Optimal Control

Wintersemester 2025/26

Dozent: Prof. Dr. Andrea Iannelli

November 24, 2025

Contents

0 Introduction

$$\dot{x} = f(t, x, u), \quad x(t_0) = x_0, \quad t \in [t_0, t_f]$$

$$f : [t_0, t_f] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$$

$$x = \text{state}, \quad u = \text{input}$$

Initial Value Problem (IVP)

Given $x_0, u(\cdot)$ we can compute $x(\cdot)$
functions of time

When is this possible? It depends on f.

Lemma 0.1 (Sufficient conditions)

Existence & Uniqueness of solutions of ODEs.
Assume that

- f is piecewise continuous in t and u
- f is globally Lipschitz in x

$$\exists\, k(t,u) \text{ s.t. } \|f(t,x_1,u) - f(t,x_2,u)\| \leq k(t,u)\|x_1 - x_2\|, \ \forall x_1, x_2 \in \mathbb{R}^{n_x}$$

Then $x(\cdot)$ exists for all t and is unique.

Remarks

- Lipschitz continuous $\Rightarrow$ continuous, but not the converse
- $\sqrt{x}$ is continuous but not Lipschitz, $\dot{x} = \sqrt{x}$ does not have a unique solution
- Continously differentiable ($\mathcal{C}^1$) $\Rightarrow$ locally Lipschitz continous $\forall x_1, x_2 \in \mathcal{X} \subset \mathbb{R}^{n_x}$
- Locally Lipschitz continuous x guarantees existence & uniqueness for small enough times

In this course we will assume $f \in \mathcal{C}^1$ and implicitely assume that t_f is chosen such that $x(\cdot)$ exists in $[t_0, t_f]$.

We do not need to worry about existence & uniqueness!

Goal in Optimal Control: Design u such that

1. $u(t) \in \underset{\uparrow}{\mathcal{U}}(t)$, $x(t) \in \underset{\uparrow}{\mathcal{X}}(t) \quad \forall t \in [t_0, t_f], \quad \mathcal{X} \subseteq \mathbb{R}^{n_x},\ \mathcal{U} \subseteq \mathbb{R}^{n_u}$
   sets defining constraints on $u \& x$
   $\Rightarrow$ Admissible input/state trajectories

2. The system behaves optimally according to

$$\underset{\uparrow}{J}(u) = \int_{t_0}^{t_f} \underset{\uparrow}{l}(t, x(t), u(t))\, dt + \underset{\uparrow}{\varphi}(t_f, x(t_f))$$

   Cost function running cost terminal cost
   $\Rightarrow$ optimal behaviour

Formally, we can state the goal as follows:
Find an admissible input $u^\star$ which causes the dynamics to follow an admissible trajectory $x^\star$ which minimizes J, that is

$$\int_{t_0}^{t_f} l(t, x^\star(t), u^\star(t))\, dt + \varphi(t_f, x^\star(t_f)) \leq \int_{t_0}^{t_f} l(t, x(t), u(t))\, dt + \varphi(t_f, x(t_f))$$

$$\forall \text{ admissible } x, u$$

Examples of cost functions

1) Minimum-time problem
   Goal: transfer the system from x_0 to a set $\mathcal{S}$ in the minimum time

$$J = t_f - t_0 = \int_{t_0}^{t_f} dt \qquad (l = 1, \varphi = 0)$$

$$x(t_f) \in \mathcal{S}$$

   Note: t_f is also a decision variable! The unknowns are (u, t_f).

2) Minimum control-effort problem

$$J = \int_{t_0}^{t_f} \|u(t)\|^2\, dt$$

$$x(t_f) \in \mathcal{S}$$

3) Tracking problem

$$J = \int_{t_0}^{t_f} (x(t) - r(t))^T Q (x(t) - r(t))\, dt$$

$Q > 0$ (positive definit matrix: symmetric & all eigenvalues positive)
$r(t)$ given signal

1 Nonlinear Programming

Nonlinear Programs (NLP) are general finite-dimensional optimization problems:

$$\min_x f(x)$$

$$\text{s.t. } g(x) \leq 0, \quad h(x) = 0$$

$f : \mathbb{R}^n \to \mathbb{R}$, objective function
$g : \mathbb{R}^n \to \mathbb{R}^{n_g}$, inequality constraints
$h : \mathbb{R}^n \to \mathbb{R}^{n_h}$, equality constraints
Feasible set:

$$D = \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0\}$$

$\bar{x} \in D$ feasible point

Definition 1.1 (Global, local Minimizers)

$x^\star \in \mathcal{D}$ Global Minimizer of the NLP if

$$f(x^\star) \leq f(x) \quad \forall x \in \mathcal{D}$$

$f(x^\star)$ is the Global Minimum (or Minimum)
Nomenclature: $x^\star$ is also called (optimal) solution, $F(x^\star)$ is optimal value
$x^\star$ is a strict global minimizer if $f(x^\star) < f(x) \quad \forall x \in \mathcal{D}$
$x^\star \in \mathcal{D}$ Local Minimizer if

$$\exists \varepsilon > 0, \text{ s.t. } f(x^\star) \leq f(x) \quad \forall x \in B_\varepsilon(x^\star) \cap \mathcal{D}$$

$$B_\varepsilon(x) := \{y \mid \|x - y\| \leq \varepsilon\} \qquad \|\cdot\| : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \text{ any norm in } \mathbb{R}^n$$

Strict local Minimizer if inequality holds strictly
Global min $\underset{\not\leftarrow}{\rightarrow}$ local min

Solving an NLP boils down to finding global or local minimizers.
Does a solution always exist? No.

Definition 1.2 (infimum)

Given $\mathcal{S} \subseteq \mathbb{R}$, $\inf(\mathcal{S})$ is the greatest lower bound of $\mathcal{S}$:

- $z \geq \inf(\mathcal{S}), \quad \forall z \in \mathcal{S} \quad$ (lower bound)
- $\forall \bar{\alpha} > \inf(\mathcal{S}) \quad \exists z \in \mathcal{S}$ s. t. $\bar{\alpha} > z \quad$ (greatest bound)

Example $\mathcal{S} = [-1, 1]$, $-50 = \inf(\mathcal{S})$? $\quad\rightarrow\quad$ No, $\inf(\mathcal{S}) = -1$

- Analogous: $\sup(\mathcal{S})$ is smallest upper bound.
- inf and sup always exist if $\mathcal{S} \neq \emptyset$
- $\inf(\mathcal{S})$ does not have to be an element of $\mathcal{S}$
- If $\mathcal{S}$ unbounded from below $\rightarrow \inf(\mathcal{S}) = -\infty$
- $\inf([a, b]) = \inf((a, b]) = a$

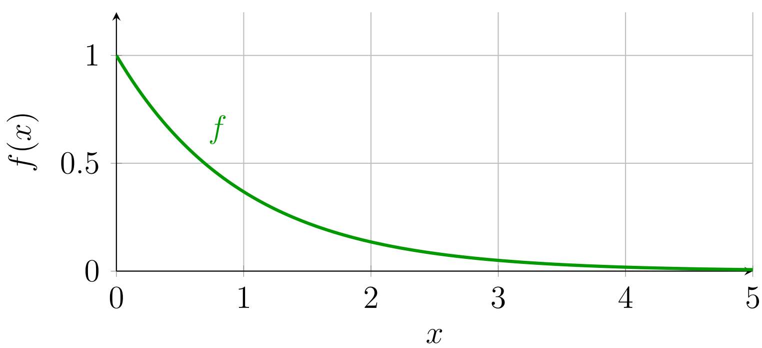
Connections with NLP?

$$f : \mathcal{D} \rightarrow \mathbb{R}$$

$$\inf(\underbrace{f(x) \,|\, x \in \mathcal{D}}_{\mathcal{S}}) := \bar{f} = \inf_{x \in \mathcal{D}} f(x) \quad \text{(similar to NLP)}$$

Whenever NLP has solution, then NLP ist equivalent to this, but $\nexists x^\star \in \mathcal{D}$ s. t. $f(x^\star) = \bar{f} \quad\rightarrow\quad$ infimum exists, but not minimum

Examples $f(x) = e^{-x}, \quad \mathcal{D} = [0, \infty), \quad \inf(\mathcal{S}) = 0$
$f(x) = x, \quad \mathcal{D} = \mathbb{R}, \quad \inf(\mathcal{S}) = -\infty$, min doesn't exist!



When does the infimum coincide with the minimum?

Theorem 1.1 (Extreme value problem) (Weierstrass Theorem)

$f : \mathcal{D} \to \mathbb{R}$, $\mathcal{D} \subseteq \mathbb{R}^n$
If:

- $f \in \mathcal{C}$ on $\mathcal{D}$
- $\mathcal{D}$ is compact
- $\mathcal{D} \neq \emptyset$

Then f attains a minimum on $\mathcal{D}$.

Definition 1.3 (Continuous function)

$f : \mathcal{D} \to \mathbb{R}$ is continuous at $x \in \mathcal{D}$ if

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ s. t. } \|x - x'\| < \delta \quad \Rightarrow \quad \|f(x) - f(x')\| < \varepsilon$$

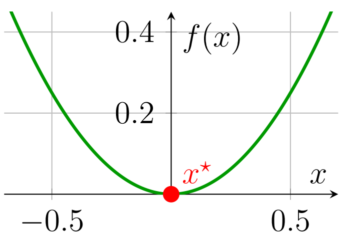
If f is continuous $\forall x \in \mathcal{D}$ then f is continuous on $\mathcal{D} \to f \in \mathcal{C}$

Implication for NLP: If f is $\mathcal{C}$ on $\mathcal{D}$ and $\mathcal{D}$ is compact and non-empty then [NLP] has a solution!

- $\mathcal{D} \subseteq \mathbb{R}^n$: in finite-dimensional spaces: compact = closed and bounded
  Not compact:
  - $(a, b]$ (not closed)
  - $(-\infty, b]$ (unbounded)

  Compact set:
  - $[a, b] \quad -\infty < a < b < \infty$

  Warning: $\mathcal{D}$ infinite dimensional (e. g. function space) then
  compact $\not\rightleftarrows$ bounded and closed



Theorem 1.1 is restrictive e. g. $f(x) = x^2, \mathcal{D} = (-\infty, \infty)$ has unique minimum

- Notation convention: Technically it is “wrong” to write

$$\min_{x \in \mathcal{D}} f(x)$$

more compact is:

$$\underset{x \in \mathcal{D}}{\text{minimize}}\ f(x) \quad \text{or} \quad \inf_{x \in \mathcal{D}} f(x)$$

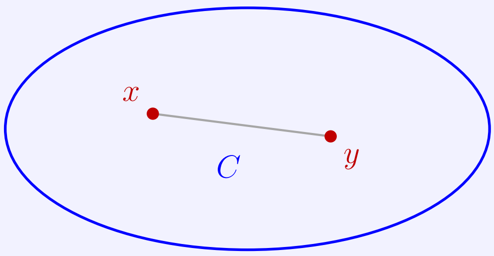
Goal of the Chapter: characterize necessary and sufficient conditions for $x^\star$ to be global minimizer of NLP.

Convexity

Definition 1.4 (Convex sets & functions)

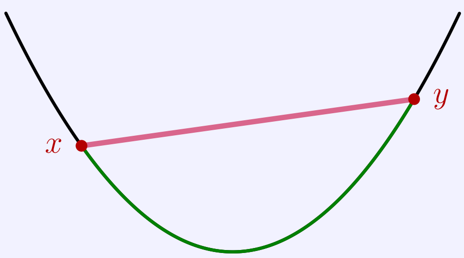
- A set $C \subseteq \mathbb{R}^n$ is <u>convex</u> (cvx) if $\forall x, y \in C$

$$\{z \mid z = \lambda x + (1 - \lambda)\lambda,\ \lambda \in [0, 1]\} \subseteq C$$



- Given a cvx set C, a function $f : C \to \mathbb{R}$ is cvx if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad \forall x, y \in C, \quad \lambda \in (0, 1)$$



- f is strictly cvx if the inequality holds strictly.

Remarks

- The definition extends to vector functions $f : C \to \mathbb{R}^n$ for convex f_i
- $f : C_1 \times C_2 \to \mathbb{R}$
  $f(x, y)$ is jointly cvx, in x, y if $z := \begin{bmatrix} x \\ y \end{bmatrix}$, $f(z)$ is in cvx in z.

Example $f(x,y) = x^2 + y^2$, $z = \begin{bmatrix} x \\ y \end{bmatrix} \to f(z) = z_1^2 + z_2^2$

Definition 1.5

An NLP is a convex program if

- f is convex function,
- $\mathcal{D}$ is convex set.

Lemma 1.1

Let $x^\star$ be a local minimizer of cvx program. Then $x^\star$ is also global minimizer.

Proof: try as an erxercise

Minimizers of convex NLP form a convex set.
This set might be empty (Convex NLPs not guaranteed to have solution).
However: Unique solution for strictly convex NLPs, if a solution exists.

Lemma 1.2 (First/Second order conditions for convexity)

1. $f : C \to \mathbb{R}$ continuously differentiable on C. Then f is cvx iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in C$$

$$(\nabla f)_i = \frac{\partial f}{\partial x_i} \text{ is gradient (sometimes } f_{x_i}\text{)}$$

2. f twice differentiable on C, then f convex iff

$$\nabla^2_{xx} f(x) \succeq 0 \quad \forall x \in C$$

$$(\nabla^2_{xx})_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad \text{(Hessian)}$$

- $A \succeq 0$ means that: $A = A^T$ and pos semi-definite, i. e. all eigenvalues non-negative
- f strictly cvx if $\nabla^2_{xx} f(x) \succ 0 \quad \forall x \in C$ with $A \succ 0$ meaning pos definite and symmetric
- Interpretation: Curvature of function should be non-negative/positive
- For exercises to check convexity, the second condition is generelly useful. First condition is useful for proofs.

For $\mathcal{D} = \{x \mid g(x) \leq 0,\, h(x) = 0\}$ the following holds: If

$$\left.\begin{array}{l} \bullet\ g \text{ are convex functions,} \\ \bullet\ h \text{ are affine functions (i.e. } h(x) = 0 \Leftrightarrow Ax = b\text{),} \end{array}\right\} \text{sufficient}$$

then $\mathcal{D}$ is a convex set.

