Convex and Non-Convex Optimization

Paul Goulart, Francesco Borrelli

Institut für Automatik
ETH Zürich

UC Berkeley

Fall Semester 2015

Table of Contents

Table of Contents

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

$$\begin{aligned} \min_{z} \quad & f(z) \\ \text{subject to:} \quad & z \in S \subseteq Z \end{aligned}$$

The problem has several ingredients:

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

This problem can be written more compactly as

$$\min_{z \in S \subseteq Z} f(z)$$

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

Describing an Optimization Problem

A more common problem format:

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{subject to:} \quad & g_i(z) \leq 0 \quad i = 1, \ldots, m \\ & h_j(z) = 0 \quad j = 1, \ldots, p \end{aligned}$$

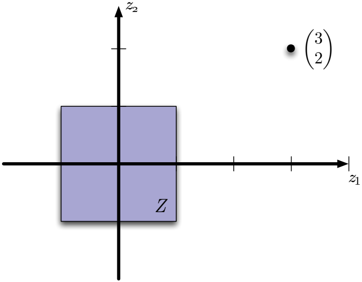
Defined by the following **problem data**:

- **Objective function** $f : Z \to \mathbb{R}$
- **Domain** $Z \subseteq \mathbb{R}^n$ of the objective function, from which the decision variable $z := (z_1; z_2; \ldots; z_n)$ must be chosen.
- Optional **inequality constraint functions** $g_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, m$
- Optional **equality constraint functions** $h_j : \mathbb{R}^n \to \mathbb{R}$, for $j = 1, \ldots, p$

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

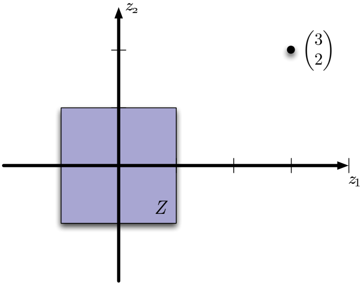
A Simple Example

Problem : In $\mathbb{R}^2$, find the point in the unit box S closest to the point $(z_1, z_2) = (3, 2)$.



A Simple Example

Problem : In $\mathbb{R}^2$, find the point in the unit box S closest to the point $(z_1, z_2) = (3, 2)$.



Same problem in standard format:

$$\min_{(z_1, z_2) \in \mathbb{R}^2} \quad (z_1 - 3)^2 + (z_2 - 2)^2$$

$$\text{subject to:} \quad \begin{aligned} z_1 &\leq 1 \\ -z_1 &\leq 1 \\ z_2 &\leq 1 \\ -z_2 &\leq 1 \end{aligned}$$

Properties of Optimization Problems

Consider the **Nonlinear Program** (NLP)

$$J^* = \min_{z \in \mathcal{S}} f(z)$$

Notation:

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

$$\arg\min_{z \in \mathcal{S}} f(z) := \{ z \in \mathcal{S} \mid f(z) = J^* \}$$

Terminology

Feasible point: A vector $z \in \mathcal{S}$ satisfying the inequality and equality constraints, i.e. $g_i(z) \leq 0$ for $i = 1, \dots, m$, $h_j(z) = 0$ for $j = 1, \dots, p$.

Strictly feasible point: A vector $z \in Z$ satisfying the inequality constraints strictly, i.e. $g_i(z) < 0$ for $i = 1, \dots, m$.

Optimal value: The lowest possible objective value, $f(z^*)$. Denoted by f^* (or p^*, or J^*).

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

$$
\begin{aligned}
\min_{z \in Z} \quad & f(z) \\
\text{subject to:} \quad & g_i(z) \leq 0 \quad i = 1, \dots, m, \\
& h_j(z) = 0 \quad j = 1, \dots, p \\
& ||z - z||_2 \leq R
\end{aligned}
$$

Terminology

Optimal solution: Any *feasible* $z^* \in Z$ such that $f(z^*) \leq f(z)$ for all *feasible* $z \in Z$.

Local optimum: a point z^*_{local} that is optimal within a neighbourhood $||z - z^*_{\text{local}}|| \leq R$.

