

Optimal Control

Wintersemester 2025/26

Dozent: Prof. Dr. Andrea Iannelli

November 14, 2025

Contents

0	Introduction	1
1	Nonlinear Programming	3
1.1	Unconstrained Problems	8
1.2	Constrained Problems	12
2	Calculus of Variations	18
2.1	Introduction to CV (Calculus of Variations)	19
2.2	Free problems of CV	27
2.2.1	$V = \mathcal{C}^1$	27
2.2.2	$V = \hat{\mathcal{C}}^1$ (Piecewise-continuously differentiable case)	33
2.3	Isoperimetric constraints	39

0 Introduction

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

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

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

Initial Value Problem (IVP)

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

\curvearrowright functions of time \curvearrowright

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

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

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
- Continuously differentiable (\mathcal{C}^1) \Rightarrow locally Lipschitz continuous $\forall x_1, x_2 \in \mathcal{X} \subset \mathbb{R}^{n_x}$
- Locally Lipschitz continuous \times guarantees existence & uniqueness for small enough times

In this course we will assume $f \in \mathcal{C}^1$ and implicitly 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. $u(t) \in \mathcal{U}(t), x(t) \in \mathcal{X}(t) \quad \forall t \in [t_0, t_f], \quad \mathcal{X} \subseteq \mathbb{R}^{n_x}, \mathcal{U} \subseteq \mathbb{R}^{n_u}$
 $\uparrow \qquad \qquad \uparrow$
 sets defining constraints on u & x

\Rightarrow Admissible input/state trajectories

2. The system behaves optimally according to

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

Cost function running cost terminal cost

\Rightarrow optimal behaviour

Formally, we can state the goal as follows:

Find an admissible input u^* which causes the dynamics to follow an admissible trajectory x^* which minimizes J , that is

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

\forall 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

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

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

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:

$$\min_x f(x)$$

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

$f : \mathbb{R}^n \rightarrow \mathbb{R}$, objective function

$g : \mathbb{R}^n \rightarrow \mathbb{R}^{n_g}$, inequality constraints

$h : \mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$, 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)

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

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

$f(x^*)$ is the Global Minimum (or Minimum)

Nomenclature: x^* is also called (optimal) solution, $F(x^*)$ is optimal value

x^* is a strict global minimizer if $f(x^*) < f(x) \quad \forall x \in \mathcal{D}$

$x^* \in \mathcal{D}$ Local Minimizer if

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

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

Strict local Minimizer if inequality holds strictly

Global min \nRightarrow 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} :

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

Example $\mathcal{S} = [-1, 1]$, $-50 = \inf(\mathcal{S})?$ \rightarrow 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 $\rightarrow \inf(\mathcal{S}) = -\infty$
- $\inf([a, b]) = \inf((a, b]) = a$

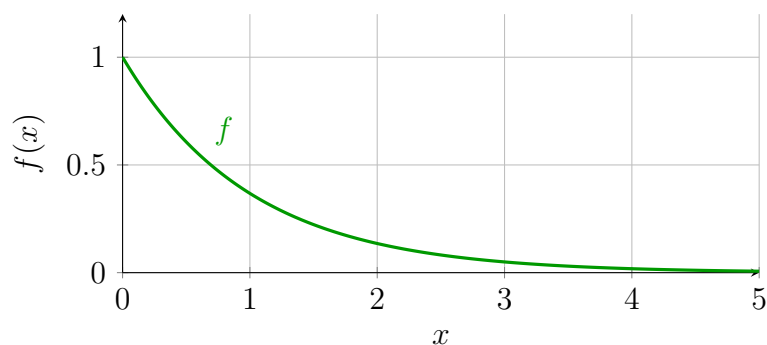
Connections with NLP?

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

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

Whenever NLP has solution, then NLP is equivalent to this, but $\nexists x^* \in \mathcal{D}$ s. t. $f(x^*) = \bar{f} \rightarrow$ 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)

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

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)

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

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

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

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

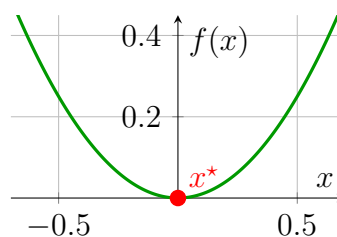
Not compact:

- $(a, b]$ (not closed)
- $(-\infty, b]$ (unbounded)

Compact set:

- $[a, b]$ – $-\infty < a < b < \infty$

Warning: \mathcal{D} infinite dimensional (e.g. function space) then
compact \nRightarrow 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:

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

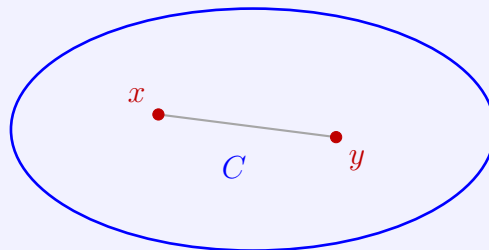
Goal of the Chapter: characterize necessary and sufficient conditions for x^* to be global minimizer of NLP.

