x_0 \end{cases}$$

Rmk: '$(P) \to (DI)$' taking $F(t,x) = f(t,x,U(t))$

but we can also have

$$F(t,x) = f(t,x,U(t,x)) \ ...$$

Classical Methods in Dynamic Optimization

In applications, optimal controls are calculated by means of **numerical schemes** based on discretization. **But continuous time optimal control has an important role**:

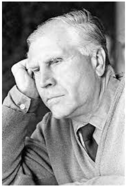
- Control problems associated with the physical world are ‘continuous’
- Theory can tell us when problems are degenerate, and computational schemes will be ill-conditioned
- Basis for high precision ‘shooting’ methods (numerical methods)
- Theory provides tests of local optimality for controls obtained by numerical methods

Classical Methods in Dynamic Optimization

1. **Dynamic Programming** (Sufficient conditions for optimality): *'Analyze minimizers via solutions (the value function) to the Hamilton Jacobi equation'*

R. Bellman
1920 - 1984

2. **Maximum Principle** (Necessary conditions for optimality): *'Analyse minimizers via solutions to a system which involves state and adjoint (costate) variables'*



L.S. Pontryagin
1908 - 1988

Hamilton Jacobi Methods (Dynamic Programming)

'Analyze minimizers via solutions to the Hamilton Jacobi equation' (**R. Bellman**)

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

Embed in family of problems, parameterized by initial data

$$P(\tau, \xi) \left\{ \begin{array}{l} \text{Minimize } g(x(T)) \\ \text{over processes } (x, u) \text{ s.t. } x(\tau) = \xi\,. \end{array} \right.$$

Define $\boxed{V(\tau, \xi) := \text{Inf}(P(\tau, \xi))}$ **Value Function**

Hamilton Jacobi Methods (Dynamic Programming)

$$V(\tau,\xi) = \text{Inf}\left(P(\tau,\xi) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over processes } (x,u) \text{ s.t. } x(\tau)=\xi \end{cases} \right)$$

Principle of Optimality: it establishes some important monotonicity properties of the Value Function:

a) the map $t \to V(t, x(t))$ is **nondecreasing** on $[\tau, T]$ for every process (x, u)

b) if the process $(\bar{x}, \bar{u})$ is optimal for $P(\tau, \xi)$, then $t \to V(t, \bar{x}(t))$ is **constant** on $[\tau, T]$

PDE of Dynamic Programming: $V(.,.)$ is a solution to

$$(HJ) \begin{cases} V_t(t,x) + \min_{u \in U(t)} V_x(t,x) \cdot f(t,x,u) = 0 \\ \qquad\qquad\qquad \text{for all } (t,x) \in (S,T) \times \mathbb{R}^n \\ V(T,x) = g(x) \quad \forall x \in \mathbb{R}^n . \end{cases}$$

Hamilton Jacobi Methods (Dynamic Programming)

Suppose that we solve the (HJ) equation, how does knowing $V(.,.)$ help?

The idea:
For each (t, x) let $(t, x) \to \chi(t, x)$ be a point in $U(t)$ (a control) such that

$$V_t(t, x) + V_x(t, x) \cdot f(t, x, \chi(t, x)) \; = \; 0$$

(a **steepest descent feedback**).
Then for any initial data (τ, ξ), the solution to

$$\begin{cases} \dot{x}(t) \; = \; f(t, x(t), \chi(t, x(t))) & \text{for a.e. } t \in [\tau, T] \\ \text{and } x(\tau) = \xi. \end{cases}$$

is **optimal**.

Hamilton Jacobi Methods (Dynamic Programming)

From the beginning some difficulties have been apparent

- $V(.,.)$ is nondifferentiable; replace $\nabla V = (V_t, V_x)$?
- Need **generalized solutions** to (HJ) equation
- Extend a **generalized solution** to (HJ), in presence of **state constraints**
- Even if $V(.,.)$ is smooth, there is no continuous $\chi(.,.)$ in general: what do we mean by a solution to $\dot{x}(t) \; = \; f(t, x(t), \chi(t, x(t)))$?

Some answers from: Non-Smooth Analysis, Viscosity Solutions Theory.

Example 1.

$$\begin{cases} \text{Minimize } x(1) \\ \text{over measurable functions } u : [0,1] \to \mathbb{R} \\ \qquad\qquad \text{and } x \in W^{1,1}([0,1];\mathbb{R}) \text{ satisfying} \\ \dot{x}(t) = xu \ \text{ a.e.,} \\ u(t) \in [-1,+1] \ \text{ a.e.,} \\ x(0) = 0. \end{cases}$$

Data: $g(x) = x$, $f(x,u) = xu$, $U = [-1,+1]$

$\Rightarrow \min_{u \in [-1,+1]} V_x(t,x) \cdot xu = -|V_x(t,x)x|$

The **Hamilton Jacobi equation** in this case takes the form

$$(HJ) \begin{cases} V_t(t,x) - |V_x(t,x)x| = 0 \quad \text{for all } (t,x) \in (0,1) \times \mathbb{R}, \\ V(1,x) = x \text{ for all } x \in \mathbb{R}. \end{cases}$$

$$(HJ) \begin{cases} V_t(t,x) - |V_x(t,x)x| = 0 & \text{for all } (t,x) \in (0,1) \times \mathbb{R}, \\ V(1,x) = x \text{ for all } x \in \mathbb{R}. \end{cases}$$

The value function is

$$V(t,x) = \begin{cases} xe^{-(1-t)} & \text{if } x \geq 0 \\ xe^{+(1-t)} & \text{if } x < 0. \end{cases}$$

Rmk 1. V satisfies the Hamilton Jacobi (HJ) equation on $\{(t,x) \in (0,1) \times \mathbb{R} \,:\, x \neq 0\}$. However **$V$ cannot be said to be a classical solution** because V is non-differentiable on the subset $\{(t,x) \in (0,1) \times \mathbb{R} \,:\, x = 0\}$.

Rmk 2. The non-differentiability of the value function encountered this example is by no means exceptional.

First Order Necessary Conditions

Take a minimizer $(\bar{x}, \bar{u})$. Define

$$\mathcal{H}(t, x, p, u) := p \cdot f(t, x, u) \quad \text{(un-maximized) Hamiltonian}$$

Maximum Principle (**L.S. Pontryagin**): There exist an arc $p \in W^{1,1}([S, T]; \mathbb{R}^n)$ (**costate arc**) and $\lambda \geq 0$, s.t.

$$(p, \lambda) \neq 0 \quad \text{(Non-trivial Lagrange Multipliers)}$$

$$-\dot{p}(t) = p(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)) \quad a.e.\ t \in [S, T]$$

(The Costate Equation)

$$\mathcal{H}(t, \bar{x}(t), p(t), \bar{u}(t)) = \max_{u \in U(t)} \mathcal{H}(t, \bar{x}(t), p(t), u) \quad a.e.\ t \in [S, T]$$

(The Weierstrass/Maximality Condition)

$$-p(T) = \lambda\, g_x(\bar{x}(T)) \quad \text{(The Transversality Condition)}$$

Widely used to solve dynamic optimization problems, either directly or via numerical methods (cf. Shooting Methods).

Enter State Constraints

Consider the state constrained dynamic optimization problem

$$(SC)\begin{cases} \text{Minimize } g(x(T)) \\ \textbf{over meas. functions } u : [S,T] \to \mathbb{R}^m\ , \\ \textbf{and arcs } x \in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) \ = \ f(t,x(t),u(t)) \quad \text{a.e. } t \in [S,T] \\ u(t) \in U(t) \subset \mathbb{R}^m \quad \text{for a.e. } t \in [S,T] \\ h(x(t)) \le 0 \quad \text{for all } t \in [S,T] \qquad (\text{state constraint}) \\ x(S) = x_0 \ \text{ and } \ x(T) \in C. \end{cases}$$

Data: $g : \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $U(t) \subset \mathbb{R}^m$, $x_0 \in \mathbb{R}^n$,
$C \subset \mathbb{R}^n$
$h : \mathbb{R}^n \to \mathbb{R}$

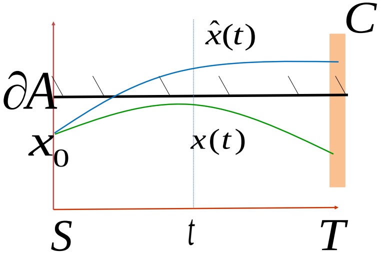
A **minimizer**: an admissible process $(\bar{x}, \bar{u})$ s.t.

$$g(\bar{x}(T)) \le g(x(T)) \quad \text{for all } \textit{admissible } (x,u)$$

Admissible trajectory/control process

A **process** (a **trajectory/control pair**) (x, u) is called **admissible** if it satisfies (for the reference dynamic optimization problem)

- the dynamic constraints: $\dot{x} = f(t, x, u)$, $u(t) \in U(t)$, a.e.
- the end-point constraints: $x(S) = x_0$, $x(T) \in C$
- the state constraint: $h(x(t)) \leq 0$ for all $t \in [S, T]$.



- x is **admissible**: $h(x(t)) \leq 0 \, \forall t$
- but $\hat{x}$ is NOT admissible: $h(\hat{x}(t)) > 0$ for some t

State Constrained Maximum Principle, a first look...

Take a minimizer $(\bar{x}, \bar{u})$.

