1. Introduction

- course introduction
- optimization examples
- vectors and matrices

Mathematical optimization

(mathematical) optimization problem

$$\begin{aligned} \text{minimize (or maximize)} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \le 0,\ i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0,\ j = 1, \ldots, p \end{aligned} \tag{1.1}$$

- *optimization variable:* $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$
- *objective function:* $f : \mathbb{R}^n \to \mathbb{R}$
- *inequality constraints functions:* $g_i : \mathbb{R}^n \to \mathbb{R}$
- *equality constraints functions:* $h_j : \mathbb{R}^n \to \mathbb{R}$

Optimal point or solution: a point $\boldsymbol{x}^\star$ is an *optimal point* or *solution* to problem (1.1) if it attains the smallest (largest) objective value among all points that satisfy the constraints

Applications

Applications

- electrical network design
- min-weight aircraft and aerospace design
- optimal space vehicle trajectories
- cost-efficient civil structure design (e.g., bridges, dams)
- min-cost material handling equipment design (e.g., trucks, cranes)
- maximizing profit investment strategies
- optimal data model fitting

Modeling: the process of identifying the objective, constraints, and variables of a given problem is called *modeling*

Finding best decision

- the variable, x, symbolizes specific *actions* such as:
  - trades in a portfolio
  - adjustments to airplane control surfaces
  - task scheduling or assignment
  - resource allocation decisions
  - transmitted signal...
- constraint functions limit the action or set conditions on outcome:
  - physical or technical limits
  - resource budgets
  - design requirements that need be satisfied...
- objective represents some criteria, we want to minimize:
  - total cost
  - deviation from desired outcome (error)
  - consumption of fuel
  - risk...

Linear and nonlinear optimization

an optimization problem is called *linear program* if it has the form

$$\begin{array}{ll} \text{minimize (or maximize)} & \sum_{i=1}^{n} c_i x_i \\ \text{subject to} & \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \quad i = 1, \ldots, m \\ & \sum_{j=1}^{n} g_{ij} x_j = h_i, \quad i = 1, \ldots, p \end{array}$$

- $\{c_i, a_{ij}, g_{ij}, h_i, b_i\}$ are given coefficients
- the objective and constraint functions are “linear”

Nonlinear program: an optimization problem is called *nonlinear program* if it is not a linear program

Other optimization classes:

- *Unconstrained optimization*: no constraints, i.e., $h_j(\boldsymbol{x}) = g_i(\boldsymbol{x}) = 0$
- *Discrete optimization*: variables take only discrete or integer values
- *Integer linear program*: a discrete optimization with linear objective and constraints
- *Mixed integer optimization*: variables can be both integer and continuous

(Note: this course focuses solely on optimization with continuous variables)

Solving optimization problems

- various methods exist to solve optimization problems
- the chosen method will typically depend on several factors (*e.g.*, problem class and structure)
- solutions guide decision-makers, who oversee, validate, and adjust the approach or problem as required.

Course topics

General course topics

- unconstrained and constrained optimization: optimality conditions
- convex optimization and duality
- solution methods: unconstrained and constrained
- modeling and applications in optimization

Prerequisites

- solid foundation in linear algebra and calculus
- programming in MATLAB: prior experience not mandatory, but self-study is expected in the course's early stages

Course objectives

- learn the basic mathematical theory of nonlinear optimization (and convex optimization) and their applications in practice
- learn and implement basic (some advanced) optimization methods
- develop the ability to identify optimization problem types and select appropriate solution methods
- equip yourself with optimization knowledge needed for research and real-world applications

Course information

Course materials: lecture slides and other course material will be posted on Moodle

Grading

- homework (20%)
- two midterm exams (50%)
- project (30%)

(these weights are approximate; we reserve the right to change them later)

refer to the syllabus on the Moodle course website for more information, such as course references, office hours, class policy, exam dates, etc.

