Model Predictive Control

Chapter 3: Introduction to Convex Optimization

Prof. Melanie Zeilinger

ETH Zurich

Spring 2020

Coauthors: Prof. Paul Goulart, University of Oxford
Prof. Colin Jones, EPFL

Learning Objectives – Lecture 3

- Learn to ‘read’ and define optimization problems
- Understand property of convexity of sets and functions
- Understand benefit of convex optimization problems
- Learn and contrast properties of LPs and QPs
- Pose the dual problem to a given primal optimization problem
- Test optimality of a primal and dual solution by means of KKT conditions
- Understand meaning of dual solution for the cost function

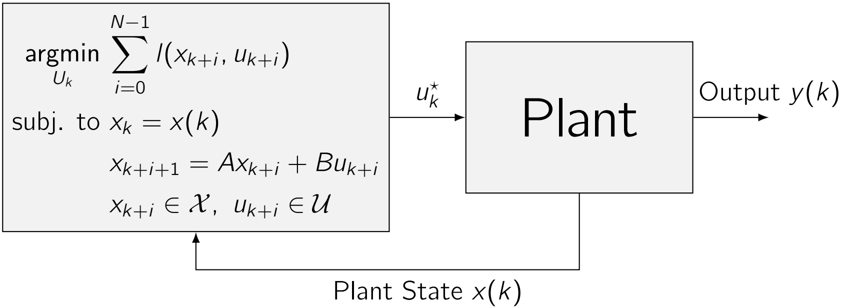
Outline

1. Optimization in MPC
2. Main Concepts
3. Convex Sets
4. Convex Functions
5. Convex Optimization Problems
6. Optimality Conditions

Outline

1. Optimization in MPC
2. Main Concepts
3. Convex Sets
4. Convex Functions
5. Convex Optimization Problems
6. Optimality Conditions

MPC: Mathematical Formulation



At each sample time:

- Measure / estimate current state $x(k)$
- **Find the optimal input sequence for the entire planning window N:**
  $U_k^\star = \{u_k^\star, u_{k+1}^\star, \ldots, u_{k+N-1}^\star\}$
- Implement only the first control action $u_k^\star$

Optimization Problems Arising in MPC

Linear Systems

- Linear system dynamics
- Continuous set of states and inputs, e.g., $x \in [x_{\min}, x_{\max}], u \in [u_{\min}, u_{\max}]$
- Example: Chemical processes

Nonlinear Systems

- Nonlinear system dynamics
- Continuous set of states and inputs, e.g., $x \in [x_{\min}, x_{\max}], u \in [u_{\min}, u_{\max}]$
- Example: Kites

Hybrid Systems

- Mixed dynamics that are both continuous and discrete, e.g.
$$\begin{cases} x_{k+1} = -c_1 & x_k \geq x_{\max} \\ x_{k+1} = c_2 - c_1 & x_k < x_{\max} \end{cases}$$
- Continuous set of states and inputs
- Example: Walking robot

Discrete Decision Variables

- Inputs and/or states can only take discrete values, e.g. $u \in \{1, 2, 3, 4, 5\}$
- Example: Internet

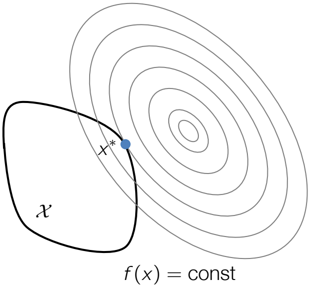
Outline

Mathematical Optimization Problem

A mathematical optimization problem is generally formulated as:

$$\begin{aligned} \min_{x \in \mathbf{dom}(f)} \quad & f(x) \\ \text{subj. to} \quad & g_i(x) \le 0 \quad i = 1, \dots, m \\ & h_i(x) = 0 \quad i = 1, \dots, p \end{aligned}$$

- Optimization variables $x := [x_1; x_2; \dots; x_n]$
- Objective function $f : \mathbf{dom}\,(f) \to \mathbb{R}$
- Domain $\mathbf{dom}\,(f) \subseteq \mathbb{R}^n$ of the objective fcn
- Optional inequality constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \dots, m$
- Optional equality constraint functions $h_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \dots, p$



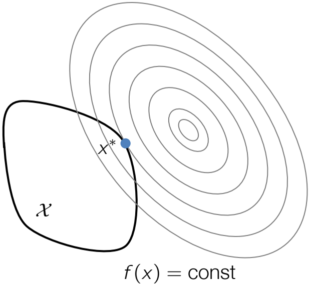
- $\mathcal{X} := \{x \in \mathbf{dom}\,(f) \mid g_i(x) \le 0,\ i = 1, \dots, m, h_i(x) = 0,\ i = 1, \dots, p\}$: set of feasible decisions, or feasible set

Terminology

Feasible point: $x \in \mathbf{dom}(f)$ satisfying the inequality and equality constraints, i.e. $g_i(x) \leq 0$ for $i = 1, \ldots, m$, $h_i(x) = 0$ for $i = 1, \ldots, p$.

Strictly feasible point: Feasible $x \in \mathbf{dom}(f)$ satisfying the inequality constraints strictly, i.e. $g_i(x) < 0$ for $i = 1, \ldots, m$.

Optimal value: Lowest possible cost value $p^* = f(x^\star) \triangleq \min_{x \in \mathcal{X}} f(x)$ also denoted by $f^\star$ or $J^\star$



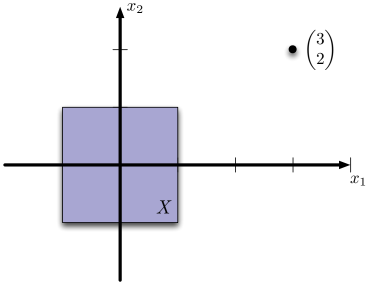
Optimizer: Any feasible $x^\star$ that achieves smallest cost p^*, i.e., $x^\star \in \mathcal{X}$ with $f(x^\star) \leq f(x)$ for all feasible $x \in \mathcal{X}$.

