IE 331 OR: Optimization **KAIST, Spring 2024**
Lecture #1: Introduction February 26, 2024
Lecturer: Dabeen Lee

1 Optimization problems from high school

Let us start with the following high school math exercise question. A pharmaceutical factory manufactures two products X and Y. To produce them, we need two kinds of base materials A and B. The following table shows the amount of each material required to produce one unit of each product. Assume that we have 150 units of material A and 70 units of material B. If our

	A	B
X	2	2
Y	3	1

Table 1: Materials required to produce products

objective is to produce as many products as possible regardless of whether they are X or Y, what is the maximum number of products we can produce with 150 units of A and 70 units of B?

How do we solve the problem? With some abuse of notation, let x denote the number of product X that we produce, and let y denote the number of product Y. As one unit of X and one unit of Y require 2 and 3 units of material A, respectively, $2x + 3y$ would be the total amount of material A used. Here, as we have 150 units of material A, we have the condition that

$$2x + 3y \leq 150.$$

We can draw the set of points (x, y) satisfying the condition in the 2-dimensional coordinate system, and we obtain the first figure in Figure 1.1.

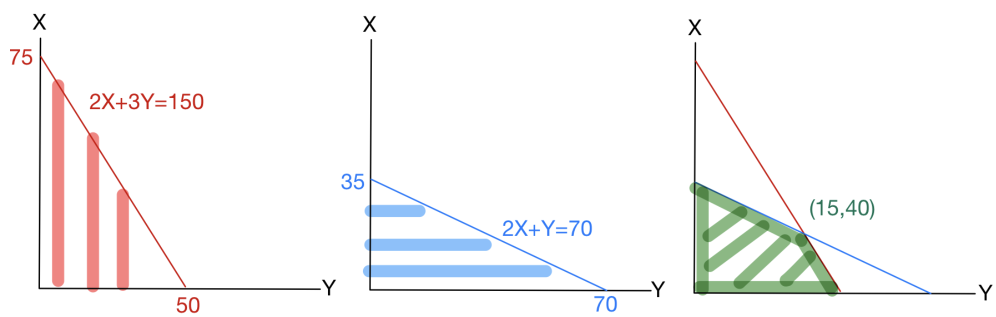

Figure 1.1: Deriving the set of possible solutions

Likewise, considering material B, we deduce the condition that

$$2x + y \leq 70.$$

The corresponding area in the xy-coordinate system is depicted as in the second figure of Figure 1.1. Here, of course, the quantities x and y should be nonnegative, so we have the condition that

$$x, y \geq 0.$$

In fact, a possible production plan for X and Y, given by the quantities (x, y), must satisfy all these three conditions. The corresponding region is given as in the third figure of Figure 1.1. By simple algebra, we may deduce that the red and blue line segments intersect at point $(15, 40)$.

We have just characterized the set of all possible production plans under the current inventory of materials. Among these production plans, what would be the one maximizing the total number of products produced? We can simply choose the point (x, y) maximizing $x + y$. We can obtain the point by drawing line segments parallel to $x+y = 0$. We move the line $x+y$ toward the upper-right corner until it does not intersect the green area. The final line is given by $x+y = 55$ going through the point $(15, 40)$ (See Figure 1.2).

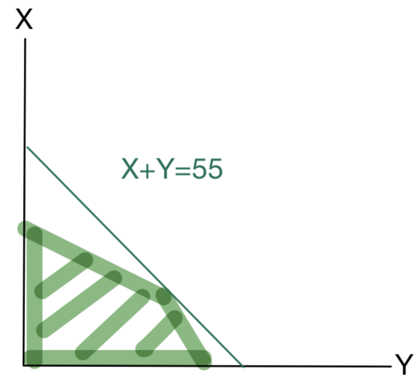

Figure 1.2: Production plan maximizing the total number of produced units

Next we consider another math problem from high school. Conisder a quadratic function $f(x) = 3x^2 + 6x + 5$. At which point x the function $f(x)$ is minimized? We can answer the question by

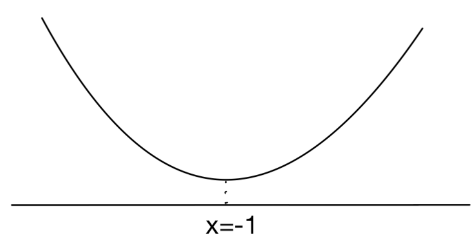

Figure 1.3: Quadratic minimization

taking the derivative of f, that is, $f'(x) = 6x + 6x$. Since the coefficient of the quadratic term is positive, we know that f is minimized precisely when $f'(x) = 0$ and $x = -1$.

2 More concrete optimization problems

As you have already guessed, the two problems from high school are some examples of **optimization** problems. Before we formally define what an optimization problem is, let us discuss a few more concrete examples.

2.1 Sports game scheduling

Major League Baseball (MLB) consists of the National League (NL) and the American League (AL), each of which has 15 teams, so there are 30 teams in total. There are 162 games for all teams over 6 months in 2024, which amounts to 2430 games for one season. For example, Figure 1.4 shows the 2024 season schedule of Los Angeles Dodgers.