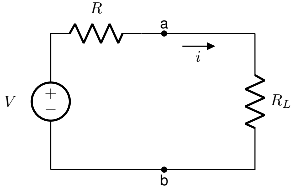
AI tools policy

- unauthorized use of AI tools, like ChatGPT, is treated as plagiarism
- AI is an aid, not a substitute for genuine understanding; reliance solely on AI without understanding can result in penalties
- suspected misuse of AI may lead to oral exams or alternative assessments

Outline

- course introduction
- **optimization examples**
- vectors and matrices

Maximum power transfer



- voltage source: V (in volts)
- line resistor: R (given value)
- objective: Determine R_L to maximize power to it

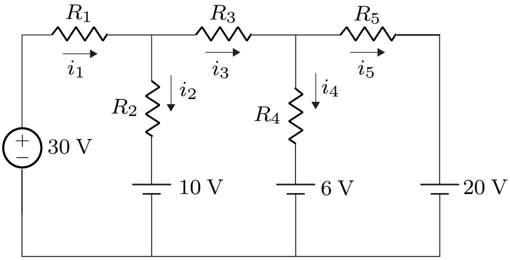
using circuit analysis, the power delivered to R_L is $p(R_L) = i^2 R_L$ and $i = V/(R + R_L)$; hence, we can formulate the problem as

$$\text{maximize} \quad \frac{V^2\, x}{(R + x)^2}$$

with variable $x = R_L$; this is an unconstrained nonlinear program

Battery charging

an electric circuit is designed to use a 30 V source to charge 10 V, 6 V, and 20 V batteries



- physical constraints limit the currents i_1, i_2, i_3, i_4, and i_5 to a maximum of 4 A, 3 A, 3 A, 2 A, and 2 A
- the batteries must not be discharged; that is, the currents i_1, i_2, i_3, i_4, and i_5 must not be negative
- we wish to find the values of the currents $i_1, i_2, \ldots, i_5$ such that the total power transferred to the batteries is maximized

using circuit analysis, the problem can be modeled as the linear program:

$$\begin{array}{ll} \text{maximize} & 10i_2 + 6i_4 + 20i_5 \\ \text{subject to} & i_1 = i_2 + i_3 \\ & i_3 = i_4 + i_5 \\ & i_1 \leq 4 \\ & i_2 \leq 3 \\ & i_3 \leq 3 \\ & i_4 \leq 2 \\ & i_5 \leq 2 \\ & i_1, i_2, i_3, i_4, i_5 \geq 0 \end{array}$$

once the currents are found, we can find the resistors $R_1, \ldots, R_5$ that draw such currents using Ohm's and Kirchhoff's laws

Concrete mixture

Concrete type 1:

- Cost: \$5/lb
- Cement: 30%
- Gravel: 40%
- Sand: 30%

Concrete type 2:

- Cost: \$1/lb
- Cement: 10%
- Gravel: 20%
- Sand: 70%

Formulate a mixture with **at least**: 5 kg of Cement, 3 kg of Gravel, 4 kg of Sand, while minimizing cost

Problem formulation:

$$\begin{array}{ll} \text{minimize} & 5x_1 + x_2 \\ \text{subject to} & 0.3x_1 + 0.1x_2 \geq 5 \\ & 0.4x_1 + 0.2x_2 \geq 3 \\ & 0.3x_1 + 0.7x_2 \geq 4 \\ & x_1 \geq 0, x_2 \geq 0 \end{array}$$

variables x_i represent the weight of concrete i that we want to buy

Knapsack problem

Description

- given n items, each with a weight and value
- determine the number of each item to include, maximizing total value while ensuring the total weight doesn't exceed a set limit

Investment example:

- Aim: invest among n opportunities
- Budget: at most d dollars
- ith investment:
    - cost: c_i dollars
    - expected profit: p_i
    - available units: b_i

how many items of each type should be bought to maximize the expected profit?

problem can be formulated as

$$
\begin{array}{ll}
\text{maximize} & \sum_{i=1}^{n} p_i x_i \\
\text{subject to} & \sum_{i=1}^{n} c_i x_i \le d, \text{ (total cost} \le \text{available amount)}, \\
 & x_i \in \{0, 1, 2, \ldots, b_i\}, \quad i = 1, \ldots, n
\end{array}
$$

- the problem is an integer optimization problem since the variables are restricted to be integers
- specifically, it is an integer linear program since the objective and constraints are “linear” and the variables are integer

Facility placement: Fermat-Weber problem

consider locations of some facilities represented by points:
$(a_1, b_1), \ldots, (a_m, b_m)$ in 2D space