Optimizer is not always unique. The set of solutions is:

$$\operatorname*{argmin}_{x \in \mathcal{X}} f(x) := \{x \in \mathcal{X} \mid f(x) = p^\star\}$$

A Simple Example

Problem : In $\mathbb{R}^2$, find the point in the unit box X closest to the point $(x_1, x_2) = (3, 2)$.



Same problem in standard format:

$$
\begin{aligned}
\min_{(x_1,x_2)\in\mathbb{R}^2} \quad & (x_1-3)^2 + (x_2-2)^2 \\
\text{subj. to} \quad & x_1 \le 1 \\
& -x_1 \le 1 \\
& x_2 \le 1 \\
& -x_2 \le 1
\end{aligned}
$$

Active, Inactive and Redundant Constraints

Consider the standard problem

$$
\begin{aligned}
\min_{x \in \mathrm{dom}(f)} \quad & f(x) \\
\text{subj. to } \quad & g_i(x) \leq 0 \quad i = 1, \dots, m \\
& h_i(x) = 0 \quad i = 1, \dots, p
\end{aligned}
$$

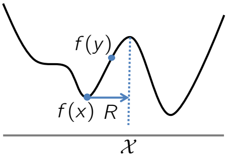
- The i^{th} inequality constraint $g_i(x) \leq 0$ is **active** at $\bar{x}$ if $g_i(\bar{x}) = 0$. Otherwise it is **inactive**.
- Equality constraints are always active.
- A **redundant** constraint does not change the feasible set. This implies that removing a redundant constraint does not change the solution. Example:

$$
\begin{aligned}
\min_{x \in \mathbb{R}} \quad & f(x) \\
\text{subj. to } \quad & x \leq 1 \\
& x \leq 2 \quad \texttt{(redundant)}
\end{aligned}
$$

Optimality

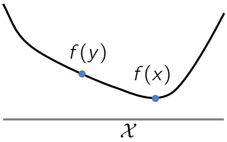
- $x \in \mathcal{X}$ is **locally optimal** if, for some $R > 0$, it satisfies

$$y \in \mathcal{X}, \|y - x\| \leq R \Rightarrow f(y) \geq f(x)$$



- $x \in \mathcal{X}$ is **globally optimal** if it satisfies

$$y \in \mathcal{X} \Rightarrow f(y) \geq f(x)$$



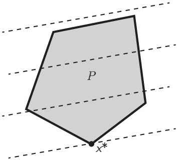
- If $p^* = -\infty$ the problem is **unbounded below**
- If $\mathcal{X}$ is empty, then the problem is said to be **infeasible** (convention: $p^* = \infty$)
- If $\mathcal{X} = \mathbb{R}^n$ the problem is said to be **unconstrained**

"Easy" and "Hard" Problems

"Easy": Linear Program (LP)

Linear cost and constraint functions.

$$
\begin{aligned}
\min_x \ & c^\top x \\
\text{subj. to } & Gx \leq h \\
& Ax = b
\end{aligned}
$$

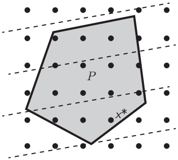


Linear optimization on a polytope.

"Hard': Mixed Integer Linear Program

Linear program with binary or integer constraints.

$$
\begin{aligned}
\min_x \ & c^\top x \\
\text{subj. to } & Gx \leq h \\
& Ax = b \\
& x \in \{0,1\}^n \text{ or } x \in \mathbb{Z}^n
\end{aligned}
$$



Linear optimization with integer constraints (dots).

Convex optimization problems can be solved efficiently and reliably.

Software Tools for Optimization

A simple optimization problem:

$$\begin{aligned} \min_{x_1, x_2} \; & |x_1 + 5| + |x_2 - 3] \\ \text{subj. to } & 2.5 \le x_1 \le 5 \\ & -1 \le x_2 \le 1 \end{aligned}$$

- This problem is equivalent to a linear program (more on this later).
- Huge variety of software tools for solving standard optimization problems:
  - **Examples**: MATLAB (linprog/quadprog), CPLEX, Gurobi, GLPK, XPRESS, qpOASES, OOQP, FORCES, SDPT3, Sedumi, MOSEK, IPOPT,...
- There is no standard interface to solvers – they are almost all different.
- General purposes modeling tools allow easy switching between solvers:
  - **Examples**: CVX, Yalmip, GAMS, AMPL

Software Tools for Optimization

A simple optimization problem:

$$\min_{x_1,x_2} |x_1 + 5| + |x_2 - 3]$$
$$\text{subj. to } 2.5 \leq x_1 \leq 5$$
$$-1 \leq x_2 \leq 1$$

The YALMIP toolbox for Matlab (from ETH / Linköping):

```
%make variables
sdpvar x1 x2;
%define cost function
f = abs(x1 + 5) + abs(x2 - 3);
%define constraints
X = set(2.5 <= x1 <= 5) + ...
    set( -1 <= x2 <= 1);
%solve
solvesdp(X,f)
```

Software Tools for Optimization

A simple optimization problem:

$$\min_{x_1, x_2} |x_1 + 5| + |x_2 - 3]$$
$$\text{subj. to } 2.5 \leq x_1 \leq 5$$
$$-1 \leq x_2 \leq 1$$

The CVX toolbox for Matlab (from Stanford):

```
cvx_begin
    %define cost function
    variables x1 x2
    %define constraints
    minimize(abs(x1 + 5) + abs(x2−3))
    subject to
        2.5 <= x1 <= 5
        −1 <= x2 <=1
evx_end     %solves automatically
```

Outline

1. Optimization in MPC
2. Main Concepts
3. Convex Sets
4. Convex Functions
5. Convex Optimization Problems
6. Optimality Conditions

