Optimization for Data Science
ETH Zürich, FS 2023 261-5110-00L

Lecture 2: Theory of Convex Functions

To a large extent based on the classical textbook

Convex Optimization by Stephen Boyd and Lieven Vandenberghe

Bernd Gärtner
Niao He

https://www.ti.inf.ethz.ch/ew/courses/ODS23/index.html

February 28, 2023

Optimization

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in X \subseteq \mathbb{R}^d \end{aligned}$$

"Find some $\mathbf{x} \in X$ with smallest possible value $f(\mathbf{x})$!"

Terminology:

- $f : \mathbf{dom}(f) \to \mathbb{R}, X \subseteq \mathbf{dom}(f) \subseteq \mathbb{R}^d$: **objective function**
- X: set of **feasible solutions** (if $X = \mathbb{R}^d$: **unconstrained** optimization)
- $\mathbf{x}^\star$: **minimum** (minimizer of f over X), if it exists
- if f is (twice) differentiable, **first-order** and **second-order** methods can be used

If f and X are **convex**, optimization error ε can be achieved through stepwise improvement:

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^\star) + \varepsilon$$

Background: The Cauchy-Schwarz inequality

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Cauchy-Schwarz inequality: $|\mathbf{u}^\top \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$.

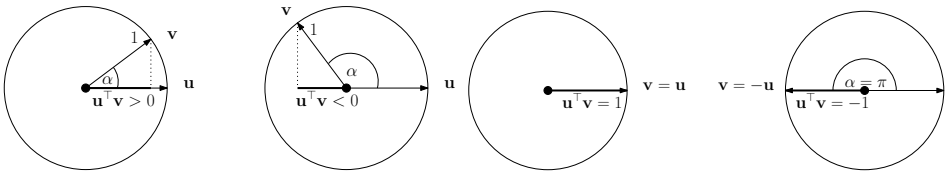
For nonzero vectors, this is equivalent to

similar to the vector multiplcation

this shows the connection between two vectors.

$$-1 \leq \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|} \leq 1.$$

Fraction can be used to define the angle α between $\mathbf{u}$ and $\mathbf{v}$: $\cos(\alpha) = \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$



Examples for unit vectors ($\|\mathbf{u}\| = \|\mathbf{v}\| = 1$)

Equality in Cauchy-Schwarz if and only if $\mathbf{u} = \mathbf{v}$ or $\mathbf{u} = -\mathbf{v}$.

Background: The spectral norm

this one is quite useful in the provement.

Let A be an $(m \times d)$-matrix. Then

$$\|A\| := \max_{\mathbf{v} \in \mathbb{R}^d, \mathbf{v} \neq 0} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} = \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

is the 2-norm (or spectral norm) of A.

$\|A\|$ is the largest factor by which a vector can be stretched in length under the mapping $\mathbf{v} \to A\mathbf{v}$.

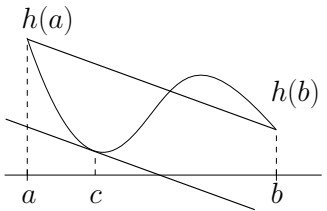
BTW, what is a norm? Read Section 2.1.3 in the notes (or wait a couple of slides)!

Background: The mean value theorem

Let $a < b$ be real numbers, and let $h : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable on (a, b); we denote the derivative by h'. Then there exists $c \in (a, b)$ such that

$$h'(c) = \frac{h(b) - h(a)}{b - a}.$$

learnt this one in high school.



Geometric interpretation:

- $(h(b) - h(a))/(b - a)$ is the slope of the line through the two points $(a, h(a))$ and $(b, h(b))$.
- The mean value theorem says that between a and b, we find a tangent to the graph of h that has the same slope.

Background: The fundamental theorem of calculus

Let $a < b$ be real numbers, and let $h : \mathbf{dom}(h) \to \mathbb{R}$ be a differentiable function on an open domain $\mathbf{dom}(h) \supset [a, b]$, and such that h' is continuous on $[a, b]$. Then

[a,b] is part of h.

$$h(b) - h(a) = \int_a^b h'(t)dt.$$

simple

This theorem is the theoretical underpinning of typical definite integral computations in high school.

For example, to evaluate $\int_2^4 x^2 dx$, we integrate x^2 (giving us $x^3/3$), and then compute

$$\int_2^4 x^2 dx = \frac{4^3}{3} - \frac{2^3}{3} = \frac{56}{3}.$$

BTW, what does differentiable mean? Read Section 2.1.6 in the notes (basics will follow on a later slide).

Background: Convex Sets

A set $C \subseteq \mathbb{R}^d$ is **convex** if the line segment between any two points of C lies in C, i.e., if for any $\mathbf{x}, \mathbf{y} \in C$ and any λ with $0 \leq \lambda \leq 1$, we have

$$\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in C.$$

this is used to define what is a convex set.

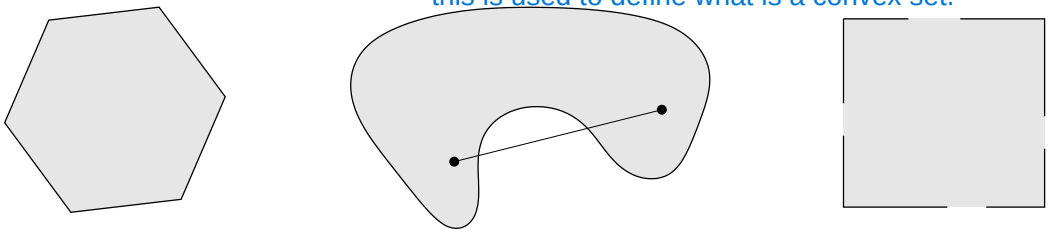


*Figure 2.2 from S. Boyd, L. Vandenberghe

Left Convex.

Middle Not convex, since line segment not in set.

Right Not convex, since some, but not all boundary points are contained in the set.

Background: Properties of Convex Sets

Intersections of convex sets are convex.