There exist **multipliers**: arc $p \in W^{1,1}$, $\lambda \geq 0$, and
a Borel measure on $[S, T]$,
a bounded Borel measurable function $\gamma : [S, T] \to \mathbb{R}^n$ s.t.

$$
\begin{aligned}
(p, \mu, \lambda) &\neq (0, 0, 0) \\
-\dot{p}(t) &= q(t) \cdot f_x(t, \bar{x}(t), \bar{u}(t)) \quad \text{a.e. } t \in [S, T] \\
\mathcal{H}(t, \bar{x}(t), q(t), \bar{u}(t)) &= \max_{u \in U(t)} \mathcal{H}(t, \bar{x}(t), q(t), u) \\
-q(T) &\in \lambda g_x(\bar{x}(T)) + N_C(\bar{x}(T)) \\
supp\{\mu\} &\subset \{t \in [S, T] \,:\, h(\bar{x}(t)) = 0\} \\
\gamma(t) &= h_x(\bar{x}(t)) \text{ for } \mu\text{-a.e. } t \in [S, T]
\end{aligned}
$$

$q \in NBV([S, T]; \mathbb{R}^n)$

$$
q(t) := \begin{cases} p(S) & \text{if } t = S \\ p(t) + \int_{[S,t]} \gamma(s) d\mu(s) & \text{if } t \in (S, T]\,. \end{cases}
$$

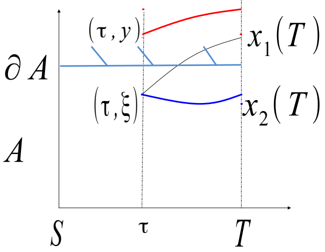
Embed in family of problems, parameterized by initial data: given any $(\tau, \xi) \in [S, T] \times \mathbb{R}^n$, $P(\tau, \xi)$ is variant on $P(S, x_0)$ when the 'initial data' (τ, ξ) replaces (S, x_0).

$$P(\tau,\xi) \begin{cases} \text{Minimize } g(x(T)) \\ \text{over admissible processes } (x(.), u(.)) \text{ s.t. } x(\tau) = \xi\,. \end{cases}$$

Define $\boxed{V(\tau,\xi) := \text{Inf}(P(\tau,\xi))}$ **Value Function**

$$V : [S, T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

(Note: $V(\tau, y) = +\infty$ since $y \notin A$.)



Dynamic Programming – State Constraints

$$P(\tau,\xi)\begin{cases}\text{Minimize } g(x(T))\\ \text{over admissible processes } (x,u) \text{ s.t.}\\ x(\tau)=\xi.\end{cases}$$

How does the state constraint affect optimality conditions?

Now, value function $V:[S,T]\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a lower semicontinuous solution to

$$\begin{cases}V_t(t,x)+\min_{u\in U(t)} V_x(t,x)\cdot f(t,x,u) = 0\\ \qquad\qquad\qquad\qquad \text{for all } (t,x)\in(S,T)\times \text{int } A\\ V(T,x) = g(x) \quad \forall x\in A\end{cases}$$

unique, in fact, in some **generalized sense** (Non-Smooth Analysis, Viscosity Solutions...)

Here $A:=\{x\in\mathbb{R}^n \;:\; h(x)\le 0\}$

Nonsmoothness

There was **a lack of suitable analytic tools** for investigating local properties of nonsmooth functions/sets are (easily) encountered in the study of dynamic optimization problems:
- Dynamic Programming (cf. Example 1)
- Necessary Optimality Conditions (for instance to take account of pathwise constraints)

Two important breakthroughs occurred in the 1970's:

1. **F. H. Clarke**'s theory of generalized gradients generalized the concept of 'subdifferentials' of convex functions to larger functions classes launched the field of **nonsmooth analysis**
2. the concept of **viscosity solutions**, due to **M. G. Crandall and P.-L. Lions**, which provides a framework for proving existence and uniqueness of generalized solutions to Hamilton Jacobi equations

Nonsmooth Analysis - Basic Definitions

Nonsmooth Analysis: provides tools for local approximations of non-differentiable functions and of sets with non-differentiable boundaries.

Key question: How should classical concepts of ‘gradients’ and ‘normals’ be adapted, to give provide useful local information about non-differentiable functions and sets with non-differentiable boundaries?

Origins - Smooth framework

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

Assume:

- boundary of C is an $n-1$ dimensional C^1 manifold
- f is continuously differentiable

The normal vector η to boundary of 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 $C = \{x \in \mathbb{R}^n \ : \ h(x) \leq 0\}$, then $\eta = \nabla h(\bar{x})$.

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

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

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

Origins - Convex Analysis

Suppose *C* **and** *f* **are smooth AND convex**

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

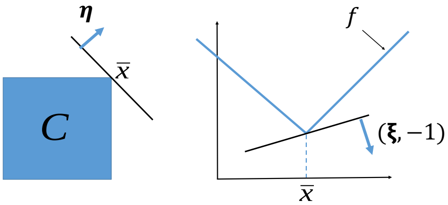
$$\eta \cdot (x - \bar{x}) \leq 0, \ \forall \ x \in C \quad \text{and} \quad \xi \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}), \ \forall \ x \in \mathbb{R}^n$$

Now assume *C* **and** *f* **are merely convex.**

set of normal vectors $\quad N_C(\bar{x}) := \{\eta \ : \ \eta \cdot (x - \bar{x}) \leq 0, \ \ \forall x \in C\}$

set of subgradients of *f*

$$\partial f(\bar{x}) := \{\xi \ : \ \xi \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}), \ \forall x \in \mathbb{R}^n\}$$



Proximal Normal Cones

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

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

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

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

Proximal Normal Cones...

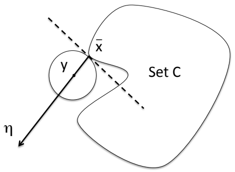


Figure: Proximal Normal Vectors

Defining property of proximal normal vectors η:

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

Equivalently:

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

Limiting Normal Cones

Take a **closed** set $C \subset \mathbb{R}^n$ and a point $\bar{x} \in C$. A vector $\eta \in \mathbb{R}^n$ is said to be a limiting normal vector to C at $\bar{x}$ if there exist $x_i \stackrel{C}{\to} x$ and $\eta_i \to \eta$ s.t.

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

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

$$N_C(\bar{x}) \;\; := \;\; \{\eta \in \mathbb{R}^n \; : \quad \exists \; x_i \stackrel{C}{\to} x \text{ and } \eta_i \to \eta \text{ s.t. } \eta_i \in N_C^P(x_i) \; \forall \; i\}.$$

$\to$ $x_i \stackrel{C}{\to} x$ indicates that $x_i \to x$ and $x_i \in C$ for all i

Limiting Normal Cones...

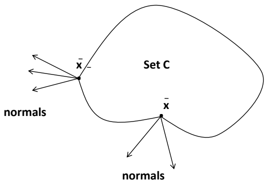


Figure: Limiting Normal Vectors at different base points

Some basic properties of $N_C(\bar{x})$:

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

Proximal Subgradients

Take an extended valued, lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x} \in \operatorname{dom}\{f\}$.
A vector $\eta \in \mathbb{R}^n$ is said to be a proximal subgradient of f at $\bar{x}$ if there exist $\epsilon > 0$ 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 $|x - \bar{x}| \leq \epsilon$.

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

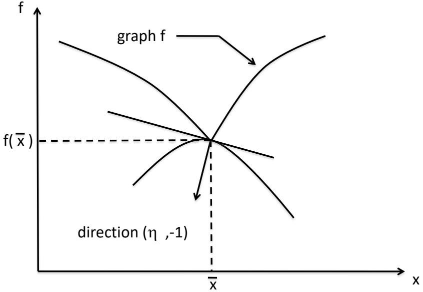
Proximal Subgradients...

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 $\partial^P f(\bar{x})$:

$$\partial^P f(\bar{x}) := \{\text{there exist } \epsilon > 0 \text{ and } M \geq 0 \text{ such that (2) is satisfied}\}.$$

$$\eta \cdot (x - \bar{x}) \leq f(x) - f(\bar{x}) + M|x - \bar{x}|^2 \quad (2)$$
$$\text{for all points } x \text{ which satisfy } |x - \bar{x}| \leq \epsilon.$$

Geometric Interpretation of Proximal Subgradients



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 on or below the graph of f on a neighbourhood of $\bar{x}$.

Limiting Subgradients

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

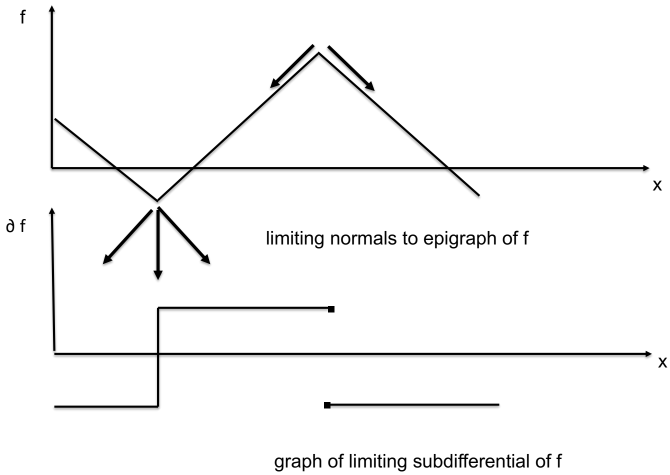
$$\eta_i \in \partial^P f(x_i) \;\text{ for all } i.$$

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

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

$\to$ $x_i \xrightarrow{f} x$ indicates that $x_i \to x$ and $f(x_i) \to f(x)$ as $i \to \infty$

Limiting Subdifferential



Limiting Subdifferential

Some basic properties of the limiting subdifferential $\partial f(\bar{x})$:

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

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

- Suppose that f is Lipschitz continuous on a neighbourhood of $\bar{x}$. Then, for any subset $\mathcal{S} \subset \mathbb{R}^n$ of zero n-dimensional Lebesgue measure, we have

$$\begin{aligned} \mathrm{co}\, \partial f(\bar{x}) &= \mathrm{co}\, \{\eta : \exists\, x_i \to x \text{ such that } \nabla f(x_i) \text{ exists and} \\ &\qquad\qquad x_i \notin \mathcal{S} \text{ for all } i \text{ and } \nabla f(x_i) \to \eta\} \\ &= \partial^C f(\bar{x}) \quad (\text{Gradient Formula}). \end{aligned}$$

$\to$ $\partial^C f(\bar{x})$ is the **Clarke subdifferential**

Subdifferentials

Other properties

- If f is of class $\mathcal{C}^1$ near $\bar{x}$, then $\partial^C f(\bar{x}) = \{\nabla f(\bar{x})\}$
- If f is of class $\mathcal{C}^1$ near $\bar{x}$ and ∇f is Lipschitz near $\bar{x}$, then $\partial^P f(\bar{x}) = \{\nabla f(\bar{x})\} = \partial^C f(\bar{x})$
- **Partial limiting subdifferential**: if $f = f(x, y)$, then $\partial_x f(\bar{x}, \bar{y})$ denotes the limiting subdifferential of $x \to f(x, \bar{y})$
- There are, in fact, a number of ways of defining subgradients and there exist equivalent ways of defining subgradients: as limits of proximal subgradients, by means of normals to epigraph sets, etc.