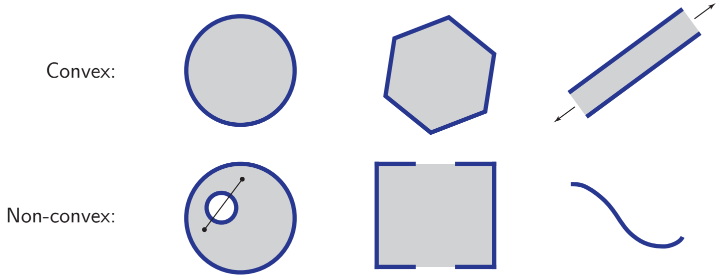
Convex Sets

Definition: Convex Set

A set $\mathcal{X}$ is convex **if and only if** for any pair of points x and y in $\mathcal{X}$:

$$\lambda x + (1 - \lambda) y \in \mathcal{X}, \; \forall \lambda \in [0, 1], \; \forall x, y \in \mathcal{X}$$

Interpretation: All line segments starting and ending in $\mathcal{X}$ stay within $\mathcal{X}$.



Convex combination of $x_1, \ldots, x_k$: Any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k \text{ with } \theta_1 + \ldots + \theta_k = 1, \theta_i \geq 0$$

Convex Sets: Hyperplanes and Halfspaces

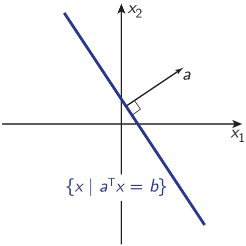
Definitions: Hyperplanes and halfspaces

A hyperplane is defined by $\left\{x \in \mathbb{R}^n \,\middle|\, a^\top x = b\right\}$ for $a \neq 0$, where $a \in \mathbb{R}^n$ is the normal vector to the hyperplane.

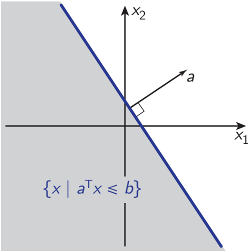
A halfspace is everything on one side of a hyperplane $\left\{x \in \mathbb{R}^n \,\middle|\, a^\top x \leq b\right\}$ for $a \neq 0$. It can either be **open** (strict inequality) or **closed** (non-strict inequality).

For $n = 2$, hyperplanes define lines. For $n = 3$, hyperplanes define planes.

Hyperplanes are affine and convex, halfspaces are convex.



A hyperplane



A closed halfspace

Convex Sets: Polyhedra and Polytopes

Definitions: Polyhedra and polytopes

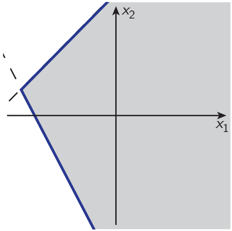
A polyhedron is the intersection of a **finite** number of closed halfspaces:

$$P := \left\{x \,\middle|\, a_i^\top x \leq b_i,\ \ i = 1, \ldots, n\right\} = \{x \,|\, Ax \leq b\}$$

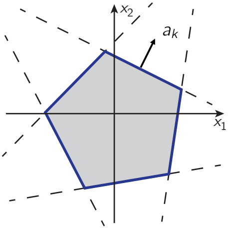
where $A := [a_1, a_2, \ldots, a_m]^\top$ and $b := [b_1, b_2, \ldots, b_m]^\top$.

A polytope is a **bounded** polyhedron.

Polyhedra and polytopes are always convex.



An (unbounded) polyhedron



A polytope

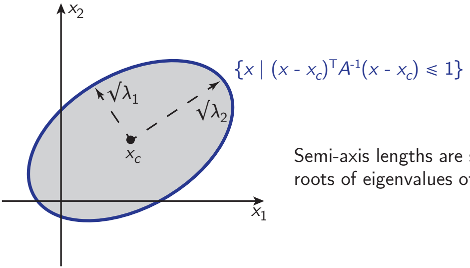
Convex Sets: Ellipsoids

Definition: Ellipsoid

An ellipsoid is a set defined as

$$\left\{x \,\middle|\, (x - x_c)^\top A^{-1}(x - x_c) \leq 1\right\},$$

where x_c is the centre of the ellipsoid, and $A \succ 0$ (i.e. A is positive definite).



Semi-axis lengths are square roots of eigenvalues of A

The **Euclidean ball** $B(x_c, r)$ is a special case of the ellipsoid, for which $A = r^2 I$, so that $B(x_c, r) := \{x \mid ||x - x_c||_2 \leq r\}$.

Convex Sets: Norm Balls

The **norm ball**, defined by $\{x \mid ||x - x_c|| \leq r\}$ where x_c is the centre of the ball and $r \geq 0$ is the radius, is always convex for any norm.

By far the most common ℓ_p norms are:

- $p = 2$ (Euclidean norm):

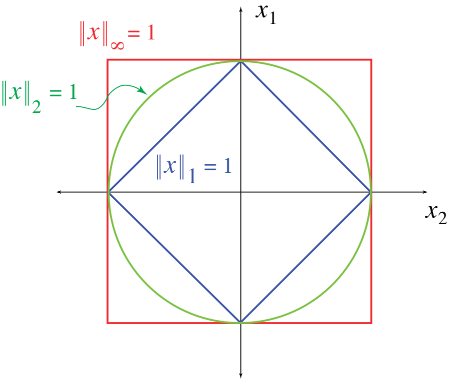
$$||x||_2 = \sqrt{\sum_i x_i^2}$$

- $p = 1$ (Sum of absolute values):

$$||x||_1 = \sum_i |x_i|$$

- $p = \infty$ (Largest absolute value):

$$||x||_\infty = \max_i |x_i|$$

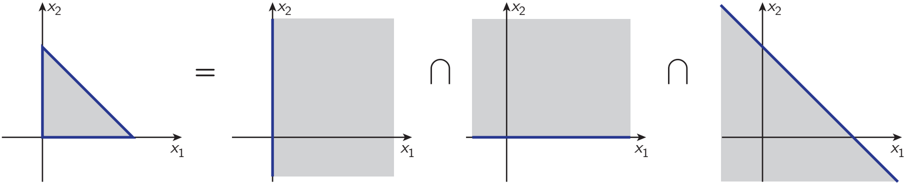


Intersection

Theorem

The intersection of two or more convex sets is itself convex.