Observation 2.9

Let $C_i, i \in I$ be convex sets, where I is a (possibly infinite) index set. Then $C = \bigcap_{i \in I} C_i$ is a convex set.

ordering the convex set will result in convex set too.

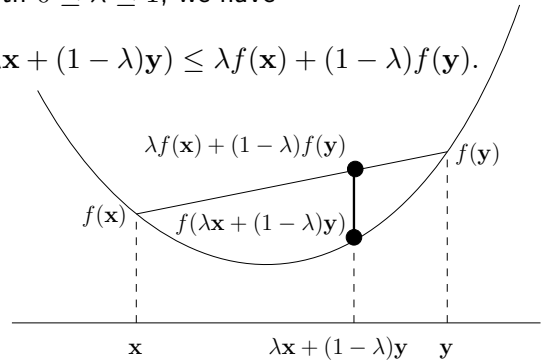
Convex Functions

two conditions, these are normally listed in the questions as the given information

Definition 2.11

A function $f : \mathbf{dom}(f) \to \mathbb{R}$ is **convex** if (i) $\mathbf{dom}(f)$ is a convex set and (ii) for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$, and λ with $0 \leq \lambda \leq 1$, we have

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$



Geometrically: The line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ lies above the graph of f.

Motivation: Convex Optimization

Convex Optimization Problems are of the form

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in X \subseteq \mathbb{R}^d \end{aligned}$$

R^d is a convex set.

where both

- f is a convex function
- $X \subseteq \mathbf{dom}(f)$ is a convex set (note: $\mathbb{R}^d$ is convex)

Crucial Property of Convex Optimization Problems

- Every local minimum is a **global minimum**, can't get stuck during stepwise improvement (see later...)

Motivation: Solving Convex Optimization - Provably

For convex optimization problems, many algorithms compute a sequence $\mathbf{x}_0, \mathbf{x}_1, \ldots$ that does **converge** to a global minimum! (assuming that f is differentiable)

local minimum is the global minimum too.

Example Theorem: The **convergence rate** is proportional to $\frac{1}{t}$, i.e. for all t,

$$f(\mathbf{x}_t) - f(\mathbf{x}^\star) \leq \frac{c}{t}$$

(where $\mathbf{x}^\star$ is some optimal solution to the problem.)

Meaning: **Absolute approximation error** converges to 0 over time.

Convex Functions and Sets

The **graph** of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\{(\mathbf{x}, f(\mathbf{x})) \,|\, \mathbf{x} \in \mathbf{dom}(f)\},$$

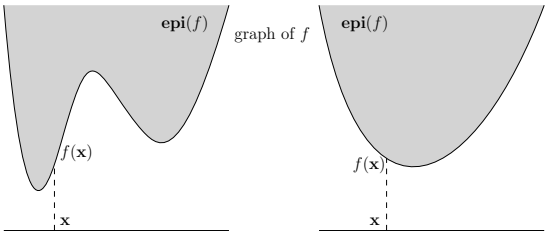
the graph definition

The **epigraph** of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathbf{epi}(f) := \{(\mathbf{x}, \alpha) \in \mathbb{R}^{d+1} \,|\, \mathbf{x} \in \mathbf{dom}(f), \alpha \geq f(\mathbf{x})\},$$

one more demension

values that are greater than f(x).



Observation 2.12

f is a convex function if and only if $\mathbf{epi}(f)$ is a convex set.

one way to prove convex function.

Convex Functions

Examples of convex functions

- Affine functions: $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$
- Squares: $f(x) = x^2$
- Exponential: $f(x) = e^{\alpha x}$
- Norms. Every norm on $\mathbb{R}^d$ is convex.
- Some more later...

these may appear in true or false questions

this one is quite important, it shows in the previous exam questions

Convexity of a norm $\|\mathbf{x}\|$

By the triangle inequality $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ and homogeneity of a norm $\|a\mathbf{x}\| = |a| \, \|\mathbf{x}\|$, $a \in \mathbb{R}$:

$$\|\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\| \leq \|\lambda \mathbf{x}\| + \|(1-\lambda)\mathbf{y}\| = \lambda \, \|\mathbf{x}\| + (1-\lambda) \, \|\mathbf{y}\| \, .$$

We used the triangle inequality for the inequality and homogeneity for the equality.

In the sequel, we use $\|\mathbf{x}\|$ for the "default" Euclidean norm of a vector.

Jensen's Inequality

this one is interesting, useful for E calculation.

Lemma 2.13

Let f be convex, $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbf{dom}(f)$, $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$ such that $\sum_{i=1}^m \lambda_i = 1$. Then

$$f\left(\sum_{i=1}^m \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^m \lambda_i f(\mathbf{x}_i).$$

For $m = 2$, this is convexity. The proof of the general case is Exercise 7.

Convex Functions are Continuous

Lemma 2.14

Let f be convex and suppose that $\mathbf{dom}(f) \subseteq \mathbb{R}^d$ is open. Then f is continuous.

Not entirely obvious (Exercise 8).

d is finite, this makes sense in the real cases.

In fact, even linear functions can be discontinuous over domains of infinite dimension (Lemma 2.15 gives a classical example).

Reminder: we are always working in finite dimension d, unless stated otherwise.

Differentiable Functions

How to check convexity? Use definition (works, but can be cumbersome).

Easier ways exist for differentiable and twice differentiable functions.

f differentiable at $\mathbf{x}$ means: f is (around $\mathbf{x}$) well-approximated by an affine function $\ell(\mathbf{y}) = f(\mathbf{x}) + A(\mathbf{y} - \mathbf{x})$, where A is some matrix (depending on $\mathbf{x}$).