Convexity

Definition 1.4 (Convex sets & functions)

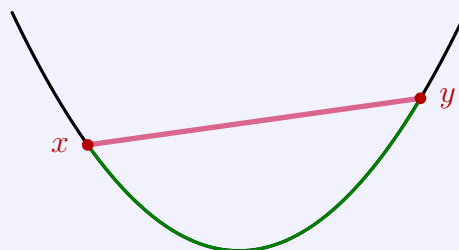
- A set $C \subseteq \mathbb{R}^n$ is convex (cvx) if $\forall x, y \in C$

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



- Given a cvx set C , a function $f : C \rightarrow \mathbb{R}$ 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 $f : C \rightarrow \mathbb{R}^n$ for convex f_i
- $f : C_1 \times C_2 \rightarrow \mathbb{R}$
 $f(x, y)$ is jointly cvx, in x, y if $z := \begin{bmatrix} x \\ y \end{bmatrix}$, $f(z)$ is in cvx in z .

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

Definition 1.5

An NLP is a convex program if

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

Lemma 1.1

Let x^* be a local minimizer of cvx program. Then x^* is also global minimizer.

Proof: try as an exercise

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

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

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

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

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

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

- $A \succeq 0$ means that: $A = A^T$ and pos semi-definite, i. e. all eigenvalues non-negative
- f strictly cvx if $\nabla_{xx}^2 f(x) \succ 0 \quad \forall x \in C$ 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 generally useful. First condition is useful for proofs.

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

- g are convex functions,
 - h are affine functions (i.e. $h(x) = 0 \Leftrightarrow Ax = b$),
- } sufficient

then \mathcal{D} is a convex set.

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

$\min_x f$		$\min_x f$
s.t. $x^3 - 1 \leq 0$	\Leftrightarrow	s.t. $x - 1 \leq 0$
$(ax + b)^2 = 0$		$ax + b = 0$
non-convex		convex
non-affine		affine

Moral to recognize convexity of NLP:

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

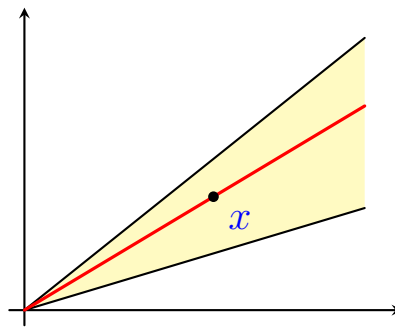
$$\exists \delta > 0 \quad \text{s.t.} \quad f(\bar{x} + \lambda d) < f(\bar{x}) \quad \forall \lambda \in (0, \delta).$$

$F(\bar{x})$: 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^* to be a local minimizer.

Lemma 1.3 (Geometric Condition for local minimum)

x^* is a local minimizer iff

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

We want an algebraic condition to be able to compute or look for x^* .

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 all descent directions.

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

If x^* is a local minimizer, then

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

Proof: Contradiction

If $\nabla f(x^*) \neq 0$, then $d = -\nabla f(x^*) \neq 0$. Therefore there exists a descent direction $d \in \mathcal{F}(x^*)$ by Lemma 1.4. Thus $\exists \delta > 0$ s.t. $f(x^* + \lambda d) < f(x^*) \quad \forall \lambda \in (0, \delta)$.

This is a contradiction with the fact, that x^* is a minimizer. \square

Why only necessary?

It can't be a sufficient condition because in case where $\nabla f(x^*) = 0$ 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 $f \in \mathcal{C}^2$

$$x^* \text{ local minimizer} \Rightarrow \nabla_{xx}^2 f(x^*) \succeq 0$$

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

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

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

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

1. If x^* is local minimzer $\Rightarrow \nabla f(x^*) = 0$

2. Divide by λ^2 :

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

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

$$\frac{f(x^* + \lambda d) - f(x^*)}{\lambda^2} \approx \frac{1}{2} d^T \nabla_{xx}^2 f(x^*) d$$

4. For x^* is a local minimizer, the left-hand-side must be ≥ 0 for any $d \in \mathbb{R}^n$

$$\Rightarrow d^T \nabla_{xx}^2 f(x^*) d \geq 0 \quad \forall d \quad \Rightarrow \quad \nabla_{xx}^2 f(x^*) \succeq 0 \quad \square$$

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

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

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

Assume f is convex.

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

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 $x = x^*$:

$$f(y) \geq f(x^*), \quad \forall y \in \mathbb{R}^n$$

(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 $\nabla f(x^*) = 0$ and $\nabla_{xx}^2 f(x^*) \succ 0 \Rightarrow x^*$ 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 $\bar{x}(s) : [0, \varepsilon) \rightarrow \mathcal{D}$ 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:

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

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

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

It 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^* is a local minimizer. Then:

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

Proof:

Item 1 \rightarrow feasibility

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

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

which would mean that p is descent direction. Contradicts x^* local minimizer.