- Goal: Determine the optimal location, $\boldsymbol{x} = (x_1, x_2)$, of a distribution center to minimize the total daily distance.
- distance between the center, $\boldsymbol{x}$, and a facility, (a_i, b_i), is:

$$d_i = \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}$$

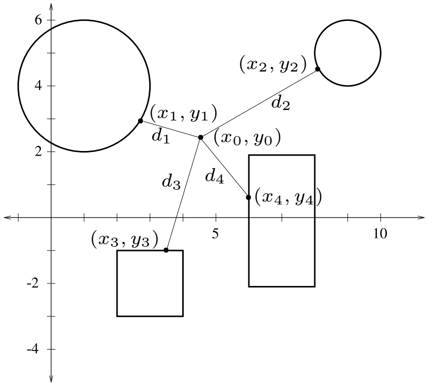
the problem can be formulated as

$$\text{minimize} \quad \sum_{i=1}^{m} w_i \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}$$

- w_i is the weight associated with distance d_i (e.g., higher values of w_i for areas with more traffic)
- this problem is known as the *Fermat-Weber problem*

Electrical wires connections

four buildings are to be connected by electrical wires



- central joining point: (x_0, y_0)
- each building i connects at (x_i, y_i) with wire length d_i
- Objective: find the positions (x_i, y_i) that minimize the total length of wires used

- building 1 (circular): center $(1, 4)$, radius 2
- building 2 (circular): center $(9, 5)$, radius 1
- building 3 (square): center $(3, -2)$, side length 2
- building 4 (rectangle): center $(7, 0)$, height 4, width 2

Problem Formulation:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{4} \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \\ \text{subject to} & (x_1 - 1)^2 + (y_1 - 4)^2 \leq 4 \\ & (x_2 - 9)^2 + (y_2 - 5)^2 \leq 1 \\ & 2 \leq x_3 \leq 4 \\ & -3 \leq y_3 \leq -1 \\ & 6 \leq x_4 \leq 8 \\ & -2 \leq y_4 \leq 2 \end{array}$$

with variables (x_i, y_i) ($i = 0, 1, \ldots, 4$)

Outline

- course introduction
- optimization examples
- **vectors and matrices**

Vectors

column vector of size n (or n-vector):

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

- a_i: ith entry (also termed as element, coefficient, or component)
- alternate notation: $\boldsymbol{a} = (a_1, \dots, a_n)$
- set of real n-vectors: $\mathbb{R}^n$
- transpose results in a row vector: $\boldsymbol{a}^T = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$

Special vectors (dimension determined from context)

- *zero vector:* $\mathbf{0} = (0, 0, \dots, 0)$
- *one vector:* $\mathbf{1} = (1, 1, \dots, 1)$
- *unit vectors:* $\boldsymbol{a} = \boldsymbol{e}_i$ has entry $a_i = 1$; all other entries are equal to zero

Vectors operations and properties

- *scalar vector multiplication*: $\boldsymbol{x} = \alpha \boldsymbol{a}$, where $x_i = \alpha a_i$
- *vector addition*: $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{b}$, where $x_i = a_i + b_i$

Vector addition properties

- *commutative:* $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$
- *associative:* $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$

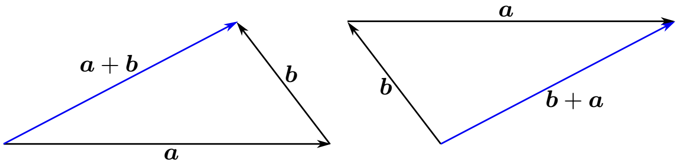
Scalar-vector multiplication properties

- *distributive:* for any real scalars α and β

$$\alpha(\boldsymbol{a} + \boldsymbol{b}) = \alpha \boldsymbol{a} + \alpha \boldsymbol{b}$$
$$(\alpha + \beta)\boldsymbol{a} = \alpha \boldsymbol{a} + \beta \boldsymbol{a}$$

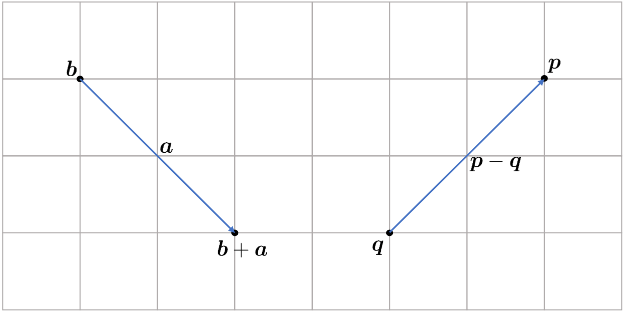
- *associative:* $\alpha(\beta \boldsymbol{a}) = (\alpha\beta)\boldsymbol{a}$, which we write as $\alpha\beta\boldsymbol{a}$

