ECH4905 - Special Topics in ChemE - Optimization

Andres Espinosa

January 16, 2025

Contents

1 Chapter 1 - Optimization Models

1.1 Tuesday 01/14/2025

1.1.1 Modeling Overview

An optimization problem is represented by degrees of freedom that inform decisions. The degrees of freedom of a problem dictate the number of decisions that can be made.

- n Decisions $> 0 \implies$ optimization problem
- n decisions $= 0 \implies$ simulation

The number of decisions is equal to the degrees of freedom, which is equivalent to the number of variables minus the number of inequalities.

1.1.2 Example Problem

This example problem will use a toy example of minimizing the perimeter of a rectangle with side lengths x and y.

$$
\begin{aligned}
\text{minimize} \quad & P && (1)\\
\text{subject to} \quad & xy = 2000 && (2)\\
& P = 2x + 2y && (3)\\
& x \geq 0, y \geq 0 && (4)
\end{aligned}
$$

with variables x, y, P. Since there are two equalities in the optimization problem above and three variables, there is one degree of freedom. This can be seen more plainly by reformulating the problem into the below problem

$$
\begin{aligned}
\text{minimize} \quad & 2(\frac{2000}{y}) + 2y && (5)\\
\text{subject to} \quad & y \geq 0 && (6)
\end{aligned}
$$

1.1.3 General Optimization Definitions

A general optimization problem can be written as

$$
\begin{aligned}
\text{minimize} \quad & F(x, y) && (7)\\
\text{subject to} \quad & h(x, y) = 0 && (8)\\
& g(x, y) \leq 0 && (9)
\end{aligned}
$$

where $x \in \mathbb{R}^n, y \in \{0, 1\}^m$. This general optimization problem can be broken down into different problems with levels of complexity.

Firstly, a **linear program** (LP) is defined as

$$
\begin{aligned}
\text{minimize} \quad & F(x) && (10)\\
\text{subject to} \quad & h(x) = 0 && (11)\\
& g(x) \leq 0 && (12)
\end{aligned}
$$

Where $x \in \mathbb{R}^n$ and F, h, g are affine functions. Linear problems typically make many approximations about the real physics/constraints that are in the world. These approximations are a tradeoff with computational tractability. The Simplex method is typically used to solve linear programs.

Secondly, a **non-linear program** (NLP) is defined as

$$\begin{aligned} \text{minimize} \quad & F(x) && (13) \\ \text{subject to} \quad & h(x) = 0 && (14) \\ & g(x) \leq 0 && (15) \end{aligned}$$

Where $x \in \mathbb{R}^n$ and any of the functions F, h, g are non-linear. A subset of NLPs are those problems that are *convex*. Convexity in a non-linear program is achieved when F, g are convex functions and h is an affine function. Some algorithms that are used to solve NLPs are SQP, IPOPT.

A higher degree of complexity, are **mixed integer linear programs** (MILP) defined as

$$\begin{aligned} \text{minimize} \quad & F(x, y) && (16) \\ \text{subject to} \quad & h(x, y) = 0 && (17) \\ & g(x, y) \leq 0 && (18) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \{0, 1\}^m$. and F, h, g are affine functions which can be represented as $F(x, y) = Ax + By$. These problems have applications in chemical engineering with examples such as biomass, waste, and fuel supply chains. In a biomass biorefinery location problem, you would have a set of locations to pick from y and you want to minimize the cost. An algorithm used to solve these problems is branch-and-bound.

With the highest degree of complexity, a **mixed integer non-linear program** (MINLP) is defined as

$$\begin{aligned} \text{minimize} \quad & F(x, y) && (19) \\ \text{subject to} \quad & h(x, y) = 0 && (20) \\ & g(x, y) \leq 0 && (21) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \{0, 1\}^m$ and any of the F, h, g functions are non-linear. BARON, SCIP, MAINGO are algorithms that have been developed in the last 10 years in response to solve MINLPs. Chemical engineering is naturally non-linear and there has been a big push from the chemical optimization community to create solutions for MINLPs. These are complex problems but also important problems.

1.1.4 Specific Model Criteria

Optimization problems can be also defined by the **number of objectives** they have.

- Single objective
- Multiple objective

Multiple objectives introduce a tradeoff between different objectives with each other. If the number of objectives is low, then it is easy to identify and visualize the tradeoffs. The concept of pareto optimal will be explored further in the course.

$$\text{minimize} \quad \langle F_1(x), F_2(x), \ldots, F_k(x) \rangle \tag{22}$$

Optimization problems can also have a level of **uncertainty** in the problem. For example, the price and cost parameters may be random and uncertain. The majority of time, the randomness and uncertainty in models are ignored by modelers. A good guess is used and substituted for the random parameters such as an average or other estimate. However, it is possible to create stochastic models that attempt to handle uncertainty.

- Deterministic models: Uncertainty is ignored and stochastic parameters are replaced by a fixed "good guess".
- Stochastic models: Uncertainty is handled and incorporated in the constraints.

Typically, stochastic models are significantly more difficult to compute than deterministic models. So, a deterministic model should be used and solved before a stochastic one is approached.

Optimization models are also specified by their **type/number of constraints**

- Unconstrained
- Constrained

1.1.5 Optimization Concepts

For general optimization problem defined above, a **feasible region** is defined as the set of $\langle x, y \rangle$ that satisfy all equalities and inequalities.

$$\begin{aligned} \text{minimize} \quad & -xy && (23)\\ \text{subject to} \quad & x \geq 0, x \leq 1 && (24)\\ & y \geq 0, y \leq 1 && (25) \end{aligned}$$

For this problem, the feasible region R is

$$R(x, y) = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\} \tag{26}$$

A **local minimum** with respect to a distance ε is defined as a point (x^*, y^*) that satisfies $F(x^*, y^*) \leq F(x, y), \|x^* - x\| \leq \varepsilon$. A **global minimum** is a point that satisfies $F(x^*, y^*) \leq F(x, y)$.