The Clarke Tangent Cone

Take a closed set $C \subset \mathbb{R}^n$ and a point $x \in C$.
The **Clarke tangent cone** to C at x is the set

$$T_C(x) := \liminf_{t \downarrow 0, y \xrightarrow{C} x} t^{-1}(C - y).$$

Rmk 1: Equivalent 'sequential' definition:

$$\begin{aligned} T_C(x) \quad = \quad & \{\xi \ : \ \forall \text{ sequences } x_i \xrightarrow{C} x \text{ and } t_i \downarrow 0 \\ & \exists \text{ a sequence } \{c_i\} \subset C \ \text{ s. t. } t_i^{-1}(c_i - x_i) \to \xi\} \end{aligned}$$

Rmk 2: The **Clarke tangent cone** $T_C(x)$ and the **limiting normal cone** $N_C(x)$ are related according to

$$T_C(x) = N_C(x)^* = \{\xi : \xi \cdot \nu \leq 0 \text{ for all } \nu \in N_C(x)\}$$

A State Constrained Problem

Consider the state constrained dynamic optimization problem

$$(SC)\begin{cases} \text{Minimize } g(x(S), x(T)) \\ \textbf{over meas. functions } u : [S,T] \to \mathbb{R}^m\,, \\ \textbf{and arcs } x \in W^{1,1}([S,T];\mathbb{R}^n) \text{ s.t.} \\ \dot{x}(t) \;=\; f(t, x(t), u(t)) \quad \text{a.e. } t \in [S,T] \\ u(t) \in U(t) \subset \mathbb{R}^m \quad \text{for a.e. } t \in [S,T] \\ h(x(t)) \le 0 \quad \text{for all } t \in [S,T] \qquad (\text{state constraint}) \\ (x(S), x(T)) \in C. \end{cases}$$

Data: $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $U(t) \subset \mathbb{R}^m$,
$C \subset \mathbb{R}^n \times \mathbb{R}^n$
$h : \mathbb{R}^n \to \mathbb{R}$

A **minimizer**: an admissible process $(\bar{x}, \bar{u})$ s.t.

$$g(\bar{x}(S), \bar{x}(T)) \le g(x(S), x(T)) \quad \text{for all } \textit{admissible } (x, u)$$

Assumptions

(H1) for fixed x, $f(.,x,.)$ is $\mathcal{L}([S,T]) \times \mathcal{B}^m$ measurable, there exists a $\mathcal{L}([S,T]) \times \mathcal{B}^m$ measurable function $k : [S,T] \times \mathbb{R}^m \to [0,\infty)$ such that $t \to k(t,\bar{u}(t))$ is integrable and, for a.e. $t \in [S,T]$,

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

for all $x, x' \in \mathbb{R}^n$ and $u \in U(t)$,

(H2): the set Gr U is $\mathcal{L}([S,T]) \times \mathcal{B}^m$ measurable,

(H3): g is Lipschitz continuous and C is a closed subset of $\mathbb{R}^{n\times n}$,

(H4): there exists $k_h > 0$ such that

$$|h(x) - h(x')| \leq k_h|x - x'| \ \text{ for all } x, x' \in \mathbb{R}^n$$

Let $(\bar{x}, \bar{u})$ be a minimizer for (SC).
Then there exist $p \in W^{1,1}([S,T];\mathbb{R}^n)$, $\lambda \geq 0$,
a Borel measure μ on $[S,T]$, a bounded Borel measurable function $\gamma : [S,T] \to \mathbb{R}^n$ s.t.

(a): $(p, \mu, \lambda) \neq (0,0,0)$,

(b): $-\dot{p}(t) \in \text{co}\, \partial_x \mathcal{H}(t, \bar{x}(t), q(t), \bar{u}(t))$ a.e. $t \in [S,T]$,

(c): $\mathcal{H}(t, \bar{x}(t), q(t), \bar{u}(t)) = \max_{u \in U(t)} \mathcal{H}(t, \bar{x}(t), q(t), u)$ a.e. ,

(d): $(q(S), -q(T)) \in \lambda \partial g(\bar{x}(S), \bar{x}(T)) + N_C(\bar{x}(S), \bar{x}(T))$,

(e): $\text{supp}\{\mu\} \subset \{t \in [S,T] \,:\, h(\bar{x}(t)) = 0\}$ and
$\gamma(t) \in \partial^{>} h(\bar{x}(t))$ for μ-a.e. $t \in [S,T]$,

where $q \in NBV([S,T];\mathbb{R}^n)$ is the function

$$q(t) := \begin{cases} p(t) & \text{if } t = S \\ p(t) + \int_{[S,t]} \gamma(s) d\mu(s) & \text{if } t \in (S,T]. \end{cases}$$

$$\partial^{>} h(\bar{x}(t)) \;:=\; \text{co} \limsup \{\partial h(y_i) \,:\, y_i \to \bar{x}(t),\; h(y_i) > 0 \;\forall i\}$$

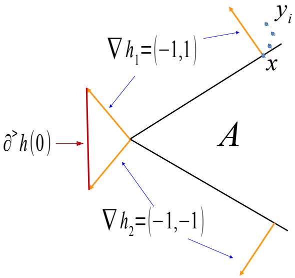
The '**hybrid subdifferential**' $\partial^{>} h(x)$ is the set

$$\partial^{>} h(x) = \text{co}\,\{\eta : \exists\, y_i \to x \text{ and } \eta_i \to \eta \text{ s. t. } \eta_i \in \partial h(y_i),\ h(y_i) > 0\ \ \forall\, i \in \mathbb{N}\}$$

$$A = \left\{(x_1, x_2) \in \mathbb{R}^2 : |x_2| - x_1 \leq 0\right\}$$

$$h(x) := |x_2| - x_1 = \max\{h_1(x), h_2(x)\}$$

$$h_1(x) = x_2 - x_1, \quad h_1(x) = -x_2 - x_1$$



Comments

- It is a state constrained version of **Clarke Nonsmooth Maximum Principle**
- **Autonomous Case:** Assume, also, that $f(t, x, u)$ and $U(t)$ are independent of t. Then, in addition to the above conditions, there exists a constant r such that

  (f): $\mathcal{H}(t, \bar{x}(t), q(t), \bar{u}(t)) = r$ a.e..

Comments...

The state constraint formulation '$h(x(t)) \leq 0$', in (SC) can be extended to '$h(t, x(t)) \leq 0$', where $h(t, x)$ is permitted to be merely **Lipschitz continuous w.r.t.** x and **upper semicontinuous w.r.t.** t. This allows to cover a number of special cases of interest.

(i): Multiple state constraints: $h_k(t, x(t)) \leq 0$ for $t \in [S, T]$, $k = 1, \ldots, M$, in which the $h_k(t, x)$'s are Lipschitz continuous w.r.t. x, can be accommodated by setting $h(t, x) := \max_k\{h_k(t, x)\}$.

(ii): Implicit state constraint: $x(t) \in A$, for $t \in [S, T]$, in which $A \subset \mathbb{R}^n$ is a given closed set. Here the necessary conditions are valid in a modified where, in condition (e), the Borel measurable function γ is now required to satisfy

$$\gamma(t) \in \text{co}\left(N_A(\bar{x}(t)) \cap \{\xi \in \mathbb{R}^n \, : \, |\xi| = 1\}\right).$$

These modified conditions can be derived by setting $h(x) = d_A(x)$ (d_A is the distance function to the set A).

Non-degeneracy and Normality of the Maximum Principle

- If $(\bar{x}, \bar{u})$ satisfies the Maximum Principle $\Longrightarrow$ *extremal*.
- If $(\bar{x}, \bar{u})$ provides the minimum $\Longrightarrow$ *optimal*.
- If $\lambda = 1 \Longrightarrow$ Normality of the Maximum Principle
  If $\lambda = 0 \Longrightarrow$ Abnormal case
- If

$$\lambda + \int_{(S,T]} d\mu(s) + \left| p(S) + h_x(\bar{x}(S))\ \mu(\{S\}) \right| \neq 0$$

$\Longrightarrow$ Non-degeneracy of the Maximum Principle

Rmk: '**Normality** $\Longrightarrow$ **Non-degeneracy**'

A degenerate situation

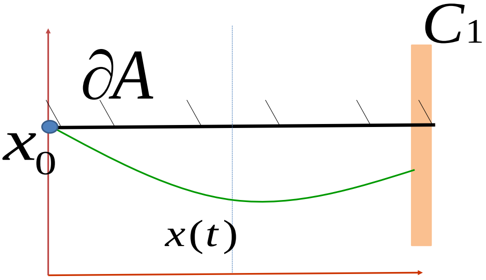
Consider a special case in which $f(t,x,u)$, $h(t,x)$ and $U(t)$ are independent of t.

Assume that $f(.,u)$ (for all $u \in \mathbb{R}^m$), g and h are of class $\mathcal{C}^1$, f is continuous, and the left end-point are fixed, i.e.

$$C = \{x_0\} \times C_1$$

To explore the degeneracy phenomenon, we suppose that

$$h(x_0) = 0.$$



A degenerate situation..

Then the necessary conditions of optimality assert the existence of an absolutely continuous arc $p \in W^{1,1}([S,T];\mathbb{R}^n)$, $\lambda \geq 0$, a measure μ s.t.

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

(ii) $-\dot{p}(t) = (p(t) + \int_{[S,t]} h_x(\bar{x}(s))d\mu(s)) \cdot f_x(\bar{x}(t), \bar{u}(t))$ a.e. $t \in [S,T]$,

(iii) $u \to (p(t) + \int_{[S,t]} h_x(\bar{x}(s))d\mu(s)) \cdot f(\bar{x}(t), u)$ is maximized over $u \in U$ at $\bar{u}(t)$, a.e. $t \in [S,T]$,

(iv) $-(p(T) + \int_{[S,T]} h_x(\bar{x}(s))d\mu(s)) \in \lambda g_x(\bar{x}(T)) + N_{C_1}(\bar{x}(T))$,

(v) $\operatorname{supp} \mu \subset \{t : h(\bar{x}(t)) = 0\}$

A degenerate situation..

Here, we find that conditions (i)-(v) above are satisfied (for some p, λ and μ) when $\bar{x}$ is *any* arc satisfying the constraints of (SC). A possible choice of multipliers is

$$(p \equiv -h_x(S),\ \mu = \delta_{\{S\}},\ \lambda = 0) \qquad (3)$$

($\delta_{\{S\}}$ denotes the unit measure concentrated at $\{S\}$.) Provided $h_x(S) \neq 0$, these multipliers are non-zero. Condition (v) is satisfied, by (3). The remaining conditions (i) – (iv) are satisfied since

$$\int_{(S,t]} h_x(\bar{x}(s))d\mu(s) \ = \ 0 \quad \text{for } t \in (S, T).$$

The fact that the necessary conditions (i) - (v) are automatically satisfied by **ALL admissible arcs** renders them useless (degenerate) as necessary conditions.