Example $a, b \in \mathbb{R}$

$$\begin{array}{ccc} \min\limits_x f & & \min\limits_x f \\ \text{s.t. } x^3 - 1 \leq 0 & \leftrightarrow & \text{s.t. } x - 1 \leq 0 \\ (ax+b)^2 = 0 & & ax + b = 0 \\ \text{non-convex} & & \text{convex} \\ \text{non-affine} & & \text{affine} \end{array}$$

<u>Moral to recognize convexity of NLP:</u>

1. Use definition of cvx NLP, cvx f, convex $\mathcal{D}$
2. If $\mathcal{D}$ written as equality/inequality-constraints, check g convex/h affine.
   If not, check further whether the feasible set is cvx or not (e.g. can be written equivalently with cvx g/affine h).

1.1 Unconstrained Problems

$$\mathcal{D} = \mathbb{R}^n$$

Assume throughout that $f \in \mathcal{C}^1$ (continuously differentiable).

Definition 1.6 (Descent Direction)

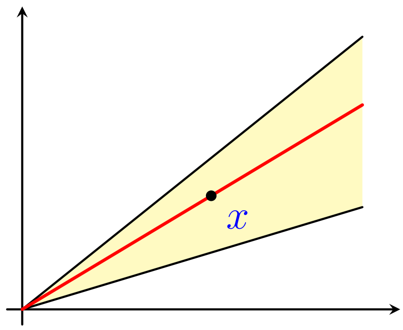
$d \in \mathbb{R}^n$ is a descent direction for f at $\bar{x} \in \mathbb{R}^n$ if

$$\exists \delta > 0 \quad \text{s.t.} \quad f(\bar{x} + \lambda d) < f(\bar{x}) \quad \forall \lambda \in (0, \delta).$$

$F(\bar{x})$: Cone of descent directions
Set of all descent directions of f at $\bar{x}$

A set $K \subseteq \mathbb{R}^n$ is a cone if it contains the full ray through any point in the set.

$$K \text{ cone if } \forall x \in K \text{ and } \rho \geq 0, \quad \rho x \in K$$



This is a geometric characterization of descent direction. It gives us a geometric condition for $x^\star$ to be a local minimizer.

Lemma 1.3 (Geometric Condition for local minimum)

$x^\star$ is a local minimizer iff

$$\mathcal{F}(x^\star) = \emptyset.$$

We want an algebraic condition to be able to compute or look for $x^\star$.

Lemma 1.4 (Algebraic first-order characterization of $\mathcal{F}$)

If $\nabla f(\bar{x}) \neq 0$, then

$$\mathcal{F}_0(\bar{x}) = \{d \mid \nabla f(\bar{x})^T d < 0\} = \mathcal{F}(\bar{x}).$$

Otherwise

$$\mathcal{F}_0(\bar{x}) \subseteq \mathcal{F}(\bar{x}).$$

Proof: try Taylor-series expansion of f at $\bar{x}$

Graphical interpretation:

∇f forms angles greater or equal than 90° with <u>all</u> descent directions.

Lemma 1.5 (First-order necessary condition for local minimum)

If $x^\star$ is a local minimizer, then

$$\underbrace{\nabla f(x^\star) = 0}_{\text{"stationary point"}} .$$

Proof: Contradiction
If $\nabla f(x^\star) \neq 0$, then $d = -\nabla f(x^\star) \neq 0$. Therefore there exists a descent direction $d \in \mathcal{F}(x^\star)$ by Lemma 1.4. Thus $\exists \delta > 0$ s.t. $f(x^\star + \lambda d) < f(x^\star) \quad \forall \lambda \in (0, \delta)$.
This is a contradiction with the fact, that $x^\star$ is a minimizer. □

Why only necessary?
It can't be a sufficient condition because in case where $\nabla f(x^\star) = 0$ we cannot use Lemma 1.4, e.g. $f_1(x) = -x^2$, $f_2(x) = x^3$, $\nabla f_1(0) = \nabla f_2(0) = 0$.

Lemma 1.6 (second order necessary condition)

Assume f is twice continuously differentiable $f \in \mathcal{C}^2$

$$x^\star \text{ local minimizer } \Rightarrow \nabla^2_{xx} f(x^\star) \succeq 0$$

Note: the condition on the Hessian of f can be interpreted as a local convexity property (around $x^\star$).

Proof: 2^{nd} order Taylor expansion around $x^\star$ in direction $d \in \mathbb{R}^n$:

$$f(x^\star + \lambda d) = f(x^\star) + \lambda \nabla f(x^\star)^T d + \frac{\lambda^2}{2} d^T \nabla^2_{xx} f(x^\star) d + \lambda^2 \|d\|^2 \alpha(\lambda d)$$

$(\rightarrow \alpha(\cdot))$ is a function that is order 1 or higher in λd)

1. If $x^\star$ is local minimzer $\Rightarrow \nabla f(x^\star) = 0$
2. Divide by λ^2:

$$\frac{f(x^\star + \lambda d) - f(x^\star)}{\lambda^2} = \frac{1}{2} d^T \nabla^2_{xx} f(x^\star) d + \|d\| \alpha(\lambda d)$$

3. $\lambda \to 0$ on the right-hand-side the first term dominates

$$\frac{f(x^\star + \lambda d) - f(x^\star)}{\lambda^2} \approx \frac{1}{2} d^T \nabla^2_{xx} f(x^\star) d$$

4. For $x^\star$ is a local minimizer, the left-hand-side must be ≥ 0 for any $d \in \mathbb{R}^n$
$$\Rightarrow \quad d^T \nabla^2_{xx} f(x^\star) d \geq 0 \quad \forall d \quad \Rightarrow \quad \nabla^2_{xx} f(x^\star) \succeq 0 \qquad \square$$

Only a necessary condition, because when $\nabla^2_{xx} f(x^\star)$ is singular, we need to use higher-order information.

Generally it is hard to get (global) sufficient conditions. $\to$ convexity to the rescue!

Lemma 1.7 (First order N&S condition for global minimizers)

Assume f is convex.

$$\exists x^\star \text{ s.t. } \nabla f(x^\star) = 0 \quad \Leftrightarrow \quad x^\star \text{ is a global minimizer}$$

If f is strictly convex, then the minimizer is unique.

Proof: ($\nabla f = 0 \Rightarrow$ global minimum)

First order condition for convexity:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \mathbb{R}^n$$

Pick $x = x^\star$:
$$f(y) \geq f(x^\star), \quad \forall y \in \mathbb{R}^n$$

(Other direction holds because of Lemma 1.5)

What if we do not have global convexity?

Lemma 1.8 (Second order sufficient condition for local minimizer)

Assume $f \in \mathcal{C}^2$.
If $\nabla f(x^\star) = 0$ and $\nabla^2_{xx} f(x^\star) \succ 0 \quad \Rightarrow \quad x^\star$ is strict local minimizer.

Proof: Taylor expansion (Similar to Lemma 1.6)

1.2 Constrained Problems

$$\mathcal{D} \subseteq \mathbb{R}^n, \quad \mathcal{D} = \{x \mid g_i(x) \leq 0,\, i = 1, \dots, n_g \quad h_j(x) = 0,\, j = 1, \dots, n_h\}$$

We assume throughout g_i, h_j are all $\mathcal{C}^1$ functions.

Definition 1.7 (Tangent vector, tangent cone)

$p \in \mathbb{R}^n$ is a tangent vector to $\mathcal{D}$ at $\bar{x} \in \mathcal{D}$ if $\exists$ differential curve $\bar{x}(s) : [0, \varepsilon) \to \mathcal{D}$ with $\varepsilon > 0$ such that $\bar{x}(0) = \bar{x}$, $\frac{d\bar{x}}{ds}\Big|_{s=0} = p$.

Tangent cone $\mathcal{T}_{\mathcal{D}}(\bar{x})$ to $\bar{x}$ is the set of all tangent vectors

$$\mathcal{T}_{\mathcal{D}}(\bar{x}) := \{p \mid p \text{ tangent vector to } \mathcal{D} \text{ at } \bar{x}\}$$

Graphical representation:

Set of directions that make us stay feasible (at least infinitesimally)

When it comes to geometric conditions for optimality in constrained problems, we now have 2 sets/2 directions:

- $d \in \mathcal{F}(x) \to$ descent direction: objective improves
- $d \in \mathcal{T}_D(\bar{x}) \to$ tangent vector: we stay feasible

Lemma 1.9 (Geometric condition for local minimizer, $\mathcal{D} \subseteq \mathbb{R}^n$)

$x^\star$ is a local minimizer iff $\mathcal{F}(x^\star) \cap \mathcal{T}_{\mathcal{D}}(x^\star) = \emptyset$

It is basically says that "any improving direction can't be feasible".

As in the unconstrained case, we want to turn geometric conditions to algebraic ones.

Lemma 1.10 (1st order Nec. condition - semi-algebraic)

If $x^\star$ is a local minimizer. Then:

1. $x^\star \in \mathcal{D}$
2. $\underbrace{\forall p \in \mathcal{T}_{\mathcal{D}}(x^\star)}_{\text{geometric}}$, it holds $\underbrace{p^T \nabla f(x^\star) \geq 0}_{\text{algebraic}}$

Proof:

Item 1 $\to$ feasibility

Item 2: Assume there is a p s.t. $p^T \nabla f(x^\star) < 0$. Then

$$\exists \text{ curve } \bar{x}(s) \in \mathcal{D} \text{ s.t. } \left.\frac{df(\bar{x})}{ds}\right|_{s=0} \underbrace{=}_{\text{chain rule}} p^T \nabla f(x^\star) < 0$$

which would mean that p is descent direction. Contradicts $x^\star$ local minimizer.

This is almost a translation of Lemma 1.9 because we replaced $\mathcal{F}(x^\star)$ with its algebraic form "$d^T \nabla f(x^\star) < 0$".

To obtain a fully algebraic test, we need a few more concepts.

Definition 1.8 (Active constraints, active set, regular points)

$\bar{x} \in \mathcal{D}$

- g_i is active at $\bar{x}$ if $g_i(\bar{x}) = 0$
- $A(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$ set of active constraints at $\bar{x}$
- $\bar{x} \in \mathcal{D}$ is a regular point if $\nabla g_i(\bar{x})$, $i \in \mathcal{A}(\bar{x})$ and $\nabla h_j(\bar{x})$, $j = 1, \dots, n_h$ are linearly independent.

Lemma 1.11 (Algebraic first-order characterization of target set)

If $\bar{x}$ is regular point. Then

$$\mathcal{T}_\mathcal{D}(\bar{x}) = \{p \mid \nabla h(\bar{x})^p = 0,\, \nabla g_i(\bar{x})^p \le 0, \quad \forall i \in \mathcal{A}(\bar{x})\} \qquad ①$$

where $\nabla h(\bar{x}) := [\nabla h_1(\bar{x}), \dots, \nabla h_{n_k}(\bar{x})] \in \mathbb{R}^{n \times n_h}$.
① can be written equivalently as $\mathcal{T}_\mathcal{D}(\bar{x}) = \{p \mid \mathcal{A}(\bar{x})p \ge 0\}$

$$A(\bar{x}) := \begin{bmatrix} \nabla h(\bar{x})^T \\ -\nabla h(\bar{x})^T \\ \vdots \\ -\nabla g_i(\bar{x})^T \\ \vdots \end{bmatrix} \in \mathbb{R}^{(2n_h + |\mathcal{A}(\bar{x})|) \times n} \quad \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} i \in \mathcal{A}(\bar{x})$$

In other words, item 2 of Lemma 1.10 can be written as follows:

$$p \in \mathbb{R}^n : \mathcal{A}(x^\star)p \ge 0,\, p^T \nabla f(x^\star) < 0$$

still not very tractable?

Farkas Lemma to the rescue:

Lemma 1.12 (Farkas Lemma)

For any matrix $A \in \mathbb{R}^{m\times n}$, vector $b \in \mathbb{R}^n$.
Exactly one of the following holds:

1. $\exists y \in \mathbb{R}^m\ y \geq 0$, such that $A^T y = b$

2. $\exists p \in \mathbb{R}^m$, such that $A^T p \geq 0$, $p^T b < 0$

Take $A \equiv A(x^\star)$ and $b \equiv \nabla f(x^\star)$

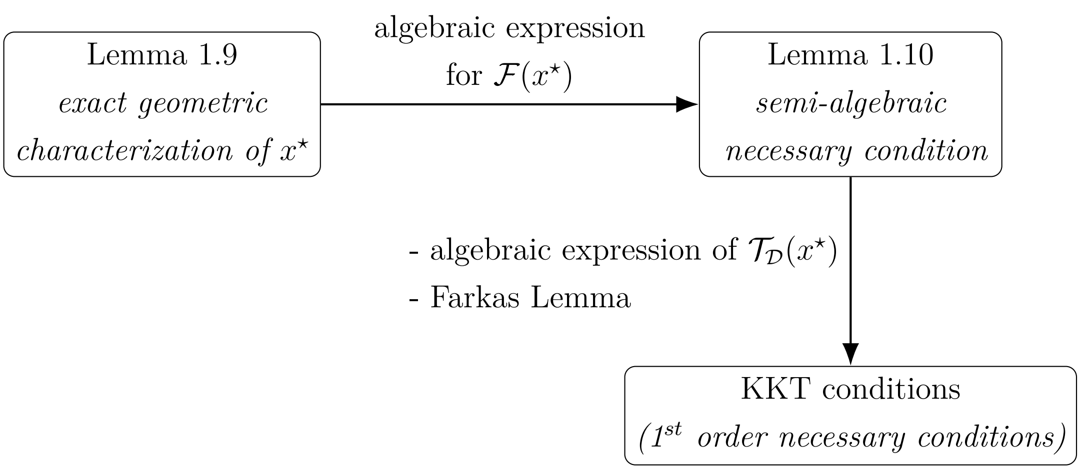
If we find y satisfying 1., then 2. can't hold $\Rightarrow$ item 2 of Lemma 1.10 is verified $\Rightarrow$ $x^\star \in \mathcal{D}$ is a local minimizer.