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

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 $\nabla g_i(\bar{x})$, $i \in A(\bar{x})$ 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

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

where $\nabla h(\bar{x}) := [\nabla h_1(\bar{x}), \dots, \nabla h_{n_h}(\bar{x})] \in \mathbb{R}^{n \times n_h}$.

$\textcircled{1}$ can be written equivalently as $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \{p \mid A(\bar{x})p \geq 0\}$

$$A(\bar{x}) := \left[\begin{array}{c} \nabla h(\bar{x})^T \\ -\nabla h(\bar{x})^T \\ \vdots \\ -\nabla g_i(\bar{x})^T \\ \vdots \end{array} \right] \left. \vphantom{\begin{array}{c} \nabla h(\bar{x})^T \\ -\nabla h(\bar{x})^T \\ \vdots \\ -\nabla g_i(\bar{x})^T \\ \vdots \end{array}} \right\} \begin{array}{l} \in \mathbb{R}^{(2n_h + |A(\bar{x})|) \times n} \\ i \in A(\bar{x}) \end{array}$$

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

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

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

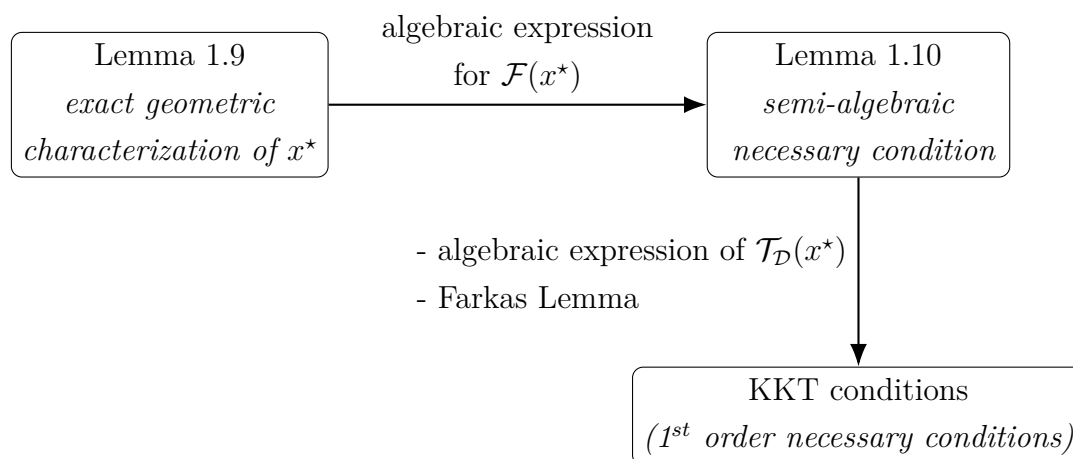
If we find y satisfying 1., then 2. can't hold \Rightarrow item 2 of Lemma 1.10 is verified $\Rightarrow x^* \in \mathcal{D}$ is a local minimizer.

KTK-conditions just follow from imposing

item 1 of Lemma 1.10 $\rightarrow x^* \in \mathcal{D}$

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

Conceptual summary:



Informal recap:

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

\Updownarrow (if x^* regular point)

$$\exists y = \left[\begin{array}{c} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ -\nu_i \\ \vdots \end{array} \right] \left. \vphantom{\begin{array}{c} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ -\nu_i \\ \vdots \end{array}} \right\} \begin{array}{l} \in \mathbb{R}^{2n_h + |\mathcal{A}(x^*)|} \\ i \in \mathcal{A}(\bar{x}) \end{array} \quad y \geq 0 \rightarrow A(x^*)^T y = \nabla f(x^*)$$

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

$$\nabla h(x^*)(\hat{\lambda}_1 - \hat{\lambda}_2) - \sum_{i \in \mathcal{A}(x^*)} \nabla g_i(x^*) \nu_i = \nabla f(x^*), \quad \hat{\lambda}_1, \hat{\lambda}_2, \nu_i \geq 0 : \text{ But } (\hat{\lambda}_1 - \hat{\lambda}_2) \not\geq 0$$

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

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

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

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

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

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

$$1. \quad \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\lambda} + \nabla g(\bar{x})\bar{\nu} = 0$$

$$2. \quad g(\bar{x}) \leq 0$$

$$3. \quad h(\bar{x}) = 0$$

$$4. \quad \bar{\nu} \geq 0$$

$$5. \quad \bar{\nu}^T g(\bar{x}) = 0$$

Lemma 1.13 (KKT necessary condition for local minimizer)

If x^* is a local minimizer **and** a regular point.
Then $\exists \lambda^*, \nu^*$ s.t. (x^*, λ^*, ν^*) 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. $\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\nu = 0$ It can be written

equivalently as

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

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

2. $\iff x^* \in \mathcal{D}$

3. $\iff x^* \in \mathcal{D}$

4. \iff non-negativity of “ y ” from Farkas Lemma

5. $\iff \nu_i = 0, i \notin \mathcal{A}(x^*)$

$$\nu^T g(x^*) = \sum_i \nu_i g_i(x^*) = 0 \quad \text{Complementary slackness}$$

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

$$\nu_i \geq 0, \forall i \quad \text{because of Farkas' lemma.}$$

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

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

Lemma 1.14 (KKT sufficient conditions for global minimizer)

Suppose f, g_i ($i = 1, \dots, n_g$) are convex functions and
 h_j ($j = 1, \dots, n_h$) are affine functions.