Technical point: The optimal value is called the **infimum**. A vector z^* 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**. Example:

$$\begin{aligned} \min_{z \in \mathbb{R}} \quad & z^2 \\ \text{subject to:} \quad & z \leq -1 \\ & z \geq 1 \end{aligned}$$

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

$$\begin{aligned} \min_{z \in \mathbb{R}} \quad & z \\ \text{subject to:} \quad & z \leq 0 \end{aligned}$$

- The value J^* might be finite, but there is no z that achieves it. Example:

$$\begin{aligned} \inf_{z \in \mathbb{R}} \quad & e^{-z} \\ \text{subject to:} \quad & z \geq 0 \end{aligned}$$

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

Active, Inactive and Redundant Constraints

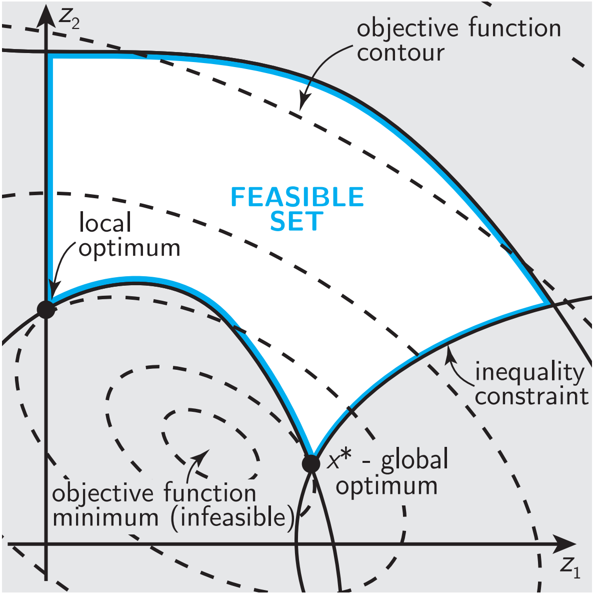
Consider the standard problem

$$\begin{array}{ll} \min_{z \in Z} & f(z) \\ \text{subject to:} & g_i(z) \leq 0 \quad i = 1, \ldots, m \\ & h_j(z) = 0 \quad j = 1, \ldots, p \end{array}$$

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

$$\begin{array}{ll} \min_{z \in \mathbb{R}} & f(z) \\ \text{subject to:} & z \leq 1 \\ & z \leq 2 \quad \texttt{(redundant)} \end{array}$$

Geometry of an Optimization Problem



Implicit and Explicit Constraints

The constraints $g_i(z) \leq 0,\ i = 1, \ldots, m$ and $h_j(z) = 0,\ j = 1, \ldots, p$ 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 z:

$$z \in \operatorname{dom}(f) \cap \bigcap_{i=1}^{m} \operatorname{dom}(g_i) \cap \bigcap_{j=1}^{p} \operatorname{dom}(h_j)$$

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:

$$\min_z\ f(z) = -\sum_{i=1}^{k} \log(a_i^\top z - b_i)$$

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

Feasibility Problem

The "constraint satisfiability" problem

$$\begin{array}{rll} \underset{z \in Z}{\text{find}} & z & \\ \text{subject to:} & g_i(z) \leq 0 & i = 1, \ldots, m \\ & h_j(z) = 0 & j = 1, \ldots, p \end{array}$$

is a special case of the general optimization problem:

$$\begin{array}{rll} \min\limits_{z \in Z} & 0 & \\ \text{subject to:} & g_i(z) \leq 0 & i = 1, \ldots, m \\ & h_j(z) = 0 & j = 1, \ldots, p \end{array}$$

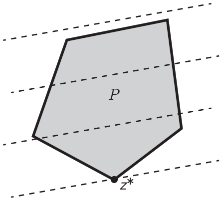
- $p^* = 0$ if the constraints are feasible (consistent). Every feasible z is optimal.
- $p^* = \infty$ otherwise.

Table of Contents

"Easier" problems: Linear and Convex Quadratic Programs

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

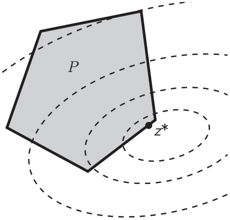
$$\begin{aligned} \min_{z} \quad & c^\top z \\ \text{subject to:} \quad & Gz \leq h \\ & Az = b \end{aligned}$$



Linear optimization on a polytope.

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

$$\begin{aligned} \min_{z} \quad & \tfrac{1}{2} z^\top P z + q^\top z \\ \text{subject to:} \quad & Gz \leq h \\ & Az = b \end{aligned}$$

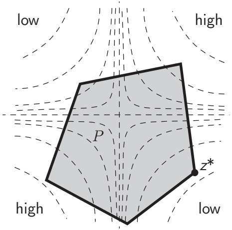


Convex quadratic optimization on a polytope.

