Optimization: Examples

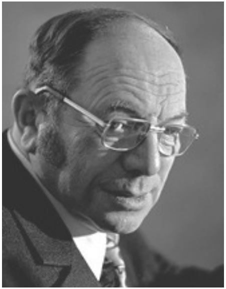
Peter Hajnal

Bolyai Istitute, University of Szeged, Hungary

2024 fall

The roots of optimization

• Optimization is the meeting point of various branches of mathematics.
• There are several applications in natural sciences, economics, and computer science, which are based on optimization results.
• Breakthroughs in the field of optimization have received significant scientific recognition due to their wide-ranging applications.



L.V. Kantorovich
(1912-1986)

• Kantorovich (Soviet mathematician) was awarded the Nobel Prize in Economics in 1975 for his contributions to optimal resource allocation.

The fundamental problem of optimization

• Let $c\colon \text{dom}\,(c) \subseteq \mathbb{R}^n \to \mathbb{R}$ be a given real-valued function of n variables ($n \in \mathbb{N}$), which we call the **objective function**.

• We use $x = (x_1, \ldots, x_n)^\mathsf{T}$ to denote a general element of the domain dom (c). NOTE! Throughout the semester, we work exclusively with column vectors.

• The fundamental task of optimization is to minimize the function c subject to predefined constraints. Hereinafter, we use the following abbreviated notation:

Minimize	$c(x)$
subject to	$x \in \mathcal{F}$

where $x \in \text{dom}\,(c)$ and $\mathcal{F} \subseteq \mathbb{R}^n$ represents the domain determined by the constraints.

Constraints

• Constraints can have diverse origins: they can be formalized as mathematical conditions or stem from physical, economic constraints.

• There are various ways to specify constraints. Here we mention two methods:

• **Implicit constraints.** An algorithm/oracle/subroutine decides whether a given $x \in \mathbb{R}^n$ satisfies the constraints:

$$x \Rightarrow \boxed{\text{ ALGORITHM }} \Rightarrow \text{good/bad } (\in \mathcal{F} / \notin \mathcal{F})$$

• **Explicit constraints.** The condition is described by a finite set of equations and/or inequalities:

$$\begin{cases} f_i(x) \le 0, & i \in [k] := \{1, 2, \ldots, k\}, \\ g_j(x) = 0, & j \in [\ell], \end{cases} \tag{1.1}$$

where f_i ($i \in [k]$) and g_j ($j \in [\ell]$) are also real-valued functions of n variables. In this case, $\mathcal{F}$ is the solution set of the above system.

Domain of Definition, Feasible Solutions

• The **domain of definition of the optimization problem** is the intersection of the domains of definition of the objective function and the constraints. For explicit constraints (1.1), this is

$$\mathcal{D} := \mathrm{dom}\,(c) \cap \left(\bigcap_{i=1}^{k} \mathrm{dom}\,(f_i) \right) \cap \left(\bigcap_{j=1}^{\ell} \mathrm{dom}\,(g_j) \right),$$

which contains those x values for which both the objective function and the constraints are defined.

• **Feasible solutions** refer to those $x \in \mathcal{D}$ vectors that satisfy the criteria, i.e., $x \in \mathcal{F}$. The set of these x values is denoted by $\mathcal{L}$. Thus,

$$\mathcal{L} := \mathcal{D} \cap \mathcal{F},$$

which, for explicit constraints (1.1), becomes

$$\mathcal{L} = \{x \in \mathcal{D} : f_i(x) \leq 0,\ g_j(x) = 0,\ i \in [k], j \in [\ell]\}.$$

Optimal Value, Optimal Location

- The **optimal value** associated with the problem is

$$p^* = \inf_{x\in\mathcal{L}} c(x) \in \mathbb{R} \cup \{-\infty\} \cup \{\infty\},$$

where the infimum is ∞ if the set of feasible solutions is empty, and $-\infty$ if the objective function c takes arbitrarily small values on the feasible solutions.

- An $x^* \in \mathcal{L}$ vector is called an **optimal location** if the objective function takes the optimal value there:

$$c(x^*) = p^*.$$

- The vector x_ℓ is a **local optimum** if for every point in its neighborhood, c is at least as large as at x_ℓ:

$$\exists\, \varepsilon > 0,\ \forall\, x : \|x - x_\ell\|_2 < \varepsilon \quad \text{implies} \quad c(x_\ell) \le c(x),$$

where $\|\ \|_2$ is the Euclidean norm defined on $\mathbb{R}^n$,
$\|\ \|_2 \colon \mathbb{R}^n \to [0,\infty)$, $y \mapsto \|y\|_2 := \sqrt{y_1^2 + \cdots + y_n^2}$.

Approximate Solutions

- We say that x_0 is an **ε-close solution** ($\varepsilon > 0$) if

$$c(x_0) \le p^* + \varepsilon.$$

- An x_0 vector is an **ε-approximate solution** ($\varepsilon > 0$) if

$$(0 < p^* \le) c(x_0) \le (1+\varepsilon)p^*.$$

- In the definition of ε-approximate solution, we include the condition that the optimal value is positive, $p^* > 0$. In problems with negative optimal values, we need to impose the condition $c(x_0) \le (1-\varepsilon)p^*$.