KTK-conditions just follow from imposing

item 1 of Lemma 1.10 $\to x^\star \in \mathcal{D}$

item 2 of Lemma 1.10 $\to \exists y \in \mathbb{R}^m$, $y \geq 0$ s.t. $\mathcal{A}(x^\star)^T y = \nabla f(x^\star)$

Conceptual summary:



Informal recap:

$x^\star$ local minimizer $\Rightarrow x^\star \in \mathcal{D}, \quad \forall p \in \mathcal{T}_D(x^\star), \quad p^T \nabla f(x^\star) \geq 0$

$\Updownarrow$ (if $x^\star$ regular point)

$$\exists y = \begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ -\nu_i \\ \vdots \end{bmatrix} \in \mathbb{R}^{2n_h + |\mathcal{A}(x^\star)|} \quad \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} i \in \mathcal{A}(\bar{x}) \qquad y \geq 0 \quad \to A(x^\star)^T y = \nabla f(x^\star)$$

Let's write down $A^T y = \nabla f$

$$\nabla h(x^\star)(\hat{\lambda}_1 - \hat{\lambda}_2) - \sum_{i \in \mathcal{A}(x^\star)} \nabla g_i(x^\star)\nu_i = \nabla f(x^\star), \quad \hat{\lambda}_1, \hat{\lambda}_2, \nu_i \geq 0 : \text{ But } (\hat{\lambda}_1 - \hat{\lambda}_2) \gtreqless 0$$

Equivalently: $\lambda := -(\hat{\lambda}_1 - \hat{\lambda}_2) \in \mathbb{R}^{n_h}$, sign undefined

$$\exists \lambda \in \mathbb{R}^{n_h}, \quad \nu \in \mathbb{R}^{n_g}, \quad \nu \geq 0, \quad \nu_i = 0, \quad i \notin \mathcal{A}(x^\star)$$

$$\text{s.t. } \nabla f(x^\star) + \nabla h(x^\star)\lambda + \nabla g(x^\star)\nu = 0$$

$$\text{with } \nabla g(x^\star) := \begin{bmatrix} \nabla g_1(x^\star) & \nabla g_2(x^\star) & \cdots & \nabla g_{n_g}(x^\star) \end{bmatrix}$$

$$\text{and } \nabla h(x^\star) := \begin{bmatrix} \nabla h_1(x^\star) & \nabla h_2(x^\star) & \cdots & \nabla h_{n_h}(x^\star) \end{bmatrix}$$

We are now ready for a fully algebraic characterization.

Definition 1.9 (Karash-Kuhn-Tucker (KKT) points)

A triplet of vectors $(\bar{x}, \bar{\lambda}, \bar{\nu}) \in \mathbb{R}^n \times \mathbb{R}^{n_h} \times \mathbb{R}^{n_g}$

$\bar{x}$: opt. variable, $\bar{\lambda}$: multiplier equality constraints,

$\bar{\nu}$: multiplier inequality constraints

is a KKT point if

1. $\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\lambda} + \nabla g(\bar{x})\bar{\nu} = 0$
2. $g(\bar{x}) \leq 0$
3. $h(\bar{x}) = 0$
4. $\bar{\nu} \geq 0$
5. $\bar{\nu}^T g(\bar{x}) = 0$

Lemma 1.13 (KKT necessary condition for local minimizer)

If $x^\star$ is a local minimizer **and** a regular point.
Then $\exists \lambda^\star, \nu^\star$ s.t. ($x^\star, \lambda^\star, \nu^\star$ is a KKT point)

Proof: Corollary of previous discussion

1. $\iff \exists \lambda \in \mathbb{R}^{n_h}$, $\nu \in \mathbb{R}^{n_g}$ s.t. $\nabla f(x^\star) + \nabla h(x^\star)\lambda + \nabla g(x^\star)\nu = 0$ It can be written

   equivalently as

$$\nabla_x \mathcal{L}(x, \lambda, \nu)|_{x=x^\star, \lambda=\lambda^\star, \nu=\nu^\star} = 0$$

   where $\mathcal{L}(x, \lambda, \nu) := f(x) + \lambda^T h(x) + \nu^T g(x)$

2. $\iff x^\star \in \mathcal{D}$
3. $\iff x^\star \in \mathcal{D}$
4. $\iff$ non-negativity of "y" from Farkas Lemma
5. $\iff \nu_i = 0,\ i \notin \mathcal{A}(x^\star)$

   $\nu^T g(x^\star) = \sum_i \nu_i g_i(x^\star) = 0$ Complementary slackness

$$g_i(x^\star) = \begin{cases} = 0, & i \in \mathcal{A}(x^\star) \\ < 0, & i \notin \mathcal{A}(x^\star) \end{cases} \quad \text{because } x^\star \in \mathcal{D}$$

   $\nu_i \geq 0,\ \forall i$ because of Farkas' lemma.

   Then $\sum_i \nu_i g_i(x^\star)$ automatically sets $\nu_i = 0$ when $i \notin \mathcal{A}(x^\star)$ or $g_i(x^\star) < 0$. $\square$

Intrestingly, if NLP is convex, KKT conditions are sufficient for global optimality:

Lemma 1.14 (KKT sufficient conditions fot global minimizer)

Suppose $f, g_i \quad (i = 1, \dots, n_g)$ are convex functions and
$h_j \quad (j = 1, \dots, n_h)$ are affine functions.
If $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point, then $x^\star$ is a local minimizer.

Proof: For $(\lambda^\star, \nu^\star)$ KKT points:

$$b(x) := \mathcal{L}(x, \lambda^\star, \nu^\star) = f(x) + \sum_{i=1}^{n_g} \nu_i^\star g_i(x) + \sum_{j=1}^{n_h} \lambda_j^\star g_j(x) \qquad \otimes$$

f, g_i, h_j are convex functions

Linear combination of cvx functions with non-negative coefficients is a convex function

$\Rightarrow b(x)$ convex

1. $b(x)$ convex
2. $\nabla b(x^\star) = 0$ because of 1., $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point $b(x) \geq b(x^\star) \quad \forall x \in \mathbb{R}^n$

$$\Updownarrow \text{ (if } x^\star \text{ regular point)}$$

$$f(x) - f(x^\star) \geq \underbrace{-\sum_{i \in \mathcal{A}(x^\star)} \nu_i^\star g_i(x)}_{\leq 0} \underbrace{- \sum_{j=1}^{n_h} \lambda_i^\star h_i(x)}_{=0} \geq 0, \qquad x \in \mathcal{D}$$

because $g(x) \leq 0$, $\nu^\star \geq 0$ because $h(x) = 0$

$\rightarrow x^\star$ is a global minimizer. $\square$

<u>Second-order conditions</u>

Similary to the unconstrained case, we can use the Hessian.

$$\nabla^2_{xx}\mathcal{L}$$

We need to check positive semi-definiteness of the Hessian only along feasible directions:

Precisel, we are interestes in this property along

$$\text{Critical Directions} = \{p \mid \underbrace{p \in \mathcal{T}_\mathcal{D}(x^\star)}_{\text{feasible directions}}, \quad \underbrace{\nabla f(x^\star)^T p = 0}_{\substack{\text{directions that cannot} \\ \text{be excluded based on} \\ \text{on first order arguments}}} \}$$

$$\begin{cases} \nabla f(x^\star)^T p < 0 & \rightarrow \quad p \text{ descent direction: } {}^{\substack{\text{already excluded by necessary} \\ \text{condition of order 1}}}, \\ \nabla f(x^\star)^T p > 0 & \rightarrow \quad p \text{ ascent direction: “not harmful”}, \\ \nabla f(x^\star)^T p = 0 & \rightarrow \quad \text{this is what is “new” compared to first order.} \end{cases}$$

Lemma 1.15 (Second order necessary condition)

- $f, g, h \in \mathcal{C}^2$ at $x^\star$
- $x^\star$ local minimizer and regular point
- $(x^\star, \lambda^\star, \nu^\star)$ KKT point (which exists by Lemma 1.13)

Then

$$p^T \nabla^2_{xx} L(x^\star, \lambda^\star, \nu^\star) p \geq 0 \quad \text{(curvature non-negative along critical directions)}$$

$\forall p \neq 0$ with

- $\nabla h(x^\star)^T p = 0$
- $\nabla g_i(x^\star)^T p \leq 0 \quad \forall i \in \mathcal{A}(x^\star)$ with $\nu_i^\star = 0 \mid p \in \mathcal{T}_{\mathcal{D}}(x^\star)$
- $\nabla g_i(x^\star)^T p = 0 \quad \forall i \in \mathcal{A}(x^\star)$ with $\nu_i^\star > 0 \mid \nabla f(x^\star)^T p = 0$

Lemma 1.16 (Second order suffiecient conditions for local minimizer)

If $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point with

$$p^T \nabla_{xx} \mathcal{L}(^\star, \lambda^\star, \nu^\star) p > 0$$

for same p as in Lemma 1.15.
Then $x^\star$ is a strict local minimizer.

2 Calculus of Variations

Goal in OC: Find a function that maximizes a functional (function of function) subject to dynamic constraints

In Chapter 1 we characterized solutions to optimization problems over vectors ($\mathbb{R}^n$)

$$\min f(x) \text{ s.t. } x \in \mathcal{D} \subseteq \mathbb{R}^n, \qquad \text{Static problem}$$

- We should introduce "time" or "stages" in the problem

$$\min_{x_1, \ldots, x_N} \sum_{k=1}^{N} f(k, x_k, x_{k-1}) \qquad \text{N coupled stages}$$

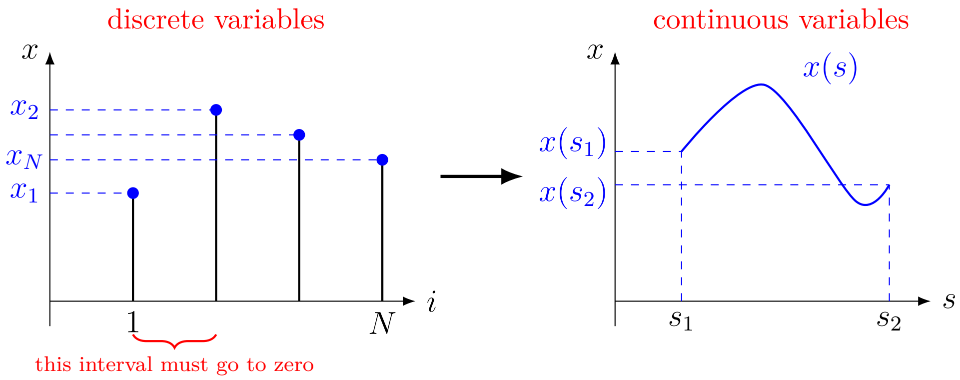
$$\text{s.t. } x_k \in \mathcal{D}_k \subseteq \mathbb{R}^n, \quad k = 1, \dots, N, \quad x_0 \text{ given}$$

equivalent to: (loses structure)

$$z = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \to \min_z \; p(z) \qquad z \in \mathcal{Z}, \quad z \in \mathbb{R}^{n \times N}$$

- Continuous-time description of dynamics

  From N stages to continuous time by taking ∞ many stages:



$$\min_{x(\cdot)} \int_{s_1}^{s_2} f\Big(s, x(s), \dot{x}(s)\Big)\, ds$$

$$\text{s.t.} \quad \begin{aligned} & x(s) \in \mathcal{X} \subseteq \mathbb{R}^n, \quad s \in [s_1, s_2], \\ & x(s_1) = x_1, \end{aligned} \quad \left\{ \begin{array}{l} \text{prototypical CV problem} \\ \bullet \text{ no ODE yet} \\ \bullet \text{ opt. variable lives in a function space} \end{array} \right.$$

2.1 Introduction to CV (Calculus of Variations)

Function CLASSES & NORMS

$$\big(\underbrace{V}_{\text{vector space}}, \underbrace{\|\cdot\|}_{\text{norm}}\big) \quad \text{normed vector space}$$

V is the set of vector functions

$$x(s), \quad s \in [s_1, s_2] \text{ taking values in } \mathbb{R}^n, \quad [s_1, s_2] \subseteq \mathbb{R}$$

Two classes:

- $V = \mathcal{C}^1([s_1, s_2], \mathbb{R}^n)$: continuously differentiable functions $x : [s_1, s_2] \to \mathbb{R}^n$
- $\hat{V} = \hat{\mathcal{C}}^1([s_1, s_2], \mathbb{R}^n)$: piecewise continuously differentiable functions

$$x : [s_1, s_2] \to \mathbb{R}^n.$$

Definition 2.1 (Piecewise continuously differentiable functions)

$x : [s_1, s_2] \to \mathbb{R}^n$ is piecewise continuously differentiable (PCD) if

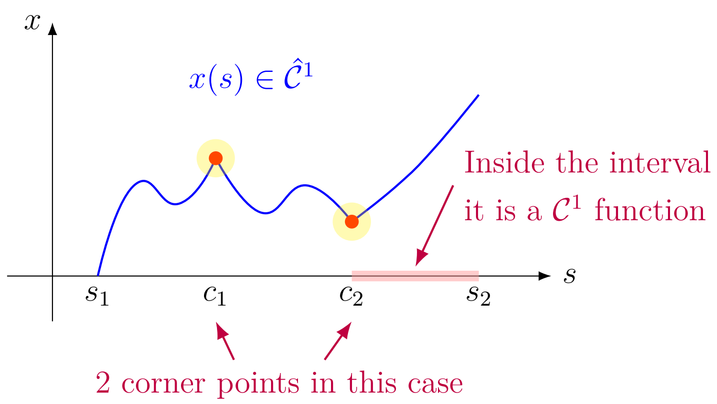
- $x \in \mathcal{C}$ on $[s_1, s_2]$,
- $\exists$ a finite partition $\{c_k\}_{k=0}^{N+1}$ with

$$s_1 = c_0 < c_1 < \cdots < c_{N+1} = s_2,$$

  such that $x : [c_k, c_{k+1}] \to \mathbb{R}^n$ is $\mathcal{C}^1$.

  That is $x \in \mathcal{C}^1([c_k, c_{k+1}], \mathbb{R}^n), \quad \forall k = 0, 1, \ldots, N.$

Example:



Norms

Definition 2.2 (Strong and weak norms)

Case $V = \mathcal{C}^1$:

- Strong norm (or ∞-norm)

$$\|x\|_\infty := \max_{s_1 \le s \le s_2} \|x(s)\|_{\leftarrow \text{ any norm in } \mathbb{R}^n}$$

- Weak norm (or 1-norm)

$$\|x\|_1 := \|x\|_\infty + \max_{s_1 \le s \le s_2} \|\dot{x}(s)\|_{\leftarrow \text{ any norm in } \mathbb{R}^n}$$

$$(\mathcal{C}^1([s_1, s_2]), \|\cdot\|) \text{ full notation}$$

Note $\forall x \in V, \|x\|_1 \ge \|x\|_\infty$

Case $V = \hat{\mathcal{C}}^1$:

- Strong norm $\to$ same as for $\mathcal{C}^1$
- weak norm

$$\|x\|_1 := \|x\|_\infty + \sup_{s \in \bigcup_{k=0}^{N} (c_k, c_{k+1})} \|\dot{x}(s)\|$$

CV problem

$$\min_{x \in V} \underset{\hookrightarrow \text{ functional } J: V \to \mathbb{R}}{\underline{J}(x)}$$

$$\text{s.t. } x \in \underset{\hookrightarrow \text{ admissible set}}{\underline{\mathcal{D}}}$$

$\bar{x} \in \mathcal{D}$ admissible curve for trajectory.