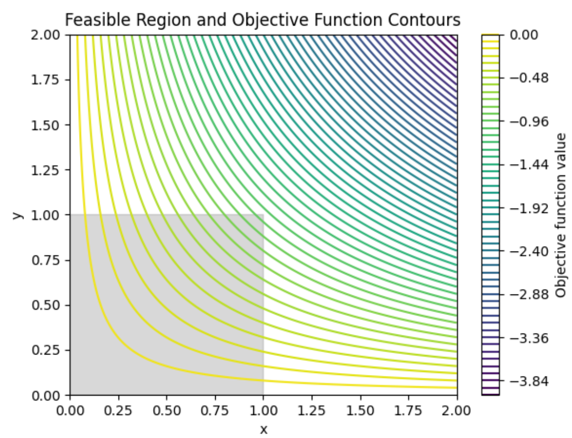

Figure 1: A plot of the optimization problem $F(x, y) = -xy$

2 Chapter 2 - Mathematics Review

2.1 Thursday 01/16/2025

2.1.1 Chapter Math Outline

The chapters over the next few lessons will include some math that will be used in different chapters.

- Chapter 1: Linear algebra
  - Matrices
  - Eigenvalues
- Chapter 2: Convex Analysis
  - Convexity
  - Quadratic forms
  - Taylor Series

2.1.2 Linear Algebra Review

A vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{27}$$

This vector represents a direction and magnitude in n dimensions.
The $l2$-norm of a vector is defned as

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots x_n^2} \tag{28}$$

Vector addition is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \tag{29}$$

Scalar multiplication of a scalar $a \in \mathbb{R}$ with a vector $\mathbf{x} \in \mathbb{R}^n$

$$a\mathbf{x} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} \tag{30}$$

The dot product of two vectors $\mathbf{x}, \mathbf{y}, \in \mathbb{R}^n$ is defined as

$$\mathbf{x}^\top \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{31}$$

With equivalent notation $\mathbf{x}^\top \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle$

$$\|\mathbf{x} - \mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos\theta \tag{32}$$
$$\|\mathbf{x} - \mathbf{y}\|_2^2 = (\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \tag{33}$$

The Cauchy-Schwarts Inequality is derived by the following

$$\cos\theta = \frac{x^\top y}{\|x\|\|y\|} \tag{34}$$
$$\frac{x^\top y}{\|x\|\|y\|} \leq 1 \tag{35}$$
$$x^\top y \leq \|x\|\|y\| \tag{36}$$

The triangle inequality is defined as

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{37}$$

A matrix $A \in \mathbb{R}^{m \times n}$ is defined as

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \tag{38}$$

The transpose of a matrix $A^\top$ flips each value for the row and column as such

$$A^\top = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} \tag{39}$$

Matrix addition is element-wise and can be shown as such between two matrices $A, B \in \mathbb{R}^{m \times n}$

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix} \tag{40}$$

Scalar multiplication of a matrix A with a scalar α is defined as

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & \dots & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{bmatrix} \tag{41}$$

Matrix multiplication between two matrices $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$ is defined as

$$AB = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1p} \\ \vdots & \dots & \vdots \\ b_{n1} & \dots & b_{np} \end{bmatrix} \tag{42}$$

With the following properties

- $AB \neq BA$
- $(ABC)^\top = C^\top B^\top A^\top$

The inverse of a matrix has the following properties

- $AA^{-1} = I$
- $(AB)^{-1} - B^{-1}A^{-1}$
- $(B^\top)^{-1} = (B^{-1})^\top$
- $(A^{-1})^{-1} = A$

Orthonormal matrices have the properties

- $O^\top O = I$
- $O^\top = O^{-1}$

Some following matrix partitions are useful

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} + \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_1 + A_1 \\ B_1 + B_2 \end{bmatrix} \tag{43}$$

$$\begin{bmatrix} A \\ B \end{bmatrix}^\top = \begin{bmatrix} A^\top & B^\top \end{bmatrix} \tag{44}$$

$$A \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}^\top = \begin{bmatrix} AB_1 & AB_2 \end{bmatrix} \tag{45}$$

The determinant of a matrix $A \in \mathbb{R}^{n \times n}$ is calculated by subtracting the product of the diagonals of a matrix recursively. The minor of a matrix is the section of a matrix that is achieved when removing a section. Simply, the determinant of a matrix can also be defined as the product of the eigenvalues.

$$\det A = \prod \lambda \tag{46}$$

2.1.3 Spaces

A vector space is defined as a set of all vectors with some properties. We define a vector space $\mathbb{V}$

$$u + v \in \mathbb{V} \tag{47}$$

$$u + (v + w) = (u + v) + w \tag{48}$$

$$u + v = v + u \tag{49}$$

A linearly dependent system of vectors $v_i \in \mathbb{R}^n$ satifies the following for a set of scalars c_i

$$c_1 v_1 + \cdots + c_m v_m = 0 \tag{50}$$

The span of a set of vectors v_i is the set of all vectors that can be created with a linear combination of those vectors.

$$\{x | c_1 v_1 + \cdots + c_m v_m = x\} \tag{51}$$

The basis of a space is the minimum number of vectors needed to span a vector space.