Non-degeneracy

How should we deal with the degeneracy phenomenon?

Extra necessary conditions or extra hypotheses are clearly required.

There are now a number of ways to do this.

We focus here on a particular analytical tool which can be used in several approaches:

$\rightarrow$ distance estimates

A useful analytical tool

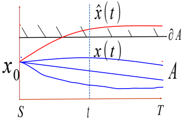
Distance estimates (Filippov-type theorems) constitute a common set of analytical tools which can be used to resolve a number of important questions in **state constrained** dynamic optimization problems.
Some applications are

- **non-degeneracy** and **normality** of the **maximum principle** (which provides necessary conditions for optimality);
- existence, characterization and regularity of the **value function** for Hamilton-Jacobi-Bellman and Hamilton-Jacobi-Isaacs equations;
- **sensitivity conditions**: adjoint variables in the Maximum Principle can be interpreted as 'gradients' of the value function;
- feedback laws, (synthesis)
- ...

A useful analytical tool...

Distance estimates consist in **constructing a admissible state trajectory x which lies 'close' to a state trajectory $\hat{x}$ that violates the state constraint**, and for which $x(S) = \hat{x}(S)$. Specifically, there exists a constant K, independent of $\hat{x}(.)$, such that

$$||x(.) - \hat{x}(.)|| \le K \times \rho(\hat{x}(.)),$$



where $||.||$ is some norm defined on the set of trajectories, for instance L^∞ or $W^{1,1}$.

Here we have a **linear** estimate w.r.t. the 'violation rate' $\rho(\hat{x}(.))$

$$||x||_{L^\infty} = \sup_{t\in[S,T]} |x(t)|, \quad ||x||_{W^{1,1}} = |x(S)| + \int_{[S,T]} |\dot{x}(t)|\, dt$$

The "constraint violation rate" of an arc $x(.)$

$\rho(x(.))$ represents the "violation rate" of an arc $x(.) : [S, T] \to \mathbb{R}^n$

- If we have a "**functional inequality representation**":

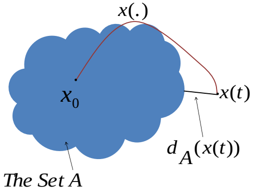
$$A = \{x \mid h(x) \leq 0\} .$$

for some (Lipschitz) function $h : \mathbb{R}^n \to \mathbb{R}$.

$$\rho(x(.)) := \max_{t \in [S,T]} \{h(x(t)) \vee 0\}$$

- If A is an **arbitrary closed set**, we can define $\rho(x(.))$ via the **distance function** to the set A, $d_A(x)$:

$$\rho(x(.)) := \max_{t \in [S,T]} d_A(x(t))$$



Distance estimates

More in general we can consider the following estimate

$$m((x(.), u(.)), (\hat{x}(.), \hat{u}(.))) \leq \theta(\rho(\hat{x}(.))) ,$$

where

- $m(.,.)$ **is a metric on the set of processes** (Strictly speaking we should say pseudo-metric, since we do not require '$m(p, p') = 0 \implies p = p'$')
- $\theta(.) : \mathbb{R}^+ \to \mathbb{R}^+$ **is a rate of convergence modulus**, i.e. a function satisfying $\lim_{\rho \downarrow 0} \theta(\rho) = 0$.

Rmk: **The stronger the metric** $m(.,.)$ and **greater the rate** at which $\theta(\rho)$ tends to zero as $\rho \to 0$, **the more the information** that is conveyed by the estimates.

Distance estimates...

$$m((x(.), u(.)), (\hat{x}(.)\hat{u}(.))) \leq \theta(h(\hat{x}(.))) \ ,$$

- **A variety of estimates** has been considered, distinguished by the choice of $m(.,.)$ and $\theta(.)$.
- At least **4 different approaches** have been employed (here, we shall see two of them).

Idea of the proof/approach

Consider the optimal control problem

$$(P1)\begin{cases} \text{Minimize } g(x(T)) \\ \text{subject to} \\ \dot{x}(t) = f(x(t), u(t)) \quad \text{a.e. } t \in [S, T], \\ u(t) \in U \quad \text{a.e. } t \in [S, T], \\ h(x(t)) \leq 0 \quad \text{for all } t \in [S, T], \\ x(S) = x_0 , \end{cases}$$

in which f and h are of class $\mathcal{C}^1$, and g is Lipschitz (of rank k_g).

Suppose that **we have at hand the distance estimate**:

$$||x(.) - \hat{x}(.)||_{L^\infty} \leq K \times \rho(\hat{x}(.))$$

Take an **optimal process** $(\bar{x}, \bar{u})$

⇒ the maximum principle applies with $\lambda = 1$ (normal case)

Distance Estimates ⇒ Normality ...

Idea of the proof/approach

CLAIM: $((\bar{z} \equiv 0, \bar{x}), \bar{u})$ is an **optimal process** for the problem

$$(P2) \begin{cases} \text{Minimize } g(x(T)) + Kk_g(z(T) \vee 0) \\ \text{subject to} \\ \dot{z}(t) = 0, \quad \dot{x}(t) \; = \; f(x(t), u(t)), \; u(t) \in U \\ h(x(t)) - z(t) \; \leq \; 0 \\ x(S) = x_0, z(S) \geq 0\,. \end{cases}$$

Indeed, suppose to the contrary that there exists a process $((z', x'), u')$ with lower cost:

$$g(x'(T)) + Kk_g \max_{t \in [S,T]} \{h(x'(t)) \vee 0\} \; < \; g(\bar{x}(T)) \quad + 0$$

Recall: $\max_{t \in [S,T]} \{h(x'(t)) \vee 0\} = \rho(x'(.))$

Distance Estimates ⇒ Normality ...

We have

$$g(x'(T)) + Kk_g \times \rho(x'(.)) \; < \; g(\bar{x}(T))$$

According to the **distance estimate** applied to (x', u'), there exists an **admissible (for (P1)) process** (x, u) s.t.

$$||x(.) - x'(.)||_{L^\infty} \; \leq \; K \times \rho(x'(.))$$

But, then (x, u) is **admissible for (P1) and satisfies**:

$$\begin{aligned} g(x(T)) \;&\leq\; g(x'(T)) + k_g ||x(.) - x'(.)||_{L^\infty} \\ &\leq\; g(x'(T)) + k_g K \times \rho(x'(.)) \\ &<\; g(\bar{x}(T)) \; . \end{aligned}$$

This contradicts the optimality of $(\bar{x}, \bar{u})$!

Distance Estimates ⇒ Normality ...

Now **apply the nonsmooth state constrained Maximum Principle** with the reference minimizer $((\bar{z} \equiv 0, \bar{x}), \bar{u})$ for $(P2)$.

Let λ and μ be the cost and 'measure' **multipliers** respectively, and let $p(.)$ and $p_z(.) \equiv -c$ be the **costate arcs** associated with the x and z variables.

We **deduce the usual Maximum Principle conditions for** $(P1)$ in relation to $(\bar{x}, \bar{u})$ and p.

BUT...

Distance Estimates $\Rightarrow$ Normality ...

BUT the **transversality conditions** in relation to $\bar{z}$ and p_z yield the additional information that

$$c \geq 0$$

and

$$c + \int_{[S,T]} d\mu(t) \;\leq\; K k_g \lambda \;.$$

If $\lambda = 0$, we would have, by the preceding condition, $\mu = 0$ and $p_z(.) \equiv 0$. But also, in consequence of the adjoint inclusion and the transversality condition for $p(.)$, we would also have $p(.) \equiv 0$. From this contradiction of the non-triviality of the multipliers.

We conclude that $\lambda = 1$.

Distance Estimates - 'Standing Hypotheses'

Recall the data of the control system:

$$\begin{cases} \dot{x}(t) &= f(t, x(t), u(t)) \text{ and } u(t) \in U(t) \\ h(x(t)) &\leq 0 \end{cases}$$

Assume that for some $c > 0$ and $k_f(.) \in L^1$

- $f(., x, .)$ is $\mathcal{L} \times \mathcal{B}^m$ (Lebesgue-Borel) meas. for each x; $U(.)$ has Borel-meas. graph; $f(t, x, U(t))$ is closed, for each t, x
- $|f(t, x, u)| \leq c(1+|x|)$ for all $u \in U(t)$, $(t, x) \in [S, T] \times \mathbb{R}^n$
- $|f(t, x, u) - f(t, x', u)| \leq k_f(t)|x - x'|$ for all $t \in [0, 1]$, $x, x' \in \mathbb{R}^n$ and $u \in U(t)$.

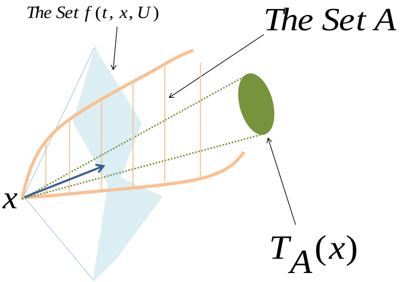
(we say 'f **is closed, meas., integr. Lipschitz with linear growth**')

and we also have the following **Constraint Qualification**

- $f(t, x, U(t)) \cap \text{int}\, T_A(x) \neq \emptyset$ for all $x \in \partial A$, $t \in [S, T]$

Inward Pointing Condition.

Inward Pointing Condition



The **Clarke tangent cone** to A at $x \in A$, $T_A(x)$, is defined by

$$T_A(x) = \liminf_{t \downarrow 0, y \xrightarrow{A} x} t^{-1}(A - y)$$

Contributions to This Area - a first (partial) List

H. M. Soner, “Optimal Control Problems with State-Space Constraints 1 & 2”, *SIAM J. Control Optim.*, 24, 1986.

F. Rampazzo and R. B. Vinter, “A Theorem on Existence of Neighbouring Trajectories Satisfying a State Constraint, with Applications to Optimal Control”, *IMA J. Math. Control Inform*, 16, 1999.

F. Forcellini and F. Rampazzo, “On Non-convex Differential Inclusions whose State is Constrained in the closure of an Open Set”, *J. Differential Integral Equations*, 12, 1999.