3 forms of J:

- Langrangian form

$$J(x) := \int_{s_1}^{s_2} \underset{\hookrightarrow \text{ running cost, } L: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}}{\underline{l}(s, x(s), \dot{x}(s)) ds}$$

- Bolza form

$$J(x) \coloneqq \varphi(s_2, x(s_2)) + \int_{s_1}^{s_2} l(s, x(s), \dot{x}(s))ds \qquad \varphi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$$

- Mayer form

$$J(x) \coloneqq \varphi(s_2, x(s_2))$$

These 3 forms are (gernerally) interchangeble, e.g. $L \Rightarrow B,\ B \Rightarrow L$.

We will see them again in Chapter 3.

Admissible sets:

- Free problems $\to$ only endpoints are constrained. Example:

$$\mathcal{D} = \{x \in V \mid x(s_1) = x_1,\ x(s_2) = x_2\}, \qquad x_1, x_2 \text{ fixed vectors}$$

- Isoperimetric constraints $\to$ level sets

$$\mathcal{D} = \bigcap_{i=1}^{n_g} \Lambda_i(k_i)$$

$$\Lambda_i(k_i) \coloneqq \{x \in V \mid \int_{s_1}^{s_2} g_i(s, x(s), \dot{x}(s))ds = k_i\}$$

Example:

$$\min_x J(x)$$

$$\text{s.t.} \quad \left.\begin{aligned} &\int_{s_1}^{s_2} x^2(s)\,ds = 1 \\ &\int_{s_1}^{s_2} x(s)\,ds = 0 \end{aligned}\right\} \qquad \begin{aligned} &n_g = 2, \\ &g_1(s) = x^2(s), \\ &g_2(s) = x(s) \end{aligned}$$

Minimizers for CV problems

Definition 2.3 (Global and local minimizers)

$x^\star \in \mathcal{D}$ is a global minimizer of [CV] if

$$J(x) \geq J(x^\star), \qquad x \in \mathcal{D}$$

$x^\star \in \mathcal{D}$ is a strong local minimizer of [CV] if

$$\exists \varepsilon > 0 \text{ s.t. } J(x) \geq J(x^\star), \quad \forall x \in B_\varepsilon^\infty(x^\star) \cap \mathcal{D}$$

$$B_\varepsilon^\infty \coloneqq \{y \in V \mid \|x - y\|_\infty \leq \varepsilon\} \quad \text{(strong ball)}$$

$x^\star \in \mathcal{D}$ is a weak local minimizer of [CV] if

$$\exists \varepsilon > 0 \text{ s.t. } J(x) \geq J(x^\star), \quad \forall x \in B_\varepsilon^1(x^\star) \cap \mathcal{D}$$

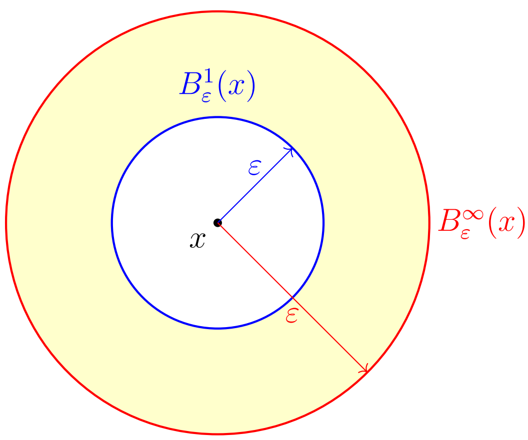
$$B_\varepsilon^1 \coloneqq \{y \in V \mid \|x - y\|_1 \leq \varepsilon\} \quad \text{(weak ball)}$$

We will often call strong/weak minimizers (without local).

Note: Every strong minimizer is a weak minimizer. In general, the converse is not true.

Why?

$\forall x \in V, \ \forall \varepsilon > 0, \ B_\varepsilon^1(x) \subseteq_\varepsilon^\infty (x)$



$$\|x\|_1 \geq \|x\|_\infty$$

Implication for CV is that a curve x that is better than all elements in $B^1_\varepsilon(x)$ is not necessarily better than all elements in $B^\infty_\varepsilon(x)$.

Example: $s_1 = 0,\ s_2 = 1$

$$J(x) = \int_0^1 \dot{x}^2(s) - \dot{x}^4(s)ds$$

$$\mathcal{D} = \{x \in \hat{\mathcal{C}}^1([0,1]) \mid x(0) = x(1) = 0\}$$

$$\bar{x}(s) = 0, \quad J(\bar{x}) = 0, \quad \bar{x} \text{ is a weak minimum but not strong}$$

Weak minimum

$B^1_\varepsilon(0)$, take $0<\varepsilon \leq 1$

$$\forall x \in B^1_\varepsilon(0),\ \|\dot{x}(s)\| \leq \varepsilon \qquad \forall s \in [0,1]$$

because x is in the weak ball with radius ε.

$$J(x) = \int_0^1 \underbrace{\dot{x}^2(s)}_{\geq 0}(\underbrace{1 - \dot{x}^2(s)}_{\geq 0 \text{ see above, } \varepsilon\leq 1})ds \geq 0, \quad \forall x \in B^1_\varepsilon$$

$J(\bar{x}) = 0 \to \bar{x}$ is a weak minimum, but not strong. Try to see why? $\to$ Find counterexamples.

Existence of solutions for problems of CV:

Weierstrass theorem still holds but compactness $\neq$ closed and bounded in function vector spaces.

A subspace $\mathcal{D}$ of a metric space V is compact, if "every sequence in $\mathcal{D}$ has a subsequence converging to some point in $\mathcal{D}$".

Example:

$$B^\infty_1(0) = \{x \in \mathcal{C}([0,1]) \mid \|x\|_\infty \leq 1\}$$

closed and bounded, not a compact set.

Bottom line: checking Weierstrass in CV can be overly restrictive because our common sets in $\mathcal{D}$ are not compact.

Existence of solutions difficult to guarantee a-priory.

Convexity is a special case where this is possible.

Variations: (extension of perturbations to the cost seen in NLP)

Definition 2.4 (First-variation of a functional)

First variation (Gateaux derivative) of J at $x \in V$ in direction $\xi \in V$ is

$$\delta J(x, \xi) := \lim_{\eta \to 0} \frac{J(x + \eta\xi) - J(x)}{\eta} = \left. \frac{\partial J(x + \eta\xi)}{\partial \eta} \right|_{\eta=0}$$

$$J : V \to \mathbb{R}$$

δJ can be interpreted as follows

$$J(x + \eta\xi) = J(x) + \eta\delta J(x; \xi) + \underline{o(\eta)}_{\text{second order term } \lim_{\eta \to 0} \frac{0(\eta)}{\eta} = 0}$$

δJ is functional associated with J and a point x mapping a perturbation ξ into a scalar, representing the variation of J in that direction $\approx$ "directional derivative" for CV.

Example:

$$J(x) = \int_{s_1}^{s_2} x^2(s) ds, \quad V = \mathcal{C}^1$$

δJ? Apply defintion:

$$\frac{J(x + \eta\xi) - J(x)}{\eta} = \frac{1}{\eta} \left[\int_{s_1}^{s_2} (x(s) + \eta\xi(s))^2 - x^2(s) ds \right]$$

$$= 2 \int_{s_1}^{s_2} x(s)\xi(s) ds + \eta \int_{s_1}^{s_2} \xi^2(s) ds$$

$$\lim_{\eta \to 0} \quad \longrightarrow \quad \delta J(x; \xi) = 2 \int_{s_1}^{s_2} x(s)\xi(s) ds$$

From the defintion, we can see that δJ is a linear operator on V.

$$\delta(J_1 + J_2)(x; \xi) = \delta J_1(x; \xi) + \delta J_2(x; \xi).$$

Moreover, it is a homogeneous operator: $\forall \alpha \in \mathbb{R}$ it holds

$$\delta J(x; \alpha\xi) = \alpha\delta J(x; \xi).$$

Definition 2.5 (Second variation of a functional)

$$\delta^2 J(x; \xi) := \left. \frac{\partial^2}{\partial \eta^2} J(x + \eta\xi) \right|_{\eta=0}$$

Interpretation $J(x+\eta\xi) = J(x) + \eta\delta J(x;\xi) + \eta^2\delta^2 J(x;\xi) + o(\eta^2)$

Fundamental Lemma in CV.

Definition 2.6 (Descent direction in CV)

Given V, $J : V \to \mathbb{R}$ Gateaux-differentiable ($\delta J(x;\xi)$ exists) at $\bar{x} \in V$, we call $\xi \in V$ a descent direction for J at $\bar{x}$ if

$$\delta J(\bar{x};\xi) < 0$$

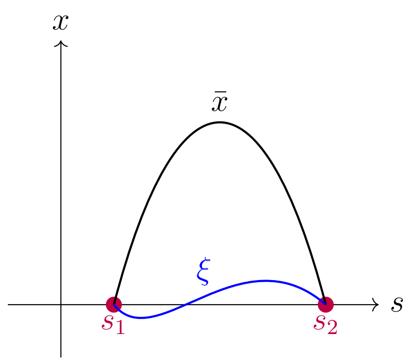
There is a close connection with "descent direction" from NLP.

Definition 2.7 ($\mathcal{D}$-admissible direction)

Given V, $\mathcal{D} \subseteq V$ and $J : V \to \mathbb{R}$.
$\xi \in V$, $\xi \neq 0$ is $\mathcal{D}$-admissible for J at $\bar{x} \in \mathcal{D}$ if

- $\delta J(\bar{x};\xi)$ exists
- $\exists \beta > 0$ s.t. $\bar{x} + \eta\xi \in \mathcal{D}, \quad \forall \eta \in (-\beta, \beta)$

Example:



$\mathcal{D}$: $\bar{x}(s_1) = \bar{x}(s_2) = 0$

ξ is D-admissible

Lemma 2.1 (Negative result for minimizers in unconstrained CV)

$(V, \|\cdot\|)$, J.
Suppose at $\bar{x}\exists$ descent direction $\bar{\xi} \in V$.
Then $\bar{x}$ cannot be a local minimizer for J (neither strong or weak).

Proos: Use definition of 1^{st} variation.

If $\delta J(\bar{x}_i\bar{\xi}) < 0$ then

$$J(\bar{x} + \eta\bar{\xi}) < J(\bar{x}) \quad \forall \eta \in (0, \beta)$$

This comes from

$$J(x + \eta\xi) = J(x) + \eta\xi J(x; \eta) + o(\eta)$$

Thus $\bar{x}$ can't be a local minimizer.

Lemma 2.2 (Geometric necessary condition for a local minimum, Fundamental Lemma of CV)

$(V, \|\cdot\|),\ \mathcal{D} \subseteq V,\ J \to \mathbb{R}.$
Suppose $x^\star \in \mathcal{D}$ is a local minimizer for J on $\mathcal{D}$, then

$$\boxed{\delta J(x^\star; \xi) = 0} \qquad \forall \mathcal{D}\text{-admissible directions at } x^\star$$

Proof: By contradiction:

Case 1: $\delta J(x^\star; \xi) < 0$

By Lemma 2.1 $x^\star$ can't be a local minimizer $\to$ contradiction

Case 2: $\delta J(x^\star; \xi) > 0$

By definition of the $\mathcal{D}$-admissible direction, if ξ is $\mathcal{D}$-admissible.

Then $-\xi$ is also $\mathcal{D}$-admissible.

$$\delta J(x^\star; -\xi) = -\delta J(x^\star; \xi) < 0 \quad \text{(because of linearity)}$$

$\to x^\star$ cannot be a local minimizer (because $\delta J(x^\star; -\xi) < 0$ and $-\xi$ is $\mathcal{D}$-admissible).
Thus $\delta J(x^\star; \xi) = 0$.

All the results in the rest of this chapter are $\underset{\substack{\text{they turn such an abstract} \\ \text{requirenment into algebraic tests}}}{\text{“merely corollaries”}}$ of this Lemma.

2.2 Free problems of CV

2.2.1 $V = \mathcal{C}^1$

$$\min_{x(\cdot)} \int_{s_1}^{s_2} l(s, x(s), \dot{x}(s)) ds \qquad [CV - P1]$$

$$\text{s.t. } x \in \{y \in \mathcal{C}^1([s_1, s_2]) \mid \underbrace{y(s_1) = x_1,\ y(s_2) = x_2}_{\substack{\text{special case of “free problems”} \\ \text{we fix the endpoint to } x_2}}\}$$

Lemma 2.3

Suppose $x^\star \in \mathcal{C}^1([s_1, s_2])$ is a weak minimum of $[CV - P1]$.
Then

$$\frac{d}{ds} l_{\dot{x}_i}(s, x^\star(s), \dot{x}^\star(s)) = l_{x_i}(s, x^\star(s), \dot{x}^\star(s)), \quad \forall s \in [s_1, s_2],\ i = 1, \ldots n$$

with $l_{\dot{x}_i} = \frac{\partial l}{\partial \dot{x}_i}$ and $l_{x_i} = \frac{\partial l}{\partial x_i}$ (Euler equations).

This is a set of nonlinear ordinary time-varying second-order differential equations. Their solutions are <u>candidate</u> local minimizers of $[CV - P1]$.

solution $x(s) : \ s \in [s_1, s_2]$

also called "stationaty solutions" of the corresponding CV problem. Why?

$$\text{Because } \delta J(x^\star; \xi) = 0 \quad \forall \xi \ \mathcal{D}\text{-admissible}$$

Proof: The idea is to derive algebraic conditions that are sufficient to guarantee $\delta J(x^\star; \xi) = 0$.