"Harder" problems: Nonconvex and Integer Programs

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

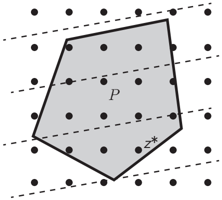
$$\begin{aligned} \min_z \quad & \tfrac{1}{2} z^\top P z + q^\top z \\ \text{subject to:} \quad & Gz \leq h \\ & Az = b \end{aligned}$$



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.

$$\begin{aligned} \min_z \quad & c^\top z \\ \text{subject to:} \quad & Gz \leq h \\ & Az = b \\ & z \in \{0,1\}^n \text{ or } z \in \mathbb{Z}^n \end{aligned}$$



Linear optimization with integer constraints (dots).

Software Tools for Optimization

A simple optimization problem:

$$\begin{aligned} \min_{z_1, z_2} \quad & |z_1 + 5| + |z_2 - 3] \\ \text{subject to:} \quad & 2.5 \leq z_1 \leq 5 \\ & -1 \leq z_2 \leq 1 \end{aligned}$$

- This problem is equivalent to a linear program (more on this later).
- 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:

$$\begin{aligned} \min_{z_1, z_2} \quad & |z_1 + 5| + |z_2 - 3] \\ \text{subject to:} \quad & 2.5 \leq z_1 \leq 5 \\ & -1 \leq z_2 \leq 1 \end{aligned}$$

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

```
%make variables
sdpvar z1 z2;
%define cost function
f = abs(z1 + 5) + abs(z2 - 3);
%define constraints
S = set(2.5 <= z1 <= 5) + ...
    set( -1 <= z2 <= 1);
%solve
solvesdp(S,f);
```

Software for Optimization

A simple optimization problem:

$$\begin{aligned} \min_{z_1, z_2} \quad & |z_1 + 5| + |z_2 - 3] \\ \text{subject to:} \quad & 2.5 \leq z_1 \leq 5 \\ & -1 \leq z_2 \leq 1 \end{aligned}$$

The CVX toolbox for Matlab (from Stanford):

```
cvx_begin
   %define cost function
   variables z1 z2
   %define constraints
   minimize(abs(z1 + 5) + abs(z2 - 3))
   subject to
      2.5 <= z1 <= 5
       -1 <= z2 <= 1
cvx_end    %solves automatically
```

Table of Contents

Table of Contents

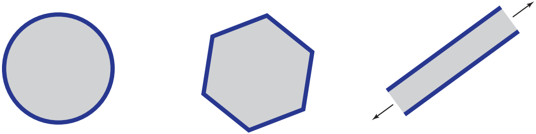
Definition (Convex Set)

A set Z is convex *if and only if* for any pair of points z and y in Z, any **convex combination** of z and y lies in Z:

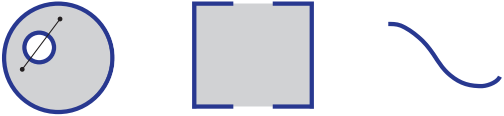
$$Z \text{ is convex} \quad \Leftrightarrow \quad \lambda z + (1-\lambda)y \in Z, \forall \lambda \in [0,1], \forall z, y \in Z$$

Interpretation: All line segments starting and ending in Z stay within Z.

Convex:



Non-convex:



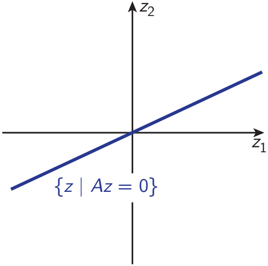
Definitions (Affine sets and Subspaces)

An affine set is a convex set defined by $Z = \{z \in \mathbb{R}^n \mid Az = b\}$. A subspace is an affine set with $b = 0$.

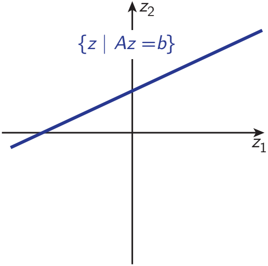
Verify convexity by definition — for all $z, y \in Z$, for all $\lambda \in [0, 1]$,

$$A(\lambda z + (1-\lambda)y) = \lambda Az + (1-\lambda)Ay = \lambda \cdot b + (1-\lambda) \cdot b = b$$

This definition encompasses lines, planes and individual points.



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



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

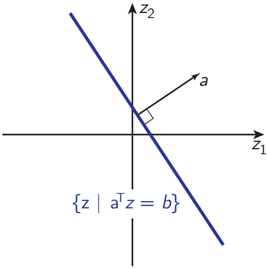
Definitions (Hyperplanes and halfspaces)

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

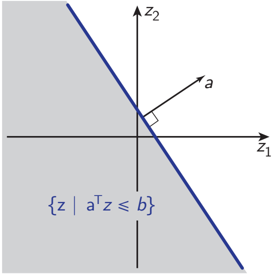
A halfspace is everything on one side of a hyperplane, i.e. $\left\{z \in \mathbb{R}^n \;\middle|\; a^\top z \leq b\right\}$ for $a \neq 0$. It can either be **open** (strict inequality) or **closed** (non-strict inequality).

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

Definitions (Polyhedra and polytopes)

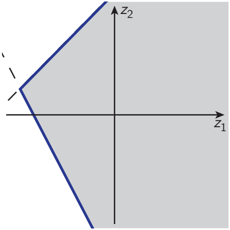
A polyhedron is the intersection of a *finite* number of closed halfspaces:

$$\begin{aligned} Z &= \{z \mid a_1^\top z \le b_1,\ a_2^\top z \le b_2, \ldots, a_m^\top \le b_m\} \\ &= \{z \mid Az \le b\} \end{aligned}$$

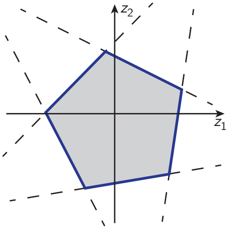
where $A := [a_1, a_2, \ldots, a_m]^\top$ and $b := [b_1, b_2, \ldots, b_m]^\top$.

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:

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

A norm is denoted $||z||_\bullet$, where a symbol in place of the dot denotes the type of norm. The notation $||z||$ refers to any arbitrary norm.

Definition (ℓ_p norm)

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

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

ℓ_p norms

By far the most common ℓ_p norms are:

- $p = 2$ (Euclidean norm):

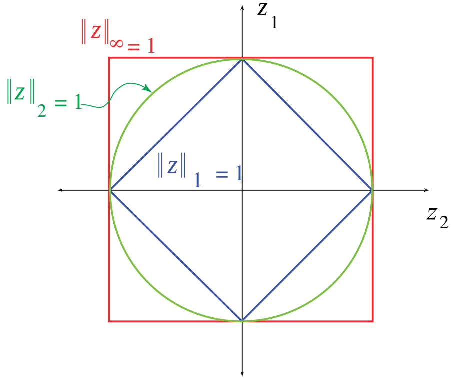
$$||z||_2 = \sqrt{\sum_i z_i^2}$$

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

$$||z||_1 = \sum_i |z_i|$$

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

$$||z||_\infty = \max_i |z_i|$$



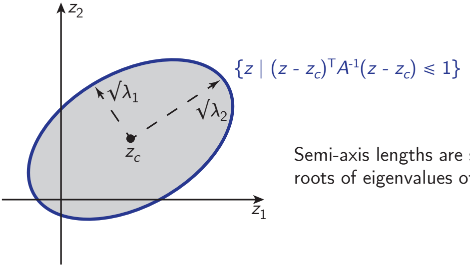
The **norm ball**, defined by $\{z \mid ||z - z_c|| \leq r\}$ where z_c is the centre of the ball and $r \geq 0$ is the radius, is always convex for any norm.

Definition (Ellipsoid)

An ellipsoid is a set defined as

$$\{z \,|\, (z - z_c)^\top A^{-1}(z - z_c) \leq 1\},$$

where z_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** $B(z_c, r)$ is a special case of the ellipsoid, for which $A = r^2 I$, so that $B(z_c, r) := \{z \,|\, ||z - z_c||_2 \leq r\}$.

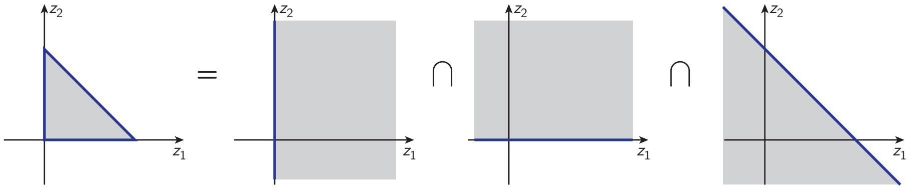
Table of Contents

Intersection $Z \cap \mathcal{Y}$

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 Z and $\mathcal{Y}$. For any $\lambda \in [0,1]$, $\lambda a + (1-\lambda)b$ is in both Z and $\mathcal{Y}$. Therefore $\lambda a + (1-\lambda)b \in Z \cap \mathcal{Y}$, $\forall \lambda \in [0,1]$. This satisfies the definition of convexity for set $Z \cap \mathcal{Y}$.