If (x^*, λ^*, ν^*) is a KKT point, then x^* is a local minimizer.

Proof: For (λ^*, ν^*) KKT points:

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

f, g_i, h_j are convex functions

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

$$\Rightarrow b(x) \text{ convex}$$

1. $b(x)$ convex

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

\Updownarrow (if x^* regular point)

$$f(x) - f(x^*) \geq - \underbrace{\sum_{i \in \mathcal{A}(x^*)} \nu_i^* g_i(x)}_{\leq 0} - \underbrace{\sum_{j=1}^{n_h} \lambda_j^* h_j(x)}_{=0} \geq 0, \quad x \in \mathcal{D}$$

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

$\rightarrow x^*$ is a global minimizer. □

Second-order conditions

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

$$\nabla_{xx}^2 \mathcal{L}$$

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

Precisely, we are interested in this property along

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

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

Lemma 1.15 (Second order necessary condition)

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

Then

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

$\forall p \neq 0$ with

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

Lemma 1.16 (Second order sufficient conditions for local minimizer)

If (x^*, λ^*, ν^*) is a KKT point with

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

for same p as in Lemma 1.15.

Then x^* 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)

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

- We should introduce “time” or “stages” in the problem

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

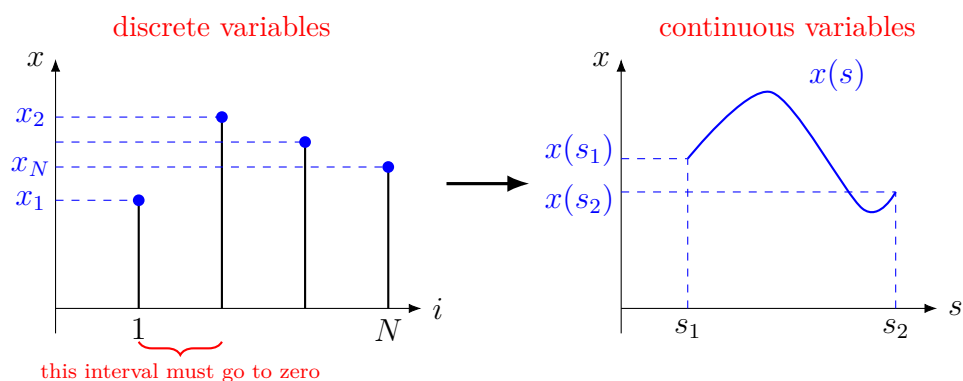
$$\text{s.t. } x_k \in \mathcal{D}_k \subseteq \mathbb{R}^n, \quad k = 1, \dots, N, \quad x_0 \text{ given}$$

equivalent to: (loses structure)

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

- Continuous-time description of dynamics

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



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

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

2.1 Introduction to CV (Calculus of Variations)

Function CLASSES & NORMS

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

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

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

Definition 2.1 (Piecewise continuously differentiable functions)

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

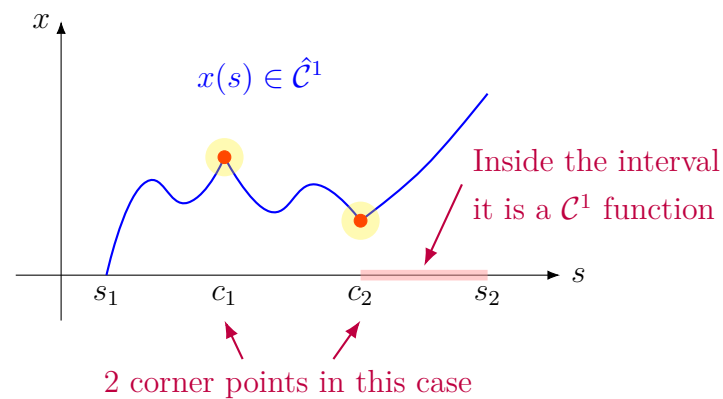
- $x \in \mathcal{C}$ on $[s_1, s_2]$,
- \exists a finite partition $\{c_k\}_{k=0}^{N+1}$ with

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

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

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

Example:



Norms

Definition 2.2 (Strong and weak norms)

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

- Strong norm (or ∞ -norm)

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

- Weak norm (or 1-norm)

