Mục lục

Chương 1

Ma trận

Trong các bài viết của về đại số tuyến tính:

- Vector sẽ được ký hiệu bởi chữ thường in đậm (ví dụ $\boldsymbol{v}, \boldsymbol{x}, \ldots$);
- Ma trận sẽ được ký hiệu bởi chữ hoa in đậm (ví dụ $\boldsymbol{A}$, $\boldsymbol{B}, \ldots$);
- Các đại lượng vô hướng (số) được ký hiệu bởi chữ thường không in đậm (ví dụ x_1, N, $t, \ldots$).

1.1 Định thức và hạng ma trận

Định thức ma trận

Định nghĩa 1. Nghịch thế

Cho tập hợp $A = \{1, 2, \cdots, n\}$ và xét hoán vị σ trên A. Ta gọi hai phần tử i và j tạo thành **nghịch thế** (inversion) nếu $i < j$ và $\sigma(i) > \sigma(j)$.

Đặt $\sigma = \{k_1, k_2, \cdots, k_n\}$ là một hoán vị của A. Ta ký hiệu

$$P\{k_1, k_2, \cdots, k_n\}$$

là số lượng nghịch thế của σ và đặt

$$(-1)^{P\{k_1,k_2,\cdots,k_n\}} = \text{sign}\{k_1, k_2, \cdots, k_n\}.$$

Ví dụ 1. Với $n = 4$, $A = \{1, 2, 3, 4\}$. Xét hoán vị $\sigma = \{4, 2, 1, 3\}$.

Ta nhận thấy các cặp nghịch thế $(4, 2)$, $(4, 1)$, $(4, 3)$, $(2, 1)$ gồm 4 cặp nghịch thế. Vậy $P\{4, 2, 1, 3\} = 4$ và $\text{sign}\{4, 2, 1, 3\} = (-1)^4 = 1$.

Định nghĩa 2. Định thức

Khi đó định thức của ma trận $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ được định nghĩa là:

$$\det(\boldsymbol{A}) = \sum_{(i_1,i_2,\cdots,i_n)} a_{1,i_1} \cdot a_{2,i_2} \cdot a_{n,i_n} \cdot \text{sign}\{i_1, i_2, \cdots, i_n\} \qquad (1.1)$$

với mọi hoán vị $(i_1, i_2, \cdots, i_n)$ của $(1, 2, \cdots, n)$. Như vậy có $n!$ phần tử cho tổng trên.

Ví dụ 2. Tính định thức ma trận $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

Xét hoán vị $\sigma_1 = \{1, 2, 3\}$. Khi đó $P\{1, 2, 3\} = 0$, $a_{11} \cdot a_{22} \cdot a_{33} \cdot (-1)^0 = 1 \cdot 5 \cdot 9 \cdot 1 = 45$.

Xét hoán vị $\sigma_2 = \{1, 3, 2\}$. Khi đó $P\{1, 3, 2\} = 1$, $a_{11} \cdot a_{23} \cdot a_{32} \cdot (-1)^1 = 1 \cdot 6 \cdot 8 \cdot (-1) = -48$.

Xét hoán vị $\sigma_3 = \{2, 1, 3\}$. Khi đó $P\{2, 1, 3\} = 1$, $a_{12} \cdot a_{21} \cdot a_{33} \cdot (-1)^1 = 2 \cdot 4 \cdot 9 \cdot (-1) = -72$.

Xét hoán vị $\sigma_4 = \{2, 3, 1\}$. Khi đó $P\{2, 3, 1\} = 2$, $a_{12} \cdot a_{23} \cdot a_{31} \cdot (-1)^2 = 2 \cdot 6 \cdot 7 \cdot 1 = 84$.

Xét hoán vị $\sigma_5 = \{3, 1, 2\}$. Khi đó $P\{3, 1, 2\} = 2$, $a_{13} \cdot a_{21} \cdot a_{32} \cdot (-1)^2 = 3 \cdot 4 \cdot 8 \cdot 1 = 96$.

Xét hoán vị $\sigma_6 = \{3, 2, 1\}$. Khi đó $P\{3, 2, 1\} = 3$, $a_{13} \cdot a_{22} \cdot a_{31} \cdot (-1)^3 = 3 \cdot 5 \cdot 7 \cdot (-1) = -105$.

Như vậy $\det(A) = 45 - 48 - 72 + 84 + 96 - 105 = 0$.

Định thức của ma trận còn được định nghĩa theo **đệ quy** như sau:

Với ma trận 1×1 là $\boldsymbol{A} = \begin{pmatrix} a_{11} \end{pmatrix}$ thì $\det(\boldsymbol{A}) = a_{11}$.

Với ma trận 2×2 là $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ thì $\det(\boldsymbol{A}) = a_{11}a_{22} - a_{21}a_{12}$.

Với ma trận $n \times n$, gọi $\boldsymbol{M}_{ij}$ là ma trận có được từ ma trận $\boldsymbol{A}$ khi bỏ đi hàng i và cột j của ma trận $\boldsymbol{A}$ và ký hiệu $A_{ij} = (-1)^{i+j} \det(\boldsymbol{M}_{ij})$. Khi đó:

Định lí 1. Định lý Laplace

Định lý Laplace cho phép ta khai triển định thức của ma trận cấp n thành tổng các ma trận cấp $n-1$.

Khai triển theo cột j:

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ij}A_{ij} = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj},\ \ j = \overline{1,n}.$$

Khai triển theo hàng i:

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} a_{ij}A_{ij} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in},\ \ i = \overline{1,n}.$$

Hạng của ma trận

Định nghĩa 3. Hạng của ma trận

Cho ma trận $\mathbf{A}_{m\times n}$. **Hạng** của ma trận là cấp của ma trận con lớn nhất có định thức khác 0. Nghĩa là một ma trận vuông mà là ma trận con (lấy 1 phần của ma trận gốc) kích thước $r \times r$ mà có định thức khác 0, thì hạng của ma trận khi đó là r. Dễ thấy do là ma trận con, và vuông, nên $r \leqslant \min(m,n)$.

Ví dụ 3. Ma trận $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 4 \end{pmatrix}$ có định thức $\det(\boldsymbol{A}) = 0$.

Nhưng ma trận con của $\boldsymbol{A}$ là $\boldsymbol{B} = \begin{pmatrix} 2 & 3 \\ 2 & 4 \end{pmatrix}$ (lấy dòng 1 và 3, lấy cột 2 và 3) có định thức $\det(\boldsymbol{B}) = 2 \neq 0$, do đó $r = \operatorname{rank}(\boldsymbol{A}) = 2$ ($\operatorname{rank}(\boldsymbol{A})$ nghĩa là hạng của $\boldsymbol{A}$).

1.2 Ma trận nghịch đảo

Ma trận $\boldsymbol{A}^{-1}$ được gọi là **ma trận nghịch đảo** của ma trận vuông $\boldsymbol{A}$ nếu $\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{I}$. Trong đó $\boldsymbol{I}$ là ma trận đơn vị cùng kích thước với $\boldsymbol{A}$.

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}[(A_{ij})_n]^T = \frac{1}{\det(\boldsymbol{A})} \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix} \tag{1.2}$$

Trong đó, A_{ij} cũng được định nghĩa tương tự như khi tính định thức bằng khai triển theo dòng hoặc cột. Gọi $\boldsymbol{M}_{ij}$ là ma trận có được từ ma trận $\boldsymbol{A}$ khi bỏ đi hàng i và cột j của ma trận $\boldsymbol{A}$ và ký hiệu $A_{ij} = (-1)^{i+j} \det(\boldsymbol{M}_{ij})$.

Như vậy, điều kiện cần và đủ để một ma trận có nghịch đảo là định thức khác 0.

Chương 2

Discrete logarithm

2.1 Các thuật toán tính discrete logarithm

Thuật toán Baby-Step-Giant-Step (BSGS) giúp tính discrete logarithm trên nhóm cyclic với order là số nguyên tố 1.

Algorithm 1 Thuật toán Baby-Step-Giant-Step

Require: Nhóm cyclic G có order n, generator g và phần tử $h \in G$.
Ensure: Số x duy nhất thuộc $\{0, 1, \ldots, n-1\}$ thỏa $g^x = h$. $m \leftarrow \lfloor \sqrt{n} \rfloor$

1: **for** $j = 0 \rightarrow m - 1$ **do**
2: Tính g^j. Lưu (j, g^j) vào bảng.
3: **end for**
4: Tính g^{-m}.
5: $\gamma \leftarrow h$.
6: **for** $i = 0 \rightarrow m - 1$ **do**
7: a) Kiểm tra điều kiện $\gamma = g^j$ với $j = 0, 1, \ldots, m-1$.
8: b) Nếu điều kiện thỏa, trả về $im + j$.
9: c) Nếu không, đặt $\gamma \leftarrow \gamma \cdot g^{-m}$.
10: **end for**

Khi order của cyclic group là lũy thừa một số nguyên tố thì ta dùng thuật toán Pohlig-Hellman 2.