Introductory example I

Minimize	$\dfrac{1}{x}$
subject to	$x \geq 0.$

• The objective function is $c(x) = x^{-1}$, a reciprocal function defined on dom $(c) = \mathbb{R} \setminus \{0\}$. The explicit constraint is a single inequality, $f_1(x) = -x \leq 0$, hence $k = 1$, $\ell = 0$. Since dom $(f_1) = \mathbb{R}$, the domain of the problem is

$$\mathcal{D} = \mathbb{R} \setminus \{0\}.$$

It can be seen that the set of feasible solutions is the open interval

$$\mathcal{L} = \mathbb{R}_{>0} := (0, \infty) =]0, \infty[$$

Introductory example I (continued)

- Since the infimum of the function values taken on the feasible solutions $x \in \mathcal{L}$ is zero, we have

$$p^* = 0.$$

- However, there does not exist an $x \in \mathcal{L}$ at which c achieves this value, hence there is no optimal location.
- It is worth noting that for any $\varepsilon > 0$, the numbers $x_0 \geq \varepsilon^{-1}$ are ε-close solutions.
- However, there are no ε-approximate solutions.

Introductory example II

Minimize	$x \log x$

- There are no constraints imposed this time, i.e., $\mathcal{F} = \mathbb{R}$.
- In this case, we consider a *global optimization problem*.
- The objective function is differentiable over its domain.

Optimization is a standard part of calculus courses: The single-variable objective function $c(x) = x \log x$ is defined on dom $(c) = \mathbb{R}_{>0}$. Since there are no constraints, we have

$$\mathcal{L} = \mathcal{D} \cap \mathcal{F} = \mathcal{D} = \text{dom } (c) = \mathbb{R}_{>0}.$$

The optimal value can be determined, for example, by elementary methods of calculus.

- As a result, we obtain

$$p^* = -\frac{1}{e}, \qquad x^* = \frac{1}{e}.$$

Transforming Our Optimization Problem

- Once we understand an optimization problem, there are several ways to formalize it.

- In other words, a formalized optimization problem can often be reformulated to describe the same problem.

- Some transformations can lead to problems that are equivalent to the original one. However, slightly different forms of the same problem can have significant differences.

- By equivalent transformation, we mean a formal rearrangement such that the optimal value/location of one problem can be easily determined based on the optimal value/location of the other problem.

- Now (without claiming completeness), let's mention a few possibilities.

Transformations: Exchange of Min/Max

Maximize	$c(x)$
subject to	$x \in \mathcal{F}$

$\equiv$

Minimize	$-c(x)$
subject to	$x \in \mathcal{F}$

• An optimal point of the minimization problem is also an optimal point of the maximization problem.

• If we know the optimal value of the minimization problem, its negative will be the optimal value of the maximization problem.

Transformations: Equivalent Reformulation of Constraints

- We have already seen equivalent transformations of formulas/inequalities/equations in high school.
- We can naturally use these techniques for our constraints.

$$\frac{x_1}{x_2^2+1} \leq 0 \quad \iff \quad x_1 \leq 0.$$

$$(x_1+x_2)^2 \leq 0 \quad \iff \quad (x_1+x_2)^2 = 0 \quad \iff \quad x_1+x_2 = 0.$$

Transformations: Substituting Inequalities with Sign Conditions

Minimize	$c(x)$
subject to	$f_i(x) \leq 0$
	$g_i(x) = 0$

$\equiv$

Minimize	$c(x)$
subject to	$f_i(x) + s_i = 0$
	$g_i(x) = 0$
	$s_i \geq 0$

- The introduced variables s_i are called *slack variables*.

Transformations: Elimination of Linear Equalities

- The solution of the linear equality system $Ax = b$ is basic algebra material. The general solution can be written in the form $x = x_0 + Fy$, where x_0 is an arbitrary solution, and the columns of F generate the linear subspace described by the equation $Ax = 0$. If F has r columns, then $y \in \mathbb{R}^r$.
- Determining x_0 and F can be done efficiently.

Minimize	$c(x)$
subject to	$f_i(x) \leq 0$
	$Ax = b$

$\equiv$

Minimize	$c(x_0 + Fy)$
subject to	$f_i(x_0 + Fy) \leq 0$

Transformations: Substituting the Objective Function with a Monotonic Function

- Let $m : \mathbb{R} \to \mathbb{R}$ be a strictly monotonic function on the range c (the set of values taken by the objective function).

- Then

Minimize	$c(x)$
subject to	$x \in \mathcal{F}$

$\equiv$

Minimize	$m(c(x))$
subject to	$x \in \mathcal{F}$

Transformations: Substituting the Objective Function with a Monotonic Function: Example

Example

Minimize	$\|x\|_2$
subject to	$x \in \mathcal{F}$

$\equiv$

Minimize	$\|x\|_2^2$
subject to	$x \in \mathcal{F}$

- In this case, range $c = \mathbb{R}_{\geq 0}$, where $m(x) = x^2$ is a monotonic function.

- We made minimal changes, but the new objective function is differentiable. We will see that such a *small* advantage can be very significant.

Transformations: Substituting the Objective Function with a Monotonic Function: A More Complex Example

Maximize	$x_1x_2 + x_2x_3 + x_3x_1.$
subject to	$x_1 + x_2 + x_3 = 100,$
	$x_1, x_2, x_3 \geq 0.$

$\equiv$

Minimize	$\sqrt{\frac{x_1^2+x_2^2+x_3^2}{3}}.$
subject to	$\frac{x_1+x_2+x_3}{3} = \frac{100}{3},$
	$x_1, x_2, x_3 \geq 0.$

- The two sets of constraints are obviously equivalent.