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

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

Note $\forall x \in V, \|x\|_1 \geq \|x\|_{\infty}$

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

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

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

CV problem

$$\begin{aligned} \min_{x \in V} \quad & \underline{J}(x) \\ & \hookrightarrow \text{functional } J: V \rightarrow \mathbb{R} \\ \text{s.t. } \quad & x \in \underline{\mathcal{D}} \\ & \hookrightarrow \text{admissible set} \end{aligned}$$

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

3 forms of J :

- Lagrangian form

$$J(x) := \int_{s_1}^{s_2} \underline{l}(s, x(s), \dot{x}(s)) ds$$

\hookrightarrow running cost, $L: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$

- Bolza form

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

- Mayer form

$$J(x) := \varphi(s_2, x(s_2))$$

These 3 forms are (generally) interchangeable, e.g. $L \Rightarrow B$, $B \Rightarrow L$.

We will see them again in Chapter 3.

Admissible sets:

- Free problems \rightarrow only endpoints are constrained. Example:

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

- Isoperimetric constraints \rightarrow level sets

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

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

Example:

$$\min_x J(x)$$

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

Minimizers for CV problems

Definition 2.3 (Global and local minimizers)

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

$$J(x) \geq J(x^*), \quad x \in \mathcal{D}$$

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

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

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

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

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

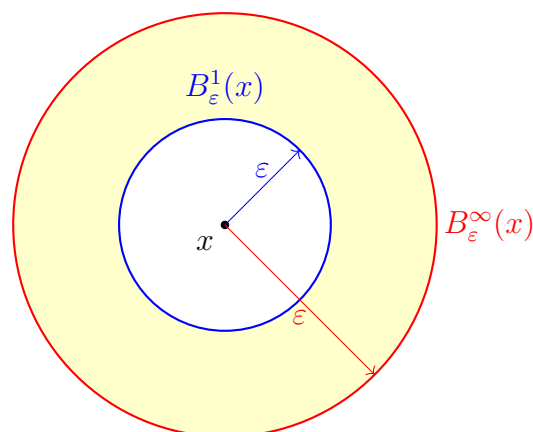
$$B_\varepsilon^1 := \{y \in V \mid \|x - y\|_1 \leq \varepsilon\} \quad (\text{weak ball})$$

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 $B_\varepsilon^1(x)$ is not necessarily better than all elements in $B_\varepsilon^\infty(x)$.

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

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

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

because x is in the weak ball with radius ε .

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

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

$$B_1^\infty(0) = \{x \in \mathcal{C}([0, 1]) \mid \|x\|_\infty \leq 1\}$$

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-priori.

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

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

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

δJ can be interpreted as follows

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

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 definition:

$$\begin{aligned} \frac{J(x + \eta\xi) - J(x)}{\eta} &= \frac{1}{\eta} \left[\int_{s_1}^{s_2} (x(s) + \eta\xi(s))^2 - x^2(s) ds \right] \\ &= 2 \int_{s_1}^{s_2} x(s)\xi(s) ds + \eta \int_{s_1}^{s_2} \xi^2(s) ds \\ \lim_{\eta \rightarrow 0} &\longrightarrow \delta J(x; \xi) = 2 \int_{s_1}^{s_2} x(s)\xi(s) ds \end{aligned}$$

From the definition, 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 $V, J : V \rightarrow \mathbb{R}$ 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.

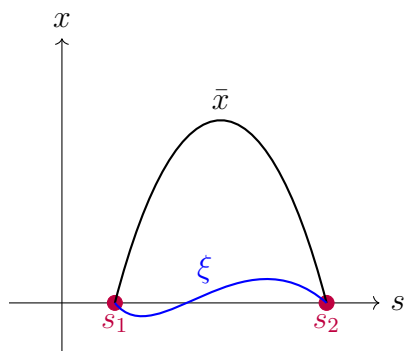
Definition 2.7 (\mathcal{D} -admissible direction)

Given $V, \mathcal{D} \subseteq V$ and $J : V \rightarrow \mathbb{R}$.

$\xi \in V, \xi \neq 0$ 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 \mathcal{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).

Proofs: Use definition of 1st variation.

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

$(V, \|\cdot\|)$, $\mathcal{D} \subseteq V$, $J \rightarrow \mathbb{R}$.

Suppose $x^* \in \mathcal{D}$ is a local minimizer for J on \mathcal{D} , then

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

Proof: By contradiction:

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

By Lemma 2.1 x^* can't be a local minimizer \rightarrow contradiction

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

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

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

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

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

Thus $\delta J(x^*; \xi) = 0$.

All the results in the rest of this chapter are “merely corollaries” of this Lemma.
they turn such an abstract requirement into algebraic tests

2.2 Free problems of CV

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

$$\begin{aligned} \min_{x(\cdot)} \int_{s_1}^{s_2} l(s, x(s), \dot{x}(s)) ds & \quad [CV - P1] \\ \text{s.t. } x \in \{y \in \mathcal{C}^1([s_1, s_2]) \mid y(s_1) = x_1, y(s_2) = x_2\} & \\ & \quad \underbrace{\hspace{10em}}_{\text{special case of “free problems”} \\ \text{we fix the endpoint to } x_2} \end{aligned}$$

Lemma 2.3

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

Then

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

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 candidate local minimizers of $[CV - P1]$.