Algorithm 2 Thuật toán Pohlig-Hellman

Require: Nhóm cyclic G có order $n = p^e$, generator g và phần tử $h \in G$.
Ensure: Số x duy nhất thuộc $\{0, 1, \ldots, n-1\}$ thỏa $g^x = h$.

1: Khởi tạo $x_0 = 0$.
2: Tính $\gamma = g^{p^{e-1}}$. Theo định lý Lagrange, γ có order là p.
3: **for** $k = 0 \to e - 1$ **do**
4: a) Tính $h_k = (g^{-x_k} \cdot h)^{e-1-k}$.
5: b) Sử dụng thuật toán baby-step-giant-step, tìm $d_k \in \{0, 1, \ldots, p-1\}$ sao cho $\gamma^{d_k} = h_k$.
6: c) Tính $x_{k+1} = x_k + p^k d_k$.
7: **end for**
8: Trả về x_e là kết quả cần tìm.

Chương 3

Số học

3.1 Thặng dư chính phương

Định nghĩa 1. Số chính phương modulo p

Xét số dương nguyên tố lẻ p. Số a được gọi là **số chính phương modulo** p nếu $(a, m) = 1$ và tồn tại số x sao cho $x^2 = a \pmod{p}$.

Nói cách khác phương trình đồng dư $x^2 \equiv a \pmod{p}$ có nghiệm.

Chúng ta sử dụng kí hiệu Legendre (Legendre's symbol) để thể hiện một số a có phải là số chính phương modulo nguyên tố p không.

Định nghĩa 2. Legendre's symbol

Xét p là số nguyên tố, a là số nguyên không chia hết cho p. Khi đó kí hiệu Legendre được định nghĩa là

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{nếu } a \text{ là số chính phương modulo } p. \\ -1, & \text{nếu ngược lại.} \end{cases} \tag{3.1}$$

Chương 4

Lattice-based crypto

4.1 Introduction

Định nghĩa 1. Lattice

Xét các vector thuộc $\mathbb{R}^n$ độc lập tuyến tính $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d\}$. **Lattice** là tập

$$L = \{a_1\boldsymbol{v}_1 + \ldots + a_d\boldsymbol{v}_d : a_i \in \mathbb{Z}\} \tag{4.1}$$

Tương tự với định nghĩa không gian vector, một **tập sinh** (hay **basis**) là bất cứ tập hợp các vector độc lập tuyến tính mà sinh ra L.

Hai tập sinh luôn có cùng số phần tử. Khi đó, số vector trong tập sinh được gọi là **số chiều** (hay **dimension**).

Giả sử $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d\}$ là một cơ sở của L. Tương tự, $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d\}$ là một cơ sở khác của L.

Ta có thể viết mỗi $\boldsymbol{w}_i$ là tổ hợp tuyến tính của các vector $\boldsymbol{v}$ như sau

$$\begin{aligned}
\boldsymbol{w}_1 &= a_{11}\boldsymbol{v}_1 + a_{12}\boldsymbol{v}_2 + \ldots + a_{1d}\boldsymbol{v}_d \\
\boldsymbol{w}_2 &= a_{21}\boldsymbol{v}_1 + a_{22}\boldsymbol{v}_2 + \ldots + a_{2d}\boldsymbol{v}_d \\
&\vdots \\
\boldsymbol{w}_d &= a_{d1}\boldsymbol{v}_1 + a_{d2}\boldsymbol{v}_2 + \ldots + a_{dd}\boldsymbol{v}_d
\end{aligned}$$

Khi đó, nếu viết các vector $\boldsymbol{w}_i$ thành hàng của ma trận $\boldsymbol{W}$ và $\boldsymbol{v}_j$ thành hàng của ma trận $\boldsymbol{V}$ thì biểu diễn trên tương đương với

$$\boldsymbol{W} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix} \cdot \boldsymbol{V}$$

Đặt

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix}$$

Do $\boldsymbol{W}$ và $\boldsymbol{V}$ là các cơ sở của L nên nếu các vector $\boldsymbol{w}_i$ có thể biểu diễn qua các vector $\boldsymbol{v}_j$ thì ngược lại, các vector $\boldsymbol{v}_j$ cũng có thể được biểu diễn qua các vector $\boldsymbol{w}_i$.