Transformations: Substituting the Objective Function with a Monotonic Function: A More Complex Example (Continued)

- The two objective functions are:

$c_1(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_3x_1$, and $c_2(x_1, x_2, x_3) = \sqrt{\frac{x_1^2+x_2^2+x_3^2}{3}}$.

- The relationship is evident (when the constraints are satisfied):

$$3c_2^2 + 2c_1 = (x_1 + x_2 + x_3)^2 = 100^2.$$

- Instead of maximizing c_1, we can minimize $100^2 - 2c_1 = 3c_2^2$. Then we can apply the strictly monotonic function $m(x) = \sqrt{\frac{x}{3}}$ (on the set of non-negative values taken by the new objective function). Thus, we obtain the second form above.

- The advantage of the new form is evident. We need to minimize the square sum given a fixed arithmetic mean. With the inequality between arithmetic and quadratic means, the optimization question can be solved with elementary mathematics.

Break

Textual Examples vs. Our Basic Problem

• Throughout the lecture series, we will often encounter problems where the formalization poses the most challenge.

• In practice, we are given informal problems. We need to communicate with the applicant. We need to understand their language, their way of thinking. Sometimes we need to learn physics, chemistry, biology to see the constrains as equations, inequalities before us.

• Often, mathematical ideas are necessary for formal description.

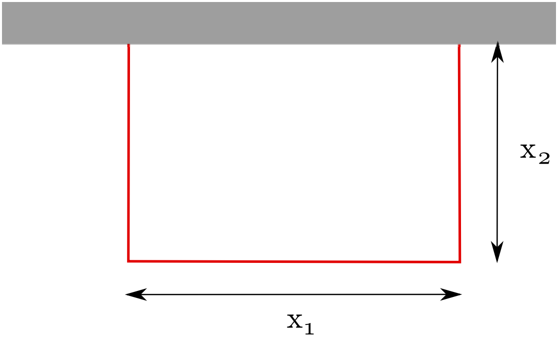
• After a good, fortunate formalization, the optimization algorithm is often ready to go.

• Below, we present some introductory examples of basic concepts and elementary formalization tricks.

Example I

Example

I want to enclose the largest possible rectangular area with a fence 100 m long, with one side being a wall. How do I need to choose the side lengths?



Example I (Continued)

- This problem can be described as the following optimization task:

Maximize	Maximize $x_1 x_2$
subject to	subject to $x_1 + 2x_2 = 100$,
	$x_1, x_2 \geq 0$.

- It is obvious that the objective function $c(x_1, x_2) = x_1 x_2$ is defined over the entire $\mathbb{R}^2$ plane and $f_1(x_1, x_2) = -x_1$, $f_2(x_1, x_2) = -x_2$, $g_1(x_1, x_2) = x_1 + 2x_2 - 100$, thus $\mathcal{D} = \mathbb{R}^2$ and $\mathcal{L} = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$.

- Applying the inequality between arithmetic and geometric means and taking the constraints into account, we get

$$50 = \frac{x_1 + 2x_2}{2} \geq \sqrt{x_1(2x_2)} \qquad (\geq 0).$$

Example I (Continued)

- Squaring both sides, we obtain

$$2500 \geq 2x_1x_2 = 2c(x_1, x_2),$$

which means the objective function is bounded from above: $c(x_1, x_2) \leq 1250$.

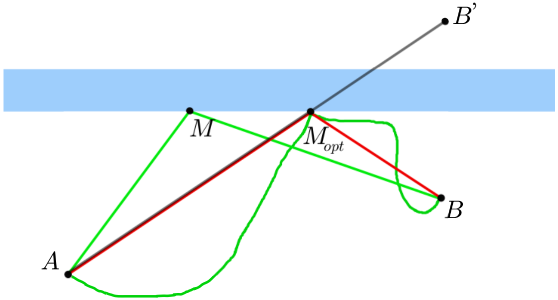
- Since $x^* = (50, 25) \in \mathcal{L}$ is a feasible solution where c reaches this bound, thus $p^* = 1250$.

- Analyzing the case of equality in the inequality between arithmetic and geometric means is beneficial. Based on this, $(50, 25)$ is not just one optimal point. $(50, 25)$ is the ONLY optimal point.

Example II

Example

Suppose a horse rider wants to reach from point A on one side of a river to point B on the same side, while also watering the horse. What is the shortest such route?



Example II: The Solution

• We might have encountered this problem in geometry class in elementary school.

• Let t be the line representing the bank on which our rider is.

• Let M be the watering point. For any possible solution, by keeping the segment AM of the rider's path and then reflecting the later segment onto t, we obtain a path leading from A to B' (the mirror image of point B on line t). The length of the modified path is the length/cost of the arbitrarily chosen possible solution.

• Among the modified paths, obviously traversing segment AB' is optimal. That is, the intersection of t and segment AB' is the *optimal watering point*.

• Here, coming straight from A and then moving straight to B will be the shortest path for the horse in the problem.