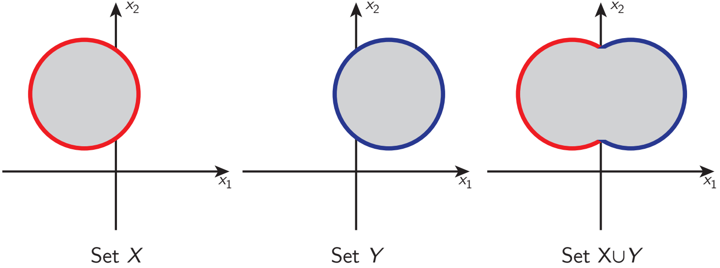
Proof (for two sets): Consider any two points a and b which **both** lie in **both** of two convex sets $\mathcal{X}$ and $\mathcal{Y}$. For any $\lambda \in [0,1]$, $\lambda a + (1-\lambda)b$ is in both $\mathcal{X}$ and $\mathcal{Y}$. Therefore $\lambda a + (1-\lambda)b \in \mathcal{X} \cap \mathcal{Y}$, $\forall \lambda \in [0,1]$. This satisfies the definition of convexity for set $\mathcal{X} \cap \mathcal{Y}$.



Many sets can be written as the intersection of convex elements, and are therefore easily shown to be convex. Any convex set can be written as a (possibly infinite) intersection of halfspaces.

Union $\mathcal{X} \cup \mathcal{Y}$

Note that the **union** of two sets is **not** convex in general, regardless of whether the original sets were convex!



Outline

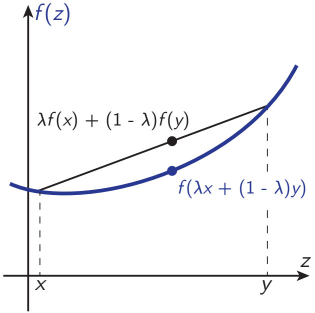
Convex Functions

Definitions: Convex Function

A function $f : \mathbf{dom}\,(f) \to \mathbb{R}$ is convex **iff** $\mathbf{dom}\,(f)$ is convex and

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in (0,1), \quad \forall x, y \in \mathbf{dom}\,(f)$$

The function $f : \mathbf{dom}\,(f) \to \mathbb{R}$ is strictly convex if this inequality is strict.

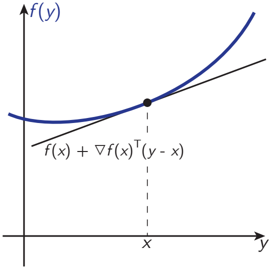


The function f is **concave** iff $\mathbf{dom}\,(f)$ is convex and $-f$ is convex.

First-order Condition for Convexity

A differentiable function $f : \mathrm{dom}\,(f) \to \mathbb{R}$ with a convex domain is **convex iff**

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y \in \mathrm{dom}\,(f)$$



$\to$ First-order approximation of f around any point x is a global underestimator of f.

The gradient $\nabla f(x)$ is given by

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^\top$$

Second-order Condition for Convexity

A twice-differentiable function $f : \mathbf{dom}\,(f) \to \mathbb{R}$ with convex domain $\mathbf{dom}\,(f)$ is **convex iff**

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom}\,(f)\ ,$$

where the Hessian $\nabla^2 f(x)$ is defined by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

If $\mathbf{dom}\,(f)$ is convex **and** $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\,(f)$, then f is **strictly convex**.

Level and sublevel sets

Definition: Level set

The **level set** L_α of a function f for value α is the set of all $x \in \mathbf{dom}(f)$ for which $f(x) = \alpha$:

$$L_\alpha := \{x \,|\, x \in \mathbf{dom}(f), f(x) = \alpha\}$$

For $f : \mathbb{R}^2 \to \mathbb{R}$ these are **contour lines** of constant "height".

Definition: Sublevel set

The **sublevel set** C_α of a function f for value α is defined by

$$C_\alpha := \{x \,|\, x \in \mathbf{dom}(f),\ f(x) \leq \alpha\}$$

Function f is convex $\Rightarrow$ sublevel sets of f are convex for all α. But not $\Leftarrow$!

Examples of Convex Functions: $\mathbb{R} \to \mathbb{R}$

The following functions are **convex** (on domain $\mathbb{R}$ unless otherwise stated):

- Affine: $ax + b$ for any $a, b \in \mathbb{R}$
- Exponential: e^{ax} for any $a \in \mathbb{R}$
- Powers: x^α on domain $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- Vector norms on $\mathbb{R}^n$: $||x||_p = (\sum_{i=1}^n |x|^p)^{1/p}$, for $p \geq 1$, $||x||_\infty = \max_i |x_i|$

The following functions are **concave** (on domain $\mathbb{R}$ unless otherwise stated):

- Affine: $ax + b$ for any $a, b \in \mathbb{R}$
- Powers: x^α on domain $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- Logarithm: $\log x$ on domain $\mathbb{R}_{++}$
- Entropy: $-x \log x$ on domain $\mathbb{R}_{++}$

Convexity-preserving Operations

Certain operations preserve the convexity of functions:

- Non-negative weighted sum
- Composition with affine function
- Pointwise maximum and supremum
- Partial minimization

and many other possibilities...

Outline

Convex Optimization Problem

A **convex** optimization problem in standard form:

$$\begin{aligned} \min_{x \in \mathbf{dom}(f)} \quad & f(x) \\ \text{subj. to} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & a_i^\top x = b_i \quad i = 1, \ldots, p \end{aligned}$$

- $f, g_1, \ldots, g_m$ are convex functions
- $\mathbf{dom}\,(f)$ is a convex set
- equality constraint functions $h_i(x) = a_i^\top x - b$ are all affine.

The affine constraints are typically gathered into matrix form:

$$\begin{aligned} \min_{x \in \mathbf{dom}(f)} \quad & f(x) \\ \text{subj. to} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & Ax = b \quad A \in \mathbb{R}^{p \times m} \end{aligned}$$

Important property: Feasible set of a convex optimization problem is convex.

Local and Global Optimality for Convex Problems

Theorem

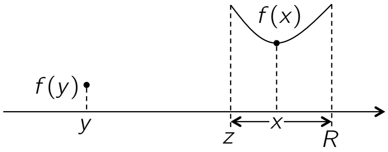
For a convex optimization problem, **any** locally optimal solution is globally optimal (local optima are global optima).

Proof:

- Assume that x is locally optimal, but not globally optimal.
- Therefore there is some other point y such that $f(y) < f(x)$.
- x locally optimal implies that there is some $R > 0$ such that

$$\|z - x\|_2 \leq R \Rightarrow f(x) \leq f(z)$$

- The problem can't be convex.



Local and Global Optimality for Convex Problems

Theorem

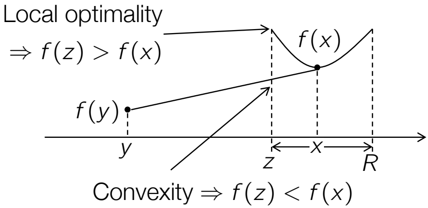
For a convex optimization problem, **any** locally optimal solution is globally optimal (local optima are global optima).

Proof:

- Assume that x is locally optimal, but not globally optimal.
- Therefore there is some other point y such that $f(y) < f(x)$.
- x locally optimal implies that there is some $R > 0$ such that

$$\|z - x\|_2 \leq R \Rightarrow f(x) \leq f(z)$$

- The problem can't be convex.



Equivalent Optimization Problems

Two problems are (informally) called **equivalent** if the solution to one can be (easily) inferred from the solution to the other, and vice versa.

- **Introducing equality constraints:**

$$\begin{aligned} \min_{x} \; & f(A_0 x + b_0) \\ \text{subj. to } & g_i(A_i x + b_i) \leq 0 \quad i = 1, \ldots, m \end{aligned}$$

  is equivalent to

$$\begin{aligned} \min_{x, y_i} \; & f(y_0) \\ \text{subj. to } \quad & g_i(y_i) \leq 0 \quad i = 1, \ldots, m \\ & A_i x + b_i = y_i \quad i = 0, 1, \ldots, m \end{aligned}$$

Equivalent Optimization Problems

Two problems are (informally) called **equivalent** if the solution to one can be (easily) inferred from the solution to the other, and vice versa.

- **Introducing slack variables for linear inequalities:**

$$\begin{aligned} \min_{x} \; & f(x) \\ \text{subj. to } & A_i x \leq b_i \quad i = 1, \ldots, m \end{aligned}$$

is equivalent to

$$\begin{aligned} \min_{x, s_i} \; & f(x) \\ \text{subj. to } & A_i x + s_i = b_i \quad i = 1, \ldots, m \\ & s_i \geq 0 \quad i = 1, \ldots, m \end{aligned}$$

Outline

5. Convex Optimization Problems

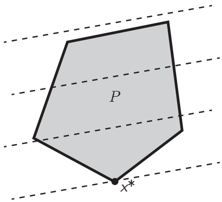
   Linear Programs

   Quadratic Programs

General Linear Program (LP)

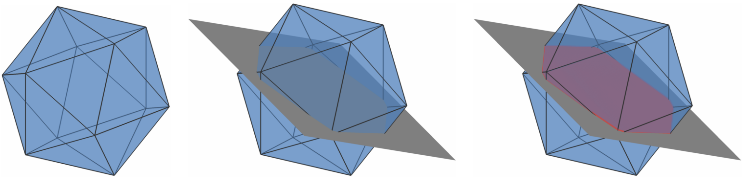
Affine cost and constraint functions:

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \ & c^\top x \\ \text{subj. to } & Gx \le h \\ & Ax = b\end{aligned}$$



Linear optimization on a polytope.

- Feasible set P is a polyhedron
- If P is empty, the problem is infeasible



(a) $Gx \le h$ (b) $A_i^\top x = b_i$ (c) $Gx \le h \cap A_i^\top x = b_i$

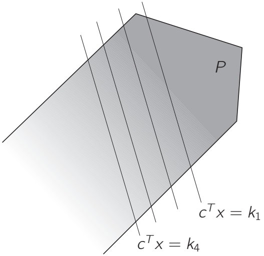
Graphical Interpretation and Solution Properties

Denote by p^* the optimal value and by X_{opt} the set of optimizers

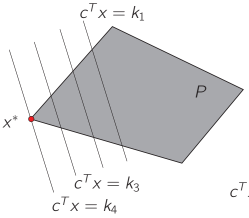
Case 1. The LP solution is unbounded, i.e., $p^* = -\infty$.

Case 2. The LP solution is bounded, i.e., $p^* > -\infty$ and the optimizer is unique. X_{opt} is a singleton.

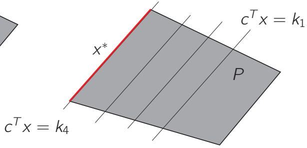
Case 3. The LP solution is bounded and there are multiple optima. X_{opt} is a subset of $\mathbb{R}^s$, which can be bounded or unbounded.



(a) Case 1



(b) Case 2



(c) Case 3

Outline

5. Convex Optimization Problems

   Linear Programs

   Quadratic Programs

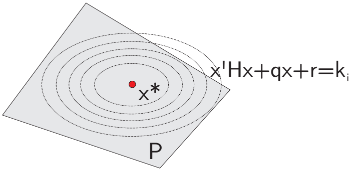
General Quadratic Program (QP)

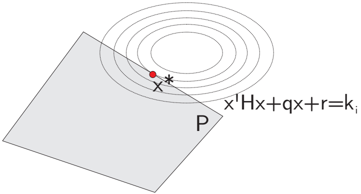
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^\top H x + q^\top x + r \\ \text{subj. to } \ & Gx \leq h \\ & Ax = b \end{aligned}$$

- Constant r can be left out, since it has no effect on the optimal solution.
- Convex if $H \succ 0$
- Problems with concave objective $H \not\succeq 0$ are quadratic programs, but hard.

Two cases can occur if feasible set P is not empty:

Case 1. The optimizer lies strictly inside the feasible polyhedron
Case 2. The optimizer lies on the boundary of the feasible polyhedron





Outline

Outline

6. Optimality Conditions

The Lagrangian Function

Recall our standard (possibly non-convex) optimization problem:

$$(P): \quad \begin{aligned} \min_{x \in \mathbf{dom}(f)} \quad & f(x) \\ \text{subj. to} \quad & g_i(x) \leq 0 \quad i = 1 \dots m \\ & h_i(x) = 0 \quad i = 1 \dots p \end{aligned}$$

with (primal) decision variable x, domain $\mathbf{dom}(f)$ and optimal value $p^\star$.

Lagrangian Function: $L : \mathbf{dom}(f) \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- λ_i : inequality Lagrange multiplier for $g_i(x) \leq 0$.
- ν_i : equality Lagrange multiplier for $h_i(x) = 0$.
- Lagrangian is a weighted sum of the objective and constraint functions.

Lagrange Dual Function

The **dual function** $d : \mathbb{R}^m \times \mathbb{R}^p$ is

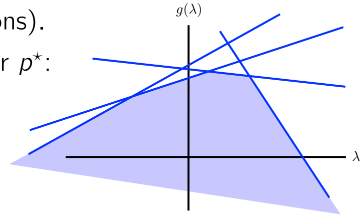
$$
\begin{aligned}
d(\lambda, \nu) &= \inf_{x \in \mathbf{dom}(f)} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathbf{dom}(f)} \left[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right]
\end{aligned}
$$

- The dual function $d(\lambda, \nu)$ is always a **concave** function (pointwise infimum of affine functions).
- The dual function generates lower bounds for $p^\star$:

$$d(\lambda, \nu) \leq p^\star, \qquad \forall (\lambda \geq 0, \nu \in \mathbb{R}^p)$$

- $d(\lambda, \nu)$ might be $-\infty$:

$$\mathbf{dom}\,(d) := \{\lambda, \nu \,|\, d(\lambda, \nu) > -\infty\}$$



The Primal and Dual Problem

A general optimization problem and its dual:

$$(P): \quad \begin{aligned} &\min_x \; f(x) \\ &\text{subj. to } g_i(x) \le 0 \quad i = 1 \dots m \\ &\qquad\quad\; h_i(x) = 0 \quad i = 1 \dots p, \end{aligned}$$

$$(D): \quad \begin{aligned} &\max_{\nu,\lambda} \; d(\nu, \lambda) \\ &\text{subj. to } \lambda \ge 0 \end{aligned}$$

- Problem (D) is **convex**, even if (P) is not.
- Problem (D) has optimal value $d^\star \le p^\star$.
- The point (λ, ν) is **dual feasible** if $\lambda \ge 0$ and $(\lambda, \nu) \in \mathbf{dom}\,(d)$.
- Can often impose the constraint $(\lambda, \nu) \in \mathbf{dom}\,(d)$ explicitly in (D).

Example : Dual of a Linear Program (LP)

$$(P): \quad \begin{aligned} \min_{x\in\mathbb{R}^n} \ & c^\top x \\ \text{subj. to } & Ax = b \\ & Cx \le e \end{aligned}$$

The **dual function** is

$$\begin{aligned} d(\lambda,\nu) &= \min_{x\in\mathbb{R}^n} \left[c^\top x + \nu^\top(Ax-b) + \lambda^\top(Cx-e)\right] \\ &= \min_{x\in\mathbb{R}^n} \left[(A^\top\nu + C^\top\lambda + c)^\top x - b^\top\nu - e^\top\lambda\right] \\ &= \begin{cases} -b^\top\nu - e^\top\lambda & \text{if } A^\top\nu + C^\top\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Lower bound property:
$-b^\top\nu - e^\top\lambda \le p^\star$ whenever $A^\top\nu + C^\top\lambda + c = 0$ and $\lambda \ge 0$.

Example : Dual of a Linear Program (LP)

$$(P): \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \; & c^\top x \\ \text{subj. to } & Ax = b \\ & Cx \leq e \end{aligned}$$

The **dual problem** is

$$(D): \quad \begin{aligned} \max_{\lambda,\nu} \; & -b^\top \nu - e^\top \lambda \\ \text{subj. to } & A^\top \nu + C^\top \lambda + c = 0 \\ & \lambda \geq 0 \end{aligned}$$

The dual of a linear program is also a linear program.