Suy ra ma trận $\boldsymbol{A}$ là ma trận khả nghịch. Do $a_{ij} \in \mathbb{Z}$ theo định nghĩa lattice, định thức của $\boldsymbol{A} \in \mathbb{Z}$.

Hơn nữa, vì

$$I = \boldsymbol{A} \cdot \boldsymbol{A}^{-1} \Rightarrow 1 = \det(\boldsymbol{A}) \cdot \det(\boldsymbol{A}^{-1})$$

nên $\det(\boldsymbol{A}) = \pm 1$.

4.2 Thuật toán GGH

Phần này tham khảo trong [2]

Trong thuật toán GGH, ta chọn số nguyên tố q làm public parameter.

Sau đó chọn hai số f và g làm secret key. Hai số này phải thỏa mãn các điều kiện

$$f < \sqrt{q/2}, \quad \sqrt{q/4} < g < \sqrt{q/2}, \quad \gcd(f, qg) = 1$$

Tính $h = f^{-1}g \pmod{q}$. Khi đó public key là h.

Encryption. Để encrypt message m với số random r thỏa mãn

$$0 < m < \sqrt{q/4}, \quad 0 < r < \sqrt{q/2}$$

Ta tính $e = rh + m \pmod{q}$ là ciphertext với $0 < e < q$.

Decryption. Để decrypt ciphertext e ta tính

$$a = fe \pmod{q}, \quad b = f^{-1}a \pmod{g}$$

Lưu ý f^{-1} là nghịch đảo modulo g. Khi đó $b \equiv m$ là message ban đầu.

Chứng minh. Để chứng minh rằng số b sau khi tính toán bằng chính xác m ban đầu ta cần xem xét điều kiện của secret key và public key.

Đầu tiên ta có

$$a \equiv fe \equiv f(rh + m) = f(rf^{-1}g + m) = rg + fm \pmod{q}$$

Từ điều kiện của f, g, r và m ta có

$$rg + fm < \sqrt{\frac{q}{2}} \cdot \sqrt{\frac{q}{2}} + \sqrt{\frac{q}{2}} \cdot \sqrt{\frac{q}{4}} < q$$

Nói cách khác $rg + fm$ giữ nguyên giá trị trong phép modulo q, hay $a \equiv rg + fm$.

Suy ra $b = f^{-1}a = f^{-1}(rg + fm) = m \pmod{g}$ (giá trị a không thay đổi khi chuyển từ modulo q sang modulo g). Do $0 < m < \sqrt{q/4}$ và $\sqrt{q/4} < g < \sqrt{q/2}$ nên $m < g$. Nói cách khác b bằng đúng m ban đầu. □

Để tấn công hệ mật mã này ta xây dựng lattice. Để ý rằng $h = f^{-1}g \pmod{q}$, hay $fh + kq = g$ với $k \in \mathbb{Z}$.

Ta thấy rằng $f \cdot (h, 1) + k \cdot (q, 0) = (g, f)$. Như vậy lattice gồm hai vector $(h, 1)$ và $(q, 0)$. Thuật toán tối giản Gauss sẽ hoạt động trong trường hợp này (số chiều bằng 2).

Chương 5

Bài toán đếm Polya

5.1 Bổ đề Burnside

Lớp tương đương

Xét nhóm G và tập hợp M. Khi đó hai phần tử m và n thuộc M được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $m = gn$.

Nhận xét 1

Quan hệ giữa các phần tử như trên là quan hệ tương đương.

Chứng minh. Để chứng minh một quan hệ là tương đương, ta cần chứng minh tính phản xạ, đối xứng và bắc cầu.

Đối với tính phản xạ, mọi vector đều có quan hệ với chính nó qua phần tử đơn vị $e \in G$.

Đối với tính đối xứng, nếu m có quan hệ với n thì tồn tại $g \in G$ sao cho $m = gn$. Theo tính chất nhóm thì tồn tại phần tử g^{-1} là nghịch đảo của g trong G. Do đó $g^{-1}m = n$. Nói cách khác n cũng có quan hệ với m. Như vậy ta có tính đối xứng.

Đối với tính bắc cầu, nếu m có quan hệ với n thì tồn tại $g_1 \in G$ sao cho $m = g_1 n$. Tiếp theo, nếu n có quan hệ với p thì tồn tại $g_2 \in G$ sao cho $n = g_2 p$. Suy ra $m = g_1 n = g_1(g_2 p) = (g_1 g_2)p$. Do $g_1, g_2 \in G$ nên $g_1 g_2 \in G$. Như vậy m cũng có quan hệ với p nên quan hệ có tính bắc cầu.