Example II: Formalization

- The formalized optimization problem can be written in several ways.
- One option is to introduce rectangular coordinates (e.g., the bank from the side of the horse as the x-axis).
- Our task is to find the point $M(x^*, 0)$ where the sum of the lengths of segments AM and MB is minimal.
- Because it is obvious that if the horse does not move in a straight line between points A and M or between M and B, then its path is not optimal.
- Therefore, given points $A(a_1, a_2)$ and $B(b_1, b_2)$, the problem can be formulated as follows:

Minimize	$\sqrt{(a_1 - x)^2 + a_2^2} + \sqrt{(b_1 - x)^2 + b_2^2}.$

Example III

Example

What is the minimal distance from the origin to any point on the plane defined by the equation $x_1 + x_2 + x_3 = 100$?

- Notice that instead of the distance, we can also consider its $\sqrt{\frac{1}{3}}$ times.

- The formalization (similar to a previous example) can be as follows:

Minimize	$\sqrt{\frac{x_1^2+x_2^2+x_3^2}{3}}$
subject to	$\frac{x_1+x_2+x_3}{3} = \frac{100}{3}$
	$x_1, x_2, x_3 > 0.$

Example III (continued)

We use the inequality between the arithmetic and quadratic means to estimate the original optimal value:

$$\sqrt{\frac{1}{3}}c(x) = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2}{3}} \geq \frac{x_1 + x_2 + x_3}{3} = \frac{100}{3}.$$

- So,

$$p^* \geq \sqrt{3}\frac{100}{3}.$$

- Moreover, the constraint is achievable if $x_1 = x_2 = x_3 = \frac{100}{3}$ (and only then). So the single optimal point of the problem is

$$x^* = \left(\frac{100}{3}, \frac{100}{3}, \frac{100}{3}\right).$$

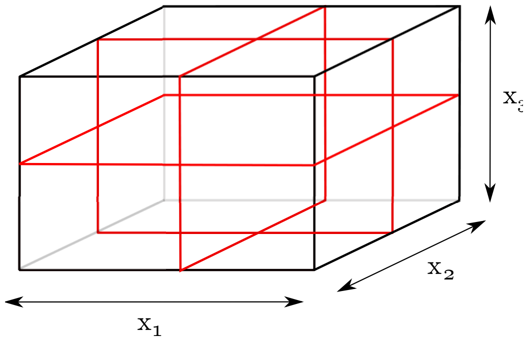
- Furthermore, its optimal value is

$$p^* = c(x^*) = \sqrt{3}\frac{100}{3}.$$

Example IV

Example

What is the largest surface area (in the shape of a rectangular prism) that can be bound by a 400 cm long string (in the manner shown in the accompanying figure)?



Example IV: Formalization

The formalization is obvious:

Maximize	$2x_1x_2 + 2x_2x_3 + 2x_3x_1$
subject to	$4x_1 + 4x_2 + 4x_3 = 400$
	$x_1, x_2, x_3 > 0.$

Based on previous considerations, this is an equivalent problem to the previous one. Rethinking the equivalence leads to the conclusion that the optimal point is the same as in the previous problem.

$$x^* = \left(\frac{100}{3}, \frac{100}{3}, \frac{100}{3}\right).$$

From this, the optimal value of the original problem is

$$p^* = c(x^*) = \frac{20\,000}{3}.$$

Example V

Example

Among the 3-part simple graphs with 100 vertices, which ones have the maximum number of edges?

- With this problem, we encountered the Turán theorem in the Combinatorics course. If we denote the sizes of the three parts as x_1, x_2, and x_3 and observe that we are looking for the maximum among complete 3-partite graphs, then the problem is as follows:

Maximize	$x_1x_2 + x_2x_3 + x_3x_1$
subject to	$x_1 + x_2 + x_3 = 100$
	$x_1, x_2, x_3 > 0$
	$x_1, x_2, x_3 \in \mathbb{Z}$.

Example V (continued)

- Since $\mathcal{F}$ is a non-empty finite set, there must be an optimal point x^* among the possible solutions, with an optimal value $p^* \in \mathbb{R}$.

- It is also evident from the condition that if (x_1, x_2, x_3) is a possible solution, then, for example, $(x_1 - 1,\, x_2 + 1, x_3)$ is also a possible solution (assuming $x_1 > 1$).

- If $x_1 \geq x_2 + 2$, then the objective function is greater in the latter case than in the former.

- This also means that at the optimal point, $|x_i^* - x_j^*| \leq 1$ for all $i, j \in \{1, 2, 3\}$.

- There are three such possible solutions where the objective function takes a common value. Thus, the optimal points are $(34, 33, 33)$, $(33, 34, 33)$, $(33, 33, 34)$.

- The optimal value is $p^* = 3333$.

Example V: Remarks

- The above problem is NOT equivalent to any of our previous examples.
- This is indicated by the difference in the sets of optimal points and possible solutions, which is due to the *essential* difference in conditions: now only integer-coordinate points are considered.
- In the *continuous* problem (previous two examples), the optimal value is $3\,333\frac{1}{3}$, which is larger than the current one.
- This is natural: in the continuous *competition*, there are more *participants*, and the maximum value is at least as much as when only integer-coordinate *competitors* are involved.
- Interestingly, but perhaps the discrete problem seems easier at first glance: We need to find the optimal one(s) among finite possibilities. Yet, it seems like more ideas are needed for the discrete case.

