CS 217: Artificial Intelligence and Machine Learning **Jan-Apr 2024**

Lecture 02: Linear and Convex Optimization

Lecturer: Swaprava Nath *Scribe(s): SG3,SG4*

Disclaimer: *These notes aggregate content from several texts and have not been subjected to the usual scrutiny deserved by formal publications. If you find errors, please bring to the notice of the Instructor.*

2.1 Graphical Approach to Linear Optimization

The graphical approach to linear programming involves representing the constraints and objective functions on a graph to find the optimal solution. By visually analyzing the graph, we can identify the maximum or minimum values for the objective function within the constraints.

Mathematical Formulation: Our aim is to maximise $\mathbf{c}^\top \mathbf{x}$ subject to the constraints $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ and all x_i's are positive, where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{bmatrix}.$$

First, we define a few terminologies.

- **Feasible region -** The common region determined by all the constraints including non-negative constraints $x, y \geq 0$ of a linear programming problem is called the feasible region (or solution region) for the problem. The region other than feasible region is called an infeasible region.
- **Feasible solutions -** Points within and on the boundary of the feasible region represent feasible solutions of the constraints. Any point outside the feasible region is called an infeasible solution.
- **Optimal (feasible) solution -** Any point in the feasible region that gives the optimal value (maximum or minimum) of the objective function is called an optimal solution.

Theorem 2.1 *Let R be the feasible region (convex polygon) for a linear programming problem and let $Z = ax+by$ be the objective function. When Z has an optimal value (maximum or minimum), where the variables x and y are subject to constraints described by linear inequalities, this optimal value must occur at a corner point (vertex) of the feasible region. [1]*

Theorem 2.2 *Let R be the feasible region for a linear programming problem, and let $Z = ax + by$ be the objective function. If R is bounded, then the objective function Z has both a maximum and a minimum value on R and each of these occurs at a corner point (vertex) of R. [1]*

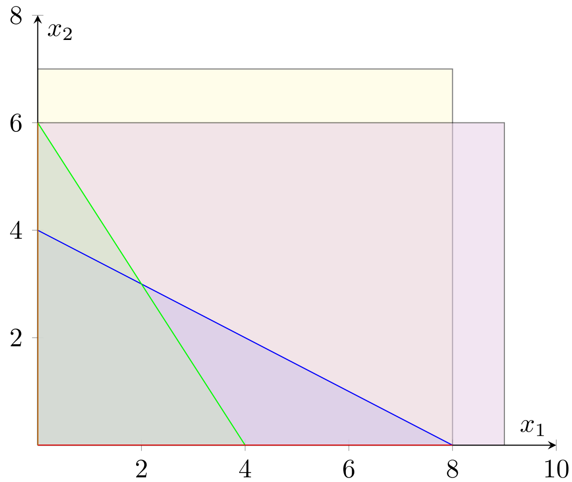
We consider an example. We want to maximize the objective function:

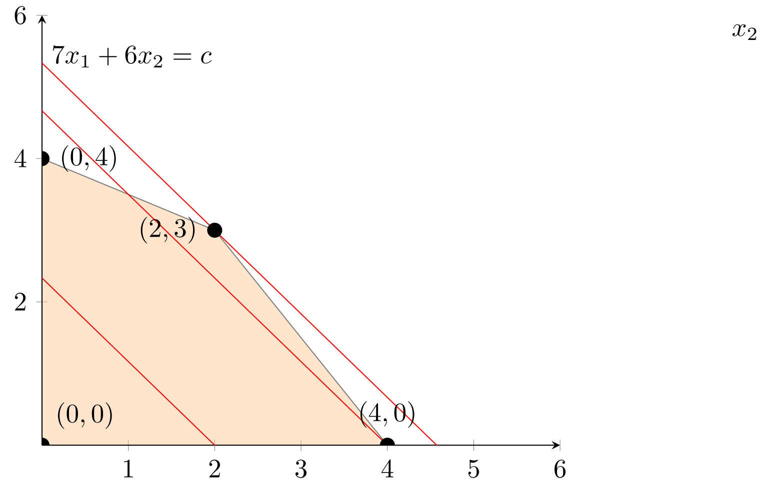
$$\max \; 7x_1 + 6x_2$$

subject to the constraints:

$$
\begin{aligned}
2x_1 + 4x_2 &\leq 16,\\
3x_1 + 2x_2 &\leq 12,\\
x_1 &\geq 0,\\
x_2 &\geq 0.
\end{aligned}
$$

We draw a graph of the inequalities and find the feasible region as shown in the accompanying figure :





In the feasible region, we draw the lines $7x_1 + 6x_2 = c$ where c $\in \mathbb{R}$ and increase c from 0. We stop at particular value of c where on further increasing, there will be no common point on line and feasible region. In above example we stop at $7x_1 + 6x_2 = 32$, and 32 is optimal value with (2,3) as optimal solution.