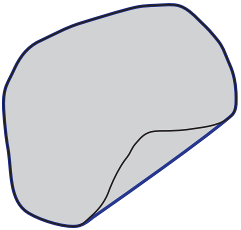
Many sets can be written as the intersection of convex elements, and are therefore easily shown to be convex. Any convex set can be written as a (possibly infinite) intersection of halfspaces.

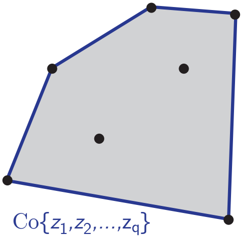
Convex Hull $\mathrm{Co}(Z)$

The **convex hull** of a set Z is the set of all convex combinations of points in Z:

$$\mathrm{Co}(Z) := \{z \,|\, z = \lambda a + (1-\lambda)b,\ \lambda \in [0,1],\ a, b \in Z\}$$

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





For a set $Z = \{z_1, z_2, \ldots, z_q\}$ comprising q points, the convex hull can be written

$$\mathrm{Co}(Z) = \left\{ \lambda_1 z_1 + \lambda_2 z_2 + \ldots + \lambda_q z_q \;\middle|\; \lambda_i \geq 0,\ i = 1, \ldots, q,\ \sum_{i=1}^{q} \lambda_i = 1 \right\}$$

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

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

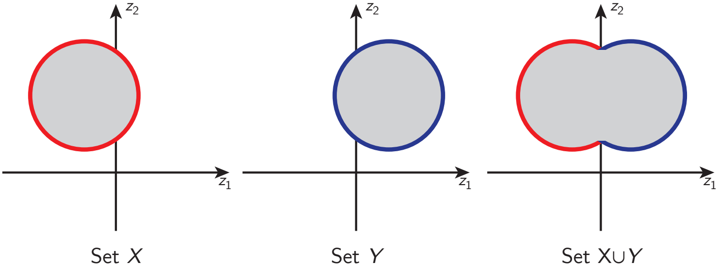


Table of Contents

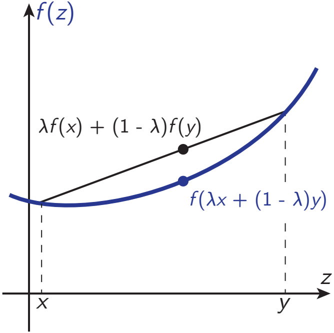
Table of Contents

Definitions (Convex Function)

A function $f : \mathrm{dom}(f) \to \mathbb{R}$ is convex *iff* [1] its domain $\mathrm{dom}(f)$ is convex *and*

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

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



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

[1] "if and only if"

1st-order condition for convexity

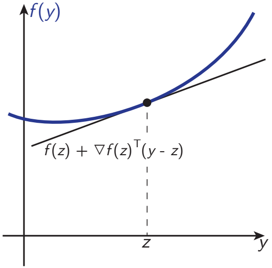
A differentiable function $f : \mathrm{dom}(f) \to \mathbb{R}$ with a convex domain is **convex** *iff*

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

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

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

$$\nabla f(z) = \left[\frac{\partial f(z)}{\partial z_1}, \frac{\partial f(z)}{\partial z_2}, \dots, \frac{\partial f(z)}{\partial z_n} \right]^\top$$



2nd-order condition for convexity

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

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

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

$$\nabla^2 f(z)_{ij} = \frac{\partial^2 f(z)}{\partial z_i \partial z_j}$$

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

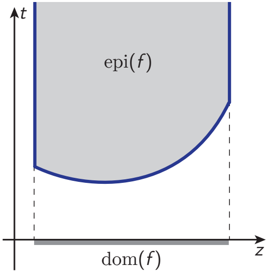
Epigraph of a Function

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

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

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 $z \in \mathrm{dom}(f)$ for which $f(z) = \alpha$:

$$L_\alpha = \{z \,|\, z \in \mathrm{dom}(f), f(z) = \alpha\}$$

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

$$C_\alpha = \{z \,|\, z \in \mathrm{dom}(f),\ f(z) \leq \alpha\}$$

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

Table of Contents

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: z^α on domain $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- Powers of absolute value: $|z|^p$, 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: z^α on domain $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- Logarithm: $\log z$ on domain $\mathbb{R}_{++}$
- Entropy: $-z \log z$ 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 $b \in \mathbb{R}$:

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

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

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

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

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 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}$:

$$f(S) = \mathrm{trace}(A^\top S) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} S_{ij} + b$$

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

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

Table of Contents

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...