Geometric interpretation: displacement



- *Left:* the displacement $\boldsymbol{a}+\boldsymbol{b}$
- *Right:* the displacement $\boldsymbol{b}+\boldsymbol{a}$

Geometric interpretation: position



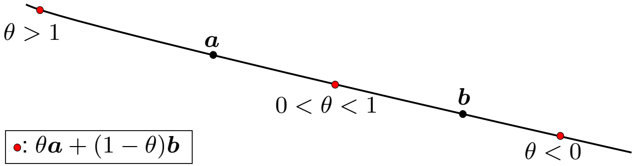
- the point $\boldsymbol{b} + \boldsymbol{a}$ is the position of the point represented by $\boldsymbol{b}$ displaced by the displacement represented by $\boldsymbol{a}$
- the vector $\boldsymbol{p} - \boldsymbol{q}$ represents the displacement from the point represented by $\boldsymbol{q}$ to the point represented by $\boldsymbol{p}$

Line segment

any point on the line passing through distinct $\boldsymbol{a}$ and $\boldsymbol{b}$ can be written as

$$\boldsymbol{c} = \theta \boldsymbol{a} + (1 - \theta)\boldsymbol{b}$$

- θ is a scalar
- for $0 \leq \theta \leq 1$, point $\boldsymbol{c}$ lie on the segment between $\boldsymbol{a}$ and $\boldsymbol{b}$



Euclidean inner product and norm

(euclidean) inner product of two n-vectors:

$$\boldsymbol{a}^T\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + \cdots + a_n b_n$$

$\boldsymbol{a}$ and $\boldsymbol{b}$ are said to be *orthogonal* if $\boldsymbol{a}^T\boldsymbol{b} = 0$

Properties

- *nonnegativity:* $\boldsymbol{a}^T\boldsymbol{a} \geq 0$ and $\boldsymbol{a}^T\boldsymbol{a} = 0$ if and only if $\boldsymbol{a} = \mathbf{0}$
- *symmetry:* $\boldsymbol{a}^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{a}$
- *additivity:* $(\boldsymbol{a} + \boldsymbol{b})^T\boldsymbol{c} = \boldsymbol{a}^T\boldsymbol{c} + \boldsymbol{b}^T\boldsymbol{c}$
- *homogeneity:* $(\alpha\boldsymbol{a})^T\boldsymbol{b} = \alpha(\boldsymbol{a}^T\boldsymbol{b})$

Euclidean norm

$$\|\boldsymbol{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} = \sqrt{\boldsymbol{a}^T\boldsymbol{a}}$$

Cauchy-Schwarz inequality

$$|\boldsymbol{a}^T\boldsymbol{b}| \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\|$$

Inner product examples

- *unit vector:* inner product $\boldsymbol{e}_i^T\boldsymbol{a} = a_i$ picks ith element of $\boldsymbol{a}$
- *sum:* $\mathbf{1}^T\boldsymbol{a} = a_1 + a_2 + \cdots + a_n$
- *average:* $(\mathbf{1}/n)^T\boldsymbol{a} = \frac{a_1+a_2+\cdots+a_n}{n}$
- *sum of squares:* $\boldsymbol{a}^T\boldsymbol{a} = a_1^2 + a_2^2 + \cdots + a_n^2 = \|\boldsymbol{a}\|^2$

Norm of block vectors

let $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{c}$, be any real n-vectors and define the block vector:

$$\boldsymbol{d} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix}$$

- the euclidean norm squared of block vector $\boldsymbol{d}$ is

$$\|\boldsymbol{d}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2$$

- the norm of $\boldsymbol{d} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ can be expressed as:

$$\|\boldsymbol{d}\| = \sqrt{\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2} = \big\|\big(\|\boldsymbol{a}\|, \|\boldsymbol{b}\|, \|\boldsymbol{c}\|\big)\big\|$$

Matrices

matrix of size (dimension) $m \times n$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- a_{ij} is the i, j element (or entry or coefficient)
- set of real $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$
- *transpose* of A: A^T is an $n \times m$ matrix with i, j entry equal to a_{ji}
- a matrix is square if $m = n$

Special matrices:

- *zero matrix:* $A = 0$ is a matrix with zeros entries $a_{ij} = 0$
- *diagonal matrix:* square matrix $A = \text{diag}(a_1, \ldots, a_n)$ with diagonal entries $a_{ii} = a_i$ and off-diagonal $a_{ij} = 0$
- *identity matrix:* diagonal matrix $A = I$, $m = n$ with diagonal entries $a_{ii} = 1$ and $a_{ij} = 0$
- *symmetric matrix:* square matrix A with $A = A^T$