Break

Example VI

Example

Let $d, n \in \mathbb{N}$ and $\ell_1(x), \ldots, \ell_n(x)\colon \mathbb{R}^d \to \mathbb{R}$ be given linear functions. Determine the minimum of the function $c(x) = \max_{1 \le i \le n} \ell_i(x)$.

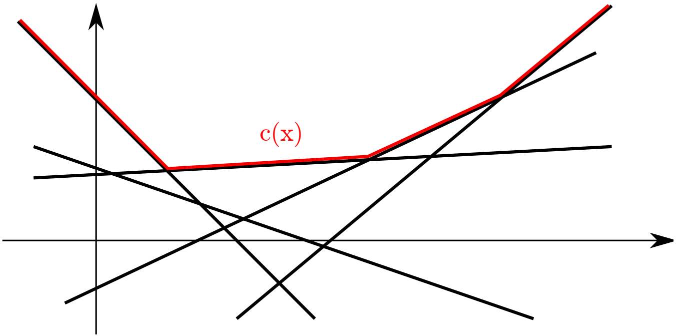
So the formalized problem is:

Minimize	$c(x)$

The problem is global (there are no constraints).

Example VI: Figure

• The accompanying figure illustrates a general configuration for the case of $d = 1$, $n = 5$.



• Even this specific choice gives a faithful representation of the objective function. The maximum of linear functions c is *piecewise* linear. For $d > 1$, this means that the domain of c can be divided into connected parts on which c is linear.

Example VI: Reformulation

• In the optimization problem equivalent to the previous one, we minimize the number m such that the definition (maximality) of the objective function in that case is built into the constraints.

Minimize	m
subject to	$\ell_1(x) \leq m$,
	$\vdots$
	$\ell_n(x) \leq m$,

where $(x, m) \in \mathbb{R}^d \times \mathbb{R}$.

Example VII: A Class of Problems

- This form of optimization problem may be familiar from the Operations Research course.

Linear Programming, LP

Let $A \in \mathbb{R}^{m \times n}$ be a given matrix, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ fixed vectors, and $x \in \mathbb{R}^n$ the unknown vector. The problem:

Minimize	$c^{\mathsf{T}} x$
subject to	$Ax \preceq b$,

where $Ax \preceq b$ denotes the *componentwise less than or equal* relation between Ax and b in $\mathbb{R}^m$.

Example VII: A Class of Problems

• The LP problem class is extensively discussed in the Operations Research course.

• It is noted that the LP problem is a very central problem. Many algorithms are considered good both in practice and in theory.

• If we formulate a problem as an LP problem, then we are *beyond the hard part*. (Just like in high school, if we reduced our work to solving a quadratic equation during solving equations).

• We finish our work with one of the general LP algorithms (such algorithms are easily accessible with public source code).

Example VIII

Example

Consider the LP problem where the linear objective function ($c^\mathsf{T}x$) is a random variable represented by the vector c. We assume that its expected value, $\overline{c} = \mathbb{E}[c]$, is known, and the constraints are no longer dependent on randomness.

Minimize the expected value of the objective function.

- Based on the linearity of expected value, $\mathbb{E}[c^\mathsf{T}x] = (\mathbb{E}[c])^\mathsf{T}x$.
- An LP problem remains our optimization question:

Minimize	$(\mathbb{E}[c])^\mathsf{T}x$
subject to	$Ax \preceq b$.

Example IX: Chebyshev Center

Definition

Sets of points in the form $\mathcal{P} = \{x \in \mathbb{R}^n : a_i^\top x \leq b, \ i = 1, \dots, k\}$ are called polyhedra. If the polyhedron $\mathcal{P}$ is bounded (i.e., compact: bounded and closed), then it is called a polytope.

Example: Chebyshev Centers of Polytopes

Given $\mathcal{P}$ polytope. What are the innermost points of $\mathcal{P}$?

- In fact, deciding whether $\mathcal{P}$ is empty is also a central problem.
- In our case, we need to measure how deeply a point in $\mathcal{P}$ is inside $\mathcal{P}$.
- There are many solutions/answers. We discuss one solution associated with Chebyshev.

Example IX: Chebyshev Center (continued)

Definition

$B(c, r) = \{x \in \mathbb{R}^n : |x - c|^2 \leq r^2\}$ is the ball centered at c with radius r.

The Chebyshev depth of a point $p \in \mathcal{P}$ is defined as

$$M(p) = \sup\{r : B(p, r) \subset \mathcal{P}\}.$$

c is a Chebyshev center of the polytope $\mathcal{P}$ if

$$M(c) = \sup_{p \in \mathcal{P}} M(p).$$

- Our fundamental problem: Given $\mathcal{P}$, find a Chebyshev center.

Example IX: Chebyshev Center (continued)

• The problem appears nonlinear at first glance. We will need some geometric knowledge.

Definition

A hyperplane in $\mathbb{R}^n$: $H = \{x : a^\mathsf{T} x = b\}$ represents a set of points.

A point r lies exactly on H if $a^\mathsf{T} r - b = 0$.

The signed distance of a point p from H is

$$\frac{a^\mathsf{T}}{|a|} p - \frac{b}{|a|}.$$

The absolute value of the signed distance is the distance, and its sign describes which side of the hyperplane the point lies on.

There are two types of signed distances. The above one is positive in the half-space $\{x : a^\mathsf{T} x > b\}$ and negative in the complementary (open) half-space.