Convexity-preserving Operations

Theorem (Non-negative weighted sum)

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

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

Theorem (Pointwise maximum)

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

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

Convexity-preserving Operations (cont'd)

Theorem (Composition with scalar functions)

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

- *g is convex, h is convex, h is non-decreasing*
- *g is concave, h is convex, h is non-increasing*

Examples

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

Convexity-preserving Operations (cont'd)

Theorem (Composition with vector functions)

For $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$, $f(z) = h(g(z)) = h(g_1(z), g_2(c), \ldots, g_k(z))$ is convex if:

- *Each g_i is convex, h is convex, h is non-decreasing in each argument*
- *Each g_i is concave, h is convex, h is non-increasing in each argument*

Examples

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

Table of Contents

Table of Contents

Convex Optimization Problem

An optimization problem:

$$
\begin{array}{ll}
\min\limits_{z \in Z} & f(z) \\
\text{subject to:} & g_i(z) \le 0 \quad i = 1, \ldots, m \\
& h_j(z) = 0 \quad j = 1, \ldots, p
\end{array}
$$

is convex if:

- The objective function f is a convex function on its domain Z.
- The feasible set S is a convex set.

Standard Form Convex Optimization Problem

A standard form **convex** optimization problem:

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{subject to:} \quad & g_i(z) \leq 0 \quad i = 1, \dots, m \\ & a_j^\top z = b_j \quad j = 1, \dots, p \end{aligned}$$

This problem is convex if:

- The domain Z is a convex set.
- The objective function f is a convex function.
- The inequality constraint functions g_i are all convex.

Standard Form Convex Optimization Problem

The affine constraints are typically gathered into matrix form:

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{subject to:} \quad & g_i(z) \leq 0 \quad i = 1, \ldots, m \\ & Az = b \quad A \in \mathbb{R}^{p \times m} \end{aligned}$$

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

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

- The problem can't be convex.

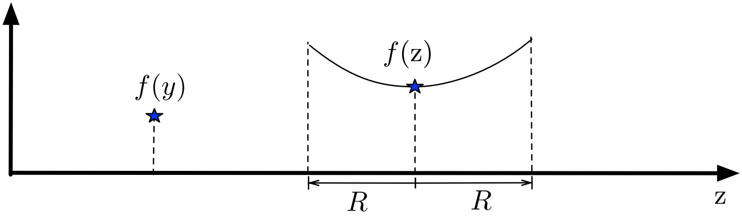


Table of Contents

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:**

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

  is equivalent to

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

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:**

$$\begin{aligned} \min_{z} \quad & f(z) \\ \text{subject to:} \quad & A_i z \leq b_i \quad i = 1, \dots, m \end{aligned}$$

is equivalent to

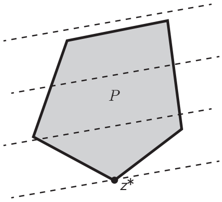
$$\begin{aligned} \min_{z, s_i} \quad & f(z) \\ \text{subject to:} \quad & A_i z + s_i = b_i \quad i = 1, \dots, m \\ & s_i \geq 0 \quad i = 1, \dots, m \end{aligned}$$

Table of Contents

General Linear Program (LP)

Affine cost and constraint functions:

$$
\begin{aligned}
\min_{z} \quad & c^\top z + d \\
\text{subject to:} \quad & Gz \leq h \\
& Az = b
\end{aligned}
$$



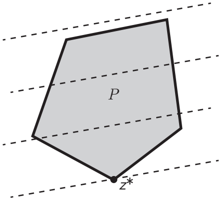
Linear optimization on a polytope.

- 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:

$$\begin{aligned} \min_{z} \quad & c^\top z \\ \text{subject to:} \quad & Az = b \\ & z \geq 0 \end{aligned}$$



Linear optimization on a polytope.

- All components of z 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 $z_1, z_2, \ldots, z_n$ 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:

$$\begin{aligned} \min_{z} \quad & c^\top z \\ \text{subject to:} \quad & Az \geq b \\ & z \geq 0 \end{aligned}$$

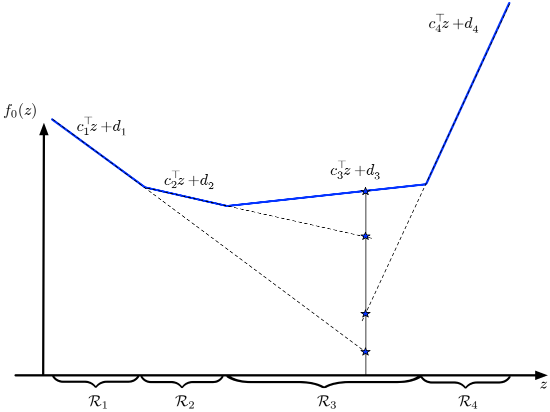
This is an example of a *resource allocation* problem.

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

$$\min_{z} \quad \left[\max_{i=1,\dots,m} \{c_i^\top z + d_i\} \right]$$
$$\text{subject to:} \quad Gz \leq h$$

- 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:

$$\begin{aligned} \min_{z} \quad & \left[\max_{i=1,\ldots,m} \{ c_i^\top z + d_i \} \right] \\ \text{subject to:} \quad & Gz \leq h \end{aligned}$$

is **equivalent** to an LP:

$$\begin{aligned} \min_{z,t} \quad & t \\ \text{subject to:} \quad & c_i^\top z + d_i \leq t \quad i = 1, \ldots, m \\ & Gz \leq h \end{aligned}$$

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

ℓ_∞ minimization

Constrained ℓ_∞ (Chebyshev) minimization:

$$\begin{aligned}\min_{z\in\mathbb{R}^n} \quad & ||z||_\infty \\ \text{subject to:} \quad & Fz \leq g\end{aligned}$$

Write this is a max of linear functions.

Equivalent to:

$$\begin{aligned}\min_{z\in\mathbb{R}^n} \quad & [\max\{z_1,\ldots,z_n,-z_1,\ldots,-z_n\}] \\ \text{subject to:} \quad & Fz \leq g\end{aligned}$$

ℓ_∞ minimization (cont'd)

Equivalent to:

$$\begin{aligned} \min_{z,t} \quad & t \\ \text{subject to:} \quad & z_i \leq t \quad i = 1, \ldots, n \\ & -z_i \leq t \quad i = 1, \ldots, n \\ & Fz \leq g \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min_{z,t} \quad & t \\ \text{subject to:} \quad & -\mathbf{1}t \leq z \leq \mathbf{1}t \\ & Fz \leq g \end{aligned}$$

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

ℓ_1 minimization

Constrained ℓ_1 minimization:

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & ||Az - b||_1 \\ \text{subject to:} \quad & Fz \leq g \end{aligned}$$

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

Equivalent to:

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & \left[\sum_{i=1}^{m} \max \left\{ (Az - b)_i, -(Az - b)_i \right\} \right] \\ \text{subject to:} \quad & Fz \leq g \end{aligned}$$

ℓ_1 minimization (cont'd)

Equivalent to:

$$\begin{aligned} \min_{z\in\mathbb{R}^n, t\in\mathbb{R}^m} \quad & t_1 + \ldots + \ldots t_m \\ \text{subject to:} \quad & (Az-b)_i \le t_i \quad i = 1,\ldots,m \\ & -(Az-b)_i \le t_i \quad i = 1,\ldots,m \\ & Fz \le g \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min_{z\in\mathbb{R}^n, t\in\mathbb{R}^m} \quad & \mathbf{1}^\top t \\ \text{subject to:} \quad & -t \le (Az-b) \le t \\ & Fz \le g \end{aligned}$$

- The notation '$\mathbf{1}$' indicates a vector of ones.
- The constraint $-t \le (Az-b) \le t$ bounds the absolute value of each component of $(Az-b)$ with a component of the vector variable t.