Columns and rows of a matrix

an $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- $\boldsymbol{a}_j = (a_{1j}, \ldots, a_{mj})$ is the jth column of A
- $\hat{\boldsymbol{a}}_i^T = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix}$ is the ith row of A
- the matrix A can represented in terms of its columns or row as

$$A = \begin{bmatrix} \boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \cdots \ \boldsymbol{a}_n \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} \hat{\boldsymbol{a}}_1^T \\ \hat{\boldsymbol{a}}_2^T \\ \vdots \\ \hat{\boldsymbol{a}}_m^T \end{bmatrix}$$

Block matrices

matrices can be partitioned into submatrices; for example,

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

denote a matrix where B, C, D, and E are matrices themselves

- B, C, D, and E are termed as *blocks* or *submatrices*
- blocks are identified by their row and column indices, for example, C is the $(1, 2)$ block of A
- block matrices must have compatible dimensions

Example

$$B = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \quad C = [-1], \quad D = \begin{bmatrix} 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}, \quad E = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

we have

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 2 & 2 & 2 & 4 \\ 3 & 3 & 3 & 4 \end{bmatrix}$$

Matrix operations

- *scalar matrix multiplication*: $X = \alpha A$, $x_{ij} = \alpha a_{ij}$
- *matrix addition*: $X = A + B$, $x_{ij} = a_{ij} + b_{ij}$

Matrix-vector product: multiplication of a matrix $A \in \mathbb{R}^{m\times n}$ with a vector of compatible size $\boldsymbol{x} \in \mathbb{R}^n$:

$$\boldsymbol{y} = A\boldsymbol{x}, \quad y_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, \dots, m$$

Matrix-matrix product: product of a matrix with a matrix $A \in \mathbb{R}^{m\times p}$ of compatible size $B \in \mathbb{R}^{p\times n}$, $C = AB$:

$$c_{ij} = \sum_{\ell=1}^{p} a_{i\ell}b_{\ell j} = a_{i1}b_{1j} + \cdots + a_{ip}b_{pj},$$

$i = 1, \dots, m, \ j = 1, \dots, n$

Matrix operations properties

Matrix addition properties

- *commutativity:* $A + B = B + A$
- *associativity:* $(A + B) + C = A + (B + C)$
- *addition with zero matrix:* $A + 0 = 0 + A = A$
- *transpose of sum:* $(A + B)^T = A^T + B^T$

Matrix-vector multiplication properties

- *distributive:* $A(\boldsymbol{u} + \boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v}$ and $(A + B)\boldsymbol{u} = A\boldsymbol{u} + B\boldsymbol{u}$ where $\boldsymbol{u}, \boldsymbol{v}$ are vectors and A, B are matrices
- *homogeneity:* $(\alpha A)\boldsymbol{u} = \alpha(A\boldsymbol{u}) = A(\alpha\boldsymbol{u})$, which we write as $\alpha A\boldsymbol{u}$

Matrix multiplication properties

- *associativity:* $(AB)C = A(BC)$, which we write it as ABC
- *distributivity with addition*

$$A(B+C) = AB + AC, \quad (A+B)C = AC + BC$$

- *transpose of product:* $(AB)^T = B^T A^T$
- for scalars α and β, we have

$$\begin{aligned} \alpha(AB) &= (\alpha A)B = A(\alpha B) \\ (\alpha A)^T &= \alpha A^T \\ (\alpha\beta)A &= \alpha(\beta A) \\ (\alpha+\beta)A &= \alpha A + \beta A \end{aligned}$$

Matrix trace

the *trace* of an $n \times n$ matrix A is the sum of its diagonal elements:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}$$

some properties of the trace are:

- $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$
- if A is an $m \times n$ matrix and B is an $n \times m$ matrix, then

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$

Functions

- $f : \mathcal{X} \to \mathcal{Y}$ to denote a *function* f that maps an element from the set $\mathcal{X}$ into the set $\mathcal{Y}$
- $f : \mathbb{R}^n \to \mathbb{R}^m$ means that f maps a real n-vector to a real m-vector:

$$f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$$

  where the entry $f_i : \mathbb{R}^n \to \mathbb{R}$ is itself a scalar-valued function of $\boldsymbol{x}$