Example IX: Chebyshev Center (continued)

- The Chebyshev center problem is equivalent to the following:

Maximize	r	
subject to	$a_i^{\mathsf{T}} x + \lvert a_i \rvert r \leq b_i,$	$i = 1, \ldots, k$
	$r \geq 0$	

- Our first type of condition is equivalent to $\frac{a_i^{\mathsf{T}}}{|a_i|} x + r \leq \frac{b}{|a_i|}$, i.e.,

$$r \leq \frac{b}{|a_i|} - \frac{a_i^{\mathsf{T}}}{|a_i|} x.$$

- On the right side, there is a signed distance, chosen so that it is positive in the half-spaces where $a_i x < b$.

- The reformulated optimization problem is an LP problem.

Break

Example X: Least Squares Problem

• The following problem and its solution methods are encountered in the field of numerical analysis.

Example: Least Squares Problem

The problem involves minimizing

$$\boxed{\text{Minimize} \qquad \|c - Ax\|}$$

or

$$\boxed{\text{Minimize} \qquad \|c - Ax\|^2}$$

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{k \times n}$ is a real matrix, and $c \in \mathbb{R}^k$.

• This is an unconstrained optimization problem. It is easily handled based on basic linear algebraic and geometric knowledge.

Example X: Least Squares Problem (continuation)

Example

Consider a series of measurements (e.g., measurements taken at different times in an experimental laboratory or measurements from a meteorological station at certain location), where $t_1, t_2, \dots t_N$ and $p_1, p_2, \dots, p_N$ denote the measurement times and the measured parameters, respectively. We are looking for a "low," d-degree polynomial that is a good "hypothesis" for the changes in p.

Example X: Least Squares Problem (continuation)

- Let

$$p(x) = x_d x^d + \cdots + x_1 x + x_0$$

be the polynomial, i.e., for the measured vector $p = (p_1, \ldots, p_N)^\mathsf{T}$, we need to minimize the distance in L_2 from

$$x_d(t_1^d, \ldots, t_N^d)^\mathsf{T} + x_{d-1}(t_1^{d-1}, \ldots, t_N^{d-1})^\mathsf{T} + \ldots$$

to p.

- In the least squares problem, the objective function is quadratic, unlike in LP where it is linear. We will now examine a common generalization of the two problems.

Example XI: Quadratic Programming

Quadratic Programming, QP

Let $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{m\times n}$ be symmetric matrices, and let $b \in \mathbb{R}^n$, $d \in \mathbb{R}^m$ be vectors.

Minimize	$x^\mathsf{T} A x + b^\mathsf{T} x$
subject to	$Cx \preceq d.$

• The constraint system of the QP basic problem is linear. However, the objective function is much more general.

• It is worth noting that the basic QP problem can also be handled (an efficient algorithm is known for it) if the objective function is convex (i.e., A is positive semidefinite). If we bring a problem into this form (convex QP), then we are "ready".

Example XII: Stochastic Linear Programming

• Consider the LP problem in which the linear objective function ($c^{\mathsf{T}}x$) is a vector of random variables c. Assume that its expected value is $\overline{c} = \mathbb{E}[c]$, and its covariance matrix is $\Sigma = \mathbb{E}[(c - \overline{c})(c - \overline{c})^{\mathsf{T}}]$. Furthermore, the constraints do not depend on randomness anymore.

• We have seen that if we only minimize the expected value of the objective function ($\mathbb{E}[c^{\mathsf{T}}x] = (\mathbb{E}[c])^{\mathsf{T}}x$), then we obtain an LP problem.

• However, in this case, we do not take into account the *risk*.

Example XII: Stochastic Linear Programming (continued)

- It is common to consider the following variation:

Stochastic LP

Minimize	$\mathbb{E}[c^{\mathsf{T}}x] + \gamma\, \mathbb{V}ar[c^{\mathsf{T}}x]$
subject to	$Ax \preceq b$
	$Dx = e$

where $\gamma \in \mathbb{R}_{>0}$ is a parameter chosen for the application.

- With simple transformations,

$$\mathbb{V}ar[c^{\mathsf{T}}x] = \mathbb{E}[(c^{\mathsf{T}}x - \mathbb{E}[c^{\mathsf{T}}x])^2] = \mathbb{E}[(c^{\mathsf{T}}x - (\mathbb{E}[c])^{\mathsf{T}}x)^2]$$
$$= \mathbb{E}[((c^{\mathsf{T}} - \mathbb{E}[c]^{\mathsf{T}})x)^2] = x^{\mathsf{T}}\mathbb{E}[(c - \mathbb{E}[c])(c^{\mathsf{T}} - \mathbb{E}[c]^{\mathsf{T}})]x = x^{\mathsf{T}}\Sigma x.$$

- The convex QP form of the question is now apparent.

Example XIII: Distance between Polyhedra

Example

Determine the distance between two given polyhedra in n-dimensional Euclidean space!

• Since the polyhedra arise as solution sets of linear inequalities, each polytope corresponds to an inequality system ($C_1 \in \mathbb{R}^{k_1 \times n}$, $C_2 \in \mathbb{R}^{k_2 \times n}$):

$$\mathcal{P}_1 = \{x_1 \in \mathbb{R}^n : C_1 x_1 \preceq d_1\} \quad \text{and} \quad \mathcal{P}_2 = \{x_2 \in \mathbb{R}^n : C_2 x_2 \preceq d_2\}.$$

• The distance between them is

$$d(\mathcal{P}_1, \mathcal{P}_2) = \inf\{d(x_1, x_2) : x_1 \in \mathcal{P}_1,\ x_2 \in \mathcal{P}_2\},$$

where $d(x_1, x_2) = \|x_1 - x_2\|_2$.