The graphical method to solve linear programs suffers from two major disadvantages :

1. **Unscalability :** It is not possible to visualise the graph of the linear program when the number of dimensions is more than 3, which is a likely scenario in the real world. It is evidently not scalable beyond 3 dimensions.

2. **Non algorithmic :** By the very nature of the method it does not provide a step by step process to optimise a given objective function. It relies on drawing the graphs of the inequalities and analysing them.

2.2 Simplex method

Let us understand the steps involved in solving the following problem using simplex method.

$$\text{Maximize } 7x_1 + 6x_2 \text{ s.t,}$$

$$2x_1 + 4x_2 \leq 16$$

$$3x_1 + 2x_2 \leq 12$$

$$x_1, x_2 \geq 0$$

- **Step 1**: Slack the constraints
  We are adding slack variables on left hand side to bring equality

$$2x_1 + 4x_2 + s_1 = 16$$
$$3x_1 + 2x_2 + s_2 = 12$$
$$s_1, s_2 \geq 0$$

  Slack variables have no role in the objective function, so their coefficients in objective function are zero

- **Step 2**: Set basic(non-zero) and non-basic(zero) variables
  We set slack variables to basic and actual variables to non-basic

$$x_1, x_2 = 0 \text{ (non-basic) and } s_1 = 16, s_2 = 12 \text{ (basic)}$$
$$\text{Objective function is } 7x_1 + 6x_2$$

  To improve this we iteratively bring one basic variables to non-basic and non-basic to basic
  Being greedy, x_1 is brought to basic since

$$\text{coefficient of } x_1 \ (7) \ > \ \text{coefficient of } x_2 \ (6)$$

  So, x_2 is still 0, any of s_1/s_2 can become greater than zero

$$2x_1 + 4x_2 + s_1 = 16 \implies 2x_1 + 4(0) + 0 \leq 16 \implies x_1 \leq 8$$
$$3x_1 + 2x_2 + s_2 = 12 \implies 3x_1 + 2(0) + 0 \leq 12 \implies x_1 \leq 4$$

  x_1 can be maximum increased to 4 that is if $x_1 > 4$, s_2 becomes negative
  so x_1 is increased to 4

- **Note:**
  - number of basic variables = number of slack variables or number of inequalities at start
  - remaining variables are non-basic variables

- Now $x_1 = 4, s_2 = 0, s_1 = 8, x_2 = 0$
  x_1, s_1 are basic variables and x_2, s_2 are non-basic variables

$$s_1 = 16 - 2x_1 - 4x_2$$
$$s_2 = 12 - 3x_1 - 2x_2$$
$$x_1 = 4 - \frac{1}{3}s_2 - \frac{2}{3}x_2$$
$$s_1 = 8 + \frac{2}{3}s_2 - \frac{8}{3}x_2$$

  Now Objective Function is $28 + \frac{4}{3}x_2 - \frac{7}{3}s_2$
  Now x_2 and s_2 are non-basic and one of them should be converted into basic
  x_2 becomes basic as it can give positive drift

- We will continue this process until the coefficients of non-basic variables are **non-positive**, as we can't increase objective function

2.3 Simplex Tableau

This is a more organized form of Simplex method. We write the coefficients of variables and slack variables present in the constraint and objective equations in the form of table. We denote the RHS values of constraint equations in the b column of the table.
As mentioned above, slack variables will be initially basic and variables already present in the objective function will be non-basic. Variables represented in the row of the table are basic variables (non-zero value).
Steps to be followed :

- **step 1:** Find the variable that has the highest positive coefficient in the objective function say x_i. Row-wise divide b column with that x_i column to obtain a fraction column.
- **step 2:** Find the basic row variable whose fraction comes out to be lowest positive. The intersection of that row and column is the pivot and should be made 1 by dividing the whole row by it's value.
- **step 3:** We need to do the gaussian elimination of that x_i column i.e. except the pivot which is made 1 rest of the column values should be made 0 using row operations.
- **step 4:** Make a new table in which the selected row variable will be moved out (basic to non-basic) and the selected column variable x_i will be moved in replacing it (non-basic to basic).
- **step 5:** Repeat the steps till we get all the values of last row (i.e. P) non-positive.