H. Frankowska and F. Rampazzo, “Filippov’s and Filippov-Wazewski’s Theorems on Closed Domains”, *JDE*, 2000.

H. Frankowska and R. B. Vinter, “Existence of Neighbouring Feasible Trajectories: Applications to Dynamic Programming for State Constrained Optimal Control Problems”, *JOTA*, 2000.

F. Rampazzo and R. B. Vinter, “Degenerate Optimal Control Problems with State Constraints”, *SIAM J. Control Optim.*, 39, 2000.

F. H. Clarke, L. Rifford and R.J. Stern, “Feedback in State Constrained Optimal Control”, *ESAIM: COCV*, 7, 2002.

Approach: use a suitable **time-delay** control argument

Assumptions: **Lipschitz continuity** set of velocities

Result: L^∞**-norm estimate** on trajectories that is **linear** w.r.t. the violation rate $\rho(\hat{x}(.))$

Idea of this approach

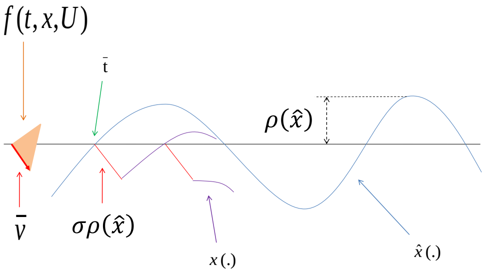


Figure: '**Time-delay control argument**': whenever the boundary is approached, use the **interior pointing vector** $\bar{v}$ to "push" inside the trajectory (candidate to be 'admissible'): apply v for a time proportional to the "violation rate", $\sigma\rho(\hat{x}(.))$.

Examples

Two examples [Bettiol-Bressan-Vinter, SICON 2010, 2011] maybe renewed some interest in this area, showing that for an arbitrary state constraint set A (merely closed):

1) Preceding **linear estimate is not valid in general** (A merely closed), when the L^∞-norm is we replaced by **stronger norms/metrics** ($W^{1,1}$, Ekeland metric).

2) **Even linear L^∞-estimates fail to hold true** in general (A merely closed) when $t \rightsquigarrow f(t, x, U(t))$ is **discontinuous**

The **Ekeland metric** $\leftrightarrow d_{\mathcal{E}}((\hat{x}, \hat{u}), (x, u)) := \text{meas}\{t \ : \ \hat{u}(t) \neq u(t)\}$

Example 1 (in $\mathbb{R}^2$) - $W^{1,1}$ Estimates

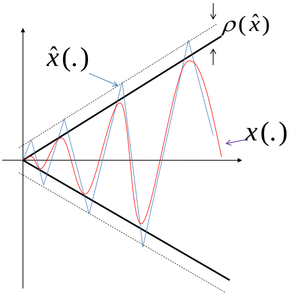


Figure: Example where Linear $W^{1,1}$ Estimate is not Valid. The trajectory $\hat{x}(.)$ approximated by a admissible trajectory $x(.)$.

Example 1: Details

$$f(t,x,u) = u, \quad \begin{array}{rcl} U & = & \mathrm{co}\left(\{(1,+2)\},\{(1,-2)\},\{(0,0)\}\right) \\ A & = & \{(x_1,x_2)\in\mathbb{R}^2 \,|\, |x_2| \le x_1\} \end{array}$$

Then $\rho(\hat{x}(.)) > 0$ and

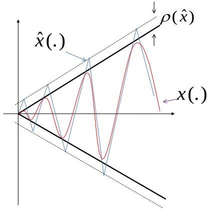
$$\begin{array}{rcl} ||\,\hat{x}(.) - x(.)\,||_{W^{1,1}} & \geq & \sum_{i=1}^{N} |(\hat{x}(t_{i+1}) - x(t_{i+1})) - (\hat{x}(t_i) - x(t_i))| \\ & \geq & 2 \times \rho(\hat{x}(.)) \times N, \end{array}$$

where N = number of switches: $3^N \geq \frac{1}{2} \times \left(\frac{1}{\rho(\hat{x}(.))} + 1\right)$.

So

$$||\,\hat{x}(.) - x(.)\,||_{W^{1,1}} \geq \text{const.} \times \rho(\hat{x}(.))\, |\log_e \rho(\hat{x}(.))|\,.$$

Example 2 (in $\mathbb{R}^3$) - L^∞ Estimates



- $f(t, x, U(t)) = U(t)$ is closed convex valued + 'inward' pointing condition
- $A = \{(x_1, x_2, x_3) \,|\, |x_2| \leq x_1\}$

But, **for any** $K > 0$ and $\varepsilon > 0$, there exists a process $(\hat{x}(.), \hat{u}(.))$ such that $\varepsilon > \rho(\hat{x}(.)) > 0$ and

$$||\hat{x}(.) - x(.)||_{L^\infty} \geq K \times \rho(\hat{x}(.)) \,.$$

In this example $t \rightsquigarrow f(t, x, U(t))$ is **discontinuous** (**measurable in time**).

More general Estimates? (L^∞, $W^{1,1}$...)

Some questions raised taking into account the examples:

- If A has a **smooth boundary**, are linear L^∞-estimates valid when $f(., x, u)$ **is measurably time-dependent**?

They can even be improved to linear $W^{1,1}$-estimates!

- And if A **is merely closed,**
  - are linear L^∞-estimates valid when $f(., ., .)$ is no longer Lipschitz?
  - what can we say about stronger norms(/metrics) than L^∞?
  - if not linear, what can we say about distance estimate regularity/behaviour?

Some motivations for stronger metrics:

- $W^{1,1}$-estimates $\rightarrow$ non-degeneracy necessary optimality conditions [Rampazzo-Vinter, SICON 2000]
- Ekeland metric $\rightarrow$ normality Maximum Principle when the dynamics and control constraint set are possibly discontinuous in and non-closed respectively.
  (cf. F. H. Clarke, The Maximum Principle Under Minimal Hypotheses, *SICON*, 1976.)

Counter-Examples

Consider now an arbitrary closed set A.

- **Eliminate "$f(.,x,u)$ is Lipschitz" assumption**. Then, for any $\alpha \in (0,1)$, the superlinear Hölder estimate

$$||\hat{x}(.) - x(.)||_{L^\infty} \ \leq \ K \times (\rho(\hat{x}(.)))^\alpha$$

  is not in general verified!

- **Replace "$f(.,x,u)$ is Lipschitz" by "$f(.,x,u)$ is continuous" assumption**. Then, the superlinear $\rho|\log(\rho)|$-estimate

$$||\hat{x}(.) - x(.)||_{L^\infty} \ \leq \ K \times \rho(\hat{x}(.))|\log(\rho(\hat{x}(.)))|$$

  is not in general valid!

(see [Bettiol, Frankowska and Vinter, JDE 2012])

$W^{1,1}$ Distance Estimates for 'smooth' A

Theorem ($W^{1,1}$ Estimates for 1 smooth State Constraint)
Assume *standing hypotheses* and

- $r = 1$ (one state constraint)
- there exist $\beta > 0$ and $\gamma > 0$ s.t., whenever $|h(x)| \leq \beta$, then

$$\inf_{u \in U(t)} \nabla h(x) \cdot f(t, x, u) < -\gamma \quad \text{(unif. "inward pointing")}.$$

Then, for any pair $(\hat{x}(.), \hat{u}(.))$ s.t. $\hat{x}(S) \in A$, there exists an *admissible* pair $(x(.), u(.))$ such that $x(S) = \hat{x}(S)$ and

$$||\hat{x}(.) - x(.)||_{W^{1,1}} \leq K \times \rho(\hat{x}(.))$$

(K does not depend on $\hat{x}(.)$)

Rmk: it is a $W^{1,1}$ estimate, **linear** w.r.t. $\rho(\hat{x}(.))$.
(This linear estimate is also valid with the '**Ekeland metric**'.)

Rmk: $W^{1,1}$ distance estimates $\Longrightarrow$ L^∞ distance estimates

(cf. [Bettiol, Bressan, Vinter, SICON 2010], [Bettiol, Vinter, IEEE TAC 2011])

Proof (in a simple case) - stronger metrics

Consider a smooth simple case ($[S, T] = [0, 1]$):

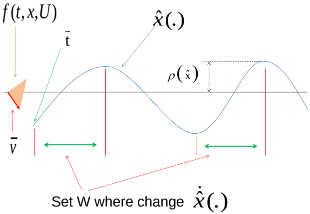
$$\begin{cases} \dot{x}(t) = u(t) \quad u(t) \in U & \text{for a.e. } t \in [0,1] \\ x(t) \in {\color{red}A} \quad \text{for all } t \in [0,1] & {\color{red}\text{state constraint}} \end{cases}$$

where $U \subset \mathbb{R}^n$ **bounded**, $b \in \mathbb{R}^n$, and $A = \{x \in \mathbb{R}^n \ : \ b \cdot x \leq 0\}$.

$$\exists\ \bar{\epsilon} > 0 \text{ and } \bar{v} \in U \text{ s.t. } b \cdot \bar{v} = -\bar{\epsilon}\,., \quad (\text{“inward pointing”})$$

Define

$$\bar{t} := \inf\{t \in [0,1] \,|\, b \cdot \hat{x}(t) > 0\}, \ W := \{t \in [\bar{t}, 1] \,|\, b \cdot \dot{\hat{x}} > 0\}$$



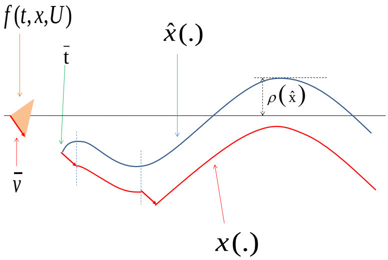
Take $\delta > 0$ the minimum number s.t.

$$\text{meas}\{W \cap [\bar{t}, \bar{t} + \delta]\} = (1/\bar{\epsilon}) \times \rho(\hat{x}(.))$$

if $meas\{W \cap [\bar{t}, \bar{t} + \delta]\} \geq (1/\bar{\epsilon})\rho(\hat{x}(.))$. Otherwise set $\bar{t} = 1$.
Now choose the trajectory $x(.)$ satisfying $x(0) = \hat{x}(0)$ and

admissible trajectory $x(.)$:

$$\dot{x}(t) = \begin{cases} \bar{v} & \text{for } t \in [\bar{t}, \bar{t} + \delta] \cap W \\ \dot{\hat{x}}(t) & \text{otherwise} \end{cases}$$



Proof: continued

Then $(x(.), u(.))$ is a process on $[0, 1]$ such that $x(0) = \hat{x}(0)$,

$$\text{meas}\{t \ : \ \dot{x}(t) \neq \dot{\hat{x}}(t)\} \ \leq \ (1/\bar{\epsilon}) \times \rho(\hat{x}(.))$$

$(\rho(\hat{x}(.)) = max_{t\in[0,1]}\{b \cdot \hat{x}(t) \vee 0\}\ )$
and for all $t \in [0, 1]$,

$$b \cdot x(t) = b \cdot \hat{x}(t) + \int_{W\cap[\bar{t},\bar{t}+\delta]} b \cdot \dot{x}(t) - \int_{W\cap[\bar{t},\bar{t}+\delta]} b \cdot \dot{\hat{x}}(t) \leq 0$$

and since U is bounded

$$||\hat{x}(.) - x(.)||_{W^{1,1}} \ \leq \ K \times \rho(\hat{x}(.)), \quad K = (\sup_{v\in U} |v|/\bar{\epsilon})$$

Bur also

$$d_{\mathcal{E}}((\hat{x}, \hat{u}), (x, u)) \ \leq \ (1/\bar{\epsilon}) \times \rho(\hat{x}(.))\ .$$

□

Contributions to This Area - 2nd (partial) list...

