Optimal Control

Wintersemester 2025/26

Dozent: Prof. Dr. Andrea Iannelli

October 28, 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

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

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 <u>finite-dimensional</u> 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)

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

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

$f(x^\star)$ is the <u>Global Minimum</u> (or Minimum)
Nomenclature: $x^\star$ is also called (optimal) solution, $F(x^\star)$ is optimal value
$x^\star$ is a strict global minimizer if $f(x^\star) < f(x) \quad \forall x \in \mathcal{D}$
$x^\star \in \mathcal{D}$ <u>Local Minimizer</u> 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 $\rightleftarrows$ 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 $\mathcal{S} = [-1, 1]$, $-50 = \inf(\mathcal{S})$? $\quad \to \quad$ 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 $\to \inf(\mathcal{S}) = -\infty$
- $\inf([a, b]) = \inf((a, b]) = a$

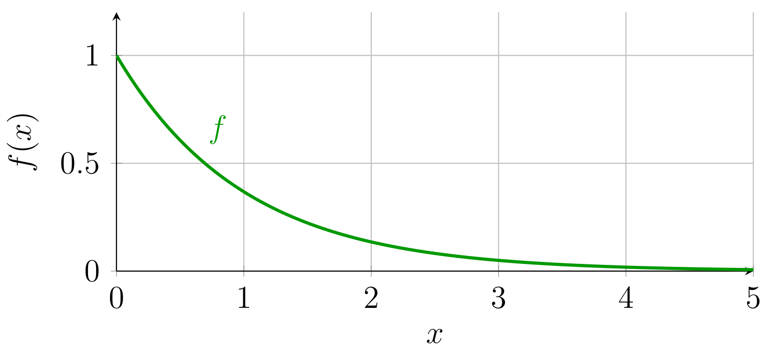
Connections with NLP?

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

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

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

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

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

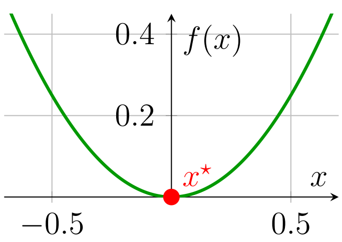
If f is continuous $\forall x \in \mathcal{D}$ then f is continuous on $\mathcal{D} \to f \in \mathcal{C}$

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:

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

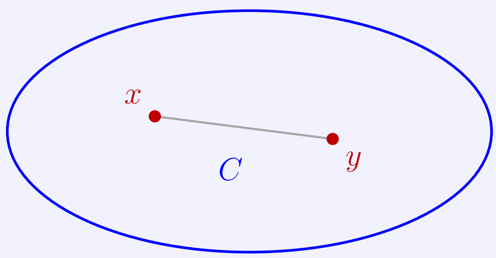
Goal of the Chapter: characterize necessary and sufficient conditions for $x^\star$ to be global minimizer of NLP.

Convexity

Definition 1.4 (Convex sets & functions)

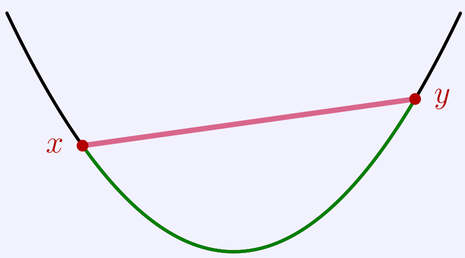
- A set $C \subseteq \mathbb{R}^n$ is <u>convex</u> (cvx) if $\forall x, y \in C$

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



- 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

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

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

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

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

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

- $A \geq 0$ means that: $A = A^T$ and pos semi-definite, i. e. all eigenvalues non-negative
- f strictly cvx if $\nabla^2_{xx} f(x) > 0 \quad \forall x \in C$ with $A > 0$ 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 $a, b \in \mathbb{R}$

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

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)

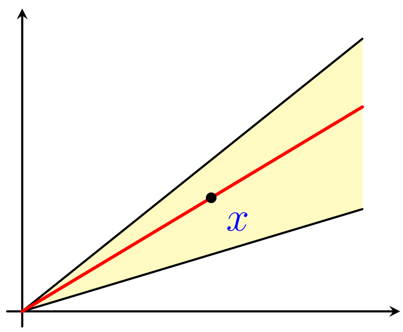
$d \in \mathbb{R}^n$ is a descent direction for f at $\bar{x} \in \mathbb{R}^n$ if

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

$F(\bar{x})$: Cone of decent 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 $f \in \mathcal{C}^2$

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

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 $d \in \mathbb{R}^n$

$$\Rightarrow \quad d^T \nabla^2_{xx} f(x^\star) \geq 0 \quad \forall d \quad \Rightarrow \quad \nabla^2_{xx} f(x^\star) \geq 0 \qquad \square$$

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 $\nabla^2_{xx} f(x^\star) > 0 \quad \Rightarrow \quad x^\star$ is strict local minimizer.