Definition 2.5

Let $f : \mathbf{dom}(f) \to \mathbb{R}^m$ where $\mathbf{dom}(f) \subseteq \mathbb{R}^d$ is open. f is called differentiable at $\mathbf{x} \in \mathbf{dom}(f)$ if there exists an $(m \times d)$-matrix A and an error function $r : \mathbb{R}^d \to \mathbb{R}^m$ defined around $\mathbf{0} \in \mathbb{R}^d$ such that for all $\mathbf{y}$ in some neighborhood of $\mathbf{x}$,

$$f(\mathbf{y}) = f(\mathbf{x}) + A(\mathbf{y} - \mathbf{x}) + r(\mathbf{y} - \mathbf{x}),$$

where

$$\lim_{\mathbf{v} \to \mathbf{0}} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} = \mathbf{0}. \quad \text{(Error } r \text{ is sublinear around } \mathbf{0})$$

A is unique and called the differential or Jacobian matrix of f at $\mathbf{x}$.

Differentiable Functions

$$f(\mathbf{y}) = f(\mathbf{x}) + A(\mathbf{y} - \mathbf{x}) + r(\mathbf{y} - \mathbf{x}),$$

$$\lim_{\mathbf{v} \to \mathbf{0}} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} = \mathbf{0}.$$

Example: $f(x) = x^2$. We know that derivative $f'(x) = 2x$. Why? For $y = x + v$,
examples are intuitive for understanding.

$$\begin{aligned} f(y) = (x+v)^2 &= x^2 + 2vx + v^2 \\ &= f(x) + 2x \cdot v + v^2 \\ &= f(x) + A(y - x) + r(y - x)), \end{aligned}$$

where $A = 2x, \quad r(y - x) = r(v) = v^2.$ $\lim_{v \to 0} \frac{|r(v)|}{|v|} = \lim_{v \to 0} |v| = 0.$

Differentiable Functions

$$f(\mathbf{y}) = f(\mathbf{x}) + A(\mathbf{y} - \mathbf{x}) + r(\mathbf{y} - \mathbf{x}), \quad \lim_{\mathbf{v} \to \mathbf{0}} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} = \mathbf{0}.$$

We denote $Df(\mathbf{x}) := A$. $Df(\mathbf{x})$ is the matrix of partial derivatives at the point $\mathbf{x}$,

$$Df(\mathbf{x})_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}).$$

f is called differentiable if f is differentiable at all $\mathbf{x} \in \mathbf{dom}(f)$.

If $f : \mathbf{dom}(f) \to \mathbb{R}$, $Df(\mathbf{x})$ is a row vector denoted by $\nabla f(\mathbf{x})^\top$.

The **gradient** $\nabla f(\mathbf{x})$ (vector of partial derivatives of f) is a column vector.

aline with the computer science, this is column vector.

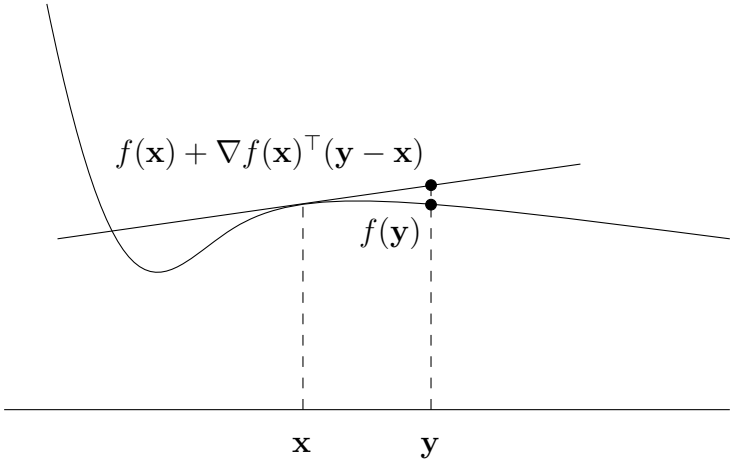
Examples:

- $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} = \sum_{j=1}^d c_j x_j \quad \Rightarrow \nabla f(\mathbf{x}) = \mathbf{c}$
- $f(\mathbf{x}) = \|\mathbf{x}\|^2 = \sum_{j=1}^d x_j^2 \quad \Rightarrow \nabla f(\mathbf{x}) = 2\mathbf{x}$

way to express vectors.

Differentiable Functions

Graph of the affine function $f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$ is a tangent hyperplane to the graph of f at $(\mathbf{x}, f(\mathbf{x}))$.



First-order Characterization of Convexity

Lemma 2.16

Suppose that $\mathbf{dom}(f)$ is open and that f is differentiable; in particular, the gradient (vector of partial derivatives)

$$\nabla f(\mathbf{x}) := \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_d}(\mathbf{x}) \right)$$

exists at every point $\mathbf{x} \in \mathbf{dom}(f)$.

Then f is convex if and only if $\mathbf{dom}(f)$ is convex and

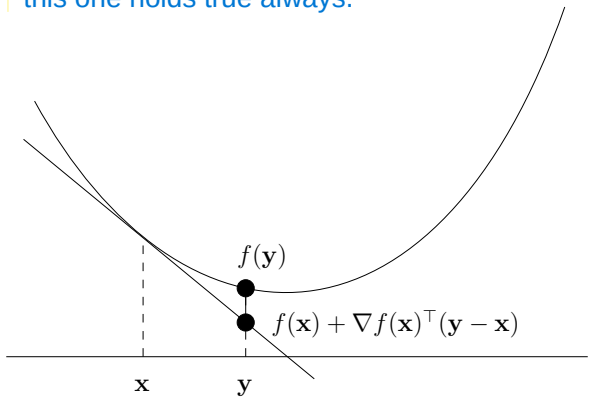
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{1}$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$.

First-order Characterization of Convexity

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

this one holds true always.



Graph of f is above all its tangent hyperplanes.

First-order Characterization of Convexity: Proof

f convex iff $\mathbf{dom}(f)$ convex and

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

$\Rightarrow$: suppose f is convex.