Example XIII: Distance between Polyhedra (continued)

• Considering the squared distance leads to an equivalent optimization problem:

Minimize	$\|x_1 - x_2\|_2^2$
subject to	$C_1 x_1 \preceq d_1,$
	$C_2 x_2 \preceq d_2.$

• To this end, consider the vectors and matrices formed from $d_1 \in \mathbb{R}^{k_1}, d_2 \in \mathbb{R}^{k_2}, x_1, x_2 \in \mathbb{R}^n$ and $C_1, C_2 \in \mathbb{R}^{(k_1+k_2)\times n}$ as follows:

$$d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^{2n}, \quad C = \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} \in \mathbb{R}^{2n\times 2n}$$

where 0 denotes appropriately sized zero matrices. Our constraint system is: $Cx \preceq d$.

Example XIII: Distance between Polyhedra (continued)

• If we further construct the $n \times n$ identity matrix, then

$$A = \begin{pmatrix} I & -I \\ -I & I \end{pmatrix} \in \mathbb{R}^{2n \times 2n}$$

our objective function can be written in the form $x^{\mathsf{T}} A x$.

• Using the above notation, we see that this is a convex QP problem. Based on the above, determining the distance between two polytopes can be efficiently accomplished.

Break

Example XIV: Clique Problem

Example

Given a simple graph G. Determine
$\omega(G) = \max\{|K| : K \subset V(G) \text{ is a clique}\}$.

• Any arbitrary vertex set can be described by its characteristic vector $\chi_U \in \{0,1\}^V \subset \mathbb{R}^V$. The size of U is exactly $1^T\chi_U$, where $1 \in \mathbb{R}^V$ is a vector with all components equal to 1.

• The elements of $\{0,1\}^V$ can be described by the conditions $0 \leq x_v \leq 1$, $x_v \in \mathbb{Z}$ for all v vertices. A 0-1 vector will be the characteristic vector of a clique if at most one of its vertices falls into every pair of disconnected vertices u and v. That is, for all $uv \notin E(G), u \neq v$ vertices, we have $x_u + x_v \leq 1$.

Example XIV: Clique Problem (continued)

- The well-known $\mathcal{N}P$-hard clique problem can be formulated as follows:

Maximize	$1^\mathsf{T} x,$
subject to	$0 \leq x_v \leq 1$, $x_v \in \mathbb{Z}$ for all v vertices
	$x_u + x_v \leq 1$
	for all disconnected vertices $uv \notin E(G), u \neq v$.

Example XV: Integer Programming

• The previous example is a general scheme element: Again, *let's intervene* in the LP basic problem. Now, modify the constraint system. Include in the constraints that *the coordinates of our competitor vectors* must be integers.

Integer Programming, IP

The basic problem:

Minimize	$c^{\mathsf{T}}x$
subject to	$Ax \preceq b$,
	$x \in \mathbb{Z}^d$.

• For LP, the set of possible solutions was a polyhedron. Now it is the intersection of a polyhedron and $\mathbb{Z}^d$ (in the case of polytopes, this is a finite set). This modification may seem *more innocent* than modifying the objective function, yet the resulting problem is difficult. This general form is capable of formalizing $\mathcal{N}P$-complete problems. A general, efficient solution cannot be expected.

Example XVI: Matching Problem

Example

Given a graph G, find the maximum matching among its edges.

- Formalizing the problem:

Maximize	$1^{\mathsf{T}}x$ (where $1^{\mathsf{T}} = (1, \ldots, 1), x \in \mathbb{R}^{E(G)}$)
subject to	x is the characteristic vector of the matching.

- Algebraically formalizing the condition yields the following equivalent optimization problem:

Maximize	$1^{\mathsf{T}}x$ (where $x \in \mathbb{R}^{E(G)}$)
subject to	$0 \le x_e \le 1$, $x_e \in \mathbb{Z}$
	$\sum_{e: v \in e} x_e \le 1$ for all v vertices.

Example XVI: Matching Problem: Relaxation

- That is, we see an IP problem. The omission of the condition $x_e \in \mathbb{Z}$ is called **LP relaxation**. With this, we *open up* the competition to *competitors* with non-integer coordinates as well. Of course, the optimal value of the maximization problem may increase.
- The relaxed problem is a manageable problem/LP problem:

Maximize	$1^\mathsf{T} x$
subject to	$0 \le x_e \le 1$
	$\sum_{e: v \in e} x_e \le 1$ for all v vertices.

Notation

$$\nu^*(G)$$

denotes the optimal value of the above LP problem.

Example XVI: Matching Problem (continued)

Theorem

If G is bipartite, then $\nu^*(G) = \nu(G)$.

That is, the optimization problem obtained by LP relaxation is equivalent to the original one.

