University of Pennsylvania, ESE 6190

Model Predictive Control

Chapter 4: Convex Optimization

Prof. Manfred Morari

Spring 2023

Coauthors: Prof. Paul Goulart, University of Oxford
Prof. Francesco Borrelli, UC Berkeley

F. Borrelli, A. Bemporad, and M. Morari, Predictive Control for Linear and Hybrid Systems, Cambridge University Press, 2017. [Ch. 1 & 2].

Outline

Outline

1. Introduction
2. Convex Sets
3. Convex Functions
4. Convex Optimization Problems
5. Duality
6. Generalized Inequalities

Outline

1. Introduction

    Motivation and Overview

    Common Types of Optimization Problems

Why Study Optimization?

Optimization is about making **good decisions** or choices in a rigorous way, often subject to **constraints**. Applications appear everywhere in science, mathematics, and business:

- Managing a share portfolio
- Scheduling public transport
- Fitting a model to measured data
- Optimizing a supply chain
- Designing electronic circuit layouts
- Choosing worker shift patterns
- Shaping aerodynamic components
- Recovering images from raw MRI data

$\Rightarrow$ Linear control design

$\Rightarrow$ Trajectory design for dynamic systems

Describing an Optimization Problem

$$\min_{x} \; f(x)$$
$$\text{subj. to } x \in \mathcal{X} \subseteq \mathbf{dom}\,(f)$$

The problem has several ingredients:

- The vector x collects the **decision variables**
- The set $\mathbf{dom}\,(f)$ is the **domain** of the decision variables
- The set $\mathcal{X} \subseteq \mathbf{dom}\,(f)$ is the **constraint set**, and describes the **feasible** decisions.
- The **objective** function $f : \mathbf{dom}\,(f) \to \mathbb{R}$ assigns a cost $f(x)$ to each decision x.

This problem can be written more compactly as

$$\min_{x \in \mathcal{X} \subseteq \mathbf{dom}(f)} f(x)$$

We call this **nonlinear mathematical program** or **nonlinear program (NLP).**

Describing an Optimization Problem

A more common problem format:

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

Defined by the following **problem data**:

- **Objective function** $f : \mathrm{dom}(f) \to \mathbb{R}$
- **Domain** $\mathrm{dom}(f) \subseteq \mathbb{R}^n$ of the objective function, from which the decision variable $x := (x_1; x_2; \ldots; x_n)$ must be chosen.
- Optional **inequality constraint functions** $g_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, m$
- Optional **equality constraint functions** $h_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, p$

NB: Any **maximization** problem can be written this way by a change of sign.

Unconstrained Optimization

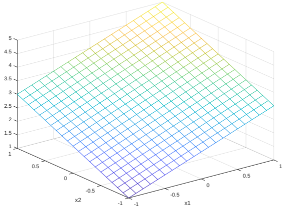
In unconstrained optimization $\mathrm{dom}\,(f) = \mathbb{R}^n$

Solving the optimization problem means to compute the least possible cost f^* and an associated optimal solution (or minimizer) x^*

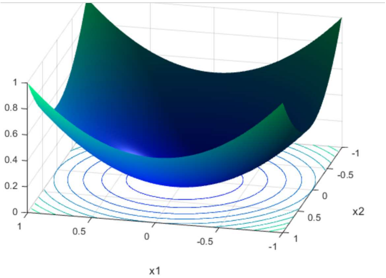
$$f(x^*) = f^*$$

Unconstrained Optimization: Examples (1/3)

Min of a linear function $c^\top x + k$

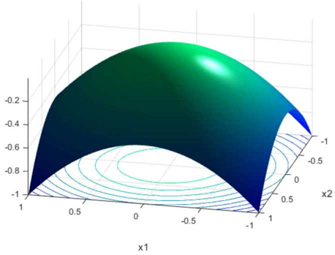


Min of a positive definite quadratic function $x^\top H x + c^\top x + k$

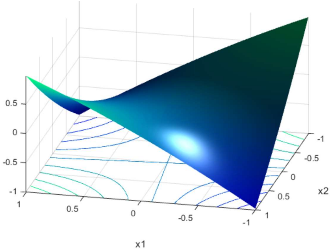


Unconstrained Optimization: Examples (2/3)

Min of a negative definite quadratic function $x^\top Hx + c^\top x + k$

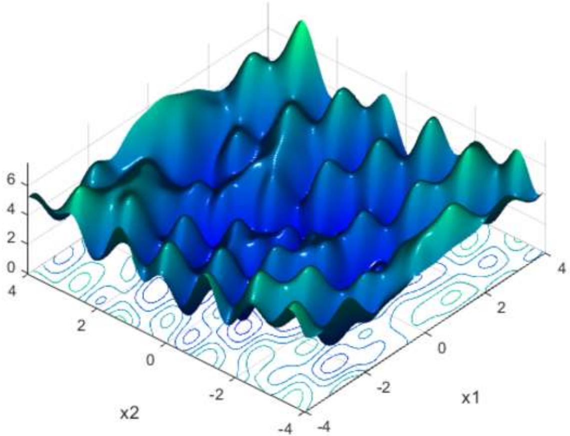


Min of an indefinite quadratic function $x^\top Hx + c^\top x + k$



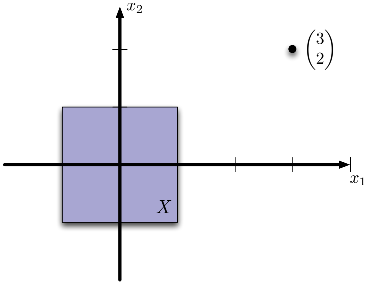
Unconstrained Optimization: Examples (3/3)

Min of **Wolfram** function $f(x) = \left\| \begin{matrix} x_1 - \sin(2x_1 + 3x_2) - \cos(3x_1 - 5x_2) \\ x_2 - \sin(x_1 - 2x_2) + \cos(x_1 + 3x_2) \end{matrix} \right\|$



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:

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

Properties of Optimization Problems

Consider the **Nonlinear Program** (NLP)

$$J^\star = \min_{x \in \mathcal{X}} f(x)$$

Notation:

- If $J^\star = -\infty$, then the problem is **unbounded below**.
- If the set $\mathcal{X}$ is empty, then the problem is **infeasible** (and we set $J^\star = +\infty$).
- If $\mathcal{X} = \mathbb{R}^n$, the problem is **unconstrained**.
- There might be more than one solution. The set of solutions is:

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

Terminology

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

Optimal value: The lowest possible objective value, $f(x^\star)$. Denoted by $f^\star$ (or $p^\star$, or $J^\star$).

Local optimality: x is locally optimal if there exists an $R > 0$ such that $z = x$ is optimal for

$$\begin{aligned} \min_{z \in \operatorname{dom}(f)} \quad & f(z) \\ \text{subj. to} \quad & g_i(z) \leq 0 \quad i = 1, \ldots, m, \\ & h_i(z) = 0 \quad i = 1, \ldots, p \\ & \|z - x\|_2 \leq R \end{aligned}$$

Terminology

Optimal solution: Any **feasible** $x^\star \in \mathrm{dom}\,(f)$ such that $f(x^\star) \le f(x)$ for all **feasible** $x \in \mathrm{dom}\,(f)$.

Local optimum: a point $x^\star_{\text{local}}$ that is optimal within a neighbourhood $||x - x^\star_{\text{local}}|| \le R$.

Technical point: The optimal value is called the **infimum**. A vector $x^\star$ that achieves the optimal value is a **minimizer** or optimal solution. There might be more than one minimizer, or none at all.

What might go wrong?

It is possible that no minimizer will exist:

- If the constraints are inconsistent, then the problem is **infeasible**.

$$\begin{aligned} \min_{x \in \mathbb{R}} \; & x^2 \\ \text{subj. to } & x \leq -1 \\ & x \geq 1 \end{aligned}$$

- It might be possible to make $f(x)$ arbitrarily negative without violating any of the constraints. Then the problem is referred to as **unbounded**.

$$\begin{aligned} \min_{x \in \mathbb{R}} \; & x \\ \text{subj. to } & x \leq 0 \end{aligned}$$

- The value $J^\star$ might be finite, but there is no x that achieves it.

$$\begin{aligned} \inf_{x \in \mathbb{R}} \; & e^{-x} \\ \text{subj. to } & x \geq 0 \end{aligned}$$

  The optimal value $J^\star = 0$ exists, but there are no optimal solutions.

Active, Inactive and Redundant Constraints

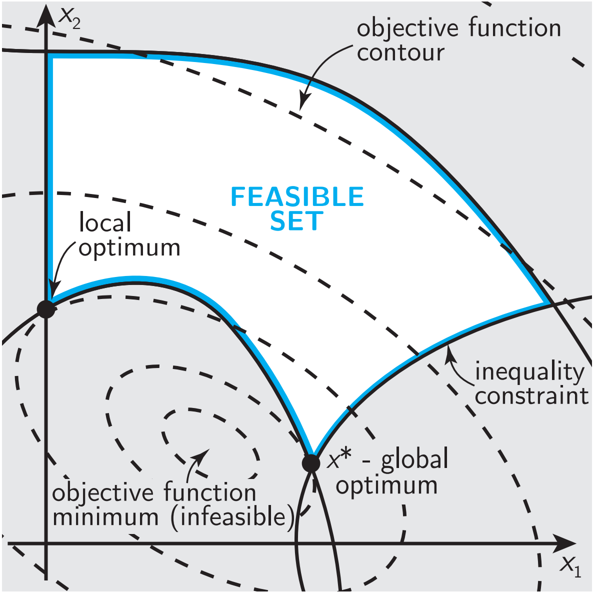
Consider the standard problem

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

- 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:

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

Geometry of an Optimization Problem



Implicit and Explicit Constraints

The constraints $g_i(x) \le 0$, $i = 1, \ldots, m$ and $h_i(x) = 0$, $i = 1, \ldots, p$ are referred to as the **explicit constraints** of the optimization problem. However, the **domains** of the objective function f and constraint functions also define an **implicit constraint** on x:

$$x \in \operatorname{dom}(f) \cap \bigcap_{i=1}^{m} \operatorname{dom}(g_i) \cap \bigcap_{i=1}^{p} \operatorname{dom}(h_i)$$

If a problem has $m = 0$ and $p = 0$, it is referred to as an **unconstrained problem**, although the limited domain of the **objective** function may still represent an implicit constraint. **Example:**

$$\min_x \; f(x) = -\sum_{i=1}^{k} \log(a_i^\top x - b_i)$$

is unconstrained but still has the implicit constraint that $a_i^\top x > b_i$ for $i = 1, \ldots, k$. In other words the constraint set $x \in \mathcal{X} = \{x \in \mathbb{R}^n \,|\, a_i^\top x > b_i,\ i = 1, \ldots, k\}$ is implied by the domain of f.

Feasibility Problem

The "constraint satisfiability" problem

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

is a special case of the general optimization problem:

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

- $p^\star = 0$ if the constraints are feasible (consistent). Every feasible x is optimal.
- $p^\star = \infty$ otherwise.

Outline

1. Introduction

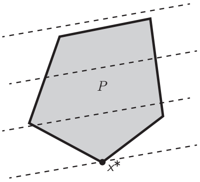
   Motivation and Overview

   Common Types of Optimization Problems

"Easier" problems

Linear Program (LP): Linear cost and constraint functions; feasible set is a polyhedron.

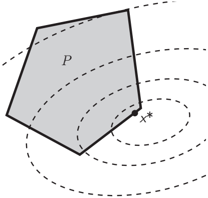
$$\begin{aligned} \min_x \;& c^\top x \\ \text{subj. to } & Gx \le h \\ & Ax = b \end{aligned}$$



Linear optimization on a polytope.

Convex Quadratic Program (QP): Quadratic cost and linear constraint functions; feasible set is a polyhedron. Convex if $P \succeq 0$.

$$\begin{aligned} \min_x \;& \frac{1}{2}x^\top Px + q^\top x \\ \text{subj. to } & Gx \le h \\ & Ax = b \end{aligned}$$

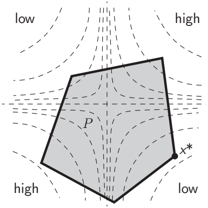


Convex quadratic optimization on a polytope.

"Harder" problems

Nonconvex Quadratic Program: $P \not\succeq 0$.

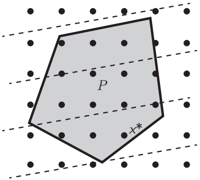
$$\begin{aligned} \min_x \ & \frac{1}{2}x^\top P x + q^\top x \\ \text{subj. to } & Gx \le h \\ & Ax = b \end{aligned}$$



Nonconvex quadratic optimization on a polytope. Contours represent a saddle-shaped objective function.

Mixed Integer Linear Program (MILP): Linear program with binary or integer constraints.

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



Linear optimization with integer constraints (dots).

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 LPs and QPs (and other standard types):
  - **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:

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

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:

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

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. Introduction

2. Convex Sets

3. Convex Functions

4. Convex Optimization Problems

5. Duality

6. Generalized Inequalities

Outline

2. Convex Sets

    Definition and Examples

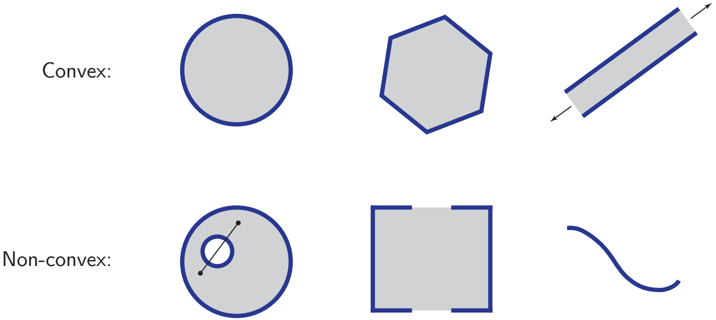
    Set Operations

Definition: Convex Set

A set $\mathcal{X}$ is convex **if and only if** for any pair of points x and y in $\mathcal{X}$, any **convex combination** $\mathrm{co}\,(x, y)$ lies in $\mathcal{X}$:

$$\mathcal{X} \text{ is convex} \quad \Leftrightarrow \quad \mathrm{co}\,(x, y) := \lambda x + (1 - \lambda) y \in \mathcal{X},\ \forall \lambda \in [0, 1],\ \forall x, y \in \mathcal{X}$$

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



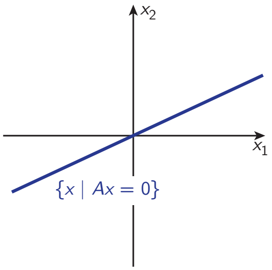
Definitions: Affine sets and Subspaces

An affine set is a convex set defined by $\mathcal{X} = \{x \in \mathbb{R}^n \mid Ax = b\}$. A subspace is an affine set with $b = 0$.

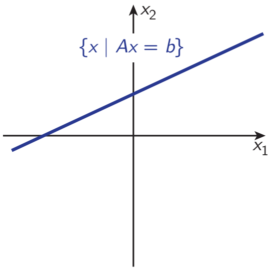
Verify convexity by definition — for all $x, y \in \mathcal{X}$, for all $\lambda \in [0, 1]$,

$$A(\lambda x + (1 - \lambda)y) = \lambda Ax + (1 - \lambda)Ay = \lambda \cdot b + (1 - \lambda) \cdot b = b$$

This definition encompasses lines, planes and individual points.



A 1D subspace in $\mathbb{R}^2$



An affine space in $\mathbb{R}^2$

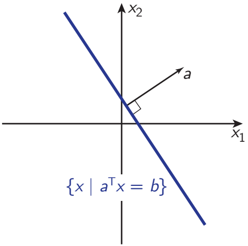
Definitions: Hyperplanes and halfspaces

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.

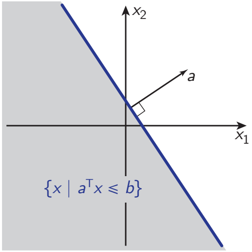
A halfspace is everything on one side of a hyperplane $\left\{x \in \mathbb{R}^n \,\middle|\, a^\top x \leq b\right\}$ for $a \neq 0$. It can either be **open** (strict inequality) or **closed** (non-strict inequality).

For $n = 2$, hyperplanes define lines. For $n = 3$, hyperplanes define planes. Compare to affine sets, which could define a line or a plane in $\mathbb{R}^3$.

Hyperplanes and halfspaces are always convex.



A hyperplane

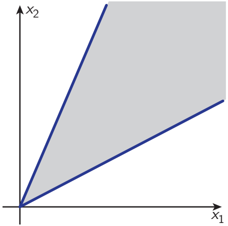


A closed halfspace

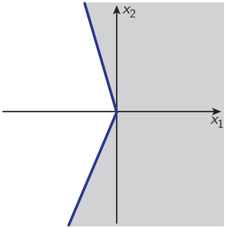
Definition: Cone

A set $\mathcal{X}$ is a cone if for all $x \in \mathcal{X}$, and for all $\theta > 0$, $\theta x \in \mathcal{X}$.
If the cone contains $x = 0$, then it is **pointed**.

A cone is **not necessarily convex**.



A convex cone



A non-convex cone

A **conic combination**, $\text{cone}\,(x_1, x_2)$, is any point that can be expressed as $\theta_1 x_1 + \theta_2 x_2$, for some $\theta_1, \theta_2 \geq 0$.

Definitions: Polyhedra and polytopes

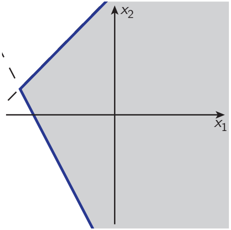
A polyhedron is the intersection of a **finite** number of closed halfspaces:

$$\begin{aligned}\mathcal{X} &= \{x \mid a_1^\top x \le b_1,\ a_2^\top x \le b_2, \ldots, a_m^\top \le b_m\} \\ &= \{x \mid Ax \le b\}\end{aligned}$$

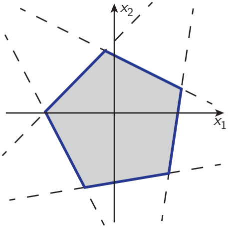
where $A := [a_1, a_2, \ldots, a_m]^\top$ and $b := [b_1, b_2, \ldots, b_m]^\top$.

A polytope is a **bounded** polyhedron.

Polyhedra and polytopes are always convex.



An (unbounded) polyhedron



A polytope

Definition: Vector norm

A norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions:

- $f(x) \geq 0$ and $f(x) = 0 \Rightarrow x = 0$.
- $f(tx) = |t| f(x)$ for scalar t.
- $f(x+y) \leq f(x) + f(y)$, for all $x, y \in \mathbb{R}^n$.

A norm is denoted $||x||_\bullet$, where the dot denotes the type of norm.
The notation $||x||$ refers to any arbitrary norm.

Definition: ℓ_p norm

The ℓ_p norm on $\mathbb{R}^n$ is denoted $||x||_p$, and is defined for any $p \geq 1$ by

$$||x||_p := \left[\sum_{i=1}^{n} |x_i|^p \right]^{1/p}$$

ℓ_p norms

By far the most common ℓ_p norms are:

- $p = 2$ (Euclidean norm):

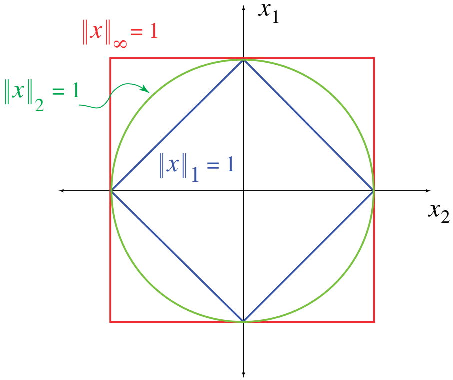
$$||x||_2 = \sqrt{\sum_i x_i^2}$$

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

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

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

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



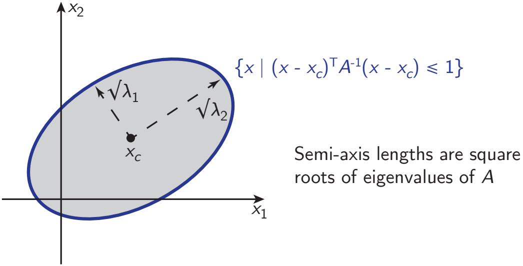
The **norm ball**, defined by $\{x \mid ||x - x_c|| \leq r\}$ where x_c is the centre of the ball and $r \geq 0$ is the radius, is always convex for any norm.

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).



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

Outline

2. Convex Sets

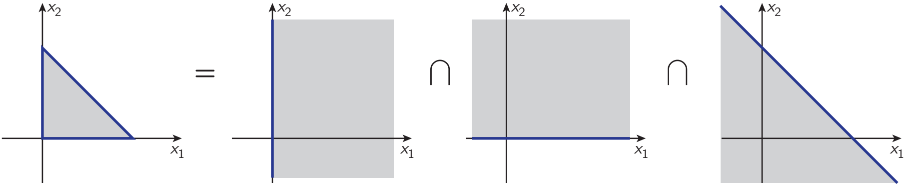
   Definition and Examples

   Set Operations

Intersection $\mathcal{X} \cap \mathcal{Y}$

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 $\mathcal{X} \cap \mathcal{Y}$.



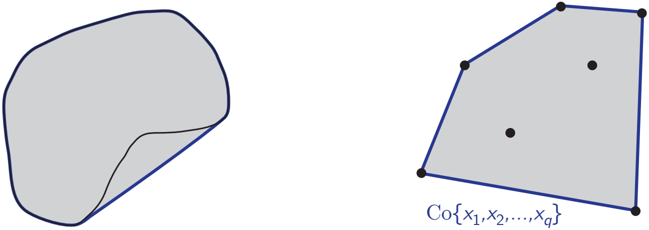
Many sets can be written as the intersection of convex elements, and are therefore easily shown to be convex. Any convex set can be written as a (possibly infinite) intersection of halfspaces.

Convex Hull $\text{conv}(\mathcal{X})$

The **convex hull** of a set $\mathcal{X}$ is the set of all convex combinations of points in $\mathcal{X}$:

$$\text{conv}(\mathcal{X}) := \{x \mid x = \lambda a + (1-\lambda)b,\ \lambda \in [0,1],\ a, b \in \mathcal{X}\}$$

It is the smallest convex set that contains $\mathcal{X}$: for all convex sets $\mathcal{Y} \supseteq \mathcal{X}$, $\text{conv}(\mathcal{X}) \subseteq \mathcal{Y}$.

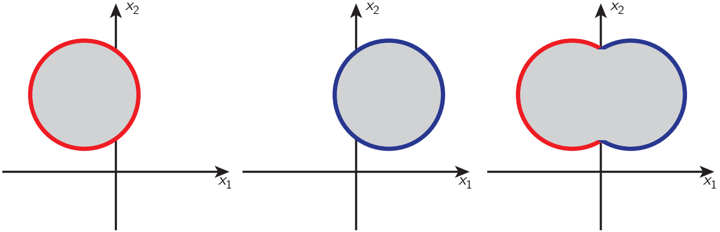


For a set $\mathcal{X} = \{x_1, x_2, \ldots, x_q\}$ with q points, the convex hull can be written

$$\text{conv}(\mathcal{X}) = \left\{ \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_q x_q \;\middle|\; \lambda_i \geq 0,\ i = 1, \ldots, q,\ \sum_{i=1}^{q} \lambda_i = 1 \right\}$$

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!



Set X Set Y Set $X \cup Y$

Outline

1. Introduction

2. Convex Sets

3. Convex Functions

4. Convex Optimization Problems

5. Duality

6. Generalized Inequalities

Outline

3. Convex Functions

    Definitions

    Examples

    Properties

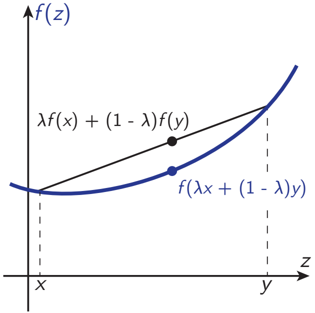
Definitions: Convex Function

A function $f : \textbf{dom}\,(f) \to \mathbb{R}$ is convex **iff**[a] its domain $\textbf{dom}\,(f)$ is convex **and**

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

The function f is strictly convex if this inequality is strict.

[a] "if and only if"



f is **concave** iff the function $-f$ is convex.

1st-order condition for convexity

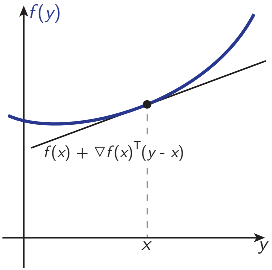
A differentiable function $f : \mathrm{dom}\,(f) \to \mathbb{R}$ with a convex domain is **convex iff**

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

i.e. a first order approximator of f around any point x is a global underestimator of f.

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

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



2nd-order condition for convexity

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

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

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**.

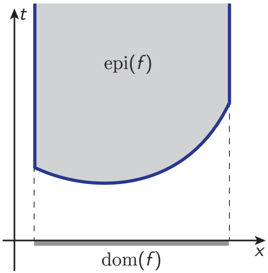
Epigraph of a Function

The **epigraph** of a function $f : \mathrm{dom}\,(f) \to \mathbb{R}$ is the **set**

$$\mathrm{epi}\,(f) := \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, x \in \mathrm{dom}\,(f)\,,\ f(x) \leq t \right\} \subseteq \mathrm{dom}\,(f) \times \mathbb{R}$$

It has dimension one higher than the domain of f.

A function is convex iff its epigraph is a convex set.



The epigraph of a convex function on a closed domain.

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

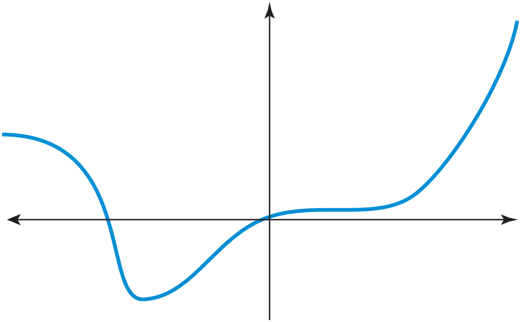
The **sublevel set** C_α of a function f for value α is defined by

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

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

Quasiconvex Functions

Function f is **quasi-convex iff** $\operatorname{dom}(f)$ is convex and all sublevel sets of f are convex.



A Quasiconvex function.

Restriction to a Line

Theorem

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex **iff** its evaluation along any line in its domain is convex.

Parameterizing the "distance along the line" by t:

$$g(t) = f(x + tv), \quad \mathbf{dom}\,(g) = \{t \,|\, x + tv \in \mathbf{dom}\,(f)\}$$

Function f is convex **iff** g is convex in t for all $x \in \mathbf{dom}\,(f)$, for all $v \in \mathbb{R}^n$.

- This means that convexity of f can be tested by checking functions of one variable.

Example: $f : \mathbb{S}^n \to \mathbb{R}$ with $f(X) = \log\det X$, and $\mathbf{dom}\,(f) = \mathbb{S}^n_{++}$. This function is concave. To test this, define $g(t) = f(X + tV)$, and check concavity of $g(t)$ for arbitrary X and V.[1]

[1] The set $\mathbb{S}^n$ denotes symmetric $n \times n$ matrices, $\mathbb{S}^n_+$ symmetric positive semidefinite matrices, and $\mathbb{S}^n_{++}$ symmetric positive definite matrices.

Extended-value Extension

A function f that is not defined everywhere can be extended to include points outside its domain by defining the **extended-value** function $\tilde{f}(x)$:

$$\tilde{f}(x) = f(x) \text{ for } x \in \mathbf{dom}\,(f)\,, \quad \tilde{f}(x) = +\infty \text{ for } x \notin \mathbf{dom}\,(f)$$

This often simplifies notation and does not change the epigraph of the function, i.e. $\mathbf{epi}\left(\tilde{f}\right) = \mathbf{epi}\,(f)$.

Also, $\tilde{f}$ is convex **iff** $\mathbf{dom}\,(f)$ is convex and f is convex.

Outline

3. Convex Functions

    Definitions

    Examples

    Properties

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 $\alpha \leq 0$
- Powers of absolute value: $|x|^p$, for $p \geq 1$

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 $\mathbb{R}_{++}$

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

Affine functions on $\mathbb{R}^n$ are both convex and concave:

- On $\mathbb{R}^n$, for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$:

$$f(x) = a^\top x + b$$

Vector Norms on $\mathbb{R}^n$ are all convex:

- On $\mathbb{R}^n$, ℓ_p norms have the form, for $p \geq 1$,

$$||x||_p = \left(\sum_{i=1}^{n} |x|^p \right)^{1/p}, \quad \text{with } ||x||_\infty = \max_i |x_i|$$

Examples of Convex Functions: $\mathbb{R}^{m \times n} \to \mathbb{R}$

Affine functions on $\mathbb{R}^{m \times n}$ are both convex and concave:

- On $\mathbb{R}^{m \times n}$, for some $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}$:

$$f(X) = \mathrm{trace}\left(A^\top X\right) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

Matrix Norms on $\mathbb{R}^{m \times n}$ are all convex:

- On $\mathbb{R}^{m \times n}$ the **spectral**, or **maximum singular value** norm is

$$||X||_2 = \sigma_{\max}(X) = [\lambda_{\max}(X^\top X)]^{1/2}.$$

Outline

3. Convex Functions

    Definitions

    Examples

    Properties

Convexity-preserving Operations (1/6)

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...

Convexity-preserving Operations (2/6)

Theorem: Non-negative weighted sum

If f is a convex function, then αf is convex for $\alpha \geq 0$. For several convex functions g_i, $\sum_i \alpha_i g_i$ is convex if all $\alpha_i \geq 0$.

Theorem: Composition with affine function

If f is a convex function, then $f(Ax + b)$ is convex.

Example: $||Ax - b||$ is convex for any norm.

Theorem: Pointwise maximum

If $f_1, \ldots, f_m$ are convex functions, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.

Example: Piecewise linear functions $\max_{i=1,\ldots,m}\{a_i^\top x + b\}$ are convex.

Convexity-preserving Operations (3/6)

Theorem: Pointwise supremum

If $f(x, y)$ is convex in x for every $y \in \mathcal{Y}$, then $g(x) = \sup_{y \in \mathcal{Y}} f(x, y)$ is convex.

Examples

- Support function of a set $\mathcal{A}$:

$$S_{\mathcal{A}}(x) = \sup_{y \in \mathcal{A}} y^\top x$$

- Distance to farthest point in a (possibly non-convex) set $\mathcal{B}$:

$$f(x) = \sup_{y \in \mathcal{B}} ||x - y||$$

- Maximum eigenvalue of a matrix $X \in \mathbb{S}^n$:

$$\lambda_{\max}(X) = \sup_{||y||_2 \leq 1} y^\top X y$$

Convexity-preserving Operations (4/6)

Theorem: Parametric Minimization

If $f(x, y)$ is convex in (x, y) and the set $\mathcal{C}$ is convex, then

$$g(x) = \min_{y \in \mathcal{C}} f(x, y)$$

is convex.

Example: $\text{dist}(x, \mathcal{S}) = \inf_{y \in \mathcal{S}} ||x - y||$ is convex if $\mathcal{S}$ is convex.

Convexity-preserving Operations (5/6)

Theorem: Composition with scalar functions

For $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$, $f(x) = h(g(x))$ is convex if:

- g is convex, h is convex, $\tilde{h}$ is non-decreasing
- g is concave, h is convex, $\tilde{h}$ is non-increasing

Examples

- $\exp g(x)$ for convex g
- $1/g(x)$ for concave positive g

Convexity-preserving Operations (6/6)

Theorem: Composition with vector functions

For $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$, $f(x) = h(g(x)) = h(g_1(x), \ldots, g_k(x))$ is convex if:

- Each g_i is convex, h is convex, $\tilde{h}$ is non-decreasing in each argument
- Each g_i is concave, h is convex, $\tilde{h}$ is non-increasing in each argument

Examples

- $\log \sum_{i=1}^{k} \exp g_i(x)$ is convex if all g_i are convex
- $\sum_{i=1}^{k} \log g_i(x)$ is concave for concave positive g_i

Outline

1. Introduction
2. Convex Sets
3. Convex Functions
4. Convex Optimization Problems
5. Duality
6. Generalized Inequalities

Outline

4. Convex Optimization Problems

   Standard Convex Optimization Problem

   Linear Programs

   Quadratic Programs

Standard Form Convex Optimization Problem

A standard form **convex** optimization problem:

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

This problem is convex if:

- The domain $\mathbf{dom}\,(f)$ is a convex set.
- The objective function f is a convex function.
- The inequality constraint functions g_i are all convex.
- The equality constraint functions $h_i(x) = a_i^\top x$ are all affine.

Standard Form Convex Optimization Problem

The affine constraints are typically gathered into matrix form:

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

Crucial Fact!

Theorem
For a convex optimization problem, **every** locally optimal solution is globally optimal.

NB: Writing or rewriting an optimization problem in convex form can be tricky, and is not always possible. It is always worth trying though.

Local and Global Optimality for Convex Problems

Theorem

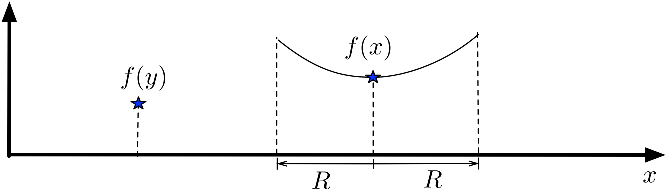
For a convex optimization problem, **every** locally optimal solution is globally optimal.

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.



Optimality Criterion for Differentiable f (1/3)

For a convex problem with a differentiable objective function f, x is optimal **iff** it is feasible, **and**

$$\nabla f(x)^\top (y - x) \geq 0, \quad \text{for all feasible } y.$$

The condition $\nabla f(x) = 0$ is the familiar sufficient first-order optimality condition.

The expression above states that the gradient may be non-zero, as long as all other feasible points are not "downhill" from the optimum.

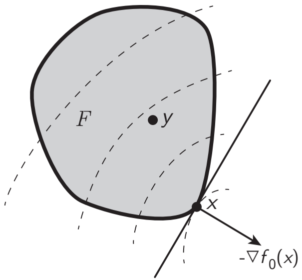


Illustration of the optimality condition for a feasible set $F \subset \mathbb{R}^2$.

Descent Direction for Differentiable f

Theorem: Descent Direction

If there exists a vector $\mathbf{d}$ such that $\nabla f(\bar{x})'\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\bar{x} + \lambda\mathbf{d}) < f(\bar{x})$ for all $\lambda \in (0, \delta)$.

- The vector $\mathbf{d}$ in the theorem above is called **descent direction**.
- The direction of **steepest descent $\mathbf{d}_s$** at $\bar{x}$ is defined as the normalized direction where $\nabla f(\bar{x})'\mathbf{d}_s < 0$ is minimized.
- The direction $\mathbf{d}_s$ of steepest descent is $\mathbf{d}_s = -\frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|}$

Optimality Criterion for Differentiable f (2/3)

Unconstrained problem:

$$\min_x \quad f(x)$$

x is optimal **iff** $\quad x \in \mathrm{dom}\,(f), \quad \nabla f(x) = 0.$

Equality constrained problem:

$$\begin{aligned} &\min_x \ f(x) \\ &\text{subj. to } Ax = b \end{aligned}$$

x is optimal **iff** $\quad x \in \mathrm{dom}\,(f), \quad Ax = b, \quad \exists \nu : \nabla f(x) + A^\top \nu = 0.$

Minimization over non-negative orthant:

$$\begin{aligned} &\min_x \ f(x) \\ &\text{subj. to } x \geq 0 \end{aligned}$$

x is optimal **iff**
$x \in \mathrm{dom}\,(f), \quad x \geq 0, \quad \nabla f(x)_i \geq 0,\ x_i = 0;\ \nabla f(x)_i = 0,\ x_i > 0.$

Optimality Criterion for Differentiable f (3/3)

Nonconvex f, Unconstrained Problem

- **Necessary condition**
  If $x^\star$ is a local minimizer, the $\nabla f(x^\star) = 0$.
- **Sufficient condition**
  Suppose that f is twice differentiable at $x^\star$. If $\nabla f(x^\star) = 0$ and the Hessian of $f(x)$ at $x^\star$ is positive definite, then $x^\star$ is a local minimizer.
- **Necessary and sufficient condition**
  Suppose that $f(x) = x^\top Hx + c^\top x + k$. If H is positive definite, then $x^\star$ is the global minimizer if and only if $\nabla f(x^\star) = 0$.

Proofs available in Ch.4 of Bazaraa, Sherali, and Shetty.
Nonlinear Programming - Theory and Algorithms. John Wiley & Sons, Inc.
New York, second edition, 1993.

A Well Known Optimization Problem - Least Squares

Least squares:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

- Analytical solution $x^\star = A^\dagger b$, if A full column rank.
- $A^\dagger = (A^\top A)^{-1} A^\top$
- $A^\dagger$ is often called the pseudo-inverse.
- In Matlab $x^\star = A \backslash b$
- Proof:

$$\min_x \|Ax - b\|_2^2 = \min_x x^\top (A^\top A) x - x^\top (2A^\top b) + b^\top b$$

  If A is full column rank then $A^\top A$ is positive definite; from previous Theorem at $x^\star$

$$\nabla f(x^\star) = 0 \Rightarrow (2A^\top A) x^\star = (2A^\top b)$$

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:**

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

  is equivalent to

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

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:**

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

is equivalent to

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

Outline

4. Convex Optimization Problems

   Standard Convex Optimization Problem

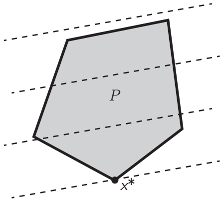
   Linear Programs

   Quadratic Programs

General Linear Program (LP)

Affine cost and constraint functions:

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



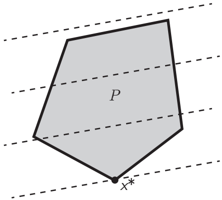
Linear optimization on a polytope.

- Feasible set is a polyhedron.
- Constant component d can be left out – it has no effect on the optimal solution.

General Linear Program (LP)

An alternative format:

$$\begin{aligned} \min_x \; & c^\top x \\ \text{subj. to } & Ax = b \\ & x \geq 0 \end{aligned}$$



Linear optimization on a polytope.

- All components of x are non-negative.
- Can easily convert previous format to this (using extra variables).

Many problems can be rewritten (with some effort!) into LPs.

Huge variety of solution methods and software are available.

Example Linear Programs

Cheapest cat-food problem:

- Choose quantities $x_1, x_2, \ldots, x_n$ of n different ingredients with unit cost c_j.
- Each ingredient j has nutritional content a_{ij} for nutrient i.
- Require for each nutrient i minimum level b_i.

In linear program form:

$$\begin{aligned} \min_x \ & c^\top x \\ \text{subj. to } & Ax \geq b \\ & x \geq 0 \end{aligned}$$

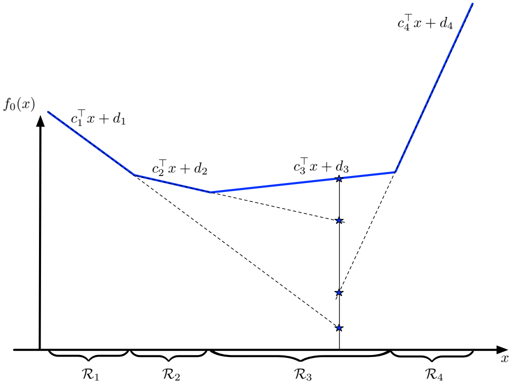
This is an example of a **resource allocation** problem.

Kantorovich and Koopmans won the Nobel Prize in Economics in 1975 for their work on this problem (and its non-cat-food variants).

Example : Piecewise Affine Minimization

$$\min_x \left[\max_{i=1,\dots,m} \{c_i^\top x + d_i\} \right]$$
$$\text{subj. to } Gx \le h$$

- The function is affine on each region $\mathcal{R}_i$.
- Any convex and piecewise affine function can be written this way.
- Can be reformulated as an LP.



A piecewise affine (PWA) convex function.

Example Linear Programs

Piecewise affine minimization:

$$\begin{aligned} \min_x \; & \left[\max_{i=1,\ldots,m} \{c_i^\top x + d_i\} \right] \\ \text{subj. to } & Gx \le h \end{aligned}$$

is **equivalent** to an LP:

$$\begin{aligned} \min_{x,t} \; & t \\ \text{subj. to } & c_i^\top x + d_i \le t \quad i = 1, \ldots, m \\ & Gx \le h \end{aligned}$$

NB: trick was to add variables and write the problem in **epigraph** form.

ℓ_∞ minimization

Constrained ℓ_∞ (Chebyshev) minimization:

$$\min_{x \in \mathbb{R}^n} \; ||x||_\infty$$
$$\text{subj. to } Fx \leq g$$

Write this is a max of linear functions.

Equivalent to:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \; [\max\{x_1, \dots, x_n, -x_1, \dots, -x_n\}] \\ &\text{subj. to } Fx \leq g \end{aligned}$$

ℓ_∞ minimization (cont'd)

Equivalent to:

$$\begin{aligned} \min_{x,t} \; & t \\ \text{subj. to} \quad & x_i \leq t \quad i = 1, \ldots, n \\ & -x_i \leq t \quad i = 1, \ldots, n \\ & Fx \leq g \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} \min_{x,t} \; & t \\ \text{subj. to} \; & -\mathbf{1}t \leq x \leq \mathbf{1}t \\ & Fx \leq g \end{aligned}$$

- The notation '$\mathbf{1}$' indicates a vector of ones.
- The constraint $-\mathbf{1}t \leq x \leq \mathbf{1}t$ bounds the absolute value of every element of x with a common scalar variable t.

ℓ_1 minimization

Constrained ℓ_1 minimization:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & ||Ax - b||_1 \\ \text{subj. to } & Fx \leq g \end{aligned}$$

Write this is a max of linear functions. Assume $A \in \mathbb{R}^{m \times n}$.

Equivalent to:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \left[\sum_{i=1}^{m} \max \left\{ (Ax - b)_i, -(Ax - b)_i \right\} \right] \\ \text{subj. to } & Fx \leq g \end{aligned}$$

ℓ_1 minimization (cont'd)

Equivalent to:

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}^m} \quad & t_1 + \ldots + \ldots t_m \\ \text{subj. to} \quad & (Ax-b)_i \leq t_i \quad i = 1,\ldots,m \\ & -(Ax-b)_i \leq t_i \quad i = 1,\ldots,m \\ & Fx \leq g \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}^m} \quad & \mathbf{1}^\top t \\ \text{subj. to} \quad & -t \leq (Ax-b) \leq t \\ & Fx \leq g \end{aligned}$$

- The notation '**1**' indicates a vector of ones.
- The constraint $-t \leq (Ax-b) \leq t$ bounds the absolute value of each component of $(Ax-b)$ with a component of the vector variable t.

Outline

4. Convex Optimization Problems

   Standard Convex Optimization Problem

   Linear Programs

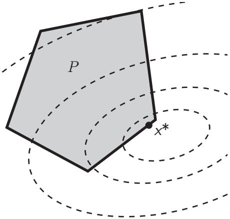
   Quadratic Programs

General Quadratic Program

Quadratic cost function with $P \in \mathbb{S}^n_+$, affine constraint functions:

$$\begin{aligned} \min_x \ & \frac{1}{2} x^\top P x + q^\top x + r \\ \text{subj. to } & Gx \leq h \\ & Ax = b \end{aligned}$$

- Constant r can be left out, since it has no effect on the optimal solution.
- Maximization problems with a concave objective function ($-P \in \mathbb{S}^n_+$) are also quadratic programs.



Optimization of a quadratic objective function over a polytopic set P.
The level sets of the objective are shown as dotted lines.

Example Quadratic Programs

Least squares:

$$\min_x ||Ax - b||_2^2$$

- Analytical solution $A^\dagger b$ ($A^\dagger$ is the pseudo-inverse).
- Extra linear constraints $l \leq x \leq u$ can be added, although the QP would no longer have an analytical solution.

Linear program with random cost:

$$\begin{aligned} \min_x \ & \mathbb{E}[c^\top x] + \gamma \ \mathrm{var}(c^\top x) = \overline{c}^\top x + \gamma x^\top \Sigma x \\ \text{subj. to } & Gx \geq h \\ & Ax = b \end{aligned}$$

- Random cost function vector c with mean $\overline{c}$ and covariance Σ, we wish to penalize expected cost plus a “risk premium” γ on the variance.
- Hence $c^\top x$ is a random variable with mean $\overline{c}^\top x$ and variance $x^\top \Sigma x$.
- Large γ means large risk aversion — we prefer a small variance to the lowest expected cost.

Example Quadratic Programs

Tikhonov Regularization: Least squares with extra penalty for nonzero terms.

$$\min_{x \in \mathbb{R}^n} ||Ax - b||_2^2 + \gamma \cdot ||x||_1$$

Equivalent to:

$$\begin{aligned} &\min_{t, x \in \mathbb{R}^n} \; ||Ax - b||_2^2 + \gamma \cdot \mathbf{1}^\top t \\ &\text{subj. to } -t \leq x \leq t \end{aligned}$$

- A larger penalty γ will tend to produce sparser solutions.
- Note that we have converted an **unconstrained** problem into a larger **constrained** one to get it into standard QP form.
- Requires $\gamma \geq 0$ for convexity.

Outline

Outline

5. Duality

   The Lagrange Dual Problem

   Weak and Strong Duality

   Optimality Conditions

   Sensitivity Analysis

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

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

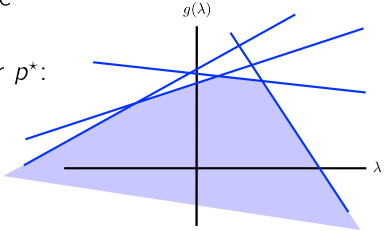
The dual function $d(\lambda, \nu)$ is always a **concave** function.

- $d(\lambda, \nu)$ is the pointwise infimum of affine functions.
- 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 $-\infty$:

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



Example : Least norm solution to a linear system

$$(P): \quad \begin{array}{l} \min\limits_{x \in \mathbb{R}} \; x^\top x \\ \text{subj. to } Ax = b \end{array}$$

The **Lagrangian** is $L(x, \nu) = x^\top x + \nu^\top (Ax - b)$.

Dual function:
Minimize the Lagrangian $L(x, \nu)$ by setting the gradient to zero:

$$\nabla_x L(x, \nu) = 2x + A^\top \nu = 0 \quad \Rightarrow \quad x = -\frac{1}{2} A^\top \nu.$$

Substitute back into L to get the dual function:

$$d(\nu) = -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu \qquad \text{[a concave function]}$$

Lower bound property: $-\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu \leq p^\star$ for every ν.

The Dual Problem

Every $\nu \in \mathbb{R}^p$, $\lambda \geq 0$ produces a lower bound for $p^\star$ using the dual function. Which is the best?

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

- Problem (D) is **concave**, even if (P) is not.
- Problem (D) has optimal value $d^\star \leq p^\star$.
- The point (λ, ν) is **dual feasible** if $\lambda \geq 0$ 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)

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

The **dual problem** is

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

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

Example : Norm minimization with equality constraint

$$(P): \quad \begin{aligned} &\min_x \ \|x\|_2 \\ &\text{subj. to } Ax = b \end{aligned}$$

The **dual function** is

$$d(\lambda) = \min_x \left[\|x\| - (A^\top \nu)^\top x + b^\top \nu\right]$$

$$= \begin{cases} b^\top \nu & \text{if } \|A^\top \nu\|_2 \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

The **dual problem** is

$$(D): \quad \begin{aligned} &\max_\nu \ b^\top \nu \\ &\text{subj. to } \|A^\top \nu\|_2 \leq 1 \end{aligned}$$

Lower bound property: $b^\top \nu \leq p^\star$ whenever $\|A^\top \nu\|_2 \leq 1$.

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,

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

NB: Dual of a QP is another QP.

Example : Dual of a Mixed-Integer Linear Program (MILP)

$$(P): \quad \begin{aligned} &\min_{x \in \mathcal{X}} \; c^\top x \\ &\text{subj. to } Ax \leq b \\ &\qquad \mathcal{X} = \{-1,\, 1\}^n \end{aligned}$$

The **dual function** is

$$d(\lambda) = \min_{x_i \in \{-1,1\}} \left[c^\top x + \lambda^\top (Ax - b) \right]$$

$$= -\|A^\top \lambda + c\|_1 - b^\top \lambda$$

The **dual problem** is

$$(D): \quad \begin{aligned} &\max_{\lambda} \; -\|A^\top \lambda + c\|_1 - b^\top \lambda \\ &\text{subj. to } \lambda \geq 0 \end{aligned}$$

The dual of a mixed-integer LP is a LP (without integers).

Outline

5. Duality

   The Lagrange Dual Problem

   Weak and Strong Duality

   Optimality Conditions

   Sensitivity Analysis

Weak and Strong Duality

- **Weak Duality** :

$$d^\star \leq p^\star$$

always holds.

- **Strong Duality**:

$$d^\star = p^\star$$

generally does not hold, even for convex problems.

- **Duality Gap**:

$$p^\star - d^\star.$$

- **Certificate of Optimality**:

$$d \leq d^\star \leq p^\star \leq p$$

The dual and primal costs bound the **optimal** dual and primal costs.

A Geometric Interpretation

Assume one inequality constraint only:

$$\mathcal{G} := \{(u,t) \mid t = f(x),\ u = g(x),\ x \in \mathcal{X}\}$$

Primal problem:

$$p^\star = \min \{t \mid (u,t) \in \mathcal{G},\ \ u \le 0\}$$

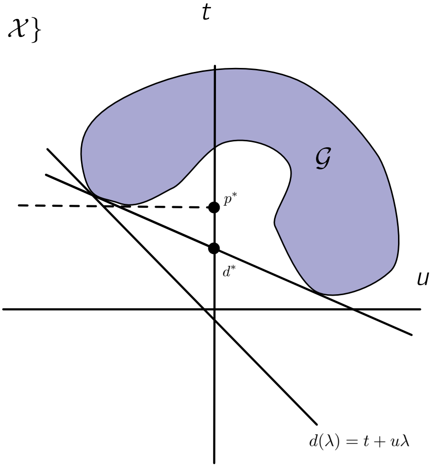
Dual function:

$$d(\lambda) = \min_{(u,t)\in\mathcal{G}} (t + \lambda u)$$

Dual problem:

$$d^\star = \max_{\lambda \ge 0} d(\lambda)$$

The quantity $p^\star - d^\star$ is the **duality gap**.



Strong Duality for Convex Problems

An optimization problem with f and all g_i convex:

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

Slater Condition

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

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

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

- Stronger version: Only $g_i(x)$ must be strictly satisfiable.
- Strong duality holds for LPs if at least one of the problems (primal or dual) is feasible.

Other **constraint qualification** conditions exist to check strong duality in convex problems.

Outline

5. Duality

   The Lagrange Dual Problem

   Weak and Strong Duality

   Optimality Conditions

   Sensitivity Analysis

Primal and Dual Solution Properties

Assume strong duality ($d^\star = p^\star$). Let $x^\star$ and $(\lambda^\star, \nu^\star)$ be primal and dual solutions. Then from the definition of the dual function:

$$
\begin{aligned}
f(x^\star) = d(\lambda^\star, \nu^\star) &= \min_x \left\{ f(x) + \sum_{i=1}^{m} \lambda_i^\star g_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right\} \\
&\leq f(x^\star) + \underbrace{\sum_{i=1}^{m} \lambda_i^\star g_i(x^\star)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^\star h_i(x^\star)}_{=0} \\
&\leq f(x^\star)
\end{aligned}
$$

$\Rightarrow$ We must have $\sum_{i=1}^{m} \lambda_i^\star g_i(x^\star) = 0$ (complementary slackness).
$\lambda^\star = 0$ for every $g_i^\star(x^\star) < 0$ and $g_i^\star(x^\star) = 0$ for every $\lambda^\star > 0$.

$\Rightarrow$ All inequalities are equalities.

$\Rightarrow$ The point $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ (stationarity condition).

Karush-Kuhn-Tucker Conditions (Necessity)

Assume that all g_i and h_i are differentiable.
If $d^\star = p^\star$, then the optimal solutions $x^\star$, $(\lambda^\star, \nu^\star)$ satisfy the KKT conditions:

1) Primal Feasibility:

$$\begin{aligned} g_i(x^\star) &\leq 0 \quad i = 1, \dots, m \\ h_i(x^\star) &= 0 \quad i = 1 \dots, p \end{aligned}$$

2) Dual Feasibility:

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

3) Complementary Slackness:

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

4) Stationarity:

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

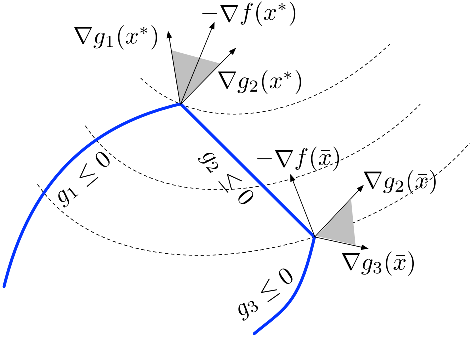
Geometric Interpretation of KKT Conditions

Assume inequality constraints only.

Rewrite stationarity condition as:

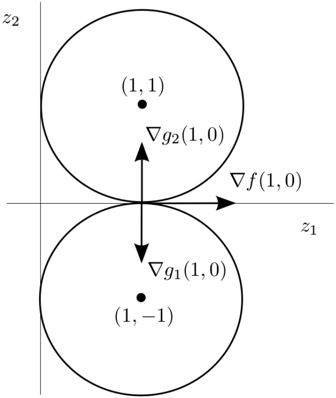
$$-\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x)$$

- Direction of steepest descent is in convex cone spanned by constraint gradients ∇g_i



Slater Condition: Example

$$\min_{z_1, z_2} \; z_1$$
$$\text{subj. to } (z_1 - 1)^2 + (z_2 - 1)^2 \leq 1$$
$$(z_1 - 1)^2 + (z_2 + 1)^2 \leq 1$$



- Single feasible point $z = (1, 0)$ does not satisfy Slater's condition.
- KKT stationarity condition does not hold for any $\lambda_i \geq 0$.

KKT Conditions for Convex Problems (Sufficiency)

For a convex optimization problem:

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

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.

Example : KKT Conditions for a QP

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

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

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

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 $0 \leq x \perp \lambda \geq 0$.

Example : Implicit vs. Explicit Constraints

It is sometimes helpful to keep some or all of the constraints in $\mathcal{X}$.

Example: Box-constrained LP (method 1)

$$(P): \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \ & c^\top x \\ \text{subj. to} \quad & Ax = b \\ & -\mathbf{1} \leq x \leq \mathbf{1} \end{aligned}$$

Dual function:

$$\begin{aligned} d(\bar{\lambda}, \underline{\lambda}, \nu) &= \min_{x \in \mathbb{R}^n} \left[c^\top x + (Ax - b)^\top \nu + (-\mathbf{1} - x)^\top \underline{\lambda} + (-\mathbf{1} + x)^\top \bar{\lambda} \right] \\ &= \begin{cases} -b^\top \nu - \mathbf{1}^\top \bar{\lambda} - \mathbf{1}^\top \underline{\lambda} & \text{if } c + A^\top \nu - \underline{\lambda} + \bar{\lambda} = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Dual Problem:

$$(D): \quad \begin{aligned} \max_{\nu, \bar{\lambda}, \underline{\lambda}} \ & -b^\top \nu - \mathbf{1}^\top \bar{\lambda} - \mathbf{1}^\top \underline{\lambda} \\ \text{subj. to} \ & c + A^\top \nu - \underline{\lambda} + \bar{\lambda} = 0 \\ & \bar{\lambda} \geq 0, \ \underline{\lambda} \geq 0 \end{aligned}$$

Example : Implicit vs. Explicit Constraints

It is sometimes helpful to keep some or all of the constraints in $\mathcal{X}$.

Example: Box-constrained LP (method 2)

$$(P): \quad \begin{aligned} \min_{\|x\|_\infty \le 1} \ & c^\top x \\ \text{subj. to } & Ax = b \end{aligned}$$

Dual function:

$$\begin{aligned} d(\nu) &= \min_{\|x\|_\infty \le 1} \left[c^\top x + (Ax - b)^\top \nu \right] \\ &= -b^\top \nu - \|A^\top \nu + c\|_1 \end{aligned}$$

Dual Problem:

(D) :

$$\begin{aligned} \max_{\nu} \ & -b^\top \nu - \|A^\top \nu + c\|_1 \\ \text{subj. to } & [\text{no constraints}] \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \max_{\nu, \lambda_+, \lambda_-} \ & -b^\top \nu - \mathbf{1}^\top \lambda_+ - \mathbf{1}^\top \lambda_- \\ \text{subj. to } & A^\top \nu + c = \lambda_+ - \lambda_- \\ & \lambda_+ \ge 0, \ \lambda_- \ge 0 \end{aligned}$$

Outline

Sensitivity Analysis

A general optimization problem and its dual:

$$\begin{aligned} \min_x \ & f(x) \\ \text{subj. to } & g_i(x) \leq 0 \quad i = 1 \dots m \\ & h_i(x) = 0 \quad i = 1 \dots p, \end{aligned} \qquad \Bigg| \qquad \begin{aligned} \max_{\nu,\lambda} \ & d(\nu, \lambda) \\ \text{subj. to } & \lambda \geq 0 \end{aligned}$$

A perturbed optimization problem and its dual:

$$\begin{aligned} \min_x \ & f(x) \\ \text{subj. to } & g_i(x) \leq u_i \quad i = 1 \dots m \\ & h_i(x) = v_i \quad i = 1 \dots p, \end{aligned} \qquad \Bigg| \qquad \begin{aligned} \max_{\nu,\lambda} \ & d(\nu, \lambda) - u^\top \lambda - v^\top \nu \\ \text{subj. to } & \lambda \geq 0 \end{aligned}$$

- 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

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

Global Sensitivity Analysis

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

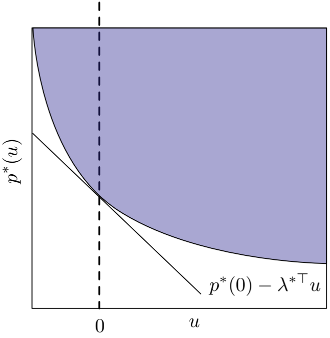
$$\begin{aligned} p^\star(u, v) &\geq d^\star(\nu^\star, \lambda^\star) - u^\top \lambda^\star - v^\top \nu^\star \\ &= p^\star(0, 0) - u^\top \lambda^\star - v^\top \nu^\star \end{aligned}$$

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.



Outline

Outline

6. Generalized Inequalities

    Convex cones

    Convex Cone Programs (CPs)

    Duality for Cone Programs

    Optimality Conditions for Cone Programs

Convex cones

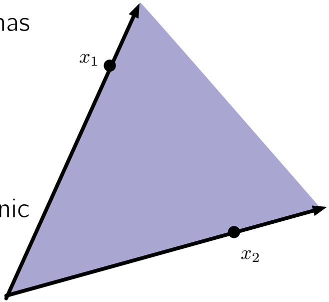
Cone

A set K is a cone if, for all $x \in K$, $\theta x \in K$ ($\theta > 0$).

Conic Combinations

A conic combination x of two points x_1 and x_2 has the form

$$x = \theta_1 x_1 + \theta_2 x_2, \qquad \theta_1 > 0,\ \theta_2 > 0.$$



Convex Cone

A convex cone is a set that contains every conic combination of points in the set.

Notation

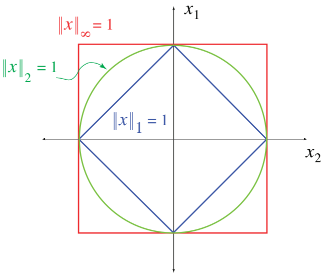
We will use the notation $x \succeq_K 0$ to mean that x is in the cone K.

Norm balls and norm cones

Norm ball

A norm ball with center x_c is radius r is the set $\{x \mid \|x - x_c\| \le r\}$.

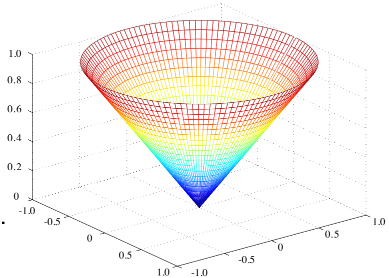
The 'shape' of the ball depends on the norm.



Norm cone

A norm cone is a set

$$\{(t, x) \mid \|x\| \le t\}$$



Norm balls and norm cones are convex sets.

Positive semidefinite cone

Positive semidefinite matrices

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n \,|\, X \succeq 0\} \qquad \text{(positive semidefinite } n \times n \text{ matrices)}$$

$$X \in \mathbb{S}^n_+ \iff z^\top X z \geq 0, \;\text{ for all } z.$$

Positive definite matrices

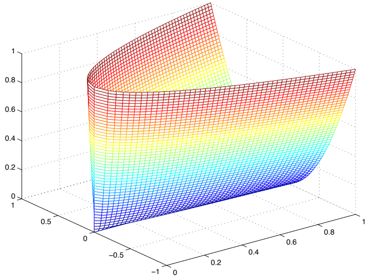
$$\mathbb{S}^n_{++} = \{X \in \mathbb{S}^n \,|\, X \succ 0\} \qquad \text{(positive definite } n \times n \text{ matrices)}$$

$$X \in \mathbb{S}^n_{++} \iff z^\top X z > 0, \;\text{ for all } z \neq 0.$$

Example

Coefficients of 2×2 positive semidefinite matrices:

$$\begin{bmatrix} x & y \\ y & z \end{bmatrix} \succeq 0.$$



Outline

Optimization using conic inequalities

A generalization of our standard problem is:

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

- Each K_i is a cone.
- Inequality constraints are equivalent to $-g_i(x) \in K_i$.

The set $\mathbb{R}^n_+$ (i.e. the positive orthant) is a convex cone.

$$x \geq y \Longleftrightarrow (x - y) \in \mathbb{R}^N_+ \Longleftrightarrow x \succeq_{\mathbb{R}^N_+} y$$

Using $K_i = \mathbb{R}^1_+$ in (P), we recover our original problem description.

Example : Semidefinite Program (SDP)

Suppose that K is the positive semidefinite cone. Consider

$$
(SDP): \quad \begin{aligned} \min_x \; & c^\top x \\ \text{subj. to } & x_1 F_1 + x_2 F_2 + \cdots x_n F_n \preceq G \\ & Ax = b \end{aligned}
$$

where the matrices $(F_1, \ldots, F_n, G)$ are all symmetric.

- The inequality constraint is called a **Linear Matrix Inequality** (LMI)
- SDPs appear in many common control design problems
- Most of the problems presented so far can be written as SDPs.

Generality of SDPs

Many common convex constraints can be remodeled as LMI constraints:

Linear constraints

$$Ax \leq b \quad \Leftrightarrow \quad \mathrm{diag}(Ax) \preceq \mathrm{diag}(\mathrm{b})$$

Quadratic constraints

$$x^\top Qx + b^\top x + c \leq 0, \quad Q \succ 0 \quad \Leftrightarrow \quad \begin{bmatrix} c + b^\top x & x^\top \\ x & -Q^{-1} \end{bmatrix} \preceq 0$$

Second-order cone constraints

$$\|Ax + b\|_2 \leq c^\top x + e \quad \Leftrightarrow \quad \begin{bmatrix} (c^\top x + e) \cdot I & Ax + b \\ (Ax + b)^\top & c^\top x + e \end{bmatrix} \preceq 0$$

Outline

6. Generalized Inequalities

   Convex cones

   Convex Cone Programs (CPs)

   Duality for Cone Programs

   Optimality Conditions for Cone Programs

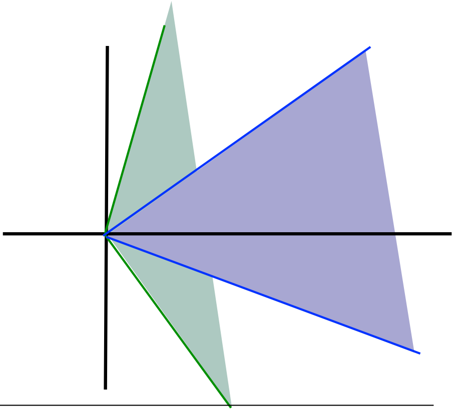
Dual cones

Assume that $K \subseteq \mathbb{R}^n$. The **dual cone** K^* is defined as

$$K^\star := \left\{x \,\middle|\, x^\top y \geq 0,\ \forall y \in K\right\}$$

$$= \{x \,|\, x \text{ is normal to a supporting hyperplane for } K\}$$

- The set K^* is always a convex cone.
- If K is closed and convex, then $K^{**} = K$.
- A cone is **self dual** if $K = K^*$.



Notation

$$\begin{aligned}
x \in K \quad &\iff \quad x \succeq_K 0 \\
-x \in K \quad &\iff \quad x \preceq_K 0 \\
x - y \in K \quad &\iff \quad x \succeq_K y \\
x \in \text{int}(K) \quad &\iff \quad x \succ_K 0
\end{aligned}$$

Dual cone examples

1) **Norm cone**: If K_p is the norm cone

$$K_p = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, \|x\|_p \leq t \right\}, \text{ then}$$

$$K_p^* = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, \|x\|_q \leq t \right\} = K_q, \text{ where } \frac{1}{p} + \frac{1}{q} = 1.$$

2) **Positive Orthant**: If $K = \{x \mid x \geq 0\}$, then

$$K^* = \left\{x \,\middle|\, x^\top y \geq 0 \ \forall y \geq 0\right\} = \{x \mid x \geq 0\}$$

3) **Positive Semidefinite Cone**: If $K = \{X \in \mathbb{S}^n \mid X \succeq 0\}$, then

$$K^* = \left\{x \in \mathbb{S}^n \,\middle|\, \mathrm{trace}\left(X^\top Y\right) \geq 0, \ \forall Y \succeq 0\right\} = \{X \in \mathbb{S}^n \mid x \succeq 0\}$$

Examples (2) and (3) above are **self-dual** cones.

Dual of a conic optimization problem

Recall the primal cone program:

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

Lagrangian:

$$
L(x, \lambda_1, \dots, \lambda_m, \nu) = f(x) + \sum_{i=1}^{m} \langle \lambda_i, g_i(x) \rangle + \sum_{i=1}^{p} \nu_i h_i(x)
$$

- Use $\langle x, y \rangle = x^\top y$ for vector problems (e.g. norm cone constraints).
- Use $\langle X, Y \rangle = \mathbf{trace}\left(X^\top Y\right)$ for matrix problems (e.g. semidefinite constraints).

Dual function:

$$
d(\lambda_1, \dots, \lambda_m, \nu) = \inf_{x \in \mathcal{X}} L(x, \lambda_1, \dots, \lambda_m, \nu)
$$

Dual of a conic optimization problem

The **dual problem** uses inequality multipliers in the **dual cones** K_i^*:

$$(D): \quad \begin{array}{rl} \max\limits_{(\lambda_1,\ldots,\lambda_m,\nu)} & d(\lambda_1,\ldots,\lambda_m,\nu) \\ \text{subj. to} & \lambda \succeq_{K_i^*} 0 \end{array}$$

Most results are parallel to our previous case:

- The dual function $-d$ is convex, as are the sets K_i^*.
- The problem (D) is convex, even when (P) is not.
- If (P) is convex and satisfies the Slater condition, then $d^\star = p^\star$.

Example : Dual of a box-constrained LP (again)

The primal problem:

$$(P): \quad \begin{aligned} \min_{x} \; & c^\top x \\ \text{subj. to} \quad & Ax = b \\ & -\begin{bmatrix} x \\ 1 \end{bmatrix} \preceq_{K_\infty} 0 \qquad (\Leftrightarrow \quad \|x\|_\infty \leq 1) \end{aligned}$$

Lagrangian:

$$L\left(x, \begin{bmatrix} \lambda \\ \sigma \end{bmatrix}, \nu\right) = c^\top x - \lambda^\top x - \sigma \cdot 1 + \nu^\top A x - \nu^\top b$$

The dual problem:

$$(D): \quad \begin{aligned} \max \; & -b^\top \nu - \sigma \\ \text{subj. to} \; & A^\top \nu + c = \lambda \\ & \begin{bmatrix} \lambda \\ \sigma \end{bmatrix} \succeq_{K_1} 0 \qquad (\Leftrightarrow \quad \|\lambda\|_1 \leq \sigma) \end{aligned}$$

Example : Dual of an SDP

The primal problem:

$$
(P): \quad \begin{array}{l} \min\ c^\top x \\ \text{subj. to } A_1 x_1 + \ldots A_n x_n \preceq B, \qquad (A_1 \ldots, A_n, B) \in \mathbb{S}^n \end{array}
$$

Lagrangian:

$$
L(x, \lambda) = c^\top x + \sum_i \langle \lambda, A_i x_i \rangle - \langle \lambda, B \rangle
$$

Dual function:

$$
g = \begin{cases} -\langle \lambda, B \rangle & \text{if } c_i + \langle \lambda, A_i \rangle = 0 \text{ for } i = 1 \ldots n \\ -\infty & \text{otherwise} \end{cases}
$$

The dual problem:

$$
(D): \quad \begin{array}{l} \max\ -\langle B, \lambda \rangle \\ \text{subj. to } \langle A_i, \lambda \rangle = -c_i \\ \qquad\qquad \lambda \succeq 0 \end{array}
$$

Outline

6. Generalized Inequalities

    Convex cones

    Convex Cone Programs (CPs)

    Duality for Cone Programs

    Optimality Conditions for Cone Programs

KKT Conditions for Cone Programs

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

1) Primal Feasibility:

$$g_i(x^\star) \preceq_{K_i} 0 \quad i = 1, \dots, m$$
$$h_i(x^\star) = 0 \quad i = 1 \dots, p$$

2) Dual Feasibility:

$$\lambda^\star \succeq_{K_i^\star} 0$$

3) Complementary Slackness:

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

4) Stationarity:

$$\nabla_x L(x^\star, \lambda^\star, \nu^\star) = \nabla f(x^\star) + \sum_{i=1}^{m} \langle Dg_i(x^\star), \lambda_i \rangle + \sum_{i=1}^{p} \nu_i \nabla h_i(x^\star) = 0,$$
$$[Df(x)]_{jk} := \frac{\partial f_j(x)}{\partial x_k}$$

KKT Conditions for an SDP

A semidefinite program:

$$\begin{aligned} \min_X \ & \operatorname{trace}\left(C^\top X\right) \\ \text{subj. to } & AX = B \\ & X \succeq 0 \end{aligned}$$

Dual of this SDP:

$$\begin{aligned} \max_Z \ & -\operatorname{trace}\left(B^\top Z\right) \\ \text{subj. to } & A^\top Z + C \succeq 0 \end{aligned}$$

KKT Conditions for this SDP:

$$\begin{aligned} A^\top Z - \lambda &= -C \\ AX &= B \\ X \succeq 0,\ \lambda &\succeq 0 \\ \langle X, \lambda \rangle &= 0 \end{aligned}$$