Then, for all $t \in (0, 1)$,

$$f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) = f((1-t)\mathbf{x} + t\mathbf{y}) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y}) = f(\mathbf{x}) + t(f(\mathbf{y}) - f(\mathbf{x})).$$

Subtracting $f(\mathbf{x})$ on both sides, dividing by t, and using differentiability at $\mathbf{x}$:

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}) + \frac{f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{t} \\ &= f(\mathbf{x}) + \frac{\nabla f(\mathbf{x})^\top t(\mathbf{y} - \mathbf{x}) + r(t(\mathbf{y} - \mathbf{x}))}{t} \\ &= f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \underbrace{\frac{r(t(\mathbf{y} - \mathbf{x}))}{t}}_{\to 0 \text{ for } t \to 0} \end{aligned}$$

First-order Characterization of Convexity: Proof

f convex iff $\mathbf{dom}(f)$ convex and

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

$\Leftarrow$: suppose the inequality holds.

$\mathbf{z} := \lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in \mathbf{dom}(f)$ for $\lambda \in [0, 1]$ (by convexity of $\mathbf{dom}(f)$)

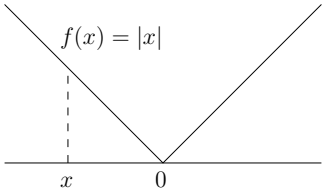
$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^\top (\mathbf{x} - \mathbf{z}) \quad | \cdot \lambda \\ f(\mathbf{y}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \quad | \cdot (1 - \lambda) \end{aligned}$$

Adding up, gradient terms cancel:

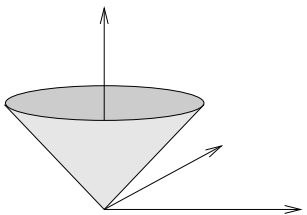
$$\lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) \geq f(\mathbf{z}) = f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}).$$

Nondifferentiable Functions...

are also relevant in practice.



More generally, $f(\mathbf{x}) = \|\mathbf{x}\|$ (Euclidean norm). For $d = 2$, graph is the ice cream cone:



Second-order Characterization of Convexity

Lemma 2.18

Suppose that $\mathbf{dom}(f)$ is open and that f is twice differentiable; in particular, the **Hessian** (matrix of second partial derivatives)

the hessian here is quite important, it will appear in the exam.

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_d \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d}(\mathbf{x}) \end{pmatrix}$$

exists at every point $\mathbf{x} \in \mathbf{dom}(f)$ and is symmetric. Then f is convex if and only if $\mathbf{dom}(f)$ is convex, and for all $\mathbf{x} \in \mathbf{dom}(f)$, we have

$$\nabla^2 f(\mathbf{x}) \succeq 0 \quad \text{(i.e. } \nabla^2 f(\mathbf{x}) \text{ is } \textbf{positive semidefinite)}$$

this is the definition of PSD, this is also quite important.

(A symmetric matrix M is positive semidefinite if $\mathbf{x}^\top M \mathbf{x} \geq 0$ for all $\mathbf{x}$, and positive definite if $\mathbf{x}^\top M \mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.)

Second-order Characterization of Convexity

this one is quite useful in the exam.

Example: $f(x_1, x_2) = x_1^2 + x_2^2$.

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succeq 0.$$

Directly checking convexity of f using the definition of convexity is more laborious.

Operations that Preserve Convexity

Exercise 10

(i) Let $f_1, f_2, \ldots, f_m$ be convex functions, $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}_+$. Then

$$f := \max_{i=1}^{m} f_i$$

as well as

this is convex

$$f := \sum_{i=1}^{m} \lambda_i f_i$$

are convex on $\mathbf{dom}(f) := \bigcap_{i=1}^{m} \mathbf{dom}(f_i)$.

(ii) Let f be a convex function with $\mathbf{dom}(f) \subseteq \mathbb{R}^d$, $g : \mathbb{R}^m \to \mathbb{R}^d$ an affine function, meaning that $g(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, for some matrix $A \in \mathbb{R}^{d \times m}$ and some vector $\mathbf{b} \in \mathbb{R}^d$. Then the function $f \circ g$ (that maps $\mathbf{x}$ to $f(A\mathbf{x} + \mathbf{b})$) is convex on $\mathbf{dom}(f \circ g) := \{\mathbf{x} \in \mathbb{R}^m : g(\mathbf{x}) \in \mathbf{dom}(f)\}$. this is also convex

Local Minima are Global Minima

Definition 2.20

A **local minimum** of $f : \mathbf{dom}(f) \to \mathbb{R}$ is a point $\mathbf{x}$ such that there exists $\varepsilon > 0$ with

$$f(\mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{dom}(f) \text{ satisfying } \|\mathbf{y} - \mathbf{x}\| < \varepsilon.$$

Meaning: in some small neighborhood, $\mathbf{x}$ is the best point.

Lemma 2.21

Let $\mathbf{x}^\star$ be a **local minimum** of a convex function $f : \mathbf{dom}(f) \to \mathbb{R}$. Then $\mathbf{x}^\star$ is a **global minimum**, meaning that $f(\mathbf{x}^\star) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{dom}(f)$.

Proof.

Suppose there exists $\mathbf{y} \in \mathbf{dom}(f)$ such that $f(\mathbf{y}) < f(\mathbf{x}^\star)$.
Define $\mathbf{y}' := \lambda \mathbf{x}^\star + (1 - \lambda)\mathbf{y}$ for $\lambda \in (0, 1)$.
From convexity, we get that that $f(\mathbf{y}') < f(\mathbf{x}^\star)$. Choosing λ so close to 1 that $\|\mathbf{y}' - \mathbf{x}^\star\| < \varepsilon$ yields a contradiction to $\mathbf{x}^\star$ being a local minimum. □

Critical Points and Global Minima

Lemma 2.23

Suppose that $f : \mathbf{dom}(f) \to \mathbb{R}$ is differentiable over an open domain $\mathbf{dom}(f) \subseteq \mathbb{R}^d$. Let $\mathbf{x} \in \mathbf{dom}(f)$. If $\mathbf{x}$ is a global minimum then $\nabla f(\mathbf{x}) = \mathbf{0}$ (a **critical point**).