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

also called “stationary solutions” of the corresponding CV problem. Why?

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

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

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

$$\begin{aligned} \frac{\partial}{\partial \eta} J(x^* + \eta \xi) &\stackrel{\text{Leibniz rule}}{=} \int_{s_1}^{s_2} \frac{\partial}{\partial \eta} l(s, x^* + \eta \xi, \dot{x}^* + \eta \dot{\xi}) ds \\ &= \int_{s_1}^{s_2} l_x [x^* + \eta \xi]^T \xi + l_{\dot{x}} [x^* + \eta \xi]^T \dot{\xi} ds \end{aligned}$$

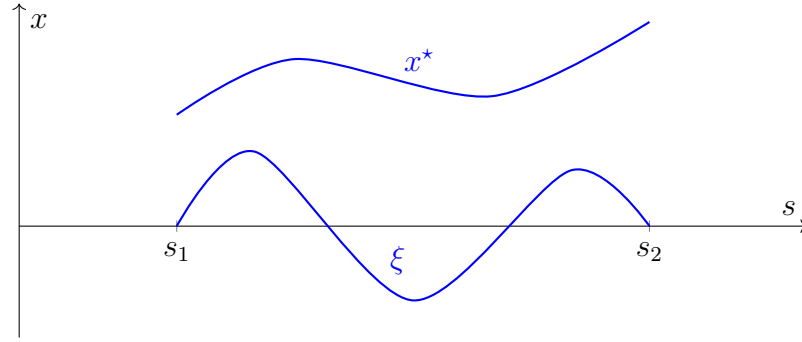
with $l_z[y] := l_z(s, y, \dot{y})$.

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

integrand is a continuous function

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

We want to obtain conditions enforcing $\delta J = 0 \quad \forall \mathcal{D}\text{-admissible } \xi$. 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 $\delta J = 0$

$$\begin{aligned} \forall i \in \{1, \dots, n\}, \quad 0 = \delta J(x^*, \xi^{(i)}) &= \int_{s_1}^{s_2} [l_{x_i}[x^*]\xi_i + l_{\dot{x}_i}[x^*]\dot{\xi}_i] ds. \\ &= \int_{s_1}^{s_2} l_{\dot{x}_i}[x^*]\dot{\xi}_i ds + \int_{s_1}^{s_2} \underbrace{\frac{d}{ds} \left[\left(\int_{s_1}^s l_{x_i}[x^*] d\sigma \right) \right]}_{v'} \underbrace{\dot{\xi}_i}_u ds \end{aligned}$$

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

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

DuBois-Reymond's Lemma if

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

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

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

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

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

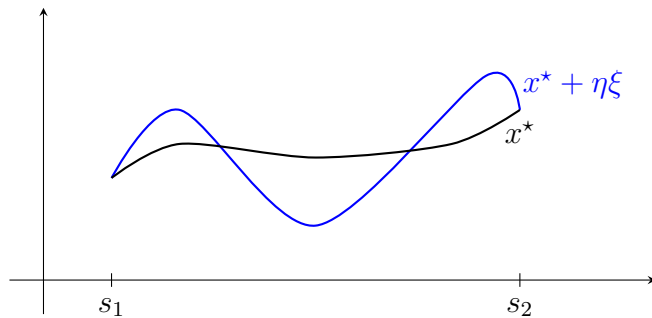
Note that this shows $l_{\dot{x}_i} \in \mathcal{C}^1$

What we obtain are the “integral” Euler equations (EE)

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

Remarks

1. Necessary conditions for local minimizers (\rightarrow weak)



as $\eta \rightarrow 0$

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

\Rightarrow “ $(x^* + \eta\xi)$ is inside the weak ball of x^* ”

EE \Rightarrow detect weak minimizers

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

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

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

Two Point Boundary Value Problem (TPBVP)

3. There exists a reformulation of EE:

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

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

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 .

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

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

[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^* is a weak minimizer of $[CV - P1]$, then

1. x^* satisfies EE
2. $\nabla_{\ddot{x}}^2 l(s, x^*, \dot{x}^*) \succeq 0, \quad \forall s \in [s_1, s_2].$
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} .

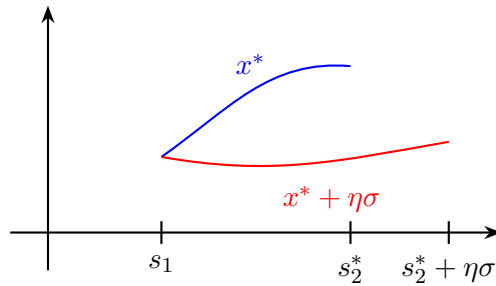
If $x^* \in \mathcal{D}$ satisfies the EE, then x^* is a global minimizer.