Example : Dual of a Quadratic Program

A quadratic program (QP) with $Q \succ 0$:

$$(P): \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^\top Q x + c^\top x \\ \text{subj. to } & Cx \leq e \end{aligned}$$

The **dual function** is

$$\begin{aligned} d(\lambda) &= \min_{x \in \mathbb{R}^n} \left[\frac{1}{2} x^\top Q x + c^\top x + \lambda^\top (Cx - e) \right] \\ &= \min_{x \in \mathbb{R}^n} \left[\frac{1}{2} x^\top Q x + (c + C^\top \lambda)^\top x - e^\top \lambda \right] \end{aligned}$$

The unconstrained minimization over x is convex for every λ. If $Q \succ 0$, then the optimal x satisfies

$$Qx + c + C^\top \lambda = 0$$

Example : Dual of a Quadratic Program (cont'd)

Substitute $x = -Q^{-1}(c + C^\top \lambda)$ into the dual function:

$$d(\lambda) = -\frac{1}{2}\left(c + C^\top \lambda\right)^\top Q^{-1}\left(c + C^\top \lambda\right) - e^\top \lambda$$

Dual of a QP:

The dual problem is to maximize $d(\lambda)$ over $\lambda \geq 0$, or equivalently,

$$(D): \quad \begin{aligned} &\min_{\lambda} \ \frac{1}{2}\lambda^\top C^\top Q^{-1} C\lambda + \left(CQ^{-1}c + e\right)^\top \lambda + \frac{1}{2}c^\top Q^{-1} c \\ &\text{subj. to } \lambda \geq 0 \end{aligned}$$

NB: Dual of a QP is another QP.

Outline

6. Optimality Conditions

Weak and Strong Duality

Weak Duality

- It is **always** true that $d^\star \leq p^\star$.

Strong Duality

- It is **sometimes** true that $d^\star = p^\star$.
- Strong duality usually does not hold for non-convex problems.
- Can impose conditions on convex problems to guarantee that $d^\star = p^\star$.
- Sometimes the dual is much easier to solve than the primal (or vice-versa).
- Example: The dual of a mixed integer linear program (difficult to solve) is a standard LP (easy to solve).

Strong Duality for Convex Problems

An optimization problem with f and all g_i convex:

$$(P): \quad \begin{aligned} \min\ & f(x) \\ \text{subj. to } & g_i(x) \leq 0 \quad i = 1 \ldots m \\ & Ax = b \quad A \in \mathbb{R}^{p \times n} \end{aligned}$$

Slater Condition

If there is at least one **strictly feasible point**, i.e.

$$\left\{ x \,\middle|\, Ax = b,\ g_i(x) < 0,\ \forall i \in \{1, \ldots, m\} \right\} \neq \varnothing$$

Then $p^\star = d^\star$.

Other **constraint qualification** conditions can also be used to check strong duality in convex problems.

Outline

6. Optimality Conditions

Karush-Kuhn-Tucker Conditions

Assume that all g_i and h_i are differentiable. **Necessary** conditions for optimality:

1) Primal Feasibility:

$$g_i(x^\star) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x^\star) = 0 \quad i = 1 \ldots, p$$

2) Dual Feasibility:

$$\lambda^\star \geq 0$$

3) Complementary Slackness:

$$\lambda_i^\star g_i(x^\star) = 0 \quad i = 1, \ldots, m$$

4) Stationarity:

$$\nabla_x L(x^\star, \lambda^\star, \nu^\star) = \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\star) + \sum_{i=1}^{p} \nu_i \nabla h_i(x^\star) = 0$$

Notes on Complementary Slackness

Assume that strong duality holds, with optimal solution $x^\star$ and $(\lambda^\star, \nu^\star)$.

1) From strong duality, $d^\star = p^\star \;\; \Rightarrow \;\; d(\lambda^\star, \nu^\star) = f(x^\star)$.

2) From the definition of the dual function:

$$\begin{aligned} f(x^\star) = d(\lambda^\star, \nu^\star) &= \min_x \left\{ f(x) + \sum_{i=1}^m \lambda_i^\star g_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right\} \\ &\leq f(x^\star) + \sum_{i=1}^m \lambda_i^\star g_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \overset{\text{[lower bound]}}{\leq} f(x^\star) \end{aligned}$$

$$\Rightarrow f(x^\star) = d(\lambda^\star, \nu^\star) = f(x^\star) + \sum_{i=1}^m \lambda_i^\star g_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star)$$

3) $$\left. \begin{array}{l} \lambda_i^\star = 0 \text{ for every } g_i(x^\star) < 0. \\ g_i(x^\star) = 0 \text{ for every } \lambda_i^\star > 0. \end{array} \right\} \textbf{Complementary slackness.}$$

KKT Conditions

For general optimization problem: Necessary condition

If $x^\star$ and $(\lambda^\star, \nu^\star)$ are primal and dual solutions, with zero duality gap, then $x^\star$ and $(\lambda^\star, \nu^\star)$ satisfy the KKT conditions.

For a convex optimization problem: Necessary and sufficient condition

If $x^\star$ and $(\lambda^\star, \nu^\star)$ satisfy the KKT conditions, then $x^\star$ and $(\lambda^\star, \nu^\star)$ are primal and dual optimal solutions.

If Slater's condition holds (i.e. strong duality holds), $x^\star$ and $(\lambda^\star, \nu^\star)$ are primal and dual optimal solutions **if and only if** they satisfy the KKT conditions.

Remark: KKT Conditions for Convex Problems

For a convex optimization problem, KKT conditions are sufficient:

If $(x^\star, \lambda^\star, \nu^\star)$ satisfy the KKT conditions, then $p^\star = d^\star$.

- $p^\star = f(x^\star) = L(x^\star, \lambda^\star, \nu^\star)$ (due to complementary slackness)
- $d^\star = g(\lambda^\star, \nu^\star) = L(x^\star, \lambda^\star, \nu^\star)$ (due to convexity of the functions and stationarity)

Example : KKT Conditions for a QP

Consider a (convex) quadratic program with $Q \succeq 0$:

$$(P): \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^\top Q x + c^\top x \\ \text{subj. to } & Ax = b \\ & x \geq 0 \end{aligned}$$

The **Lagrangian** is $L(x, \lambda, \nu) = \frac{1}{2} x^\top Q x + c^\top x + \nu^\top (Ax - b) - \lambda^\top x$.