- **Linear Estimates with $W^{1,1}$ and Ekeland metric for one ('smooth') state constraint with 'standing hypotheses' + counter-examples**:

  P. Bettiol, A. Bressan and R. Vinter, *SIAM J. Control and Optim.* 2010.

  P. Bettiol and R. Vinter, *IEEE TAC* 2011.

- **Linear and Superlinear-$\rho \times |\log(\rho)|$, $W^{1,1}$ Estimates in the Cone, for constant set of velocities** (here the **strict convexity** has an important role in establishing **linear** estimates):

  P. Bettiol, A. Bressan and R. Vinter, *SIAM J. Control and Optim.*, 2011.

- **Linear $W^{1,1}$ Estimates for closed state constraint with 'standing hypothesis' + stronger inward pointing conditions**:

  H. Frankowska and M. Mazzola, Calculus Var. Partial Differ. Equ., 2013

  H. Frankowska and M. Mazzola, Nonlinear Differ. Equ. Appl., 2013.

- **Superlinear-$\rho \times |\log(\rho)|$, $W^{1,1}$ + Linear L^∞ Estimates in a convex set (convexity argument)**

  A. Bressan and G. Facchi, *J. Differential Eq.*, 2011

  J. Bernis, P. Bettiol, R. Vinter, J. Differential Eq., 2022.

- **Linear Estimates L^∞-estimate, $t \mapsto f(t, x, U(t))$ is absolutely continuous, has bonded variation (time-delay argument)**

  P. Bettiol, H. Frankowska and R. Vinter, *J. Diff. Eq.*, 2012

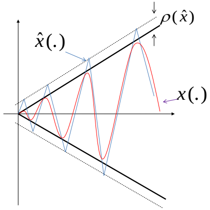
  P. Bettiol and R. Vinter, Math. Program., Ser. B, 2018

- **$\sqrt{\rho}$-Estimates (higher order inward pointing condition)**

  G. Colombo, N. Khalil, F. Rampazzo, *SIAM J. Control and Optim.* 2022

- ...

Counter-Example for L^∞ Estimates



- $f(t, x, U(t)) = U(t)$ is closed convex valued + 'inward' pointing condition
- $A = \{(x_1, x_2, x_3) \,|\, |x_2| \leq x_1\}$

But, **for any** $K > 0$ and $\varepsilon > 0$, there exists a process $(\hat{x}(.), \hat{u}(.))$ such that $\varepsilon > \rho(\hat{x}(.)) > 0$ and

$$||\hat{x}(.) - x(.)||_{L^\infty} \geq K \times \rho(\hat{x}(.)) \,.$$

In this counter-example $t \rightsquigarrow f(t, x, U(t))$ is **discontinuous**.
Ref.: [Bettiol, Bressan and Vinter, SICON 2010]

A positive answer for arbitrary closed sets

Theorem ("Linear" L^∞-estimates for arbitrary closed sets)
Assume *standing hypotheses* and

- $t \mapsto f(t, x, U(t))$ has **bounded variation**, uniformly over a neighbourhood of ∂A.
- For each $(t, x) \in [S, T] \times \partial A$,

$$\mathrm{co}\, f(t, x, U(t)) \cap \mathrm{int}\, T_A(x) \neq \emptyset, \quad (\text{"inward pointing"}) .$$

Then, for any pair $(\hat{x}(.), \hat{u}(.))$ s.t. $\hat{x}(S) \in A$, there exists a (strictly) *admissible* process $(x(.), u(.))$ such that $x(S) = \hat{x}(S)$ and

$$||\hat{x}(.) - x(.)||_{L^\infty} \leq K \times \rho(\hat{x}(.)) .$$

(**Bettiol and Vinter, Math Prog. 2018**)

Rmk: This allows data when the time-dependence is governed by a fractional power modulus of absolute continuity. $\Rightarrow$ can apply Maximum Principle in the **normal** form.

Definition. (Bounded variation)
$t \rightsquigarrow F(t,x)(= f(t,x,U(t)))$ has bounded variation uniformly over $x \in X_0 \subset \mathbb{R}^n$ if there exists a **non-decreasing bounded variation function** $\eta : [S,T] \to \mathbb{R}$ (called a 'modulus of variation of $F(.,x)$') such that, for every $[s,t] \subset [S,T]$ and $x \in X_0$,

$$d_H(F(s,x), F(t,x)) \leq \eta(t) - \eta(s).$$

$d_H(A,B)$ is the Hausdorff distance between two arbitrary non-empty closed sets in $\mathbb{R}^n$ A and B:

$$d_H(A,B) := \max\left\{\sup_{a\in A} d_B(a), \sup_{b\in B} d_A(b)\right\}.$$

Bounded variation multifunctions

$t \rightsquigarrow F(t,x)(= f(t,x,U(t)))$ has bounded variation if for every $[s,t] \subset [S,T]$ and $x \in X_0$,

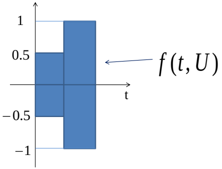
$$d_H(F(s,x), F(t,x)) \leq \eta(t) - \eta(s).$$

Example. Consider the control system for $t \in [0,1]$

$$\begin{cases} \dot{x}(t) &= b(t)u(t) \text{ a.e.} \\ u(t) \in U = [-1,1] \end{cases}$$

where

$$b(t) = \begin{cases} 0.5 & \text{if} \quad t \in [0,0.5] \\ 1 & \text{if} \quad t \in (0.5,1] \, . \end{cases}$$



$t \rightsquigarrow f(t,U)$ has boun. var. (It is **discontinuous**.)

Comments and further developments

Rmk. Examples show that if ('coupled') hypotheses (Bounded Variation)-(Inward Ponting Condition) are not satisfied, than we might have '**very bad**' behaviour of distance estimates (cf. examples in [Bettiol and Vinter, Math Prog. 2018]).

Rmk. In **Differential Games** theory, one can define two value functions for the game via **non-anticipative strategies** (or Varayia-Roxin-Elliot-Kalton strategies).

Distance estimates constructs can be used to build up non-anticipative strategies, obtaining linear/super-estimates w.r.t. Ekeland/$W^{1,1}/L^\infty$ metrics.

It follows that (under appropriate assumptions) the (lower/upper) **value function is Lipschitz continuous**.

Dynamic Programming – State Constraints

$$P(\tau,\xi)\left\{\begin{array}{l}\text{Minimize } g(x(T))\\ \text{over admissible processes } (x,u) \text{ s.t. } x(\tau)=\xi.\end{array}\right.$$

$\rightarrow g:\mathbb{R}^n\rightarrow\mathbb{R}\cup\{+\infty\}$ is **extended valued**; incorporates an implicit **terminal constraint**

$$x(T)\in C\,,$$

where $C:=\{x\in\mathbb{R}^n\mid g(x)<+\infty\}$ is a closed set.

$\Rightarrow$ It is necessary to consider lower semicontinuous solutions (lsc) to (HJ)
$\rightarrow$ we impose the condition in addition to the 'standing hypotheses':

($*$) *the multifunction $(t,x)\rightsquigarrow f(t,x,U(t))$ is* ***convex*** *and (to simplify)* ***continuous***

Dynamic Programming – State Constraints

$$P(\tau,\xi)\begin{cases}\text{Minimize } g(x(T))\\ \text{over admissible processes } (x,u) \text{ s.t. } x(\tau)=\xi.\end{cases}$$

Define $\boxed{V(\tau,\xi) := \text{Inf}(P(\tau,\xi))}$ **Value Function**

The goal: represent the value function as the unique solution, appropriately defined, of the (HJ). Various, equivalent, definitions of 'solution' of (HJ) are involved: **Dini solution**, **proximal solution** (of Clarke), viscosity solution.

Two different classical paths:

- **viscosity solutions**: it is possible to show directly (without consideration of state trajectories) that the Hamilton Jacobi equation has a unique solution.
- **system theoretic**: it is intimately connected with properties of state trajectories; invariance (viability) theorems are employed to show that a solution to the Hamilton Jacobi equation provides a lower bound to the cost of an arbitrary state trajectory and this lower bound is achieved by some state trajectory. (**Nonsmooth Analysis**)

Theorem (Characterization of Value Functions for State Constrained Problems (I): Outward-Pointing Condition)
Assume the 'standing hypotheses'. Suppose in addition that

$(\text{CQ})_{\text{outward}}$: for each $s \in [S, T)$, $t \in (S, T]$ and $x \in \partial A$,

$$f(t, x, U(t)) \cap \big(- \operatorname{int} T_A(x) \big) \neq \emptyset$$

Take a function $V : [S, T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Then assertions (a)–(c) below are equivalent:

(a) V is the value function for (SC).