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:

$$\begin{aligned} \delta J(x, s_2; \xi, \sigma) &:= \lim_{\eta \rightarrow 0} \frac{J(x + \eta\xi, s_2 + \eta\sigma) - J(x, s_2)}{\eta} \\ &= \left. \frac{\partial}{\partial \eta} J(x + \eta\xi, s_2 + \eta\sigma) \right|_{\eta=0} \end{aligned}$$



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

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

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

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

1. x^* solves EE on $[s_1, s_2^*]$

2. The transversal conditions

A) $[l_{\dot{x}} + \varphi_x] \Big|_{x=x^*, s=s_2^*} = 0$

B) $[l - \dot{x}^T l_{\dot{x}} + \varphi_s] \Big|_{x=x^*, s=s_2^*} = 0$

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 $\psi : \mathbb{R} \rightarrow \mathbb{R}^n$, $x(s_2) = \psi(s_2)$

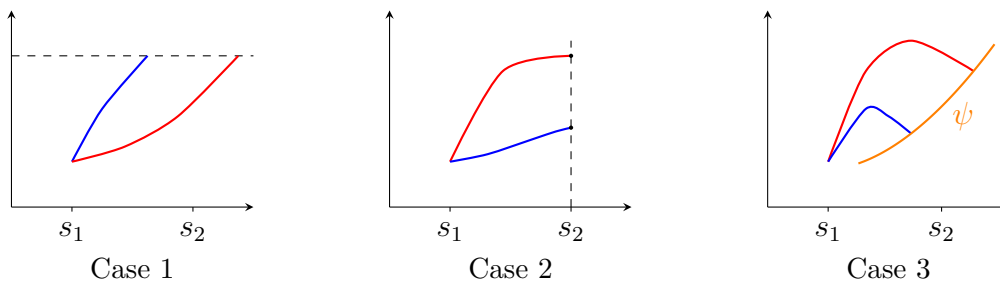
We do not need 2A \rightarrow it is replaced by $x(s_2) = \psi(s_2)$.

We need an extension to 2B.

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

Note $\psi(s_2) = x_2$

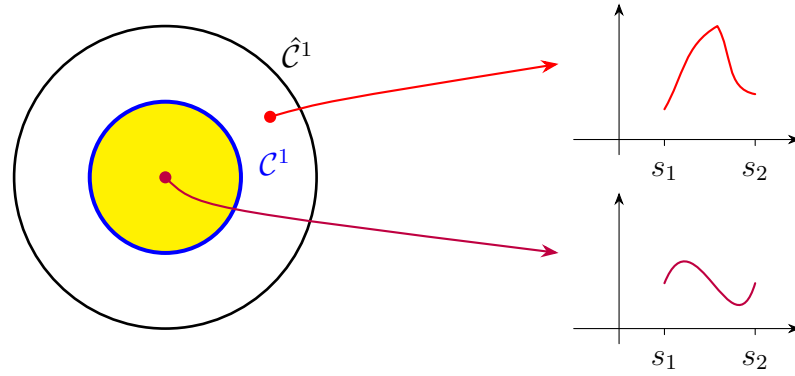
When $\dot{\psi} = 0$, 2B' collapses to 2B.



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)

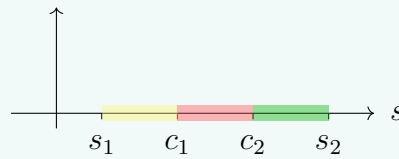
Lemma 2.7 (First-order necessary conditions for strong minimizers)

Consider $[CV - P1]$ with $x \in \hat{\mathcal{C}}^1([s_1, s_2])$.

Suppose x^* with corner points $\{c_i\}_{i=1}^N$ is a strong minimizer.

Then:

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



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

$$[A] \quad l_{\dot{x}}(c_i^-) = l_{\dot{x}}(c_i^+)$$

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

$$[B] \quad [-l(c_i^-) + \dot{x}^*(c_i^-)^T l_{\dot{x}}(c_i^-)] = [-l(c_i^+) + \dot{x}^*(c_i^+)^T l_{\dot{x}}(c_i^+)]$$

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

[A] and [B] are also called Weierstrass-Erdmann-conditions (WE).

[A] is the 1st WE-condition and is also necessary for weak minimizers.

[B] is the 2nd WE-condition and is only necessary for strong minimizers.

[A] prescribes continuity of p (momentum)

[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.