Vậy quan hệ được định nghĩa như trên là quan hệ tương đương. □

Tác động nhóm lên vector

Xét nhóm G và không gian vector $\mathbb{F}_2^n$, $n \in \mathbb{N}$. Khi đó hai vector $\boldsymbol{x}$ và $\boldsymbol{y}$ thuộc $\mathbb{F}_2^n$ được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $\boldsymbol{x} = g\boldsymbol{y}$.

Ví dụ, xét nhóm hoán vị $\mathcal{S}_3$. Giả sử các vector trong $\mathbb{F}_2^3$ có dạng

$$\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{F}_2^3.$$

Khi đó vector $(1,0,0)$ có quan hệ với $(0,0,1)$ với hoán vị $(1,3)(2)$. Cụ thể là $(x_1, x_2, x_3) \xrightarrow{(1,3)(2)} (x_3, x_2, x_1)$.

Tương tự, vector $(1,0,0)$ cũng có quan hệ với $(0,1,0)$ với hoán vị $(1,2)(3)$. Thêm nữa, vector $(1,0,0)$ có quan hệ với chính nó qua hoán vị đồng nhất $(1)(2)(3)$.

Trong môn toán rời rạc ta đã biết, nếu một tập có quan hệ tương đương thì ta có thể phân các phần tử của tập đó vào các lớp tương đương rời nhau. Nghĩa là nếu hai phần tử có quan hệ với nhau thì vào cùng một lớp tương đương. Từ phần trên ta đã biết rằng dưới tác động của nhóm, các phần tử trong tập hợp bất kì sẽ phân bổ thành các lớp tương đương.

Câu hỏi đặt ra là, có bao nhiêu lớp tương đương như vậy?

Để giải quyết vấn đề này ta sử dụng bổ đề Burnside.

Nhóm $\mathcal{S}_3$ có các hoán vị

$$\mathcal{S}_3 = \{(1)(2)(3), (1,2)(3), (1,3)(2), (2,3)(1), (1,3,2), (1,2,3)\}$$

Lần lượt xét từng hoán vị. Đầu tiên, với $(1)(2)(3)$ thì các phần tử trong vector đứng yên. Do đó dưới tác động của hoán vị này, x_1 biến thành x_1, x_2 biến thành x_2 và x_3 biến thành x_3. Số cách chọn cho mỗi x_i là 2 nên theo quy tắc nhân ta có $2^3 = 8$ cách.

Tiếp theo, với hoán vị $(1,2)(3)$ thì $x_1 \to x_2$, $x_2 \to x_1$ và $x_3 \to x_3$. Do đó x_1 và x_2 có cùng giá trị nên có 2 cách chọn, x_3 cũng có 2 cách chọn nên tổng số cách là $2 \cdot 2 = 4$. Hoán vị $(1,3)(2)$ và $(2,3)(1)$ tương tự.

Với hoán vị $(1,2,3)$ thì $x_1 \to x_2$, $x_2 \to x_3$ và $x_3 \to x_1$ nên $x_1 = x_2 = x_3$, có 2 cách chọn trong trường hợp này. Hoán vị $(1,3,2)$ tương tự.

Như vậy, theo bổ đề Burnside, số lớp tương đương các vector trong $\mathbb{F}_2^3$ là

$$t(\mathcal{S}_3) = \frac{1}{6}(1 \cdot 2^3 + 3 \cdot 2^2 + 2 \cdot 2) = 4$$

Thật vậy, ta có thể chia các vector thành 4 lớp tương đương là $\{000\}$, $\{001, 010, 011\}$, $\{011, 101, 110\}$, $\{111\}$.

Ngoài nhóm $\mathcal{S}_3$ ra còn các nhóm khác cũng tác động lên các vector. Một số nhóm hay được sử dụng là:

1. Nhóm general linear: gồm các ma trận khả nghịch $n \times n$ trên $\mathbb{F}_2$. Tác động nhóm lúc này là phép nhân ma trận $\boldsymbol{A} \in GL(n,2)$ với vector $\boldsymbol{x} \in \mathbb{F}_2^n$, hay $\boldsymbol{A} \cdot \boldsymbol{x}$.
2. Nhóm general affine: gồm các ma trận khả nghịch $n \times n$ trên $\mathbb{F}_2$ và vector bất kì trong $\mathbb{F}_2^n$. Tác động nhóm lúc này là biến đổi affine $\boldsymbol{A} \cdot \boldsymbol{x} + \boldsymbol{b}$ với $\boldsymbol{A} \in GL(n,2)$ và $\boldsymbol{b} \in \mathbb{F}_2^n$.