Let's solve the previous problem using Tableau method

	x_1	x_2	s_1	s_2	b	Fraction
s_1	2	4	1	0	16	$16/2 = 8$
s_2	3	2	0	1	12	$12/3 = 4$
P	7	6	0	0		

$s_1 = 16$	$s_2 = 12$	$x_1 = 0$	$x_2 = 0$

In the above table, 7 is the largest so we will divide b column with the x_1 column to get the fractions. In the fraction column 4 is the smallest positive value for s_2 row. Here, the intersection of the s_2 row and x_1 column i.e. the value 3 is the pivot.So, x_1 column will undergo gaussian elimination and s_2 will be moved out and x_1 will be moved in.

	x_1	x_2	s_1	s_2	b	Fraction
s_1	0	8/3	1	-1/3	8	3
x_1	1	2/3	0	1/3	4	6
P	0	4/3	0	-7/3		

$s_1 = 8$	$s_2 = 0$	$x_1 = 4$	$x_2 = 0$

We got 4/3 as the largest positive value in P row, so we will repeat the steps again.In this case x_2 column will undergo gaussian elimination . According to the fraction column obtained after repetition of previous steps, s_1 will be moved out as 3 is the lowest positive value in the fraction column and x_2 will be moved in.

	x_1	x_2	s_1	s_2	b	Fraction
x_2	0	1	1	3/8	-1/4	3
x_1	1	0	0	-1/4	1/2	2
P	0	0	-1/2	-2		

$s_1 = 0$	$s_2 = 0$	$x_1 = 2$	$x_2 = 3$

As we got all the values in P row as non-positive, so we should stop the process.Above are the final values of all the variables.
Point to Note:
Simplex always goes from **one corner point to another** in the feasible region as (x_1, x_2) changes from (0,0) to (4,0) to (2,3) after every iteration

Hence the maximum value of the objective function is $7 \times 2 + 6 \times 3 = 32$

2.4 Convex optimization

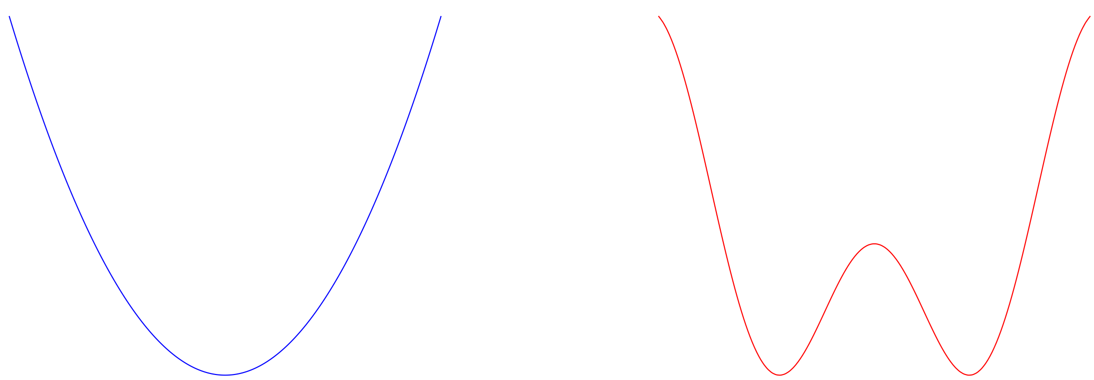
Before learning about convex optimization we need to define **Convex functions** and **Convex sets**

- **Convex Functions :**

  A function f:$\mathbb{R}^n \to \mathbb{R}$ is said to be a convex function if

  $$\forall x, y \in \mathbb{R}^n \quad \forall \lambda \in [0,1] \quad f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y) \tag{2.1}$$

  Examples for convex functions in 1 variable is f(x) = x^2 and in 2 variables f(x_1, x_2) = $x_1 x_2$