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

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

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

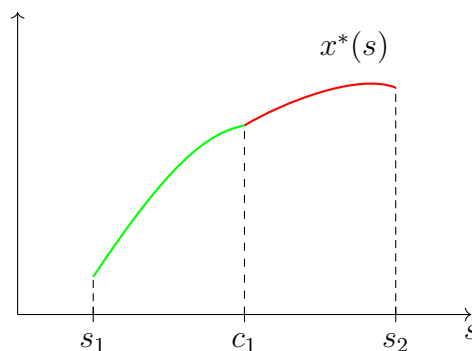
We have N corner points. At each c_i we can enforce:

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

→ The problem is closed

Proof:

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



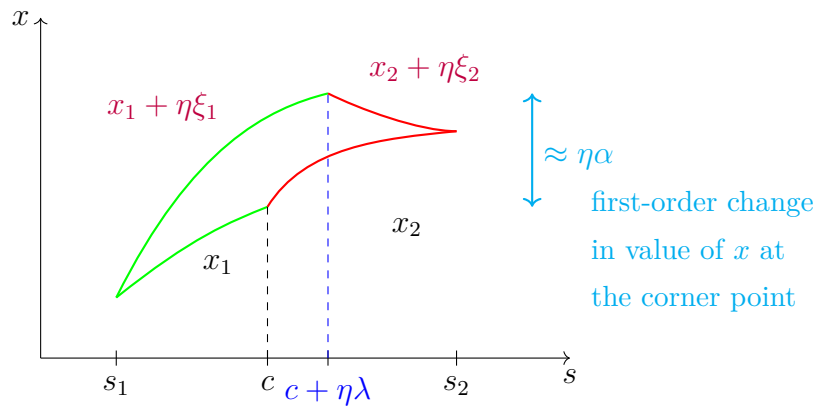
I will drop $*$ from x .

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

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

The location of c can also be perturbed

$$c \rightarrow c + \eta\lambda$$



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

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

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

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

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

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

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

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 $\|(x + \eta\xi) - x\|_1 \approx \eta$.

→ 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

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

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

Definition 2.8 (Weierstrass excess function)

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

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 $[CV - P1]$ $x \in \hat{\mathcal{C}}^1$.

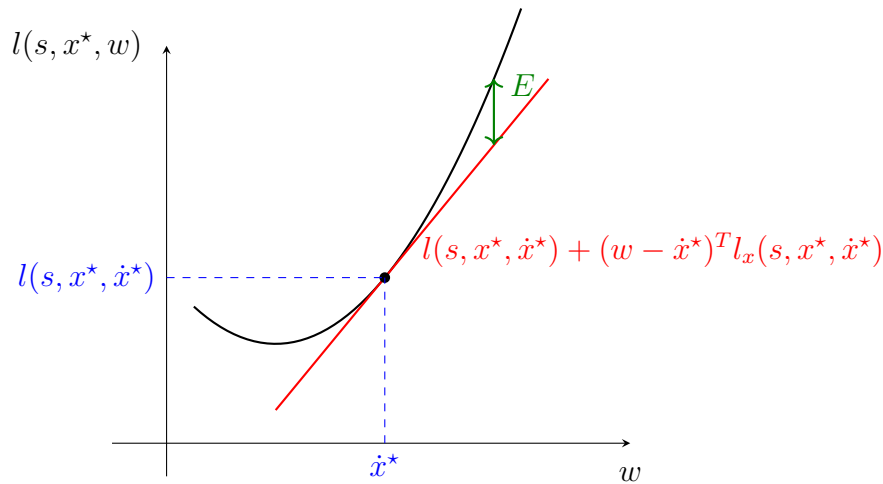
Suppose x^* with corner point $\{c_i\}_{i=1}^N$ is a strong minimizer.

Then

- x^* solves EE inside the intervals (c_i, c_{i+1}) , $i = 0, 1, \dots, N$
- $E(s, x, \dot{x}, w) \geq 0 \quad \forall w \in \mathbb{R}^n, \quad \forall s \in [s_1, s_2]$ 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^* , we draw for fixed $s \in [s_1, s_2]$ the following curve:



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

So this is a weaker requirement than joint convexity of l

→ we only need to check that at x^*

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

$$E(s, x^*, \dot{x}^*, w) \geq 0 \quad \Leftrightarrow \quad H(s, x^*, \dot{x}^*, p^*) - H(s, x^*, w, p^*) \geq 0 \quad \forall w \in \mathbb{R}^n$$

substitute definition of E in $E \geq 0$

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

2.3 Isoperimetric constraints

$$\begin{aligned}
 [CV - P3] \quad & \min_{x(\cdot)} \int_{s_1}^{s_2} l(s, x, \dot{x}) ds \\
 \text{s.t. } x \in & \left\{ y \in C^1 \mid \underbrace{\int_{s_1}^{s_2} g_i(s, y(s), \dot{y}(s)) ds}_{G_i(y): V \rightarrow \mathbb{R}, \quad k_i \in \mathbb{R}} = k_i, \quad i = 1, \dots, n_g; \ y(s_1) = x_1; \ y(s_2) = x_2 \right\}
 \end{aligned}$$

Lemma 2.9

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

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

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

for n_g independent directions $\{\xi_i\}_{i=1}^{n_g} \in V$.

If x^* is a local minimizer of J , then $\exists \lambda \in \mathbb{R}^{n_g}$ such that

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

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

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