Proof: Taylor expansion (Similar to Lemma 1.6)

1.2 Constrained Problems

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

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 $\bar{x}(s) : [0, \epsilon) \to \mathcal{D}$ with $\epsilon > 0$ 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 $\mathcal{F}(\bar{x}) \cap \mathcal{T}_\mathcal{D}(\bar{x}) = \emptyset$

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

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

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 $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \{p \mid \mathcal{A}(\bar{x})p \geq 0\}$

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

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

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

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 $b \equiv \nabla f(x^\star)$

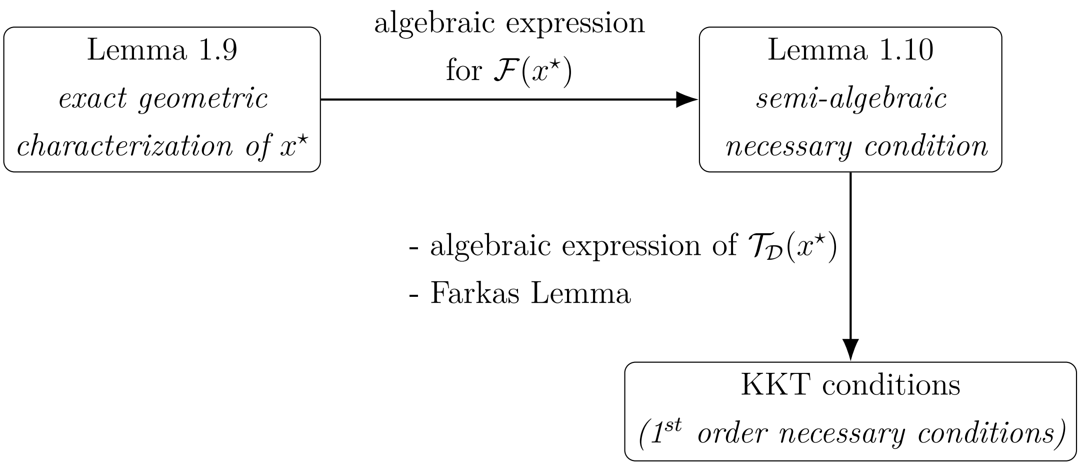
If we find y satisfying 1., then 2. can't hold $\Rightarrow$ item 2 of Lemma 1.10 is verified $\Rightarrow$ $x^\star \in \mathcal{D}$ is a local minimizer.

KTK-conditions just follow from imposing

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

item 2 of Lemma 1.10 $\rightarrow \exists y \in \mathbb{R}^m$, $y \geq 0$ 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)

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

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

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

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

   $\nu_i \geq 0,\ \forall i \quad$ 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 $g_i(x^\star) < 0$. □

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

Lemma 1.14 (KKT sufficient conditions fot global minimizer)

Suppose $f, g_i \quad (i = 1, \ldots, n_g)$ are convex functions and
$h_j \quad (j = 1, \ldots, n_h)$ 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

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

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

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

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

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

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

$\forall p \neq 0$ 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 $\nu_i^* > 0 \mid \nabla f(x^\star)^T p = 0$

Lemma 1.16 (Second order suffiecient conditions for local minimizer)

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

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

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

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

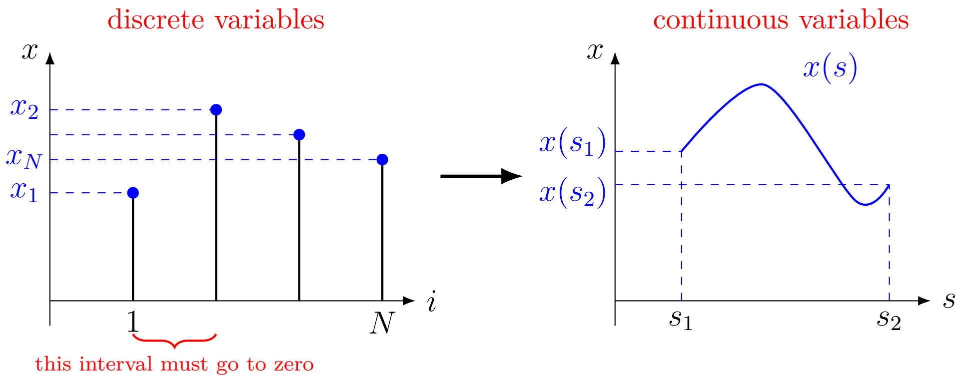
$$\text{s.t. } x_k \in \mathcal{D}_k \subseteq \mathbb{R}^n, \quad k = 1, \dots, N, \quad x_0 \text{ given}$$

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:



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

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

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: