Optimization

Pieter Vanderschueren

Year 2024-2025

Contents

Fundamental Concepts

1 Fundamental Concepts of Optimization

Theorem 1.1: Optimization problem in standard form

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0, \\ & h(x) \geq 0. \end{aligned}$$

Theorem 1.2: Level Set

The set

$$\{\ x \in \mathbb{R}^n \mid f(x) = c\ \}$$

is the level set of f for the value c, i.e. the set of all points that map to the same value c.

Theorem 1.3: Feasible set

The set

$$\{\ x \in \mathbb{R}^n \mid g(x) = 0 \ \wedge\ h(x) \geq 0\ \}$$

is the feasible set Ω, i.e. the set of all points that satisfy the constraints.

Theorem 1.4: Global minimizer

The point x^* is a global minimizer if and only if

$$x^* \in \Omega,\ \forall x \in \Omega :\ f(x^*) \leq f(x).$$

Theorem 1.5: Strict global minimizer

The point x^* is a strict global minimzer if and only if

$$x^* \in \Omega,\ \forall x \in \Omega \setminus \{x^*\} :\ f(x^*) < f(x).$$

Theorem 1.6: Strict local minimizer

The point x^* is a strict local minimizer if and only if $x^* \in \Omega$ and there exists a neighborhood $\mathcal{N}$ of x^* such that

$$\forall x \in (\Omega \cap \mathcal{N}) \setminus \{x^*\} :\ f(x^*) < f(x).$$

Theorem 1.7: Weierstrass

If $\Omega \in \mathbb{R}^n$ is compact, i.e. limited and closed, and $f : \Omega \to \mathbb{R}$ is continuous, then there exists a global minimizer (a solution) of the optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega. \end{aligned}$$

Proof 1.1: Weierstrass

Regard the graph of f, $G = \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \Omega\}$. G is a compact set, and so is the projection of G onto its last coordinate, the set $G = \{s \in \mathbb{R} \mid \exists x \text{ such that } (x, s) \in G\}$, which is a compact interval $[f_{\min}, f_{\max}] \subset \mathbb{R}$. By construction, there must be at least one x^* such that $(x^*, f(x^*)) \in G$.

$\square$

2 Types of Optimization Problems

Theorem 2.1: Convex set

A set $\Omega \subset \mathbb{R}^n$ is convex if

$$\forall x, y \in \Omega,\ \forall \lambda \in [0, 1] :\ x + \lambda(y - x) \in \Omega.$$

or if "all connecting lines lie inside the set".

Theorem 2.2: Convex function

A function $f : \Omega \to \mathbb{R}$ is convex if

$$\forall x, y \in \Omega,\ \forall \lambda \in [0, 1] :\ f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x)).$$

or if "all secants (i.e. a line segment between two points on the graph) are above graph". This definition is equivalent to saying that the Epigraph of f, i.e. the set $\{(x, s) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \Omega,\ s \geq f(x)\}$, is a convex set.

Note: a concave function is the same but then with $\geq$ instead of $\leq$.

Property 2.1: Convex function

If $f :\ D \to \mathbb{R}$ and $\Omega_f = \{(x, y) \mid x \in D, y \geq f(x)\}$ then the following holds:

$$f \text{ is convex } \Leftrightarrow\ \Omega_f \text{ is convex.}$$

Theorem 2.3: Convex optimization problem

An optimization problem with convex feasible set Ω and convex objective function $f:\ \Omega \to \mathbb{R}$ is called a convex optimization problem.

Property 2.2: Globality of local minima of convex optimization problem

For a convex optimization problem, a local minimum is also a global one.

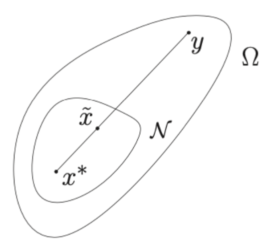
Proof 2.1: Globality of local minima of convex optimization problem

Regard a local minimum x^* of the convex optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega. \end{aligned}$$

We will show that for any point $y \in \Omega$ holds $f(y) \geq f(x^*)$, i.e. x^* is a global minimum.

First we choose, using local optimality, a neighborhood $\mathcal{N}$ of x^* such that for all $\tilde{x} \in \Omega \cap \mathcal{N}$ holds $f(\tilde{x}) \geq f(x^*)$. Second, we regard the connecting line between x^* and y. This line is completely contained in Ω due to the convexity of Ω. Now we choose a point $\tilde{x}$ on this line such that it is in the neighborhood, but not equal to x^*, i.e. $\tilde{x} = x^* + \lambda(y - x^*)$ for some $\lambda \in (0,1)$, and $\tilde{x} \in \Omega \cap \mathcal{N}$. Due to local optimality, we have $f(\tilde{x}) \geq f(x^*)$, and due to convexity we have



$$f(\tilde{x}) = f(x^* + \lambda(y - x^*)) \leq f(x^*) + \lambda(f(y) - f(x^*)).$$

It follows that $\lambda(f(y) - f(x^*)) \geq 0$, and since $\lambda \in (0,1)$, we have $f(y) \geq f(x^*)$, as desired.

□

Theorem 2.4: Unconstrained optimization problem

An optimization problem with no constraints, i.e. $g(x) = 0$ and $h(x) \geq 0$ are empty, is called an unconstrained optimization problem.

Example 2.1: Nonlinear Programming (NLP)

Nonlinear Programming Problems are poblems of the following form:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0, \\ & h(x) \geq 0, \end{aligned}$$

where $f:\ \mathbb{R}^n \to \mathbb{R}$, $g:\ \mathbb{R}^n \to \mathbb{R}^p$, and $h:\ \mathbb{R}^n \to \mathbb{R}^q$, are assumed to be continuously differentiable at least once.

Example 2.2: Linear Programming (LP)

When the functions f, g, h are affine (i.e., they can be expressed in the form $f(x) = a^T x + b$ for some vector a and scalar b) in the general formulation (see Theorem 2.1), the general NLP gets something easier to solve, namely a Linear Program (LP), which can be written as follows:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax - b = 0, \\ & Cx - d \geq 0, \end{aligned}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, $C \in \mathbb{R}^{q \times n}$, and $d \in \mathbb{R}^q$.

Example 2.3: Quadratic Programming (QP)

When the functions g, h are affine (as for an LP in Theorem 2.2), and the objective function f is a linear-quadratic function, the NLP becomes a Quadratic Program (QP), which can be written as follows:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x + \frac{1}{2} x^T B x \\ \text{subject to} \quad & Ax - b = 0, \\ & Cx - d \geq 0, \end{aligned}$$

where, in addition to the LP parameters, $B \in \mathbb{R}^{n \times n}$ is the Hessian matrix. Specifically, for the objective function $f(x) = c^T x + \frac{1}{2} x^T B x$, the Hessian is given by:

$$\nabla^2 f(x) = B.$$

Theorem 2.5: Convex QP

If the Hessian matrix B is positive semi-definite (i.e. if $\forall z \in \mathbb{R}^n \setminus \{0\} : z^T B z \geq 0$) we call the QP a convex QP. Convex QPs are tremendously easier to solve globally than "non-convex QPs" (i.e., where the Hessian B is not positive semi-definite), which might have different local minima (i.e. have a non-convex solution set).

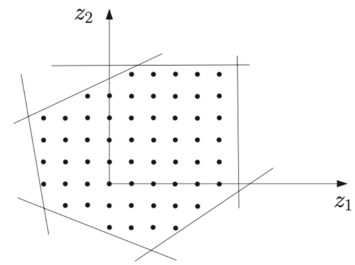
Theorem 2.6: Strictly convex QP

If the Hessian matrix B is positive definite (i.e. if $\forall z \in \mathbb{R}^n : z^T B z > 0$) we call the QP a strictly convex QP. Strictly convex QPs are a subset of convex QPs (see Theorem 2.5).