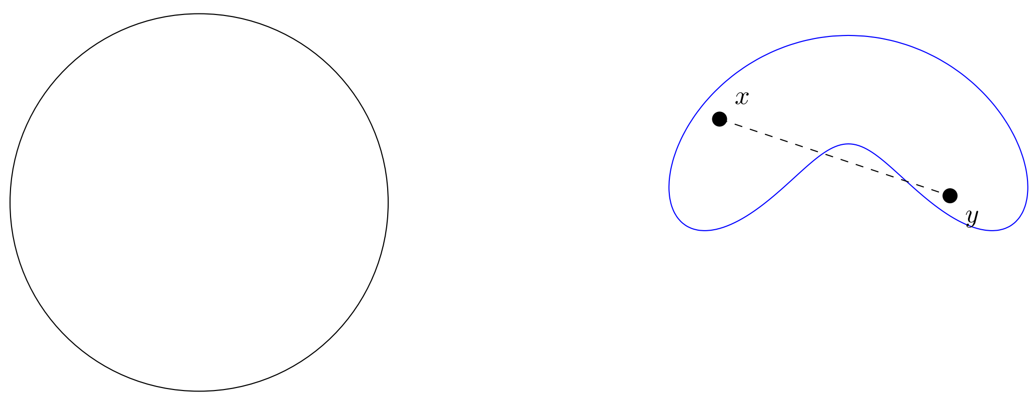
(a) Convex curve

(b) Non-convex curve

Figure 2.1: Convex and non-convex curves.

- **Convex Sets :**

  A set C $\subseteq \mathbb{R}^n$ is a convex set if $\forall$ x,y $\in$ C , $\forall$ $\lambda \in$ [0,1] $\big(\lambda x + (1-\lambda)y\big) \in$ C
  Clearly in figure 2.2(b) some points on the line joining points x and y are outside the set, so the set is not convex.



(a) convex set (b) non-convex set

Figure 2.2: Convex and non-convex sets

Intuition: We have learnt in one-variable calculus that a convex function is one which has a positive second derivative. Let us see how that applies to multivariable functions.

However, convexity depends on the comparison of values in the range, which is not defined for multivariable points. There is simply no such thing as 'less than' over in a domain of two or more variables (two or more dimensions). For us to extend the definition, we need the range to be of one variable. In other words, we speak only of functions of the form $f : R^n \to R$.

An example with $n = 2$ would be $f(x_1, x_2) = x_1 \cdot x_2$. However, this is not a convex function.

For one dimension or linear function, the constraint would be would be

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \quad \forall \lambda \in 0, 1 \quad \forall (x, y) \in R$$

We extend it to multiple dimensions. A set or region $C \in R^n$ is said to be convex iff

$$\forall(\vec{x}, \vec{y}) \in C \quad \forall \lambda \in 0, 1 \quad \lambda\vec{x} + (1-\lambda)\vec{y} \in C$$

For convex regions in n dimensions, the domain $D \subseteq R^n$ should be convex and secondly,

$$(\lambda\vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda)f(\vec{y}) \quad \forall \lambda \in 0, 1 \quad \forall(\vec{x}, \vec{y}) \in D$$

2.4.1 Analysis of Convexity of Functions and domains

$C_1 \cap C_2$ and $C_1\hat{C}_2$, where C_1 and C_2 are convex domains

$C_1 \cap C_2$ is not a convex domain. A very simple counterexample would be two disjoint convex domains.

On the other hand, $C_1\hat{C}_2$ is definitely convex. If any points $\vec{x}$ and *vecy* lie both in C_1 and C_2, so does $\lambda\vec{x} + (1-\lambda)\vec{y}$ lie in both and therefore in $C_1\hat{C}_2$, regardless of choice of $\lambda, \vec{x}, \vec{y}$ within constraints.

Set $S = \{x | x \in R^2_{>=0}, x_1, x_2 \geq 1\}$

This is exactly one of the four quadrants, shifted one unit right and one unit upward.

Clearly, this is convex. If it is not obvious, just consider $\vec{x} - (1, 1)$ and $\vec{y} - (1, 1)$. All choices within this region have resultant point vectors with both coordinates positive. Any linear combination of those two with positive coefficients must necessarily follow this property.

Function $f : R^2_{\geq 0} \to R; \quad f(\vec{x}) = x_1 \cdot x_2$

Clearly, this is not convex. A simple counterexample would be $\vec{x} = (1, -1)$ and $\vec{y} = (-1, 1)$ and $\lambda = 0.5$. This means that the point in between the two points would be the origin itself.

It is obvious that $f(\vec{x}) = f(\vec{y}) = -1$ and that $f(\vec{0}) = 0$. This blatantly violates the condition of convexity.

Moreover, it has no global minima (think of the second and fourth quadrants in the domains), which is a must for any convex function. We shall look into the reason as to why this is so late.

Function $f : R^2 \to R; \quad f(\vec{x}) = x_1^2 + x_2^2 + x_1 \cdot x_2$

We can rewrite this as $f(\vec{x}) = \frac{(x_1+x_2)^2}{2} + \frac{x_1^2+x_2^2}{2} = \frac{T(\vec{x})^2}{2} + \frac{\|x\|_2^2}{2}$.

The second term is a very standard convex function i.e. the square of the distance from the origin. It is as convex as a parabola is, except we express it as the sum of two independent parabolas on two independent variables.

Regarding the first, again, the term T that is squared is a linear transformation of the vector onto one variable i.e. $T(\lambda\vec{x} + \mu\vec{y}) = \lambda T(\vec{x}) + \mu T(\vec{y})$. Here we just place $\mu - 1 - \lambda$ and add constraints for λ. It is as convex as a parabola is.

2.4.2 Convex Optimization Problem

Given that the objective function of a linear programming problem is convex, we are supposed to minimize the function. Obviously, the exact opposite of convex functions can be defined with the inequality reversed, where we maximize the value of a concave objective function.

2.4.3 Examples of Convex Functions

$$
\begin{aligned}
f(x) &= e^{ax} \\
f(x) &= -ln(x) \\
f(\vec{x}) &= \vec{x}^T\vec{x} = \|\vec{x}\|_2^2 = \sum_{i=1}^{n} x_i^2 \\
f(\vec{x}) &= \|\vec{x}\|_1 = \sum_{i=1}^{n} \|x_i\|
\end{aligned}
$$

2.4.4 Some Important Definitions

Global optimal point

A point $\vec{x}$ is said to be globally optimal if $\vec{x}$ is feasible and there does not exist any other $\vec{y}$ with the objective function $f(\vec{y}) \leq f(\vec{x})$.

Local optimal point

A point $\vec{x}$ is said to be locally optimal if $\vec{x}$ is feasible and there exists a constant R_1 such that for all feasible $\vec{y}$ with $\|\vec{y} - \vec{x}\|_2 \leq R_1$, the objective function $f(\vec{y}) \leq f(\vec{x})$

Theorem 2.3 *For a convex optimization problem all locally optimal points are globally optimal points*

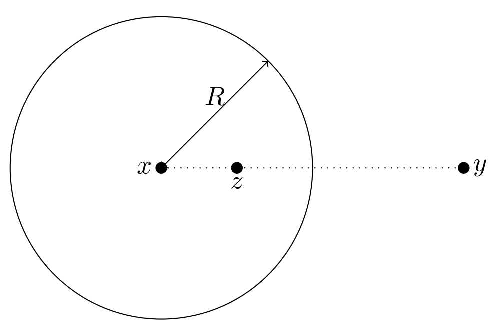
Proof: A high level idea of proof will be discussed without mathematical rigour
The proof is by contradiction, Let $\exists$ a local optimum (x) which is not global optimum (y)

$$\implies f(y) < f(x) \tag{2.2}$$

Let $\mathcal{C}$ be the region of radius R such that $\forall$ feasible k with $\|k - x\|_2 <$ R,

$$f(x) \leq f(k)$$

Now let us find a point z $= \lambda x + (1-\lambda)y$ such that $\lambda \in [0,1]$ and z $\in \mathcal{C}$



$$\implies f(z) = f(\lambda x + (1-\lambda)y)$$
$$\implies f(z) \leq \lambda f(x) + (1-\lambda)f(y)$$

From equation 2.2 we can say

$$\implies f(z) < \lambda f(x) + (1-\lambda)f(x)$$
$$\implies f(z) < f(x)$$

This violates the definition of local optimality.
Therefore, contradiction.

■

References

[1] National Council of Educational Research and Training. *Mathematics Class 12*. Chapter 12: Linear Programming, Lecture Notes. NCERT, Year.