Project Plan: Hill Ciphers in Encryption

Linear Algebra Course Project

Team Leader: [Ngo Binh Nguyen]

December 15, 2025

Objective

To explain the mathematical theory behind the Hill Cipher, implement it using Python on a real text dataset, and analyze the results to demonstrate how Linear Algebra secures information.

Topic Overview: What is the Hill Cipher?

The Core Concept:
Our project explores the intersection of **Linear Algebra** and **Cryptography**. While simple codes replace one letter with another (e.g., A becomes Z), the Hill Cipher is more advanced: it uses **Matrix Multiplication** to encrypt blocks of letters (polygraphic substitution) simultaneously.

How it Works (The Math):

1. **Convert:** We turn text (e.g., "TOS") into column vectors based on the alphabet ($A = 0, B = 1, \dots$).
2. **Encrypt:** We multiply these vectors by our project's **Key Matrix (K)**. This transforms the numbers linearly.
3. **Modulus:** We apply modulo 26 (the size of the alphabet) to keep the results within the range of A-Z.
4. **Decrypt:** To read the message, we multiply the encrypted vector by the **Inverse Matrix (K^{-1})**.

Our Mission: We are proving that Linear Algebra works for security.

- **The Theory Team** will explain the math and manually verify the encryption of the string "TOS" using our specific 3×3 matrix to prove the logic holds up.
- **The Implementation Team** will scale this up using Python to encrypt the entire text of *Sherlock Holmes* and analyze how the letter frequencies change to hide information.

---

1 Report Structure (Table of Contents)

The final report will adhere to the following structure:

- **ABSTRACT** (Executive Summary)
- **CHAPTER 1: OVERVIEW**
  - 1.1 History of Cryptography (Lester S. Hill, 1929).
  - 1.2 Problem Definition: Why Simple Substitution Ciphers fail (Frequency Analysis).
  - 1.3 The Solution: Polygraphic Substitution via Linear Algebra.
- **CHAPTER 2: MATHEMATICAL FOUNDATIONS**
  - 2.1 Modular Arithmetic ($\mathbb{Z}_{26}$).
  - 2.2 Matrix Operations as Linear Transformations.
  - 2.3 Invertibility & The Key Matrix (Conditions for $\det(K)$).
- **CHAPTER 3: THE HILL CIPHER ALGORITHM**
  - 3.1 Key Generation.
  - 3.2 Encryption Process ($C = K \cdot P \pmod{26}$).
  - 3.3 Decryption Process ($P = K^{-1} \cdot C \pmod{26}$).
  - 3.4 Manual Verification (Hand calculation of subset "TOS").
- **CHAPTER 4: CRYPTANALYSIS (VULNERABILITIES)**
  - 4.1 Known Plaintext Attack (Solving $Y = KX$).
  - 4.2 The "Linearity" Weakness in modern cryptography.
- **CHAPTER 5: CASE STUDY - PYTHON IMPLEMENTATION**
  - 5.1 Dataset Description (*Sherlock Holmes*).
  - 5.2 Methodology (Data Cleaning & Code).
  - 5.3 Results & Visualization (Histograms).
  - 5.4 Interpretation (Analysis of flattened distribution).
- **CHAPTER 6: CONCLUSION**
- **REFERENCES**

2 Role Assignments

Team 1: Theory & Mathematics (3 Members)

Focus: Writing the technical content for Chapters 1, 2, 3, and 4.

- **Member 1: The Theorist (Context & Foundations)**
  - **Chapter 1 (Overview):** Write the history of Hill Cipher and the "Problem Definition" (Frequency Analysis).
  - **Chapter 2 (Foundations):** Define the "Alphabet Space" ($\mathbb{Z}_{26}$) and Linear Transformations.
- **Member 2: The Algorithmist (The Core Logic)**
  - **Chapter 3.1 - 3.3:** Explain the Key Matrix rules ($\det(K) \neq 0$ and $\gcd(\det(K), 26) = 1$).
  - Present formal equations for Encryption ($C = KP$) and Decryption ($P = K^{-1}C$).
- **Member 3: The Analyst (Verification & Security)**
  - **Chapter 3.4 (Manual Verification):** Take the first three letters of the dataset (**"TOS"**), convert to vectors, and encrypt them **by hand** using the project Key Matrix. Show every step.
  - **Chapter 4 (Cryptanalysis):** Write the section on "Vulnerabilities" (Known Plaintext Attacks).

Team 2: Implementation & Reporting (2 Members)

Focus: Coding the solution, analyzing results, and compiling the final report.

- **Member 4: The Leader (Implementation & Formatting)**
  - **Chapter 5.1 - 5.3:** Write Python script to encrypt the *Sherlock Holmes* dataset. Generate Histograms (Original vs. Encrypted).
  - **Compilation:** Compile all members' work into the final LaTeX report.
- **Member 5: The Storyteller (Interpretation & Conclusion)**
  - **Abstract:** Write the executive summary.
  - **Chapter 5.4 (Interpretation):** Analyze the histograms generated by Member 4. Explain *why* the flattened distribution proves security.
  - **Chapter 6 (Conclusion):** Summarize findings and limitations.

3 Project Resources

1. The Key Matrix (K)

This matrix must be used for all manual calculations and Python code. Do not change it.

$$K = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}$$

Why must we use this specific matrix?

- **Consistency:** The manual math result (Chapter 3) must match the Python result (Chapter 5) exactly. If we use different matrices, the project fails verification.
- **Simplicity:** This matrix was chosen because its Determinant is 1. This makes the manual calculation of the Inverse Matrix (K^{-1}) much easier for Member 3, as it avoids complex modular fractions.
- **Validity:** This matrix is guaranteed to be invertible modulo 26 (Safe to use).

2. The Dataset

- **Source:** Project Gutenberg
- **Title:** *The Adventures of Sherlock Holmes*
- **Link:** `https://www.gutenberg.org/files/1661/1661-0.txt`