First step is to write $\delta J : \ \eta \in \mathbb{R},\ \xi \in \mathcal{C}^1$

$$\frac{\partial}{\partial \eta} J(x^\star + \eta\xi) \underset{\text{Leibniz rule}}{=} \int_{s_1}^{s_2} \frac{\partial}{\partial \eta} l(s, x^\star + \eta\xi, \dot{x}^\star + \eta\dot{\xi})\, ds$$

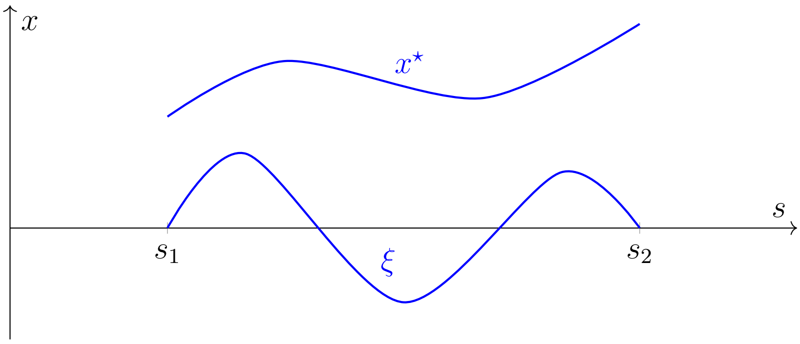
$$= \int_{s_1}^{s_2} l_x \left[x^\star + \eta\xi\right]^T \xi + l_{\dot{x}} \left[x^\star + \eta\xi\right]^T \dot{\xi}\, ds$$

with $l_z[y] := l_z(s, y, \dot{y})$.

Take $\eta \to 0 \quad \delta J(x^\star; \xi) = \int_{s_1}^{s_2} \underbrace{l_x \left[x^\star\right]^T + l_x \left[x^\star\right]^T \dot{\xi}}_{\text{integrand is a continuous function}}\, ds$

$\to$ first-variation exists $\forall \xi \to J$ is Gateaux-differentiable

We want to obtain conditions enforcing $\delta J = 0\ \forall \mathcal{D}$-admissible ξ. This means $\xi(s_1) = \xi(s_2) = 0$ and $\xi \in \mathcal{C}^1$.



To do this, we select n perturbations $\xi^{(i)}(i = 1, \dots n)$ defined as follows

$$\xi^{(i)} = \begin{bmatrix} \xi_1^{(i)} \\ \vdots \\ \xi_n^{(i)} \end{bmatrix}$$

- $\xi_j^{(i)} = 0, \quad j \neq i$
- $\xi_i^{(i)}$ arbitrary but not identically zero with $\xi_i^{(i)}(s_1) = \xi_i^{(i)}(s_2) = 0$.

We replace those n perturbations in the equation $\delta J = 0$

$$\forall i \in \{1, \dots, n\}, \quad 0 = \delta J(x^\star, \xi^{(i)}) = \int_{s_1}^{s_2} \left[l_{x_i}[x^\star]\xi_i + l_{\dot{x}_i}[x^\star]\dot{\xi}_i \right] ds.$$

$$= \int_{s_1}^{s_2} l_{\dot{x}_i}[x^\star]\dot{\xi}_i \, ds + \int_{s_1}^{s_2} \underbrace{\frac{d}{ds}\left[\left(\int_{s_1}^{s} l_{\dot{x}_i}[x^\star] d\sigma\right)\right]}_{v'} \underbrace{\dot{\xi}_i}_{u} \, ds$$

integral by parts $\int_a^b uv' = [uv]_a^b - \int_a^b u'v$

$$= \int_{s_1}^{s_2} l_{\dot{x}_i}[x^\star]\dot{\xi}_i \, ds + \underbrace{\left[\xi_i \int_{s_1}^{s} l_{x_i}[x^\star] d\sigma\right]_{s_1}^{s_2}}_{=0 \text{ because } \mathcal{D}\text{-admissible}} - \int_{s_1}^{s_2} \left[\int_{s_1}^{s} l_{x_i}[x^\star] d\sigma\right] \dot{\xi}_i \, ds$$

$$= \int_{s_1}^{s_2} \left[l_{\dot{x}_i}[x^\star] - \int_{s_1}^{s_2} l_{x_i}[x^\star] d\sigma \right] \dot{\xi}_i \, ds$$

DuBois-Reymond's Lemma if

- $h(s)$ continuous in $[s_1, s_2]$
- $\displaystyle\int_{s_1}^{s_2} h(s)\, \dot{y}(s)\, ds = 0$
- $y(s_1) = y(s_2) = 0$

$\Rightarrow \quad h(s)$ constant in $[s_1, s_2]$

This applies to our problem $\quad y \equiv \dot{\xi}_i$

$$h \equiv L_{\dot{x}_i}[x^\star] - \int_{s_1}^{s} C_{x_i}[x^\star]\, d\sigma$$

$$\Rightarrow \quad l_{\dot{x}_i}[x^\star] - \int_{s_1}^{s} l_{x_i}[x^\star]\, d\sigma = c_i, \qquad \forall s \in [s_1, s_2],\ i = 1, \dots, n,\ c_i:\ \text{constants}$$

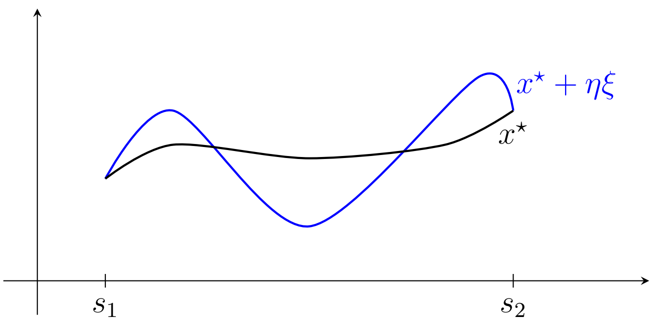
Note that this shows $l_{\dot{x}_i} \in \mathcal{C}^1$

What we obtain are the "integral" Euler equations (EE)

If we take $\dfrac{d}{ds}$ we obtain the EE.

Remarks

1. Necessary conditions for local minimizers ($\to$ weak)

   as $\eta \to 0$

   $(x^\star + \eta\xi)$ and $x^\star$ differentiable both in magnitude and in derivative.

   $\Rightarrow$ "$(x^\star + \eta\xi)$ is inside the weak ball of $x^\star$"

   EE $\Rightarrow$ detect weak minimizers

2. To solve ODE we need Boundary conditions (BC).

   BC come from the admissible set $x(s_1) = x_1,\ x(s_2) = x_2$

   $\to 2n$ equations for a 2^{nd} order ODE in n unknowns.

   Two Point Boundary Value Problem (TPBVP)

3. There exists a reformulation of EE:

$$p(s) \coloneqq l_{\dot{x}}(s, x, \dot{x}) \quad \text{Momentum associated with a given } x.$$

$$H(s, x, \dot{x}, p) \coloneqq -l(s, x, \dot{x}) + \dot{x}^T p. \quad \text{Hamiltonian.}$$

EE can be rewritten as:

$$\dot{x} = H_p(s, x, \dot{x}, p).$$

$$\dot{p} = -H_x(s, x, \dot{x}, p).$$

Canonical equation and x, p canonical Variables.

An immediate benefit of this reformulation is that we easily see the following special cases.

[A] $l(x, \dot{x})$. l does not depend on s.

$$\frac{d}{ds}H = -l_x^T \dot{x} - l_{\dot{x}}^T \ddot{x} + \dot{x}^T \dot{p} + \ddot{x}^T p = \dot{x}^T \underbrace{\left(\underbrace{\frac{dl_{\dot{x}}}{ds}}_{\dot{p}} - l_x \right)}_{=0 \text{ because of EE}} = 0$$

$$H = \text{const.} \coloneqq c_1 \quad \text{on stationary solutions.}$$

[B] $l(s, \dot{x})$ no dependence on x.

$$\frac{d}{ds}p = \dot{p} = 0.$$

$$p = c_2.$$

Lemma 2.4 (Second-order necessary conditions for $(CV - P1)$)

Assume $l \in \mathcal{C}^2$.
If $x^\star$ is a weak minimizer of $[CV - P1]$, then

1. $x^\star$ satisfies EE
2. $\nabla^2_{\dot{x}\dot{x}} l(s, x^\star, \dot{x}^\star) \succeq 0, \quad \forall s \in [s_1, s_2]$.
   Legendre condition

Lemma 2.5 (First-order sufficient condition for global minimizers)

Assume that $l(s, x, \dot{x})$ is jointly convex in x and $\dot{x}$.
If $x^\star \in \mathcal{D}$ satisfies the EE, then $x^\star$ is a global minimizer.

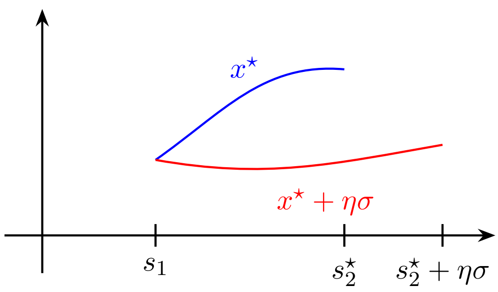
Free end-point problems

$$\mathcal{D} = \{(x, s_2) \in \mathcal{C}^1 ([s_1, \infty) \times (s_1, \infty)) \mid x(s_1) = x_1, x(s_2) \text{ free}, s_2 \text{ free}\}$$

First-variation also suitably redefined:

$$\delta J(x, s_2; \xi, \sigma) := \lim_{\eta \to 0} \frac{J(x + \eta\xi, s_2 + \eta\sigma) - J(x, s_2)}{\eta}$$

$$= \frac{\partial}{\partial \eta} J(x + \eta\xi, s_2 + \eta\sigma)\Big|_{\eta=0}$$



[CV-P2] $\quad \min_{x(\cdot), s_2} \varphi(s_2, x(s_2)) + \int_{s_1}^{s_2} l(s, x(s), \dot{x}(s))\, ds$

$$\text{s.t.} \quad x \in \{y \in \mathcal{C}^1([s_1, s_2]) \mid y(s_1) = x_1\}, \quad s_2 \in (s_1, \infty)$$

Lemma 2.6 (First order necessary conditions for local minimizers of $(CV - P2)$)

Suppose $(x^\star, s_2^\star)$ is a weak minimizer of $[CV - P2]$, then

1. $x^\star$ solves EE on $[s_1, s_2^\star]$
2. The transversal conditions
   A) $[l_{\dot{x}} + \varphi_x]\Big|_{x=x^\star, s=s_2^\star} = 0$
   B) $[l - \dot{x}^T l_{\dot{x}} + \varphi_s]\Big|_{x=x^\star, s=s_2^\star} = 0$

The transversal conditions "replace" the BC at s_2 because in $[CV - P2]$ we have none.

- 2A is a vector equation with n components $\rightarrow$ it replaces "$x(s_2) = x_2$"
- 2B is a scalar equation $\rightarrow$ it provides an equation to find s_2

"Partially" free end-point problems

Case 1 s_2 free, $x(s_2) = x_2$ given

Same as Lemma 2.6, but we only use 2B

(2A not needed bc the BC on $x(s_2)$ is given)

Case 2 s_2 fixed, $x(s_2)$ free

Same as Lemma 2.6, but we only use 2A

Case 3 $s_2, x(s_2)$ are free but related through $\psi : \mathbb{R} \rightarrow \mathbb{R}^n, \quad x(s_2) = \psi(s_2)$

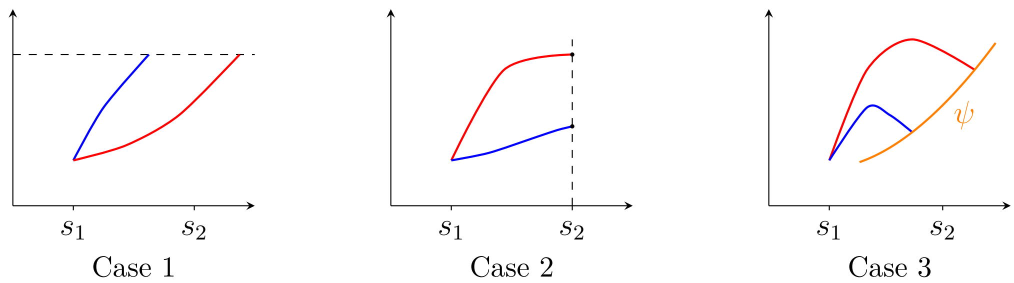
We do not need 2A $\rightarrow$ it is replaced by $x(s_2) = \psi(s_2)$.

We need an extension to 2B.

$$2B' : \left[l + l_{\dot{x}}^T(\dot{\psi} - \dot{x}) + \varphi_s + \varphi_x^T \dot{\psi}\right]_{x=x^\star, s=s_2^\star} = 0$$

Note $\psi(s_2) = x_2$

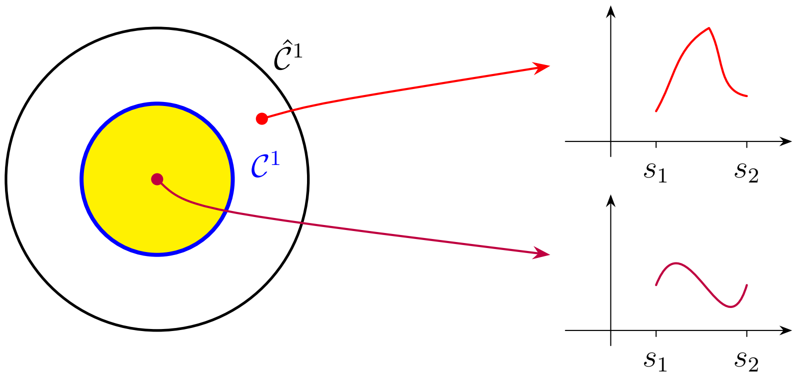
When $\dot{\psi} = 0$, 2B' collapses to 2B.



2.2.2 $V = \hat{\mathcal{C}}^1$ (Piecewise-continuously differentiable case)

Why?

1. "$\mathcal{C}^1 \subseteq \hat{\mathcal{C}}^1$" by enlarging our search space we can achieve better costs

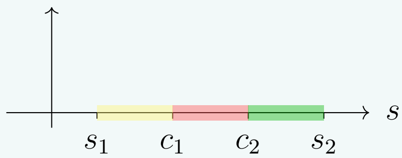


2. We can study conditions that are necessary for strong minimizers (only)

Lemma 2.7 (First-order necessary conditions for strong minimizers)

Consider $[CV-P1]$ with $x \in \hat{\mathcal{C}}^1([s_1, s_2])$.
Suppose $x^\star$ with corner points $\{c_i\}_{i=1}^N$ is a strong minimizer.
Then:

- $x^\star$ solves the EE inside the intervals (c_i, c_{i+1}), $\quad i = 0, 1, \ldots, N$
  $c_0 = s_1, \; c_{N+1} = s_2$



- At every corner point c_i the following continuity condition holds:

  $\boxed{A}$ $l_{\dot{x}}(c_i^-) = l_{\dot{x}}(c_i^+)$

  with $l(c_i^\pm) := l(c_i, x^\star(c_i), \dot{x}^\star(c_i^\pm))$

  $\boxed{B}$ $\left[-l(c_i^-) + \dot{x}^\star(c_i^-)^T l_{\dot{x}}(c_i^-)\right] = \left[-l(c_i^+) + \dot{x}^\star(c_i^+)^T l_{\dot{x}}(c_i^+)\right]$

  with $l_{\dot{x}}(c_i^\pm) := l_{\dot{x}}(c_i, x^\star(c_i), \dot{x}^\star(c_i^\pm))$

$\boxed{A}$ and $\boxed{B}$ are also called Weierstrass-Erdmann-conditions (WE).

$\boxed{A}$ is the 1^{st} WE-condition and is also necessary for weak minimizers.

$\boxed{B}$ is the 2^{nd} WE-condition and is only necessary for strong minimizers.

$\boxed{A}$ prescribes continuity of p (momentum)

$\boxed{B}$ prescribes continuity of H (Hamiltonian)

In the $\mathcal{C}^1$ case we need (and have) $2n$ boundary conditions.

In the $\hat{\mathcal{C}}^1$ case we need $2(N+1)n$ boundary conditions.

$$\left.\begin{array}{l} n \text{ at } c_0 = s_1 \\ n \text{ at } c_1^- \\ n \text{ at } c_1^+ \\ n \text{ at } c_2^- \\ \quad\vdots \\ n \text{ at } c_N^+ \\ n \text{ at } c_{N+1} = s_2 \end{array}\right\} 2(N+1)n \text{ in total.}$$

$\rightarrow$ The Problem $[CV-P1]$ still gives you only $2n$ conditions $x(s_1) = x_1,\ x(s_2) = x_2$.

$$\underbrace{2(N+1)n}_{\text{required}} - \underbrace{2n}_{\text{given by the BC}} = 2Nn$$

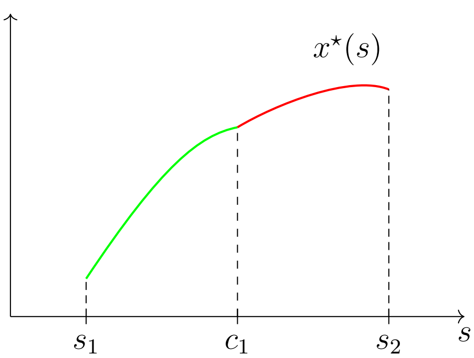
We have N corner points. At each c_i we can enforce:

- continuity of x $\quad$ (n conditions)
- $\boxed{A}$: continuity of p $\quad$ (n conditions)

$\rightarrow$ The problem is closed

Proof:

$x^\star \in \hat{\mathcal{C}}^1$ with 1 corner point $c_1 \in (s_1, s_2)$.

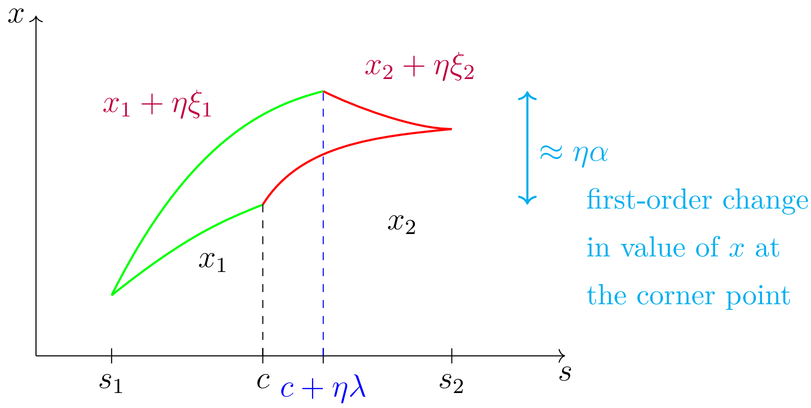


I will drop * from x.

$$x_1 : [s_1, c] \to \mathbb{R}^n \quad \xi_1 \text{ is the perturbation to } x_1$$
$$x_2 : [c, s_2] \to \mathbb{R}^n \quad \xi_2 \text{ is the perturbation to } x_2$$

The location of c can also be perturbed

$$c \to c + \eta\lambda$$



$$\begin{aligned} J(x + \eta\xi; c + \eta\lambda) &= \int_{s_1}^{s_2} l[x + \eta\xi]ds \\ &= \underbrace{\int_{s_1}^{c+\eta\lambda} l[x + \eta\xi]ds}_{J_1} + \underbrace{\int_{c+\eta\lambda}^{s_2} l[x + \eta\xi]ds}_{J_2} \end{aligned}$$

To show that (x, c) is optimal, δJ must be zero.

In fact $\delta J_1 = \delta J_2 = 0 \to x_1$ and x_2 are optimal in their intervals.

Enforcing $\delta J = 0$ results in this expression. (as well as continuity in the new corners $c + \eta\lambda$)

$$\forall\alpha, \forall\lambda : \underbrace{\left[l_{\dot{x}}(c_i^-) - l_{\dot{x}}(c_i^+)\right]}_{\boxed{A}}\alpha - \underbrace{\left[-l(c_i^-) - \dot{x}^\star(c_i^-)^T l_{\dot{x}}(c_i^-) + l(c_i^+) + \dot{x}^\star(c_i^+)^T l_{\dot{x}}(c_i^+)\right]}_{\boxed{B}}\lambda = 0$$

with α related to change in function value at the corner point and λ as the change of location of the corner points.

1. Because α, λ are independent, this means that $\boxed{A}$, $\boxed{B}$ must hold. The proof shows why only $\boxed{B}$ is necessary for strong minima.

   When c can change the perturbed curve

   $$(x + \eta\xi)$$

   is not in the weak ball of x.

   In the interval $[c, c + \eta\lambda]$ we have that

   $$\|(x + \eta\xi) - x\|_1 \approx \underbrace{\left\|\dot{x}(c^-) - \dot{x}(c^+)\right\|}_{\neq 0 \quad \forall \eta \neq 0}.$$

   This shows that the perturbation is outside of the weak ball around x. It is instead in the strong ball around x because $\|\cdot\|_\infty$ does not look at derivatives of the functions.

   If instead $\boxed{\lambda = 0}$ (no perturbation to c) then $\|(x + \eta\xi) - x\|_1 \approx \eta$.

   $\rightarrow$ The perturbed area is inside the weak ball of x.

2. Condition $\boxed{A}$ is not surprising after all.

   Recall the proof of EE, you can see, that the integral version

   $$l_{\dot{x}_i}[x^\star] - \int_{s_1}^{s} l_{x_i}[x^\star] d\sigma = c_i$$

   This shows already that $l_{\dot{x}}$ is always continuous if $x^\star$ satisfies the EE.

Definition 2.8 (Weierstrass excess function)

$$E(s, x, \dot{x}, w) := l(s, x, w) - \left[l(s, x, \dot{x}) + (w - \dot{x})^T l_{\dot{x}}(s, x, \dot{x})\right]$$

interpretation: difference between $l(s, x, w)$ and its first order approximation around $w = \dot{x}$.

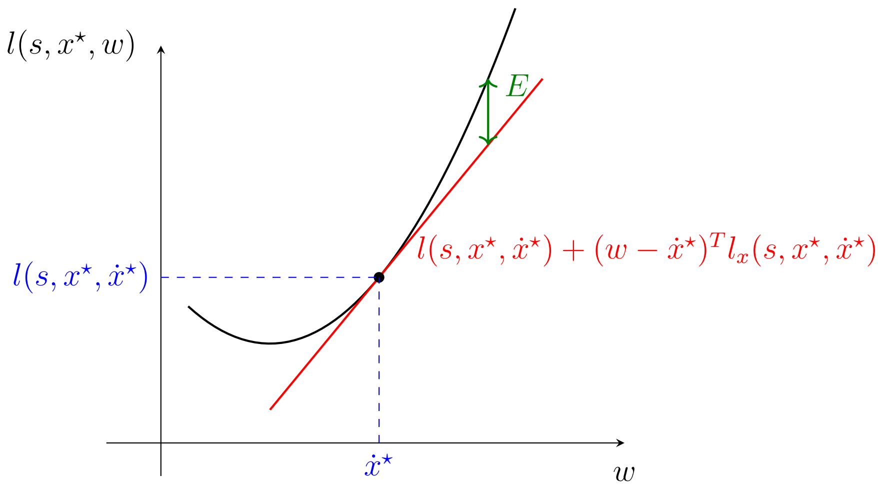
Lemma 2.8 (First-order necessary conditions for strong minimizers - the Weierstrass condition)

Consider $[CV-P1]$ $x \in \hat{\mathcal{C}}^1$.
Suppose $x^\star$ with corner point $\{c_i\}_{i=1}^N$ is a strong minimizer.
Then

- $x^\star$ solves EE inside the intervals (c_i, c_{i+1}), $\quad i = 0, 1, \ldots, N$
- $E(s, x, \dot{x}, w) \geq 0 \quad \forall w \in \mathbb{R}^n, \quad \forall s \in [s_1, s_2]$ except corner points

w condition implies $\boxed{A}$ and $\boxed{B}$ from previous Lemma (it is a stronger condition).

interpretation: for a given solution $x^\star$, we draw for fixed $s \in [s_1, s_2]$ the following curve:



This can be interpreted as a convexity requirement of the function $l(s, x^\star, \cdot)$ for fixed $s, x^\star$.

So this is a weaker requirement than joint convexity of l
$\quad \rightarrow$ we only need to check that at $x^\star$

The Weierstrass condition can equivalently be written as a maximization condition on H:

$$\underbrace{E(s, x^\star, \dot{x}^\star, w) \geq 0 \quad \Leftrightarrow \quad H(s, x^\star, \dot{x}^\star, p^\star) - H(s, x^\star, w, p^\star) \geq 0}_{\text{substitude definition of } E \text{ in } E \geq 0} \qquad \forall w \in \mathbb{R}^n$$

$\rightarrow H(s, x^\star, \cdot, p^\star)$ has a maximum at $w = \dot{x}^\star$

2.3 Isoperimetric constraints

$[CV - P3]$ $\min_{x(\cdot)} \int_{s_1}^{s_2} l(s, x, \dot{x}) ds$

$$\text{s.t. } x \in \left\{ y \in \mathcal{C}^1 \mid \underbrace{\int_{s_1}^{s_2} g_i(s, y(s), \dot{y}(s)) ds}_{G_i(y): V \to \mathbb{R}, \quad k_i \in \mathbb{R}} = k_i, \quad i = 1, \ldots, n_g;\ y(s_1) = x_1;\ y(s_2) = x_2 \right\}$$

Lemma 2.9

Consider $[CV - P3]$ and assume the following regularity condition

$$\text{Det}\left(G(\{\xi_i\}_{i=1}^{n_g})\right) \neq 0$$

$$(G(\{\xi_i\}_{i=1}^{n_g})) := \begin{bmatrix} \delta G_1(x^\star, \xi_1) & \cdots & \delta G_1(x^\star, \xi_{n_g}) \\ \vdots & \ddots & \vdots \\ \delta G_{n_g}(x^\star, \xi_1) & \cdots & \delta G_{n_g}(x^\star, \xi_{n_g}) \end{bmatrix}$$

for n_g independent directions $\{\xi_i\}_{i=1}^{n_g} \in V$.
If $x^\star$ is is a local minimizer of J, then $\exists \lambda \in \mathbb{R}^{n_g}$ such that

$$\frac{d}{ds} \mathcal{L}_{\dot{x}_i}(s, x^\star(s), \dot{x}^\star(s)) = \mathcal{L}_{x_i}(s, x^\star(s), \dot{x}^\star(s)), \quad \forall s \in [s_1, s_2], \quad i = 1, \ldots, n$$

$\mathcal{L}(s, x, \dot{x}) := l(s, x, \dot{x}) + \lambda^T g(s, x, \dot{x}),$ Lagrangian

$$g := \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{n_g} \end{bmatrix}^T$$

3 The CV approach to optimal control

3.1 Intro to CV problems

$$\min_{u \in V} J(u)$$

$$[OC] \qquad \text{s.t. } \dot{x} = f(t, x, u), \quad x(t_0) = x_0, \quad t \in [t_0, t_f]$$

$$(t_f, x(t_f)) \in S \subseteq (t_0, \infty) \times \mathbb{R}^{n_x}$$

Remarks:

- t_0, t_f: initial and final time: $\begin{array}{l} t_0 \in (-\infty, \infty) \\ t_f \in (t_0, \infty) \text{ finite horizon problem} \\ t_f = +\infty \text{ infinite horizon problem} \end{array}$

  t_0 always given

  t_f can be given or free variable (like s_2 in CV)

- $f : [t_0, t_f] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$

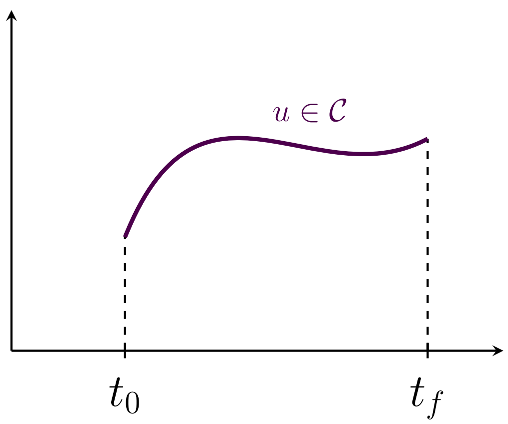
- $(V, \|\cdot\|)$ normed vector space where u belongs

$$u : [t_0, t_f] \to \mathcal{U} \subseteq \mathbb{R}^{n_u}, \qquad \mathcal{U} \subset \mathbb{R}^{n_u} \text{ if we have input contraints}$$

2 classes of functions:

- $V = \mathcal{C}([t_0, t_f], \mathcal{U})$ continuous

- $V = \hat{\mathcal{C}}([t_0, t_f], \mathcal{U})$ piecewise continuous

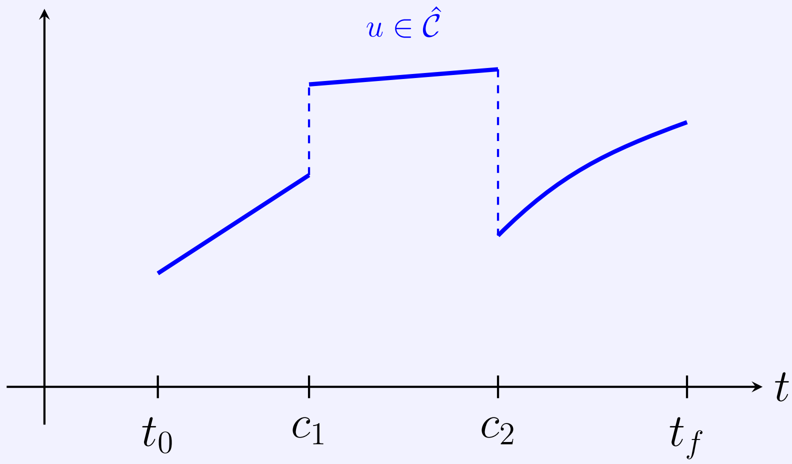


Definition 3.1 (Piecewise continuous function $\hat{\mathcal{C}}$)

$u \in \hat{\mathcal{C}}$ if there is a finite partition $\{c_k\}_{k=0}^{N+1}$ with $t_0 = c_0 < c_1 < \cdots < c_N < c_{N+1} = t_f$ such that

$$u : [c_k, c_{k+1}] \to \mathbb{R}^{n_u} \text{ is continuous}$$

$$\{c_k\}_{k=1}^{N} : \text{ corner points}$$



- $J : V \to \mathbb{R}$
  - Lagrangian form
  $$J(u) = \int_{t_0}^{t_f} l(t, x, u)\, dt, \qquad l\text{: running cost}$$
  - Bolza form
  $$J(u) = \varphi(t_f, x(t_f)) + \int_{t_0}^{t_f} l(t, x, u)\, dt, \qquad \varphi\text{: terminal cost}$$
  - Mayer form
  $$J(u) = \varphi(t_f, x(t_f))$$

  These terms are fully interchangeable: We can go from one to the others by reformulating the problem.

  $L \Rightarrow M$

- Introduce fictitious states: x_l evolving according to $\begin{aligned} \dot{x}_l &= l(t, x, u) \\ x_l(t_0) &= 0 \end{aligned}$

Our new system $\tilde{x} = \begin{bmatrix} x \\ x_l \end{bmatrix}$

$$J(u) = \int_{t_0}^{t_f} l(t, x, u)\, dt = \varphi(t_f, \tilde{x}(t_f)) = x_l(t_f)$$

$M \Rightarrow L$?

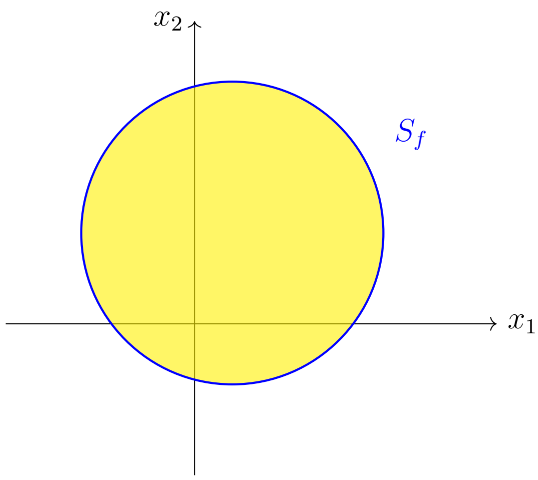
- S is the target set. Examples:
  - free-time free-end point case:

$$S = \underbrace{(t_0, \infty)}_{t_f} \times \underbrace{\mathbb{R}^{n_x}}_{x(t_f)}$$

  - free-time constrained endpoint case:

$$S = (t_0, \infty) \times S_f, \qquad S_f \subseteq \mathbb{R}^{n_x}$$

  E.g. $S_f = \{x_f\}$



  - fixed-time fixed-end point case:

$$S = \{t_f\} \times \{x_f\}$$

- Standing assumptions:
  - l is $\mathcal{C}$ in (t, x, u) and $\mathcal{C}^1$ in x
  - f is $\mathcal{C}$ in (t, x, u) and $\mathcal{C}^1$ in x

In the first part of section 3 we also assume f, l are $\mathcal{C}^1$ in u.

Definition 3.2 (Strong and weak norms)

$V = \mathcal{C}([t_0, t_f], U)$

- $\|\cdot\|_\infty$ strong norm

$$\|u\|_\infty := \max_{t_0 \le t \le t_f} \|u(t)\|$$

- $\|\cdot\|_1$ weak norm

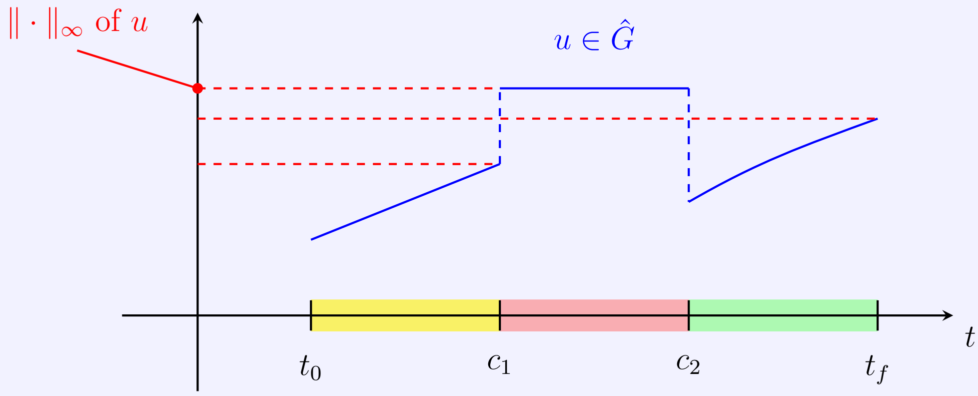
$$\|u\|_1 := \begin{cases} \|u\|_\infty + \max\limits_{t_0 \le t \le t_f} \|\dot{u}(t)\| & \text{if } u \in \mathcal{C}^1 \\ \|u\|_\infty + \sup\limits_{t \in \bigcup_{k=0}^{N} (\hat{c}_k, \hat{c}_{k+1})} \|\dot{u}(t)\| & \text{if } u \in \hat{\mathcal{C}}^1 \end{cases}$$

where $\{\hat{c}_k\}$ are corner points of $\dot{u}$.

$V = \hat{\mathcal{C}}([t_0, t_f], U)$

- $\|\cdot\|_\infty$ strong norm

$$\|u\|_\infty := \sup_{t \in \bigcup_{k=0}^{N} (c_k, c_{k+1})} \|u(t)\|$$



- $\|\cdot\|_1$ weak norm: same rationale

Constraints

- point constraints: Constraints on x/u on specific time points

$$\Psi_1(t, x(t), u(t)) \begin{cases} = 0 \\ \ge 0 \end{cases} \text{at } t = \bar{t}$$

- path constraints:

$$\Psi_2(t, x(t), u(t)) \begin{cases} = 0 \\ \geq 0 \end{cases} \quad \forall t \in [t_1, t_2]$$

- isoperimetric constraints:

$$\int_{t_0}^{t_f} g(t, x, u)\, dt \leq G$$

Ensuring constraint satisfaction in OC problems is hard.

In this course:

- point constraints only at $t = t_f$
- path constraints only on $u \to u : [t_0, t_f] \to \mathcal{U}, \qquad \mathcal{U} = \{u : \Psi_2(t, u) = 0\}$
- isoperimetric constraints

If u satisfies constraints it is called "admissible control" and is denoted by $u \in \mathcal{D}$

Definition 3.3 (Global and local minima)

Admissible control $u^\star(\cdot)$ is a

global minimizer of $[OC]$ if

$$J(u) \geq J(u^\star), \qquad \forall u \in \mathcal{D}$$

strong local minimizer if

$$\exists \epsilon > 0 \text{ s.t. } J(u) \geq J(u^\star), \qquad \forall u \in \mathcal{D} \cap B_\epsilon^\infty(u^\star)$$

is a weak local minimizer if

$$\exists \epsilon > 0 \text{ s.t. } J(u) \geq J(u^\star), \qquad \forall u \in \mathcal{D} \cap B_\epsilon^1(u^\star)$$

Reflection: $u^\star$ is an open-loop controller. We see this from how we defined the OC problem here ("optimizer over $u(\cdot) : [t_0, t_f] \to U$") which are just functions of time. In **??** we change viewport and study closed-loop optimal controller $u(\underset{t}{\cdot}, \underset{x}{\cdot})$ with Dynamic Programming.

3.2 Unconstrained problems and weak minima

$[OC-P1]$ $$\min_{u(\cdot)\in V} \int_{t_0}^{t_f} l(t,x,u)\,dt$$

$$\dot{x} = f(t,x,u), \quad x(t_0) = x_0, \quad S = \{t_f\} \times \mathbb{R}^{n_x}$$

3.2.1 $V = \mathcal{C}([t_0,t_f])$

Note on regularity of solution of $\dot{x} = f(t,x,u)$:

if $u \in \mathcal{C}, f \in \mathcal{C}$, then $\dot{x} \in \mathcal{C} \Leftrightarrow x \in \mathcal{C}^1$

if $u \in \hat{\mathcal{C}}, f \in \mathcal{C}$, then $\dot{x} \in \hat{\mathcal{C}} \Leftrightarrow x \in \hat{\mathcal{C}}^1$

Lemma 3.1 (First-Order Necessary Conditions for unconstrained OC)

Suppose $u^\star$ is a weak minimum of $[OC-P1]$ and $x^\star \in \mathcal{C}^1$ the associated response. Then $\exists \lambda^\star \in \mathcal{C}^1$ (called adjoint or costate) such that $(x^\star, u^\star, \lambda^\star)$ satisfies:

$$\left.\begin{array}{ll} n_x & \dot{x}^\star = f(t,x^\star,u^\star), \quad x^\star(t_0) = x_0 \\ n_x & \dot{\lambda}^\star = l_x(t,x^\star,u^\star) - f(t,x^\star,u^\star)\lambda^\star, \quad \lambda^\star(t_f) = 0 \\ n_u & 0 = -l_u(t,x^\star,u^\star) + f_u(t,x^\star,u^\star)\lambda^\star, \quad \forall t \in [t_0,t_f] \end{array}\right\} \text{Euler Lagrange equations (ELE)}$$

where $[f_x]_{i,j} = \frac{\partial f_j}{\partial x_i}$, $\quad [f_u]_{i,j} = \frac{\partial f_j}{\partial u_i}$.

$\underset{x}{n_x} + \underset{\lambda}{n_x} + \underset{u}{n_u}$ unknowns and $2n_x ODE + n_u$ algebraic equations $\Rightarrow$ Problem is closed

Proof: Notation: I will drop * from x, u, λ.

Rationale: Cast problem $[OC-P1]$ as a CV problem.

Then use the fundamental Lemma of CV.

What is the difference between OC and CV? $\dot{x} = f(t,x,u)$

Conceptional step: given u (candidate), x is fixed. In other words, think of x as $x(t,u)$.

$\rightarrow$ Rewrite J of $[OC-P1]$ as follows:

$$J(u) = \int_{t_0}^{t_f} l(t, x(t,u), u(t))\,dt$$

$$= \int_{t_0}^{t_f} l(t, x(t,u), u(t)) + \underbrace{\lambda^T(t)}_{\mathcal{C}^1 \text{ function}} \Big(\underbrace{\dot{x}(t,u) - f(t, x(t,u), u(t))}_{=0} \Big)\, dt$$

This functional is equivalent to the one in $[OC - P1]$ and it is only a function of u. "There's no more x".

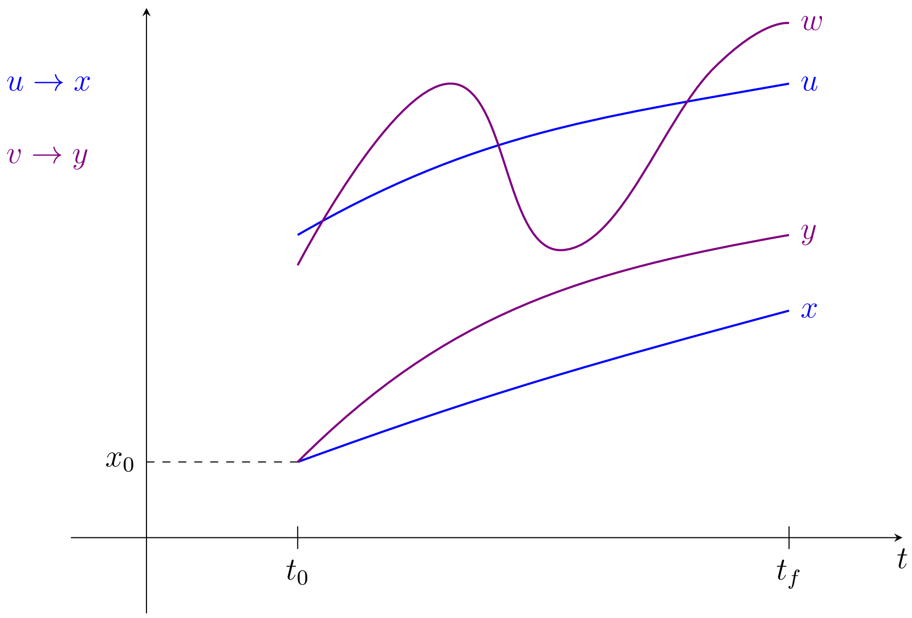
Thus we now have the following CV problem

$$\min_{u(\cdot)} \int_{t_0}^{t_f} l(t, x(t,u), u) + \lambda^T(t) \dots\, dt, \quad \text{There is only } u \text{ inside } J(u)$$

Step 2: enforce $\delta J = 0$ for this CV problem:

Perturb candidate u:

$$\nu(t, \eta) := u(t) + \eta\omega(t), \quad \omega(t) \in \mathcal{C}([t_0, t_f])$$



$y(t, \eta) \in \mathcal{C}^1$ is the state response of the system under ν.
$x(t) \in \mathcal{C}^1$ is the state response of the system under u.