For convex functions, the converse is also true:

Lemma 2.22

Suppose that f is convex and differentiable over an open domain $\mathbf{dom}(f)$. Let $\mathbf{x} \in \mathbf{dom}(f)$. If $\nabla f(\mathbf{x}) = \mathbf{0}$ (a critical point), then $\mathbf{x}$ is a global minimum.

Proof.

Suppose that $\nabla f(\mathbf{x}) = \mathbf{0}$. According to the first-order characterization of convexity, we have

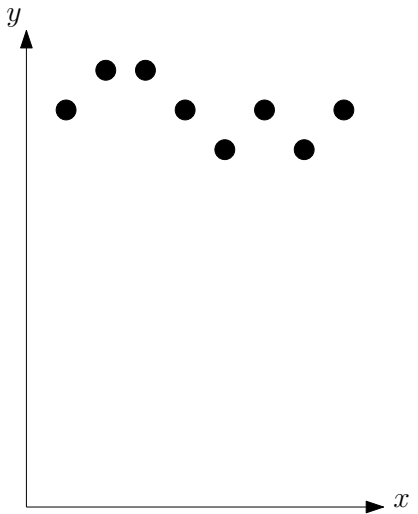
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) = f(\mathbf{x})$$

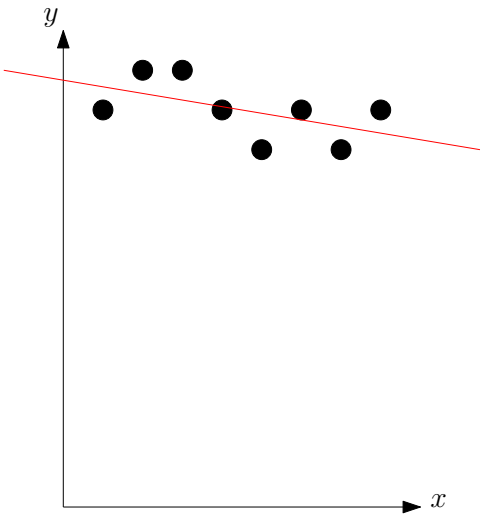
for all $\mathbf{y} \in \mathbf{dom}(f)$, so $\mathbf{x}$ is a global minimum. □

Critical point: tangent hyperplane is horizontal at $\mathbf{x}$.

Example: Least Squares

Problem: Fit a line to a set of points





Example: Least Squares

Points $(x_i, y_i), i = 1, \ldots, 8$:

$$(1, 10), (2, 11), (3, 11), (4, 10), (5, 9), (6, 10), (7, 9), (8, 10)$$

Line: $y = w_0 + w_1 x$

Fitting error (sum of squared vertical distances of points to the line):

$$\begin{aligned} f(w_0, w_1) &= \sum_{i=1}^{8} (w_1 x_i + w_0 - y_i)^2 \\ &= 204 w_1^2 + 72 w_1 w_0 - 706 w_1 + 8 w_0^2 - 160 w_0 + 804 \end{aligned}$$

Function is convex:

$$\nabla^2 (w_0, w_1) = \begin{pmatrix} 16 & 72 \\ 72 & 408 \end{pmatrix} \succeq 0.$$

Another proof of convexity: f is the sum of (more easily seen to be) convex functions $(w_1 x_i + w_0 - y_i)^2, i = 1, \ldots, 8$.

Example: Least Squares

Global minimum (solve for critical point):

$$\nabla f(w_0, w_1) = (72w_1 + 16w_0 - 160, 408w_1 + 72w_0 - 706) = (0, 0).$$

System of linear equations.

$$(w_0^\star, w_1^\star) = \left(\frac{43}{4}, -\frac{1}{6}\right).$$

Hence, the optimal line is

$$y = -\frac{1}{6}x + \frac{43}{4}.$$

Fact

Convex quadratic functions can be minimized by solving a system of linear equations, no need to run any optimization algorithm.

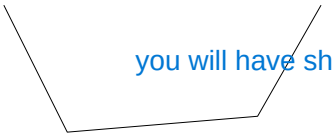
Solving for a critical point is always a system of equations, but these are typically nonlinear and therefore hard to solve analytically $\rightarrow$ optimization!

Strictly Convex Functions (no flat parts)

Definition 2.24

A function $f : \mathbf{dom}(f) \to \mathbb{R}$ is **strictly convex** if (i) $\mathbf{dom}(f)$ is convex and (ii) for all $\mathbf{x} \neq \mathbf{y} \in \mathbf{dom}(f)$ and all $\lambda \in (0, 1)$, we have

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$



convex, but not strictly convex

strictly convex

Lemma 2.26

Let $f : \mathbf{dom}(f) \to \mathbb{R}$ be strictly convex. Then f has at most one global minimum.

Constrained Minimization

Definition 2.27

Let $f : \mathbf{dom}(f) \to \mathbb{R}$ be convex and let $X \subseteq \mathbf{dom}(f)$ be a convex set. A point $\mathbf{x} \in X$ is a minimizer of f over X if

f(x) is the lowest value among all the points of f.

$$f(\mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in X.$$

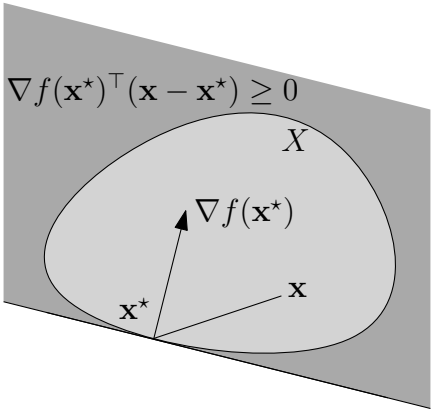
Lemma ([BV04, 4.2.3])

Suppose that $f : \mathbf{dom}(f) \to \mathbb{R}$ is convex and differentiable over an open domain $\mathbf{dom}(f) \subseteq \mathbb{R}^d$, and let $X \subseteq \mathbf{dom}(f)$ be a convex set. Point $\mathbf{x}^\star \in X$ is a minimizer of f over X if and only if

this one might be quite important.

$$\nabla f(\mathbf{x}^\star)^\top (\mathbf{x} - \mathbf{x}^\star) \geq 0 \quad \forall \mathbf{x} \in X.$$

Constrained Minimization



this meaning is very intuitive.

Geometric meaning: X is contained in the halfspace $\{\mathbf{x} \in \mathbb{R}^d : \nabla f(\mathbf{x}^\star)^\top(\mathbf{x}-\mathbf{x}^\star) \geq 0\}$ (normal vector $\nabla f(\mathbf{x}^\star)$ at $\mathbf{x}^\star$ pointing into the halfspace).

Existence of a minimizer

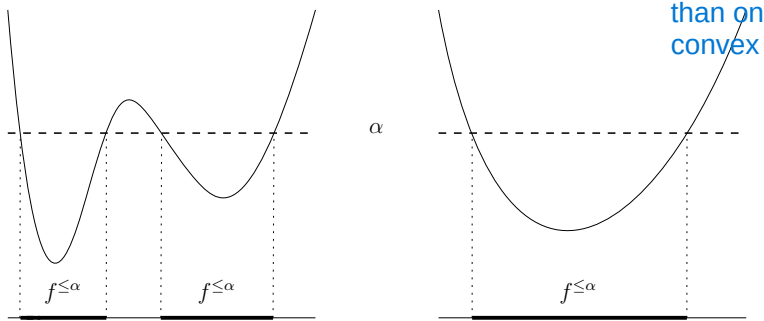
How do we know that a global minimum exists?

Not necessarily the case, even if f bounded from below (example: $f(x) = e^x$)

Definition 2.29

$f : \mathbb{R}^d \to \mathbb{R}$, $\alpha \in \mathbb{R}$. The set $f^{\leq \alpha} := \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \leq \alpha\}$ is the α-sublevel set of f.

this could be more than one if f is not a convex function.



The Weierstrass Theorem

Theorem 2.30
Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function, and suppose there is a nonempty and bounded sublevel set $f^{\leq \alpha}$. Then f has a global minimum. this is true.

Proof:

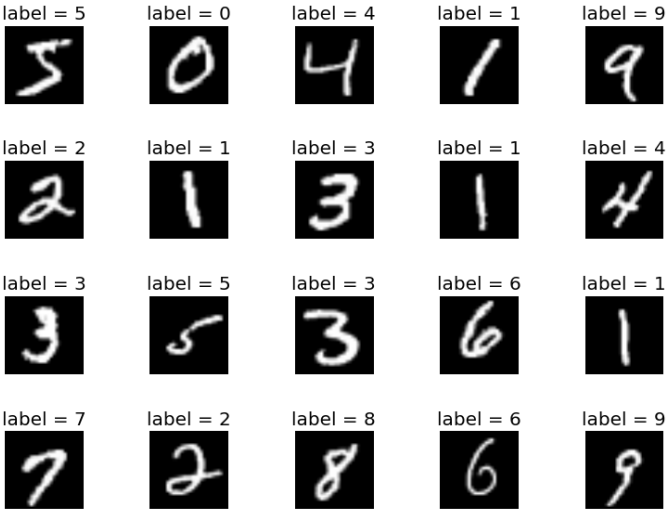
We know that f—as a continuous function—attains a minimum over the closed and bounded (= compact) set $f^{\leq \alpha}$ at some $\mathbf{x}^\star$. This $\mathbf{x}^\star$ is also a global minimum as it has value $f(\mathbf{x}^\star) \leq \alpha$, while any $\mathbf{x} \notin f^{\leq \alpha}$ has value $f(\mathbf{x}) > \alpha \geq f(\mathbf{x}^\star)$.

Applies to convex functions (they are continuous by Exercise 8).

Generalizes to suitable domains $\mathbf{dom}(f) \neq \mathbb{R}^d$.

Example: Handwritten Digit Recognition (MNIST database)

Task: recognize handwritten decimal digits $0, 1, \ldots, 9$



Example: Handwritten Digit Recognition

Training data:

- set P of grayscale images (28×28 pixels)
- for each $\mathbf{x} \in P$, the correct digit $d(\mathbf{x}) \in \{0, \ldots, 9\}$

Approach:

- represent image as feature vector $\mathbf{x} \in \mathbb{R}^{28 \cdot 28} = \mathbb{R}^{784}$, where x_i is the gray value of the i-th pixel
- fit a matrix $W \in \mathbb{R}^{10 \times 784}$ to the training data
- use vector $\mathbf{y} = W\mathbf{x} \in \mathbb{R}^{10}$ to predict the digit seen in an arbitrary image $\mathbf{x}$
- idea: $y_j, j = 0, \ldots, 9$ should tell us the probability of the digit being j
- For example, use probabilities $z_j = z_j(\mathbf{y}) = \frac{e^{y_j}}{\sum_{k=0}^{9} e^{y_k}}$

Example: Handwritten Digit Recognition

Matrix W should minimize the recognition error on the training data.

Measure recognition error by a loss function (there are many choices).

For example,

$$\ell(W) = -\sum_{\mathbf{x}\in P} \ln\left(z_{d(\mathbf{x})}(W\mathbf{x})\right) = \sum_{\mathbf{x}\in P}\left(\ln\left(\sum_{k=0}^{9} e^{(W\mathbf{x})_k}\right) - (W\mathbf{x})_{d(\mathbf{x})}\right).$$

- $z_{d(\mathbf{x})}(W\mathbf{x}) \in (0,1)$: probability of predicting the correct digit $d(\mathbf{x})$ on training image $\mathbf{x}$
- $-\ln\left(z_{d(\mathbf{x})}(W\mathbf{x})\right) > 0$
- tends to ∞ for probability tending to 0 (punishes small probability)
- tends to 0 for probability tending to 1 (rewards large probability)

Example: Handwritten Digit Recognition

Exercise 12

The function $\ell : \mathbb{R}^{10 \cdot 784} \to \mathbb{R}$ given by

$$\ell(W) = -\sum_{\mathbf{x} \in P} \ln\left(z_{d(\mathbf{x})}(W\mathbf{x})\right) = \sum_{\mathbf{x} \in P} \left(\ln\left(\sum_{k=0}^{9} e^{(W\mathbf{x})_k} \right) - (W\mathbf{x})_{d(\mathbf{x})} \right)$$

is convex.

The function ℓ does not necessarily have a global minimum, but one can characterize the training sets for which it does. This needs material on weakly coercive functions, see notes (Exercise 13).

Convex Programming under (In)equality Constraints

Optimization problem in standard form [BV04, 4.1.1]:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

Problem domain $\mathcal{D} = \cap_{i=0}^{m} \mathbf{dom}(f_i) \cap \cap_{i=1}^{p} \mathbf{dom}(h_i)$ (assumed to be open)

Convex program:

- all f_i are convex functions, and all h_i are affine functions with domain $\mathbb{R}^d$.
- $\Rightarrow$ constrained minimization as in Section 2.4.3, with feasible region X induced by finitely many (in)equality constraints.

In the following, we do not need to assume convexity, unless explicitly stated.

Lagrange Duality

A powerful tool that (under suitable conditions) allows us to express an optimization problem differently.

This dual view often provides new insights and may help us in solving the original problem.

Example: linear programming duality with its many applications is a special case of Lagrange duality.

Lagrangian and Lagrange dual

Definition 2.45

Given an optimization problem in standard form, its Lagrangian is the function $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ given by

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}).$$

The λ_i, ν_i are called Lagrange multipliers.

The Lagrange dual function is the function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ defined by

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

get the inf via adjusting lambda and mu.

The Lagrange dual function g can assume value $-\infty$. : Not a pathology but typical!

"Interesting" $(\boldsymbol{\lambda}, \boldsymbol{\nu})$: the ones for which $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.

Example: Linear Programming

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

change the form to <= 0

This is a (convex) optimization problem in standard form ($\mathcal{D} = \mathbb{R}^d$, as all functions are defined everywhere):

- $f_0(\mathbf{x}) := \mathbf{c}^\top \mathbf{x}$
- $f_i(\mathbf{x}) := -x_i, \quad i = 1, \ldots, m$
- $h_i(\mathbf{x}) := \mathbf{a}_i^\top \mathbf{x} - b_i, \quad i = 1, \ldots, p$ (vector $\mathbf{a}_i$ generates i-th row of A)

Lagrangian: $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{c}^\top \mathbf{x} - \boldsymbol{\lambda}^\top \mathbf{x} + \boldsymbol{\nu}^\top (A\mathbf{x} - \mathbf{b}) = -\mathbf{b}^\top \boldsymbol{\nu} + (\mathbf{c}^\top - \boldsymbol{\lambda}^\top + \boldsymbol{\nu}^\top A)\mathbf{x}$.

inequality equality

Lagrange dual function:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^\top \boldsymbol{\nu} & \text{if } \mathbf{c}^\top - \boldsymbol{\lambda}^\top + \boldsymbol{\nu}^\top A = \mathbf{0}, \\ -\infty & \text{otherwise.} \end{cases}$$

this one is not very clear, why

Weak Lagrange duality

Lagrange dual function values are lower bounds on primal function values $f_0(\mathbf{x})$.

Lemma 2.46

Let $\mathbf{x}$ be a feasible solution, meaning that $f_i(\mathbf{x}) \leq 0$ for $i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1, \ldots, p$. Let g be the Lagrange dual function of and $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^p$ such that $\boldsymbol{\lambda} \geq \mathbf{0}$. Then

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x}).$$

Proof.

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x})}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i h_i(\mathbf{x})}_{=0} \leq f_0(\mathbf{x}).$$

□

What is the best lower bound we can get in this way?

The Lagrange dual

Choose $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\nu}$ such that $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is maximized!

Definition 2.47

The Lagrange dual of an optimization problem in standard form is the optimization problem

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

By weak duality, the supremum value of the Lagrange dual is a lower bound for the infimum value of the primal problem.

Exercise 18: the Lagrange dual is a convex program even if the primal is not!

What we mean: $-g$ is convex, so the equivalent problem "minimize $-g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ subject to $\boldsymbol{\lambda} \geq \mathbf{0}$" is a convex program. we need to put - in front of g.

(Requires proper handling of $-\infty$ values.)

Example: Linear Programming (continued)

Linear program:

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Lagrange dual function:

how to arrange a lagrange dual function, pay attention to htis part too.

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^\top \boldsymbol{\nu} & \text{if } \mathbf{c}^\top - \boldsymbol{\lambda}^\top + \boldsymbol{\nu}^\top A = \mathbf{0}, \\ -\infty & \text{otherwise.} \end{cases}$$

arrange all the terms ralted to x.

Lagrange dual (inequalities are equivalent to $\boldsymbol{\lambda} \geq \mathbf{0}$):

since we want to find the lower bound of the primal question.

$$\begin{array}{ll} \text{maximize} & -\mathbf{b}^\top \boldsymbol{\nu} \\ \text{subject to} & \mathbf{c}^\top + \boldsymbol{\nu}^\top A \geq \mathbf{0} \end{array}$$

this is from lambda >= 0

Renaming $-\boldsymbol{\nu}$ to $\mathbf{y}$, transposing the constraints $\Rightarrow$ "standard" dual linear program:

$$\begin{array}{ll} \text{maximize} & \mathbf{b}^\top \mathbf{y} \\ \text{subject to} & A^\top \mathbf{y} \leq \mathbf{c} \end{array}$$

make another transpose, then change the symbol.

Strong Lagrange Duality

this means that there is no duality gap.

Linear programming: $\inf \mathbf{c}^\top \mathbf{x} = \sup \mathbf{b}^\top \mathbf{y} \in \mathbb{R} \cup \{-\infty, \infty\}$.