(b) V is lsc on $[S, T] \times \mathbb{R}^n$, $V(t, x) = +\infty$ if $x \notin A$, and

(i) for all $(t, x) \in ([S, T) \times A) \cap \operatorname{dom} V$

$$\inf_{u \in U(t)} D_{\uparrow} V((t, x); (1, f(t, x, u))) \leq 0,$$

(ii) for all $(t, x) \in ((S, T] \times \operatorname{int} A) \cap \operatorname{dom} V$

$$\sup_{u \in U(t)} D_{\uparrow} V((t, x); (-1, -f(t, x, u))) \leq 0,$$

(iii) for all $x \in A$

$$\liminf_{\{(t',x') \to (T,x) : t' < T, x' \in \operatorname{int} A\}} V(t', x') = V(T, x) = g(x).$$

The lower Dini (directional) derivative

Definition.
Take a function $\varphi : \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$, a point $x \in \operatorname{dom} \varphi$ and a vector $d \in \mathbb{R}^k$. The lower Dini (directional) derivative of φ at x in the direction $d \in \mathbb{R}^k$ is defined to be:

$$D_{\uparrow}\varphi(x; d) \; := \; \liminf_{h \downarrow 0,\, e \to d} \; h^{-1} \left[\varphi(x + he) - \varphi(x) \right] \, .$$

(c) V is lsc on $[S, T] \times \mathbb{R}^n$, $V(t, x) = +\infty$ if $x \notin A$, and

(i) for all $(t, x) \in ((S, T) \times A) \cap \mathrm{dom}\, V$, $(\xi^0, \xi^1) \in \partial_P V(t, x)$

$$\xi^0 + \inf_{u \in U(t)} \xi^1 \cdot f(t, x, u) \leq 0,$$

(ii) $(t, x) \in ((S, T) \times \mathrm{int}\, A) \cap \mathrm{dom}\, V$, $(\xi^0, \xi^1) \in \partial_P V(t, x)$

$$\xi^0 + \inf_{u \in U(t)} \xi^1 \cdot f(t, x, u) \geq 0,$$

(iii) for all $x \in A$,

$$\liminf_{\{(t',x') \to (S,x) : t' > S\}} V(t', x') = V(S, x)$$

and

$$\liminf_{\{(t',x') \to (T,x) : t' < T,\ x' \in \mathrm{int}\, A\}} V(t', x') = V(T, x) = g(x).$$

Theorem (Characterization of Value Functions for State Constrained Problems (II): Inward-Pointing Condition)
Assume the 'standing hypotheses'. Suppose in addition that $g(.)$ is continuous on A and

$(CQ)_{\text{inward}}$: for each $s \in [S, T)$, $t \in (S, T]$ and $x \in \partial A$,

$$f(t, x, U(t)) \cap \operatorname{int} T_A(x) \neq \emptyset$$

Take a function $V : [S, T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Then assertions (a)–(c) below are equivalent:

(a) V is the value function for (SC).

(b) V is lsc on $[S, T] \times \mathbb{R}^n$, $V(t, x) = +\infty$ if $x \notin A$, and

- **(i)** for all $(t, x) \in ([S, T) \times A) \cap \operatorname{dom} V$

$$\inf_{u \in U(t)} D_{\uparrow} V((t, x); (1, f(t, x, u))) \leq 0,$$

- **(ii)** for all $(t, x) \in ((S, T] \times \operatorname{int} A) \cap \operatorname{dom} V$

$$\sup_{u \in U(t)} D_{\uparrow} V((t, x); (-1, -f(t, x, u))) \leq 0,$$

- **(iii)** for all $x \in A$, $\quad V(T, x) = g(x)$.

(c) V is lsc on $[S,T]\times\mathbb{R}^n$, $V(t,x)=+\infty$ if $x\notin A$, and

(i) for all $(t,x)\in((S,T)\times A)\cap\operatorname{dom}V$, $(\xi^0,\xi^1)\in\partial_P V(t,x)$

$$\xi^0+\inf_{u\in U(t)}\xi^1\cdot f(t,x,u)\ \le\ 0,$$

(ii) $(t,x)\in((S,T)\times\operatorname{int}A)\cap\operatorname{dom}V$, $(\xi^0,\xi^1)\in\partial_P V(t,x)$

$$\xi^0+\inf_{u\in U(t)}\xi^1\cdot f(t,x,u)\ge 0,$$

(iii) for all $x\in A$,

$$\liminf_{\{(t',x')\to(S,x):t'>S\}} V(t',x')=V(S,x)$$

and

$$V(T,x)=g(x).$$

Theorem (Viscosity solution characterization of Value Functions for State Constrained Problems - Inward/Outward-pointing Condition)

Assume the 'standing hypotheses' and, in addition, $(CQ)_{\text{outward}}$ and $(CQ)_{\text{inward}}$, and that $g_{|A}$ is locally bounded and satisfies $((g_{|A})^*)_* = g_{|A}$. Take a lower semicontinuous, locally bounded function $V : [S, T] \times \mathbb{R}^n \to \mathbb{R}$ such that $V(t, x) = +\infty$ when $x \notin A$.

Then V is the value function for (SC) if and only if V is a locally bounded function on $[S, T] \times A$, lower semicontinuous constrained viscosity solution of (HJ).

W_* and W^* (referred to as the upper envelope and the lower envelope of W, respectively) are the functions:

$$W^*(y) := \limsup_{y' \to y} W(y') \text{ and } W_*(y) := \liminf_{y' \to y} W(y') .$$

Constrained viscosity solution of (HJ):

(i) (V is a viscosity supersolution) for any point $(t,x) \in (S,T) \times A$ and any C^1 function $\psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ such that

$$(t',x') \to V(t',x') - \psi(t',x')$$

has a local minimum at (t,x) (relative to $[S,T] \times A$) we have

$$-\psi_t(t,x) + H(t,x,-\psi_x(t,x)) \geq 0\,,$$

(ii) (V is a viscosity subsolution) for any point $(t,x) \in (S,T) \times \operatorname{int} A$ and any C^1 function $\psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ such that

$$(t',x') \to V^*(t',x') - \psi(t',x')$$

has a local maximum at (t,x) (relative to $[S,T] \times A$) we have

$$-\psi_t(t,x) + H(t,x,-\psi_x(t,x)) \leq 0\,,$$

(iii) for all $x \in A$

$$\liminf_{\{(t',x')\to(S,x) | t'>S\}} V(t',x') = V(S,x),$$

$$(V_{|[S,T]\times A})^*(T,x) = (g_{|A})^*(x) \quad \text{and} \quad V(T,x) = g(x).$$

H is, as usual, the **Hamiltonian** function

$$H(t,x,p) := \max_{u\in U(t)} p \cdot f(t,x,u)\,.$$

The solution to the Growth/Consumption problem

Techniques of **dynamic programming** provide.
The **state feedback** function $\chi : [0, T] \times (0, \infty) \to [0, 1]$:

$$\chi(t, x) := \begin{cases} 0 & \text{if } x > \bar{y}(t) \\ 1 & \text{if } x < \bar{y}(t) \\ \alpha & \text{if } x = \bar{y}(t) \text{ and } t \leq T - \Delta \\ 0 & \text{if } x = \bar{y}(t) \text{ and } t > T - \Delta \end{cases}$$

in which $\bar{y} : (-\infty, T] \to (0, \infty)$ is the function

$$\bar{y}(t) := \begin{cases} \hat{x} & \text{if } t \leq T - \Delta \\ \left[\frac{b}{a}(1 - e^{-a\alpha(T-t)}\right]^{\frac{1}{1-\alpha}} & \text{if } t > T - \Delta \end{cases}$$

$$\hat{x} := \left(\frac{\alpha b}{a}\right)^{\frac{1}{1-\alpha}} \text{ and } \Delta := \frac{1}{a\alpha} \ln\left(\frac{1}{1-\alpha}\right)$$

The solution to the Growth/Consumption pb...

Given initial data $(t_0, x_0) \in [0, T] \times (0, \infty)$, the **optimal output** x^* is the unique solution in the space of Lipschitz continuous functions on $[t_0, T]$ of the differential equation

$$\begin{cases} \dot{x}^*(t) = -ax^*(t) + b(x^*)^\alpha(t)\chi(t, x^*(t)) \text{ a.e. } t \in [t_0, T], \\ x(t_0) = x_0 \, . \end{cases}$$

The optimal proportion of financial return for investment u^* is unique (w.r.t. the equivalence class of almost everywhere equal functions) and is given by

$$u^*(t) = \chi(t, x^*(t)), \text{ for a.e. } t \in [t_0, T] \, .$$

Rmk: the solution is expressed in **state feedback** form: the optimal control u^* is expressed as a function of the current state.

The solution to the Growth/Consumption pb...

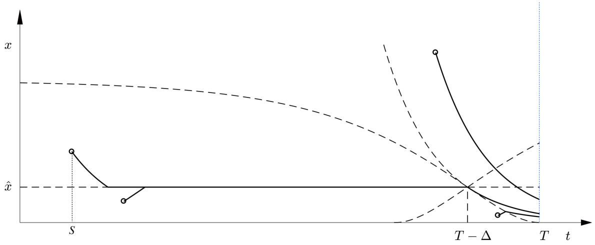


Figure: Optimal Trajectories for the Consumption/Growth Problem

References:

→ K. Miao and R. Vinter, OCAM 2021 (solution of the problem) see also for the state constrained (HJ) eq. solution interpretation:

J. Bernis, P. Bettiol, R. Vinter, JDE 2022

J. Bernis and P. Bettiol, JCA 2023

Proximal solution

Write $V : [0, T] \times (0, \infty) \to \mathbb{R}$ the value function for (GC).
Let $\psi : [0, \infty) \to [0, \infty)$ be the mapping

$$\psi(x) := x^{1-\alpha} \text{ for } x \in [0, \infty).$$

Then

$$V(t, x) = (W \circ (Id, \psi))(t, x), \quad \text{for all } (t, x) \in [0, T] \times [0, \infty),$$

where $W : [0, T] \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is the unique upper semicontinuous function s.t. $W(t, y) = -\infty$ whenever $y < 0$,

(i) for all $(t, y) \in (0, T) \times [0, \infty)$, $(\xi^0, \xi^1) \in \partial^P W(t, y)$

$$\xi^0 + \sup_{u \in [0,1]} \left(\xi^1 \cdot (-a(1-\alpha)y + (1-\alpha)bu) + (1-u)y^{\frac{\alpha}{1-\alpha}} \right) \geq 0;$$

(ii) for all $(t, y) \in (0, T) \times (0, \infty)$, $(\xi^0, \xi^1) \in \partial^P W(t, y)$

$$\xi^0 + \sup_{u \in [0,1]} \left(\xi^1 \cdot (-a(1-\alpha)y + (1-\alpha)bu) + (1-u)y^{\frac{\alpha}{1-\alpha}} \right) \leq 0;$$

$\partial^P W(t, y) = -\partial_P(-W)(t, y)$: **proximal superdifferential** of W

Proximal solution...