Function domain: the *domain* of function f, symbolized by $\operatorname{dom} f \subseteq \mathcal{X}$, represents the set of points where f is defined and finite; for example, the functions

$$f_1(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ \infty & \text{otherwise} \end{cases}, \quad f_2(x) = \begin{cases} 1/x & \text{if } x > 0 \\ \infty & \text{otherwise} \end{cases}$$

are different since they have different domains

Examples

defined everywhere ($\mathrm{dom}\, f = \mathbb{R}^n$)

- $f : \mathbb{R} \to \mathbb{R}$: $f(x) = x^2 + x + 1$ maps a scalar x to a scalar $f(x)$
- $f : \mathbb{R}^3 \to \mathbb{R}$: $f(x_1, x_2, x_3) = x_1 + x_2 + x_3^2$
- $f : \mathbb{R}^n \to \mathbb{R}^m$: $f(\boldsymbol{x}) = A\boldsymbol{x}$ where $\boldsymbol{x} \in \mathbb{R}^n$ and A is an $m \times n$ matrix
- $f : \mathbb{R}^2 \to \mathbb{R}^3$: $f(x_1, x_2) = (x_1, x_2, x_1 + x_2^2)$

undefined everywhere

- $f(x) = \log x$ ($f : \mathbb{R} \to \mathbb{R}$) is a function that takes a real number and outputs a real number and it is valid only for $x > 0$, hence $\mathrm{dom}\, f = \{x \mid x > 0\}$
- $f(x_1, x_2) = x_1/x_1 + x_2$ ($f : \mathbb{R}^2 \to \mathbb{R}$) with $\mathrm{dom}\, f = \{\boldsymbol{x} \mid x_1 + x_2 \neq 0\}$ where $\boldsymbol{x} = (x_1, x_2)$

Linear functions

a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if it satisfies the *superposition* property:

$$f\big(\alpha \boldsymbol{x} + \beta \boldsymbol{y}\big) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})$$

for any an n-vectors $\boldsymbol{x}, \boldsymbol{y}$ and any scalars α, β

- a linear function $f : \mathbb{R}^n \to \mathbb{R}$ can always be expressed as: $f(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}$ for some n-vector $\boldsymbol{a}$.
- similarly, a linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented as: $f(\boldsymbol{x}) = A\boldsymbol{x}$ for some $m \times n$ matrix A
- to see this, using the linear property of f, we have:

$$f(\boldsymbol{x}) = f(x_1 \boldsymbol{e}_1 + \cdots + x_n \boldsymbol{e}_n) = x_1 f(\boldsymbol{e}_1) + \cdots + x_n f(\boldsymbol{e}_n) = A\boldsymbol{x}$$

  where the matrix A has columns:

$$A = [f(\boldsymbol{e}_1) \ \cdots \ f(\boldsymbol{e}_n)]$$

Affine functions

a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *affine* if it can be expressed as

$$f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$$

for some $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$

- an affine function f satisfies the superposition

  $$f(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})$$

  for any affine combination $\alpha + \beta = 1$
- $f : \mathbb{R}^n \to \mathbb{R}$ is affine if we can write it as

  $$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$$

  for some n-vector $\boldsymbol{a}$ and scalar b (linear function plus a constant)

Quadratic functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is *quadratic* if it can be expressed as

$$f(\boldsymbol{x}) = \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{r} + c$$

where

- Q is an $n \times n$ matrix
- $\boldsymbol{r}$ is an n-vector
- c is a scalar

Quadratic Form:

- a quadratic form is a special case: $\boldsymbol{x}^T Q \boldsymbol{x}$ where Q is symmetric
- we can always assume Q is symmetric because:

$$\boldsymbol{x}^T Q \boldsymbol{x} = (1/2)\boldsymbol{x}^T (Q + Q^T)\boldsymbol{x}$$

hence, $\boldsymbol{x}^T Q \boldsymbol{x} = \boldsymbol{x}^T P \boldsymbol{x}$ with $P = \frac{1}{2}(Q + Q^T)$ being symmetric

References

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares,* Cambridge University Press, 2018.
- E. KP. Chong and S. H. Zak. *An Introduction to Optimization,* John Wiley & Sons, 2013.
- S. Boyd and L. Vandenberghe. *Convex Optimization,* Cambridge University Press, 2004.
- G. C. Calafiore and L. El Ghaoui. *Optimization Models,* Cambridge University Press, 2014.
- I. Griva and S. G. Nash and A. Sofer. *Linear and Nonlinear Optimization,* SIAM, 2009.