Figure 1.4: LA Dodgers game schedule 2024

There are many factors and requirements when scheduling matches over the season.

- The number of home games and that of away games over the season should be balanced.
- The number of home games and that of away games over the weekends in a month should be balanced.
- Games are placed in series: 3–4 consecutive games with the same team.
- 2–3 consecutive series for a home stand.
- Moderate traveling for away games.
- Requirements for interleague games (NL vs AL) and inter-division games (NL East vs NL West).
- Lots of other requirements: rival matches (Yankees vs Mets, Cups vs White Sox).

Let alone the number of games that we have to schedule, these requirements and restrictions placed on teams are what make the scheduling problem extremely difficult. At the same time, coming up with right schedules is crucial in terms of profits. MLB is indeed a huge business recording \$11 billion revenue in 2022. Attentance and league fairness heavily depend on a schedule.

In fact, finding a feasible schedule is not so trivial. To see this, let us play with a toy example.

1. There are 6 teams (A,B,C,D,E,F), and all teams play each other once.
2. There 5 days, and we schedule 3 games per day.
3. Each team plays exactly once a day.

The following table is a partial schedule of the league. Let us complete it.

	Game 1	Game 2	Game 3
Day 1		(C,D)	(E,F)
Day 2	(A,C)		(D,F)
Day 3	(A,D)	(B,F)	(C,E)
Day 4	(A,E)		
Day 5			

Table 2: Partial schedule of games

First, Game 1 for Day 1 has to be (A,B), and Game 2 for Day 2 should be (B,E). Next, Game 2 on Day 4 can be (B,C), which forces Game 2 on Day 4 to be (D,F). However, teams D and F have already played on Day 2, a conflict. Therefore, we'd better have (B,D) for Day 4. Then we may complete the rest. This example demonstrates that for sports scheduling problems, finding a

	Game 1	Game 2	Game 3
Day 1	(A,B)	(C,D)	(E,F)
Day 2	(A,C)	(B,E)	(D,F)
Day 3	(A,D)	(B,F)	(C,E)
Day 4	(A,E)	**(B,D)**	(C,F)
Day 5	(A,F)	(B,C)	(D,E)

Table 3: Complete schedule

single schedule that satisfies all the requirements is not easy. (Note: the contents and materials in this section were directly or indirectly taken from Michael Trick's talk on sports scheduling given at Bucknell University [1].)

2.2 Least squares

Next we consider a linear regression problem in statistics. For simplicity, let us assume that there is a single feature a as well as the response b. Our hypothesis is that there exists some x such that

$$b = ax,$$

which means that the response b is simply the product of the feature a and the number x. We infer the value of x based on the data of pair (a, b) as in Figure 1.5.

Suppose that we have n data points

$$(a_1, b_1), \ldots, (a_n, b_n).$$

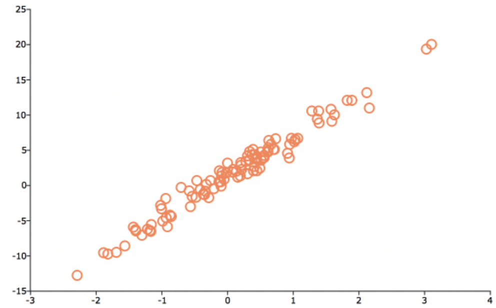

Figure 1.5: Linear regression

The method of least squares proposes a value for x as the one minimizing the following expression.

$$\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i x)^2.$$

For this optimization problem, any number $x \in \mathbb{R}$ can be a valid solution, in contrast to the previous example of sports scheduling. In fact, taking the derivative of the function with respect to x, we obtain that

$$x = \frac{\sum_{i=1}^{n} a_i b_i}{\sum_{i=1}^{n} a_i^2}$$

minimizes the quadratic function. However, for regression, we often deal with millions of data points, in which case the number n is huge. Hence, the focus here is to develop an algorithm that iteratively finds good solutions fast.

3 What we cover and focus on throughout the course

We have discussed the production planning problem with inequalities, the sports scheduling problem, the problem of finding the minimum of a parabola, and the linear regression problem. The first two problems involve nontrivial constraints, coming up with a solution satisfying which is not always easy. This is the case in many general classes of optimization problems that arise in operations research. For example,

- Linear programming,
- Integer programming,
- Stochastic programming,
- Network flow models.

The main objective in these problem classes is to find a solution satisfying possibly complex functional constraints while optimizing the given objective function.

On the other hand, we face the quadratic minimization problems in applications with large data sets, although they are typically unconstrained or the constraint structure is easy to deal with. As coming up with a solution is relatively easy, many solution methods for solving such problems are iterative, in that we keep updating solutions until we reach a certain stopping criteria. We learn these topics in IE539 Convex Optimization, but they are not the main focus of this course. Instead, we focus on the first kind of optimization problems. At the end of this course, we will briefly touch on optimization methods that have modern applications such as machine learning and artificial intelligence.

References

[1] Michael Trick, Sports Scheduling: Optimization and Analytics, Analytics and Operations Management Seminar at Bucknell University (October 28, 2021) 2.1