$y_\eta(t; \eta) := \frac{\partial y}{\partial \eta}$: $y(t, 0) = x(t), \quad \forall t \in [t_0, t_f]$ by definition

Write J for the perturbed u:

$$J(\nu(\cdot; \eta)) = \int_{t_0}^{t_f} l\Big(t, y(t;\eta), \nu(t;\eta)\Big) + \lambda^T(t)[\dot{y}(t;\eta) - f\Big(t, y(t;\eta), \nu(t;\eta)\Big)]\, dt$$

$$= \int_{t_0}^{t_f} \Big(l(t,y,\nu) - \dot\lambda^T y - \lambda^T f(t,y,\nu) \Big)\, dt + \lambda^T(t_f) y(t_f;\eta) - \lambda^T(t_0) y(t_0;\eta)$$

↳integral by parts applied to $\lambda^T \dot y$

Now we write the derivative $\frac{\partial J}{\partial \eta}$ and set $\eta \to 0$.

$$\frac{\partial J}{\partial \eta}(\nu) = \int_{t_0}^{t_f} \underline{[l_u(t,y,\nu) - f_u(t,y,\nu)\lambda]^T \omega} + \underline{[l_x(t,y,\nu) - f_x(t,y,\nu)\lambda - \dot\lambda]^T y_\eta}\, dt$$

$$+ \underline{\lambda^T(t_f) y_\eta(t_f;\eta)} - \underbrace{\lambda^T(t_0) y_\eta(t_0;\eta)}_{=0 \text{ because } y(t_0;\eta)=x_0\ \forall \eta}$$

Take $\eta \to 0$:

$$\delta J(u;\omega) = \int_{t_0}^{t_f} \underline{[l_u(t,x,u) - f_u(t,x,u)\lambda]^T \omega} + \underline{[l_x(t,x,u) - f_x(t,x,u)\lambda - \dot\lambda]^T y_\eta}\, dt$$

$$+ \underline{\lambda^T(t_f) y_\eta(t_f;0)}$$

$\delta J = 0 \quad \forall \omega\ \forall \lambda$ for u to be a candidate minimizer.

We are free to choose any ω, λ we want:

1. We can choose λ as follows:

$$\dot\lambda = -l_x - f_x^T \lambda \quad \text{with BC } \lambda(t_f) = 0 \quad \text{2nd ELE + BC}$$

   This eliminates the green and blue terms.

2. To set the red term to 0, we can choose n_u "special" perturbations $\omega^{(i)}$ defined as follows: $\begin{cases} \omega_i^{(i)} = l_{u_i} - f_{u_i}^T \lambda \\ \omega_j^{(i)} = 0, & \forall j \neq i \end{cases}$ where $\omega^{(i)} = \begin{bmatrix} \omega_1^{(i)} \\ \vdots \\ \omega_{n_u}^{(i)} \end{bmatrix}, \quad i = 1, \dots, n_u$

   This yields:

$$0 = \int_{t_0}^{t_f} \left[l_{u_i} - f_{u_i}^T \lambda \right]^2 dt, \quad \forall i = 1, \dots, n_u$$

   Which is only possible if:

$$l_{u_i} - f_{u_i}^T \lambda = 0, \quad i = 1, 2, \dots, n_u \quad \text{3rd ELE eq.}$$

1^{st} ELE is just the dynamic equation for x. □

Remark:

- we always have $\begin{matrix} n_x \text{ BC at } t_0 \\ n_x \text{ BC at } t_f \end{matrix}$ $\rightarrow$ This is a TPBVP

- Typically we extract from the 3^{rd} equation a relationship between u and x, λ:

$$\rightarrow u(x, \lambda).$$

  In this case we can replace $u(x, \lambda)$ in the ODEs. These ODEs then are a system of equations in x, λ. We solve for x, λ and we find $u(x, \lambda)$.

- What happens to the ELE when $f(t, x, u) = u \rightarrow \dot{x} = u$

$$\begin{cases} \dot{x} = u & x(t_0) = x_0 \\ \dot{\lambda} = l_x(t, x, \dot{x}) & \lambda(t_f) = 0 \\ \lambda = l_u(t, x, \dot{x}) \end{cases}$$

$$\Leftrightarrow \begin{cases} \dot{x} = u & x(t_0) = x_0 \\ \frac{d}{dt} l_u(t, x, \dot{x}) = l_x(t, x, \dot{x}) & [l_u(t, x, \dot{x})]|_{t=t_f} = 0 \\ \lambda = l_u(t, x, \dot{x}) \end{cases}$$

  These are the EE!

  $\frac{dl_u}{dt} = l_x, \quad [l_u]|_{t=t_f} = 0$ is the transversality condition for the case s_2 fixed and $x(s_2)$ free in the CV problem. **Surprising?**

  If we write $[OC - P1]$ for $\dot{x} = u$ we get

$$\min_{x(\cdot)} \int_{t_0}^{t_f} l(t, x, \underbrace{\dot{x}}_{=u}) \, dt$$

$$\text{s.t. } x(t_0) = x_0, \quad t_f \text{ given}, \quad x(t_f) \text{ free}$$

  This problem is equivalent to finding a curve $x(s)$, $s \in [s_1, s_2]$ where $s_1 = t_0$, $s_2 = t_f$.

  In this case: $\lambda(t) = l_u(t, x, \dot{x})|_{x(t)} = p(t)$

  Therefore we can think of the adjoint as the momentum for the OC problem.

- We can introduce the Hamiltonian for OC problems:

In CV $\rightarrow \mathcal{H}(s, x, \dot{x}, p) = -l(s, x, \dot{x}) + \dot{x}^T p$

In OC $\rightarrow \mathcal{H}(t, x, u, \lambda) = -l(t, x, u) + f^T(t, x, u)\lambda$

Once we have defined $\mathcal{H}$, we can rewrite ELE compactly:

H	
A	D
M	Y
I	N
L	A
T	M
O	I
N	C
I	S
A	
N	

$$\begin{aligned}
\dot{x}^\star &= \mathcal{H}_\lambda(t, x^\star, u^\star, \lambda^\star), & x^\star(t_0) &= x_0 \\
\dot{\lambda}^\star &= -\mathcal{H}_x(t, x^\star, u^\star, \lambda^\star), & \lambda^\star(t_0) &= 0 \\
0 &= \mathcal{H}_u(t, x^\star, u^\star, \lambda^\star), & \forall t &\in [t_0, t_f]
\end{aligned}$$

Candidate optimal controllers
are stationary points for $\mathcal{H}$

- $\left.\frac{d\mathcal{H}}{dt}\right|_{\substack{\text{evaluated at} \\ x^\star, \lambda^\star, u^\star \text{ solutions} \\ \text{of ELE}}} = \mathcal{H}_t + \mathcal{H}_x^T \dot{x} + \mathcal{H}_u^T \dot{u} + f^T \dot{\lambda}$

$$= \mathcal{H}_t + \underbrace{\mathcal{H}_u^T \dot{u}}_{=0} + f^T (\underbrace{\mathcal{H}_x + \dot{\lambda}}_{=0}) = \mathcal{H}_t = \frac{\partial \mathcal{H}(t, x, u, \lambda)}{\partial t}$$

$\mathcal{H}_t \neq 0$ only if

$f(t, x, u)$ depends on time and/or

$l(t, x, u)$ depends on time

Time-invariant problems (dynamics do not depend on t and running cost neither) have $\mathcal{H}_t = 0$.

Thus: $\mathcal{H}$ is constant over ELE solutions in time-invariant problems.

- Consider the case where $S = \{t_f\} \times \underbrace{\{x_f\}}_{1}$ 1 fixed end-point

What changes in Lemma 3.1?

We remove $\lambda^\star(t_f) = 0$ and add $x(t_f) = x_f$. All the rest stay the same.

If we have 2^{nd} order regularity properties on f, l, we can derive 2^{nd} order necessary conditions for weak minima.

Lemma 3.2 (Second order necessary conditions for opitmal control problems)

Consider $[OC - P1]$ with the standing assumptions and $u \in \mathcal{C}$.
Assume also that f and l have continuous second-order derivative in u.

$$(\nabla_{uu} l,\ \nabla_{uu} f \text{ exist})$$

Suppose $u^\star$ is a weak minimum and $x^\star$ its response.
Then $(x^\star, u^\star, \lambda^\star)$ satisfies ELE and also

$$-\nabla_{uu} l(t, x^\star, u^\star) + \nabla_{uu}(f^T(t, x^\star, u^\star)\lambda^\star) \preceq 0 \qquad \forall t \in [t_0, t_f]$$

(Legendre-Clebsch condition).
Equivalently:

$$\mathcal{H}_{uu}(t, x^\star, u^\star, \lambda^\star) \preceq 0 \qquad \forall t \in [t_0, t_f]$$

Proof: skipped: It proceeds similary to the proof of Legendre condition in CV by deriving $\delta^2 J = 0$.

Note: If $S = \{t_f\} \times \{x_f\}$, this Lemma still applies. The cange of S only affects the boundary conditions of the ELE.

What about sufficient conditions?

Lemma 3.3 (First order sufficient conditions for unconstrained OC - Mangasarian conditions)

Consider $[OC - P1]$ with standard assumptions and $u \in \mathcal{C}$.
Assume also that

$$l(t, x, u) \text{ and } f(t, x, u) \text{ are jointly cvx in } x \text{ and } u \ \forall t \in [t_0, t_f]$$

If

- $(u^\star, x^\star, \lambda^\star)$ satisfies ELE,
- $\lambda^\star(t) \leq 0 \quad \forall t \in [t_0, t_f]$.

Then $u^\star$ is a global minimizer.

Proof: The goal is to show that

$$J(u) - J(u^\star) \geq 0 \quad \forall u \in \mathcal{D}$$

$$J(u) - J(u^\star) = \int_{t_0}^{t_f} \left(l(t,x,u) - l(t,x^\star,u^\star)\right) dt \qquad \Big| \text{ jointly convex } l \text{ wrt } x, u$$

$$\geq \int_{t_0}^{t_f} \underbrace{l_x^\star}_{l_x(t,x^\star,u^\star)} (x - x^\star) + \underbrace{l_u^{\star T}}_{l_u(t,x^\star,u^\star)} (u - u^\star)\, dt \qquad \Big| \; x^\star, u^\star, \lambda^\star \text{ satisfy ELE}$$

$$= \int_{t_0}^{t_f} [f_x^\star \lambda^\star - \dot{\lambda}^\star]^T (x - x^\star) + [f_u^\star \lambda^\star]^T (u - u^\star)\, dt \qquad \Big| \text{ Integration by parts for } \dot{\lambda}^{\star T}(x - x^\star)$$

$$= \int_{t_0}^{t_f} \lambda^{\star T} \left[f_x^{\star T}(x - x^\star) + f_u^{\star T}(u - u^\star) - \Big(f(t,x,u) - f(t,x^\star,u^\star) \Big) \right] dt$$

$$+ \underbrace{\lambda^{\star T}(t_f)}_{\text{=0 because BC of ELE}} (x(t_f) - x^\star(t_f)) - \lambda^{\star T}(t_0) \Big(\underbrace{x(t_0) - x^\star(t_0)}_{\substack{\text{=0 because } x(t_0)=x^\star(t_0)=x_0 \text{ BC of ELE} \\ \text{(all responses must start at } x_0)}} \Big)$$

In summary:

$$J(u) - J(u^\star) = \int_{t_0}^{t_f} \underbrace{\lambda^{\star T}}_{\boxed{A}} \underbrace{\left[f_x^{\star T}(x - x^\star) + f_u^{\star T}(u - u^\star) - \Big(f(t,x,u) - f(t,x^\star,u^\star) \Big) \right]}_{\boxed{B}} dt$$

By joint convexity of f we have

$$f(t,x,u) \geq f(t,x^\star,u^\star) + f_x^{\star T}(x - x^\star) + f_u^{\star T}(u - u^\star) \quad \forall x, u$$

$$\Leftrightarrow \underbrace{f_x^{\star T}(x - x^\star) + f_u^{\star T}(u - u^\star) - \Big(f(t,x,u) - f(t,x^\star,u^\star) \Big)}_{\boxed{B}} \leq 0 \quad \forall x, u$$

$\boxed{B} \leq 0 \rightarrow \text{If } \lambda^\star \leq 0 \underset{\text{assumption}}{\rightarrow} \int\limits_{t_0}^{t_f} \boxed{A} \cdot \boxed{B} dt \geq 0$

Therefore: $J(u) - J(u^\star) \geq 0 \quad \forall u \in \mathcal{D}$ $\square$

Remarks:

- When f is linear, can we relax the assumptions of the Lemma?

$$f \text{ linear} \rightarrow \boxed{B} = 0$$

  Thus, we do not need to require $\lambda^\star(t) \leq 0 \quad \forall t$

- What happens if $S = \{t_f\} \times \{x_f\}$?

  The Lemma holds exactly the same. In the proof instead of having $\lambda^\star(t_f) = 0$ we have that $x^\star(t_f) = x(t_f) = x_f$.

- When f is jointly concave ($\leftrightarrow -f$ is convex).

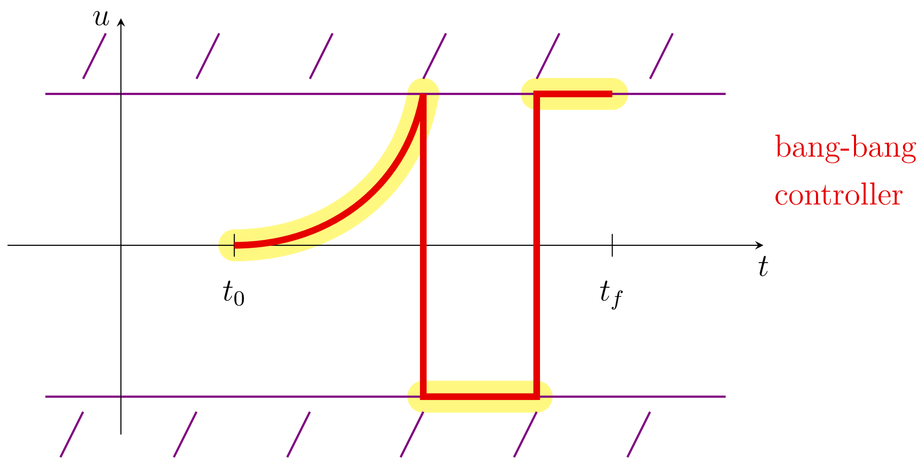
  Then we require $\lambda^\star \geq 0$. The rest ist still valid.

3.2.2 Piecewise continuous control

$$u \in \hat{\mathcal{C}}, \quad S = \{t_f\} \times \mathbb{R}^{n_x}$$
$$S = \{t_f\} \times \{x_f\}$$

Why?

- For some OC problems, a continuous solution to ELE might not exist.
  - This happens very frequently when we have input constraints.



  - Usually u is obtained by solving

$$0 = \mathcal{H}_u(t, x, u, \lambda)$$

$\rightarrow u = u_1(x, \lambda),\ u = u_2(x, \lambda),\ \ldots$ might have more than 1 solution

even without constraints, u might jump from u_1 to u_2.

- To improve the cost J.

  $u \in \hat{\mathcal{C}}$ is a larger class than $u \in \mathcal{C}$.