• The initial form of our problem can also be equivalently rewritten in LP form. Originally, we had finitely many *competitors*, the characteristic vectors of the matchings. By replacing this set of competitors with their convex hulls, we obtain an equivalent problem (our objective function is linear!):

Maximize	$1^{\mathsf{T}} x$
subject to	$x \in \mathrm{conv}\{\chi_M : M \text{ is a matching}\}$

Example XVI: Matching Problem (continued)

- The condition describes a polytope. This can be described as the intersection of finitely many half-spaces, that is, algebraically the solution set of a finite system of linear inequalities.

Theorem*

Let G be a bipartite graph. Then

$$\mathrm{conv}\{\chi_M : M \text{ is a matching}\} =$$
$$=\{x \in \mathbb{R}^E : 0 \leq x_e \leq 1, \sum_{e:v\in e} x_e \leq 1 \text{ for all } v \text{ vertices.}\}$$

- The case of non-bipartite graphs will be studied further later.

Example XIV: Clique Problem Revisited

• Finally, let's formalize/algebraize an $\mathcal{N}P$-hard problem. The resulting form is that of a QP basic problem. However, it will certainly not be a convex problem.

Example

Given a graph G. Determine the clique parameter ($\omega(G)$), i.e., the size of the largest clique.

• We saw that the question can also be formulated as an IP problem. The following theorem gives an alternative, far from obvious, formalization of determining $\omega(G)$.

Example XIV: Clique Problem Revisited: Theorem

Theorem

Let $A \in \mathbb{R}^{V\times V}$ be the adjacency matrix of graph G, and consider the following problem $(x, 0, 1 \in \mathbb{R}^V)$:

Maximize	$x^T A x$
subject to	$x \succeq 0$
	$1^T x = 1.$

Then

$$p^* = 1 - \frac{1}{\omega(G)}.$$

Example XIV: Clique Problem Revisited: Proof

(1): $p^* \geq 1 - \frac{1}{\omega(G)}$.

- Let K be an optimal (maximal size) clique. Then at

$$x_K = (0, \ldots, 0, \underbrace{\frac{1}{\omega}, \ldots, \frac{1}{\omega}}_{K}, 0, \ldots, 0)$$

feasible location the value of

$$c(x_K) = x_K^{\mathsf{T}} A x_K = \frac{1}{\omega^2}\omega(\omega - 1) = 1 - \frac{1}{\omega}.$$

Example XIV: Clique Problem Revisited: Proof

(2): $p^* \leq 1 - \frac{1}{\omega(G)}$.

- Consider the

$$x = \left(\frac{\overset{v_1}{n_1}}{N}, \ldots, \frac{\overset{v_k}{n_k}}{N} \right) \in \mathcal{L} \cap \mathbb{Q}^{V(G)}$$

vector, where $k = |V(G)|$ and $n_1, \ldots, n_k \in \mathbb{N}$ numbers form a partition of $N \in \mathbb{N}$.

- Then assign a graph $\widetilde{G}$ to G as follows:

  ◦ Each point v_i of G corresponds to an independent set V_i of n_i points in $\widetilde{G}$ ($i = 1, \ldots, k$).

  ◦ If points $v_i, v_j \in V(G)$ are adjacent, then a complete bipartite graph K_{n_i,n_j} is formed between the respective V_i, V_j sets (all possible edges are drawn).

  ◦ If points $v_i, v_j \in V(G)$ are not adjacent, then there is no edge between the corresponding V_i, V_j sets.

Example XIV: Clique Problem Revisited: Proof

- By definition, the number of points of $\widetilde{G}$ is

$$|V(\widetilde{G})| = \sum_{i=1}^{k} n_i = N \sum_{i=1}^{k} \frac{n_i}{N} = N$$

- Based on the construction, the number of edges of $\widetilde{G}$ is

$$|E(\widetilde{G})| = \sum_{\{v_i,v_j\}\in E(G)} n_i n_j = \frac{1}{2} N^2 \sum_{v_i v_j \in E(G)} \frac{n_i}{N}\frac{n_j}{N} = \frac{N^2}{2} x^{\mathsf{T}} A x = \frac{N^2}{2} c(x).$$

- The largest clique of $\widetilde{G}$ has size $\omega(G)$. Applying Turán's theorem, its number of edges is at most the number of edges of the $\omega(G)$-class/N-point Turán graph ($\omega(G)|N$):

$$|E(\widetilde{G})| \leq |E(T_{N,\omega(G)})| = \binom{\omega(G)}{2}\left(\frac{N}{\omega(G)}\right)^2 = \left(1 - \frac{1}{\omega(G)}\right)\frac{N^2}{2}.$$

Example XIV: Clique Problem Revisited: Proof

- Summarizing the above, we obtain

$$c(x) = \frac{2}{N^2}|E(\widetilde{G})| \leq 1 - \frac{1}{\omega(G)},$$

if $x \in \mathcal{L} \cap \mathbb{Q}^{V(G)}$.

- $\mathbb{Q}^{V(G)}$ is a dense set in $\mathcal{L}$, and the objective function is continuous.
- This proves the missing inequality. Thus, $p^* = 1 - \frac{1}{\omega(G)}$.

Example XIV: Clique Problem Revisited: Remarks

• In the formalization provided by the theorem, a quadratic form needs to be maximized.

• Among the conditions of the convex QP problem, **only** the positive semidefiniteness of the quadratic form matrix is missing.

• Omitting this condition results in a form capable of formalizing hopeless optimization problems.

This is the End!

Thank you for your attention!