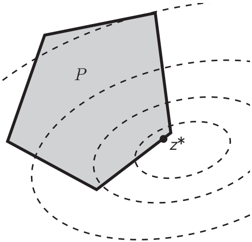
Table of Contents

General Quadratic Program

Quadratic cost function with $P \in \mathbb{S}_+^n$, affine constraint functions. Feasible set is a polyhedron:

$$\begin{aligned} \min_z \quad & \tfrac{1}{2} z^\top P z + q^\top z + r \\ \text{subject to:} \quad & Gz \leq h \\ & Az = b \end{aligned}$$

- Constant component 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

Least squares:

$$\min_z ||Az - b||_2^2$$

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

Example Quadratic Programs

Linear program with random cost:

$$\begin{aligned} \min_{z} \quad & \mathbb{E}[c^\top z] + \gamma \mathrm{var}(c^\top z) = \overline{c}^\top z + \gamma z^\top \Sigma z \\ \text{subject to:} \quad & Gz \geq h \\ & Az = b \end{aligned}$$

- 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 $c^\top z$ is a random variable with mean $\overline{c}^\top z$ and variance $z^\top \Sigma z$.
- 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.

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

Equivalent to:

$$\begin{aligned} \min_{z \in \mathbb{R}^n, t \in \mathbb{R}^n} \quad & ||Az - b||_2^2 + \gamma \cdot \mathbf{1}^\top t \\ \text{subject to:} \quad & -t \leq z \leq t \end{aligned}$$

- 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.

Table of Contents

Table of Contents

Decent Direction for Differentiable f

Theorem (Descent Direction)

$f : \mathbb{R}^s \to \mathbb{R}$ differentiable at $\bar{z}$. If there exists a vector $\mathbf{d}$ such that $\nabla f(\bar{z})'\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\bar{z} + \lambda \mathbf{d}) < f(\bar{z})$ 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 $\bar{z}$ is defined as the normalized direction where $\nabla f(\bar{z})'\mathbf{d}_s < 0$ is minimized.
- The direction $\mathbf{d}_s$ of steepest descent is $\mathbf{d}_s = -\frac{\nabla f(\bar{z})}{\|\nabla f(\bar{z})\|}$.

Optimality Criterion for Differentiable f

Theorem (Necessary condition*)

$f : \mathbb{R}^s \to \mathbb{R}$ is differentiable at $\bar{z}$. If $\bar{z}$ is a local minimizer, then $\nabla f(\bar{z}) = 0$.

Theorem (Sufficient condition*)

Suppose that $f : \mathbb{R}^s \to \mathbb{R}$ is twice differentiable at $\bar{z}$. If $\nabla f(\bar{z}) = 0$ and the Hessian of $f(z)$ at $\bar{z}$ is positive definite, then $\bar{z}$ is a local minimizer.

Theorem (Necessary and sufficient condition*)

Suppose that $f : \mathbb{R}^s \to \mathbb{R}$ is differentiable at $\bar{z}$. If f is convex, then $\bar{z}$ is a global minimizer if and only if $\nabla f(\bar{z}) = 0$.

*Proofs available in Chapter 4 of M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. Nonlinear Programming - Theory and Algorithms. John Wiley & Sons, Inc., New York, second edition, 1993.

Table of Contents

Optimality Conditions

Consider the problem

$$
\begin{array}{lll}
\min_z & f(z) & \\
\text{subject to:} & g_i(z) \leq 0 & \text{for } i = 1, \ldots, m \\
& h_j(z) = 0 & \text{for } j = 1, \ldots, p
\end{array}
$$

- In general, an analytical solution does not exist.
- Solutions are usually computed by recursive algorithms which start from an initial guess z_0 and at step k generate a point z_k such that $\{f(z_k)\}_{k=0,1,2,\ldots}$ converges to f^*.
- These algorithms recursively use and/or solve analytical **conditions for optimality**

KKT optimality conditions

z^*, (u^*, v^*) of an optimization problem, with differentiable cost and constraints and zero duality gap, have to satisfy the following conditions:

$$
\begin{aligned}
0 &= \nabla f(z^*) + \sum_{i=1}^{m} u_i^* \nabla g_i(z^*) + \sum_{j=1}^{p} v_j^* \nabla h_i(z^*), && \text{(1a)}\\
0 &= u_i^* g_i(z^*), \quad i = 1, \dots, m && \text{(1b)}\\
0 &\leq u_i^*, \quad i = 1, \dots, m && \text{(1c)}\\
0 &\geq g_i(z^*), \quad i = 1, \dots, m && \text{(1d)}\\
0 &= h_j(z^*) \quad j = 1, \dots, p && \text{(1e)}
\end{aligned}
$$

Conditions (1a)-(1e) are called the *Karush-Kuhn-Tucker* (KKT) conditions.