Example 2.4: Mixed-Integer Programming (MIP)

A mixed-integer programming problem or mixed-integer program (MIP) is a problem with both real and integer decision variables. A MIP can be formulated as follows:

$$\begin{aligned} \underset{\substack{x\in\mathbb{R}^n \\ z\in\mathbb{Z}^m}}{\text{minimize}} \quad & f(x,z) \\ \text{subject to} \quad & g(x,z) = 0, \\ & h(x,z) \geq 0, \end{aligned}$$



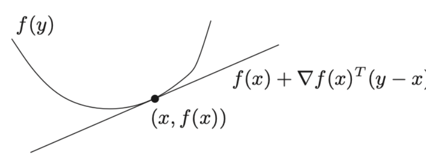
where $x \in \mathbb{R}^n$ are the continuous variables, and $z \in \mathbb{Z}^m$ are the integer variables.

3 Convex Optimization

Theorem 3.1: Convexity for C^1 functions

Assume that $f : \Omega \to \mathbb{R}$ is continuously differentiable and Ω is convex. Then holds that f is convex if and only if

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



i.e. tangents lie below the graph.

Proof 3.1: Convexity for C^1 functions

"$\Rightarrow$": Due to convexity of f holds for given $x, y \in \Omega$ and for any $\lambda \in [0,1]$ that

$$f(x + \lambda(y - x)) - f(x) \leq \lambda(f(y) - f(x)).$$

and therefore that

$$\nabla f(x)^T (y - x) = \lim_{\lambda \to 0} \frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq \frac{f(y) - f(x)}{y - x}.$$

"$\Leftarrow$": To prove that $z = x + \lambda(y - x) = (1-\lambda)x + \lambda y$ holds that $f(z) \leq (1-\lambda)f(x) + \lambda f(y)$, we can use the equation from Theorem 3.1 twice to get

$$f(x) \geq f(z) + \nabla f(z)^T (x - z) \quad \text{and} \quad f(y) \geq f(z) + \nabla f(z)^T (y - z),$$

which yield, when weighted with $(1-\lambda)$ and λ respectively, that

$$(1-\lambda)f(x) + \lambda f(y) \geq f(z) + \nabla f(z)^T \underbrace{[(1-\lambda)(x-z) + \lambda(y-z)]}_{=0}$$

□

Theorem 3.2: Generalized Inequality for Symmetric Matrices

We write for a symmetric matrix $B = B^T$, $B \in \mathbb{R}^{n\times n}$ that "$B \succeq 0$" if and only if B is positive semi-definite, i.e.,

$$\forall z \in \mathbb{R}^n : \; z^T B z \geq 0,$$

or, equivalently, if all (real) eigenvalues of B are non-negative.

Theorem 3.3: $O(\cdot)$

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we write

$$f(x) = O(g(x))$$

if and only if there exists a constant $C > 0$ and a neighborhood $\mathcal{N}$ of 0 such that

$$\forall x \in \mathcal{N} : \; \|f(x)\| \leq C g(x),$$

i.e. "f shrinks as fast as g".

Theorem 3.4: $o(\cdot)$

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we write

$$f(x) = o(g(x))$$

if and only if there exists a neighborhood $\mathcal{N}$ of 0 and a function $c : \mathcal{N} \to \mathbb{R}$ with $\lim_{x\to 0} c(x) = 0$ such that

$$\forall x \in \mathcal{N} : \; \|f(x)\| \leq c(x) g(x),$$

i.e. "f shrinks faster than g".

Theorem 3.5: Convexity for C^2 functions

Assume that $f : \; \Omega \to \mathbb{R}$ is twice continuously differentiable and Ω is convex and open. Then holds that f is convex if and only if

$$\forall x \in \Omega : \; \nabla^2 f(x) \succeq 0,$$

i.e. the Hessian of f is positive semi-definite.

Proof 3.2: Convexity for C^2 functions

To prove that the Theorem 3.1 implies the Theorem 3.5, we can use a second order Taylor expansion of y at x in an arbitrary direction p:

$$f(x + tp) = f(x) + t\nabla f(x)^T p + \frac{t^2}{2} p^T \nabla^2 f(x) p + o(t^2\|p\|).$$

From this we obtain that

$$p^T \nabla^2 f(x) p = \lim_{t \to 0} \frac{2}{t^2} \left(\underbrace{f(x+tp) - f(x) - t\nabla f(x)^T p}_{(3.1):\ \geq 0} \right) \geq 0.$$

Conversely, to prove the other direction, we use the Taylor rest term formula with some arbitrary $\theta \in (0, 1)$:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \underbrace{\frac{1}{2} t^2 (y-x)^T \nabla^2 f(x + \theta(y-x))(y-x)}_{(3.5):\ \geq 0}.$$

□

Property 3.1: Convexity perserving operations on convex functions

The following operations preserve the convexity of a function:

1. Affine input transformation: If $f: \ \Omega \to \mathbb{R}$ is convex, then also

$$A \in \mathbb{R}^{n \times m}: \ \tilde{f}(x) = f(Ax + b)$$

   is convex on the domain $\tilde{\Omega} = \{x \mid Ax + b \in \Omega\}$.

2. Concatenation with monotone convex function: If $f: \ \Omega \to \mathbb{R}$ is convex and $g: \ \mathbb{R} \to \mathbb{R}$ is convex and monotonely increasing, then the composition $g \circ f$ is convex.

3. The supremum over a set of convex functions $f_i(x)$, $i \in I$, i.e.,

$$f(x) = \sup_{i \in I} f_i(x)$$

   is convex.

Proof 3.3: Convexity perserving operations on convex functions

1. Not seen in this course.

2. Recall that g is a convex and monotonely increasing function, then:

$$\nabla^2 (g \circ f) = \underbrace{g''(f(x))}_{\geq 0} \underbrace{\nabla f(x) \nabla f(x)^T}_{\succeq 0} + \underbrace{g'(f(x))}_{\geq 0} \underbrace{\nabla^2 f(x)}_{\succeq 0} \succeq 0,$$

   i.e. $g \circ f$ is convex, since the Hessian is positive semi-definite.

3. Epigraph of f is the intersection of the epigraphs of f_i, which are convex.

□

Theorem 3.6: Concave function

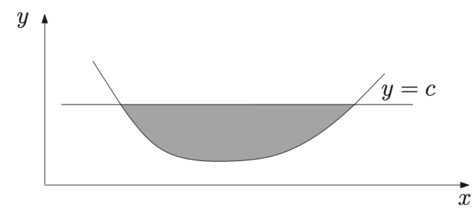
A function $f: \ \Omega \to \mathbb{R}$ is concave if and only if $-f$ is convex.

Theorem 3.7: Convexity of Sublevel subsets

The sublevel set

$$\{ \ x \in \Omega \mid f(x) \leq c \ \}$$

of a convex function $f : \Omega \to \mathbb{R}$ with respect to any constant $c \in \mathbb{R}$ is convex.



Proof 3.4: Convexity of Sublevel subsets

If $f(x) \leq c$ and $f(y) \leq c$, then for any $\lambda \in [0,1]$ holds also

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \leq \underbrace{(1-\lambda)c + \lambda c}_{=c}.$$

□

Property 3.2: Convexity perserving operations on convex sets

The following operations preserve the convexity of a set:

1. The intersection of finitely or infinitely many convex sets is convex.
2. Affine image: if Ω is convex, then for $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ also the set $A\Omega + b = \{y \in \mathbb{R}^m \mid \exists x \in \Omega : y = Ax + b\}$ is convex.
3. Affine pre-image: if Ω is convex, then for $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$ also the set $\{z \in \mathbb{R}^m \mid Az + b \in \Omega\}$ is convex.

Example 3.1: Convex Feasible set

If $\forall i \in [1, m] : f_i : \ \mathbb{R}^n \to \mathbb{R}$ are convex functions, then the set

$$\Omega = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0, \ i \in [1, m]\}$$

is a convex set, because it is the intersection of sublevel sets Ω_i of convex functions f_i, i.e.

$$\begin{aligned} \Omega &= \bigcap_{i=1}^{m} \Omega_i \\ &= \bigcap_{i=1}^{m} \{x \in \mathbb{R}^n \mid f_i(x) \leq 0\}. \end{aligned}$$

4 The Lagrangian Function and Duality

Unconstrained Optimization Algorithms

5 Optimality Conditions

6 Estimation and Fitting Problems

7 Newton Type Optimization

8 Globalisation Strategies

9 Calculating Derivatives

Constrained Optimization Algorithms

10 Optimality Conditions for Constrained Optimization

11 Equality Constrained Optimization Algorithms

12 Inequality Constrained Optimization Algorithms

13 Optimal Control Problems