(iii) for all $y \in [0, \infty)$

$$\limsup_{\{(t',y')\to(0,y):t'>0\}} W(t',y') = W(0,y)$$

and

$$\limsup_{\{(t',y')\to(T,x):t'<T,\ y'>0\}} W(t',y') = W(T,y) = 0.$$

Viscosity solution

$W : [0, T] \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is the unique upper semicontinuous function such that W is continuous on $[0, T] \times [0, \infty)$, $W(t, y) = -\infty$ whenever $y < 0$ and

(i) for all $(t, y) \in (0, T) \times [0, \infty)$, $(\xi^0, \xi^1) \in \partial_+ W(t, y)$

$$\xi^0 + \sup_{u \in [0,1]} \left(\xi^1 \cdot (-a(1-\alpha)y + (1-\alpha)bu) + (1-u)y^{\frac{\alpha}{1-\alpha}} \right) \geq 0;$$

(ii) for all $(t, y) \in (0, T) \times (0, \infty)$, $(\xi^0, \xi^1) \in \partial_- W(t, y)$

$$\xi^0 + \sup_{u \in [0,1]} \left(\xi^1 \cdot (-a(1-\alpha)y + (1-\alpha)bu) + (1-u)y^{\frac{\alpha}{1-\alpha}} \right) \leq 0;$$

(iii) for all $y \in [0, \infty)$

$$\limsup_{\{(t',y') \to (0,y),\, t' > 0\}} W(t', y') = W(0, y)$$

and

$$W(T, y) = 0.$$

The Fréchet subdifferential

The **Fréchet subdifferential** (also called *strict subdifferential*) of φ at $\bar{x} \in \operatorname{dom}\varphi$ is defined by

$$\partial_{-}\varphi(\bar{x}) := \{\xi \,|\, (\xi, -1) \in \hat{N}_{\text{epi }\varphi}(\bar{x}, \varphi(\bar{x}))\}.$$

We recall also that, if $\varphi : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ is an upper semicontinuous function and $\bar{x} \in \operatorname{dom}\varphi$, then the *Fréchet superdifferential* of φ at $\bar{x}$ is defined as $\partial_{+}\varphi(\bar{x}) := -\partial_{-}(-\varphi)(\bar{x})$.

$$\hat{N}_C(x) \;:=\; \left\{ \xi \in \mathbb{R}^m \,|\, \limsup_{y \xrightarrow{C} x} |y-x|^{-1}\, \xi \cdot (y-x) \;\leq 0 \right\}.$$

Well known properties are:
$\hat{N}_C(x) = \{\xi \in \mathbb{R}^m \,|\, \xi \cdot v \leq 0, \forall v \in T_C(x)\}$ (i.e. $\hat{N}_C(x)$ is the polar cone to $T_C(x)$) and

$$N^P_C(x) \subset \hat{N}_C(x).$$

Sensitivity Results with state constraints

Theorem

Assume that $A = \{h(x) \leq 0\}$, $h \in \mathcal{C}^{1+}$, and 'standing hypotheses'.

Let $(\bar{x}, \bar{u})$ be a minimizer for problem (SC). Then there exists a function of bounded variation q, right continuous on (S, T), and a Radon measure μ on $[S, T]$ s.t.

- **(i):** the conditions of the state constrained Maximum Principle are satisfied
- **(ii):** $(\mathcal{H}(t, \bar{x}(t), q(t)), -q(t)) \in \partial^0 V(t, \bar{x}(t))$ **a.e.** $[S, T]$
- **(iii):** $p(S) \in \partial_x (-V)^+(S, \bar{x}(S))$

Notation: $(-V)^+(.,.)$ is the extended valued function on $\mathbb{R} \times \mathbb{R}^n$

$$(-V)^+(t,x) \ := \ \begin{cases} -V(t,x) & \text{if } t \in [S,T] \text{ and } x \in A \\ +\infty & \text{otherwise .} \end{cases}$$

$\partial^0 V$ is the 'hybrid' (from the interior) subdifferential:

$$\partial^0 V(t,x) \quad := \quad \text{co}\ \limsup \left\{ \partial V(t',x') \mid (t',x') \xrightarrow{A^0} (t,x) \right\} \ ,$$

$$A^0 \ := \ \{x \mid h(x) < 0\} \ .$$

Theorem

Assume A is (nonempty) closed, 'standing assumptions' on $\dot{x} \in F$, $(CQ)_{\text{inward}}$ and F is BV w.r.t. t. Then $V(.,.)$ is locally Lipschitz continuous on $[S,T] \times A$.

Then there exists $p(.) \in W^{1,1}([S,T];\mathbb{R}^n)$ and a function of bounded variation $\eta(.) : [S,T] \to \mathbb{R}^n$, continuous from the right on (S,T), such that

(i): for some finite positive Borel measure μ on $[S,T]$ and Borel measurable selection

$$\gamma(t) \in (\overline{\text{co}}\, N_A(\bar{x}(t))) \cap \mathbb{B} \qquad \mu - \text{a.e.} \quad t \in [S,T]$$

we have

$$\eta(t) = \int_{[S,t]} \gamma(s) d\mu(s), \qquad \text{for all} \quad t \in (S,T]\,,$$

(ii): $\dot{p}(t) \in \text{co}\{r : (r, q(t)) \in N_{\text{Gr}\{F(t,.)\}}(\bar{x}(t), \dot{\bar{x}}(t))\}$ a.e.

(iii): $-q(T) \in \partial g(\bar{x}(T))$, $q(S) \in \partial(-V)^+(S, \bar{x}(S))$ and

(iv): $q(t) \cdot \dot{\bar{x}}(t) = \max_{v \in F(t,\bar{x}(t))} q(t) \cdot v$ a.e.,

where $q(t) := p(t) + \eta(t)$ for $t \in (S,T]$.

Theorem (continue...)

Furthermore $p(.)$ **and** $\eta(.)$ **can be chosen also to satisfy the 'partial and the full sensitivity relations':**

(v) $-q(t) \in \partial_x^0 V(t, \bar{x}(t))$ a.e. $t \in (S, T]$,

where, for $(t, x) \in [S, T] \times A$

$$\partial_x^0 V(t,x) := \cap_{\epsilon>0} \overline{\mathrm{co}} \cup_{\{x' \in (x+\epsilon\mathbb{B}) \cap \mathrm{int}\, A\}} \partial V(t,x') ;$$

vi) $(H(t, \bar{x}(t), q(t)), -q(t)) \in \partial^0 V(t, \bar{x}(t))$ a.e. $t \in (S, T]$,

where, for $(t, x) \in [S, T] \times A$

$$\partial^0 V(t,x) :=$$
$$\cap_{\epsilon>0} \overline{\mathrm{co}} \cup_{\{(t',x') \in ((t,x)+\epsilon\mathbb{B}) \cap [S,T]\times \mathrm{int}\, A\}} \partial V(t',x') .$$

Exercises

Ex 1. Let V be the value function for (SC). Assume that the 'standing hypotheses' are satisfied, and that $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous.

(a) Then $V(t,x) > -\infty$ for all $(t,x) \in [S,T] \times \mathbb{R}^n$.

(b) If in addition $f(t,x,U(t))$ takes convex values, then V is lower semicontinuous and $V(t,x) > -\infty$ for all $(t,x) \in [S,T] \times \mathbb{R}^n$.

(c) If in addition to the 'standing hypotheses' also hypotheses (BV) and (Inward Pointing) are satisfied and g is locally Lipschitz continuous on A (resp. continuous on A), then V is locally Lipschitz continuous on $[S,T] \times A$ (resp. continuous on $[S,T] \times A$).

Exercises

Ex 2 (State constrained maximum principle in Gamkrelidze form.) Let $(\bar{x}, \bar{u})$ be a minimizer for the state constrained problem

$$\begin{cases} \text{Minimize } g(x(S), x(T)) \\ \text{subject to } \dot{x}(t) = f(x(t), u(t)), u(t) \in U \text{ a.e.} \\ h(x(t)) \leq 0 \text{ for all } t \in [S, T] \\ (x(S), x(T)) \in C\,. \end{cases}$$

with data functions $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}$ and sets $U \subset \mathbb{R}^m$ and $C \subset \mathbb{R}^n \times \mathbb{R}^n$.

Exercises

Assume that standing hypotheses are satisfied. Assume further that g is C^1, $f(.,\bar{u}(t))$ is C^1 a.e. and h is C^2. Show that there exist $p \in W^{1,1}([S,T];\mathbb{R}^n)$, a BOrel measure μ on $[S,T]$ and $\lambda \geq 0$ such that

(i)): $(p,\mu,\lambda) \neq (0,0,0)$,

(ii): $-\dot{p}(t) = \big(p(t) + \int_{[S,t]} d\mu(s) h_x(\bar{x}(t))\big) \cdot f_x(\bar{x}(t),\bar{u}(t)) + \int_{[S,t]} d\mu(s) h_{xx}(\bar{x}(t)) \cdot f(\bar{x}(t),\bar{u}(t))$,

(iii): $u \to (p(t) + \int_{[S,t]} d\mu(s) h_x(\bar{x}(t))) \cdot f(\bar{x}(t),u)$ is maximized over U at $u = \bar{u}(t)$. a.e.,

(iv): supp $\{\mu\} \subset \{t \,:\, h(\bar{x}(t)) = 0\}$,

(v): $(p(S), -(p(T) + \int_{[S,T]} d\mu(t) h_x(\bar{x}(T))) = \lambda \nabla g(\bar{x}(S),\bar{x}(T)) + N_C(\bar{x}(S),\bar{x}(T))$.

Some References

- Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, SIAM, 1990.
- Clarke, Ledyaev, Stern, Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, Graduate Texts in Mathematics vol. 178, 1998.
- Vinter, *Optimal Control*, Birkhäuser, Boston, 2010.
- Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer, Graduate Texts in Mathematics 264, 2013.
- Bettiol and Vinter, *Principles of Dynamic Optimization*, Springer, Monographs in Mathematics, to appear.