Linear programming with finite value: $\min \mathbf{c}^\top \mathbf{x} = \max \mathbf{b}^\top \mathbf{y} \in \mathbb{R}$.

Theorem 2.48

Suppose that a convex program has a feasible solution $\tilde{\mathbf{x}}$ that in addition satisfies $f_i(\tilde{\mathbf{x}}) < 0, i = 1, \ldots, m$ (a Slater point). Then the infimum value of the primal equals the supremum value of its Lagrange dual. Moreover, if this value is finite, it is attained by a feasible solution of the dual.

this one here is quite important, it will shift the question into another one.

Convex programming with Slater point and finite value: $\inf f_0(\mathbf{x}) = \max g(\boldsymbol{\lambda}, \boldsymbol{\nu})$.

Exercise 19: an illustration of Theorem 2.48 in a simple example, also showing that a finite value is not necessarily attained in the primal.

Application of Lagrange Duality

Turn "hard" constraints into "soft" ones by moving them to the objective function.

Constrained problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

note, < = 0, change the form if not.

Unconstrained problem (for fixed $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\nu}$):

minimize the Lagrangian

$$f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

Infimum of unconstrained problem: $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$

$\Rightarrow$ lower bound for the infimum of the constrained problem (weak Lagrange duality).

If $\inf f_0(\mathbf{x}) = \max g(\boldsymbol{\lambda}, \boldsymbol{\nu})$, there are $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star \geq \mathbf{0}$ and $\boldsymbol{\nu} = \boldsymbol{\nu}^\star$ such that the constrained and the unconstrained problem have the same infimum.

In practice, we often search for some $\boldsymbol{\lambda}, \boldsymbol{\nu}$ such that the lower bound is useful.

Zero Duality Gap

Case of particular interest: $\min f_0(\mathbf{x}) = \max g(\boldsymbol{\lambda}, \boldsymbol{\nu})$

Definition 2.49

Let $\tilde{\mathbf{x}}$ be feasible for the primal and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ feasible for the Lagrange dual. The primal and dual solutions $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are said to have zero duality gap if $f_0(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$.

Consequence: **Master Equation**

$$
\begin{aligned}
f_0(\tilde{\mathbf{x}}) &= g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\
&= \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\mathbf{x}) \right) \\
&\leq f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \underbrace{\tilde{\lambda}_i f_i(\tilde{\mathbf{x}})}_{\leq 0} + \sum_{i=1}^{p} \underbrace{\tilde{\nu}_i h_i(\tilde{\mathbf{x}})}_{0} \\
&\leq f_0(\tilde{\mathbf{x}}).
\end{aligned}
$$

All inequalities are equalities!

Complementary Slackness

Lemma 2.50
If $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, then

$$\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, \ldots, m.$$

Proof.
Follows from

$$\underbrace{\tilde{\lambda}_i f_i(\tilde{\mathbf{x}})}_{=0}$$

in the Master Equation. □

Complementarity: If there is slack in the i-th inequality of the primal ($f_i(\tilde{\mathbf{x}}) < 0$), then there is no slack in the i-th inequality of the dual ($\tilde{\lambda}_i = 0$); and vice versa.

Vanishing Lagrangian gradient

Lemma 2.51

If $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, and if all f_i and h_i are differentiable, then

$$\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = \mathbf{0}.$$

Proof.

By equality in the third line of the Master Equation, $\tilde{\mathbf{x}}$ minimizes the differentiable function

$$f_0(\mathbf{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\mathbf{x}).$$

Hence its gradient vanishes by Lemma 2.23. □

Karush-Kuhn-Tucker necessary conditions

Follow directly from complementary slackness, vanishing Lagrangian gradient.

Theorem 2.52

Let $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ be feasible solutions of the primal optimization problem and its Lagrange dual, respectively, with zero duality gap. If all f_i and h_i in (2.20) are differentiable, then

$$\begin{aligned} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \ldots, m, \\ \nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) &= \mathbf{0}. \end{aligned}$$

Karush-Kuhn-Tucker conditions:

- primal and dual feasibilty
- complementary slackness
- vanishing Lagrangian gradient

Karush-Kuhn-Tucker sufficient conditions

Suppose that all f_i and h_i are differentiable, all f_i are convex, all h_i are affine.

Let $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ be such that the Karush-Kuhn-Tucker conditions hold.

Then $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap and hence are primal and dual optimal solutions (Theorem 2.53).

"Solving" the Karush-Kuhn-Tucker conditions may be easier than solving the primal optimization problem.

But we cannot always count on the Karush-Kuhn-Tucker conditions being solvable!

Theorem 2.52 guarantees them only if there are primal and dual solutions of zero duality gap.

But if the primal has a Slater point, then the Karush-Kuhn-Tucker conditions are indeed equivalent to the existence of a primal minimizer.

Complexity of solving convex programs

A general algorithm (**interior-point method**) exists [NN94].

Can be analyzed under suitable conditions.

First phase (find a feasible solution): runtime inversely depends on how close the problem is to being infeasible.

Second phase (find an almost optimal solution): number of iterations is of the order

$$O\left(\sqrt{m}\log\left(\frac{M-p^\star}{\varepsilon}\right)\right),$$

where $p^\star$ is the infimum value, and M some known upper bound for $p^\star$.

Number of variables d and number of equality constraints p enter the complexity of individual iterations.

Problem with interior point methods: each iteration is very costly (too costly for most applications).

Starting from next week: simple algorithms, with low cost per iteration, but possibly a higher number of iterations.

Bibliography


Stephen Boyd and Lieven Vandenberghe.
Convex Optimization.
Cambridge University Press, New York, NY, USA, 2004.
https://web.stanford.edu/~boyd/cvxbook/.


Y. Nesterov and A. Nemirovskii.
Interior-Point Polynomial Methods in Con- vex Programming.
Society for Industrial and Applied Mathematics, 1994.