The KKT conditions are:

$$\begin{aligned} \nabla_x L(x, \lambda, \nu) = Qx + A^\top \nu - \lambda + c &= 0 & & \text{[stationarity]} \\ Ax &= b & & \text{[primal feasibility]} \\ x &\geq 0 & & \text{[primal feasibility]} \\ \lambda &\geq 0 & & \text{[dual feasibility]} \\ x_i \lambda_i &= 0 \quad i = 1 \ldots n & & \text{[complementarity]} \end{aligned}$$

The final three conditions are often written together as $0 \leq x \perp \lambda \geq 0$.

Outline

6. Optimality Conditions

   The Lagrange Dual Problem

   Weak and Strong Duality

   Optimality Conditions

   Sensitivity Analysis

Sensitivity Analysis

A general optimization problem and its dual:

$$\begin{aligned} \min_x \ & f(x) \\ \text{subj. to } & g_i(x) \le 0 \quad i = 1 \dots m \\ & h_i(x) = 0 \quad i = 1 \dots p, \end{aligned}$$

$$\begin{aligned} \max_{\nu,\lambda} \ & d(\nu, \lambda) \\ \text{subj. to } & \lambda \ge 0 \end{aligned}$$

A perturbed optimization problem and its dual:

$$\begin{aligned} \min_x \ & f(x) \\ \text{subj. to } & g_i(x) \le u_i \quad i = 1 \dots m \\ & h_i(x) = v_i \quad i = 1 \dots p, \end{aligned}$$

$$\begin{aligned} \max_{\nu,\lambda} \ & d(\nu, \lambda) - u^\top \lambda - v^\top \nu \\ \text{subj. to } & \lambda \ge 0 \end{aligned}$$

- x is the primal decision variable. (λ, ν) are the dual decision variables.
- u and v are parameters representing perturbations to the constraints.
- $p^\star(u, v)$ is the optimal value as a function of (u, v).

Sensitivity and Lagrange Multipliers

Assume strong duality for the unperturbed problem with $(\nu^\star, \lambda^\star)$ dual optimal.

Weak duality for the perturbed problem implies

$$\begin{aligned} p^\star(u,v) &\geq g^\star(\nu^\star,\lambda^\star) - u^\top\lambda^\star - v^\top\nu^\star \\ &= p^\star(0,0) - u^\top\lambda^\star - v^\top\nu^\star \end{aligned}$$

Global Sensitivity Analysis

- $\lambda_i^\star$ large and $u_i < 0$ $\quad\Rightarrow$ $p^\star(u,v)$ increases greatly.
- $\lambda_i^\star$ small and $u_i > 0$ $\quad\Rightarrow$ $p^\star(u,v)$ does not decrease much.
- $\left\{\begin{array}{l} \nu^\star \text{ large and positive and } v_i < 0 \\ \nu^\star \text{ large and negative and } v_i > 0 \end{array}\right\}$ $\quad\Rightarrow$ $p^\star(u,v)$ increases greatly.
- $\left\{\begin{array}{l} \nu^\star \text{ small and positive and } v_i > 0 \\ \nu^\star \text{ small and negative and } v_i < 0 \end{array}\right\}$ $\quad\Rightarrow$ $p^\star(u,v)$ does not decrease much.

Note: Results are **not** symmetrical. We only have a lower bound on $p^\star(u,v)$.

Sensitivity and Lagrange Multipliers

Assume strong duality for the unperturbed problem with $(\nu^\star, \lambda^\star)$ dual optimal.

Weak duality for the perturbed problem implies

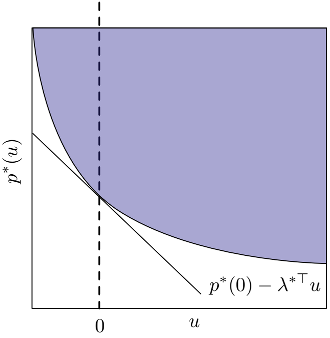
$$\begin{aligned} p^\star(u, v) &\geq g^\star(\nu^\star, \lambda^\star) - u^\top \lambda^\star - v^\top \nu^\star \\ &= p^\star(0, 0) - u^\top \lambda^\star - v^\top \nu^\star \end{aligned}$$

Local Sensitivity Analysis

If in addition $p^\star(u, v)$ is differentiable at $(0, 0)$, then

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

- $\lambda_i^\star$ is sensitivity of $p^\star$ relative to i^{th} inequality.
- $\nu_i^\star$ is sensitivity of $p^\star$ relative to i^{th} equality.



Summary: Convex Optimization

- Convex optimization problem:
    - Convex cost function
    - Convex inequality constraints
    - Affine equality constraints
- Benefit of convex problems: Local = Global optimality
- Only need to find one minimum, it is the global minimum!
- For convex optimization problem: If slater condition holds, x^* optimal iff $\exists(\lambda^*, \nu^*)$ satisfying KKT conditions
- Convex optimization problems can be solved efficiently
- Many problems can be written as convex opt. problems (with some effort)

Note: Duality and optimality conditions similarly extend to Convex Cone Programs

Summary: Why did we need the dual problem?

- The dual problem is convex, even if the primal is not
  –> can be 'easier' to solve than primal

- The dual problem provides a lower bound for the primal problem:

$$d^* \le p^* \quad \text{(and } d(\lambda, \nu) \le p(x) \text{ for all feasible } x, \lambda, \nu)$$

  (provides suboptimality bound)

- The dual provides a certificate of optimality via the KKT conditions for convex problems

- KKT conditions lead to efficient optimization algorithms

- Lagrange multipliers provide information about active constraints at the optimal solution: if $\lambda_i^* > 0$, then $g_i(x^*) = 0$

- Lagrange multipliers provide information about sensitivity of optimal cost: if λ_i^* karge, then tightening constraint will significantly increase cost