Cần nhắc lại một chút, số lượng phần tử của nhóm $GL(n,2)$ là

$$(2^n-1)\cdot(2^n-2)\cdots(2^n-2^{n-1})$$

Khi $n=3$ thì $|GL(3,2)| = (2^3-1)\cdot(2^3-2)\cdot(2^3-4) = 168$ ma trận.

Tác động nhóm lên hàm boolean

Ta tiếp tục xét nhóm G và không gian vector $\mathbb{F}_2^n$, $n \in \mathbb{N}$. Khi đó hai hàm boolean n biến $f(x_1,\ldots,x_n)$ và $g(x_1,\ldots,x_n)$ được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $g(\boldsymbol{x}) = f(g\boldsymbol{x})$ với mọi $\boldsymbol{x} \in \mathbb{F}_2^n$.

Ta cũng xét hoán vị $\mathcal{S}_3$. Ta cũng lần lượt xét các phần tử của nhóm.

Đặt $f_0, f_1, \ldots, f_7$ lần lượt là các giá trị hàm f với các vector $\boldsymbol{x} \in \mathbb{F}_2^3$.

Đầu tiên, với (1)(2)(3), ta có bảng chuyển vector như hình 5.1.

x_1	x_2	x_3	f	(1)(2)(3)	x_1	x_2	x_3	f
0	0	0	f_0	→	0	0	0	f_0
0	0	1	f_1		0	0	1	f_1
0	1	0	f_2		0	1	0	f_2
0	1	1	f_3		0	1	1	f_3
1	0	0	f_4		1	0	0	f_4
1	0	1	f_5		1	0	1	f_5
1	1	0	f_6		1	1	0	f_6
1	1	1	f_7		1	1	1	f_7

Hình 5.1: Hoán vị (1)(2)(3)

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_1$, ..., $f_7 \to f_7$ nên có 8 chu trình. Vậy số lượng cách chọn là 2^8.

Tiếp theo, xét các hoán vị dạng $(1)(2,3)$, ta có bảng chuyển vector như hình 5.2.

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_2 \to f_1$, $f_3 \to f_3$, $f_4 \to f_4$, $f_5 \to f_6 \to f_5$, $f_7 \to f_7$. Ở đây có 6 chu trình nên số cách chọn là 2^6.

Tiếp theo ta xét các hoán vị dạng $(1,2,3)$ (hình 5.3).

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_2 \to f_4 \to f_1$, $f_3 \to f_6 \to f_5 \to f_3$, $f_7 \to f_7$ nên ở đây có 4 chu trình. Số cách chọn là 2^4.

Như vậy theo bổ đề Burnside, số lớp hàm bool tương đương dưới tác động của nhóm $\mathcal{S}_3$ là

$$t(\mathcal{S}_3) = \frac{1}{6}(2^8 + 3\cdot 2^6 + 2\cdot 2^4) = 80.$$

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

$\xrightarrow{(1)(2,3)}$

x_1	x_3	x_2	f
0	0	0	f_0
0	1	0	f_2
0	0	1	f_1
0	1	1	f_3
1	0	0	f_4
1	1	0	f_6
1	0	1	f_5
1	1	1	f_7

Hình 5.2: Hoán vị $(1)(2,3)$

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

$\xrightarrow{(1,2,3)}$

x_2	x_3	x_1	f
0	0	0	f_0
0	1	0	f_2
1	0	0	f_4
1	1	0	f_6
0	0	1	f_1
0	1	1	f_3
1	0	1	f_5
1	1	1	f_7

Hình 5.3: Hoán vị $(1,2,3)$

Chương 6

Giải tích

Định nghĩa 1. Dãy Cauchy

Dãy (x_n) được gọi là dãy Cauchy nếu với mọi $\varepsilon > 0$, tồn tại $N_0 \in \mathbb{N}$ sao cho, với mọi $m, n > N_0$ thì $|x_m - x_n| < \varepsilon$.

Định lí 1. Tiêu chuẩn Cauchy

Dãy số (x_n) có giới hạn hữu hạn khi và chỉ khi nó là dãy Cauchy.

Định lí 2. Bổ đề Fermat

Cho f là một hàm số có đạo hàm trên (a, b). Nếu $x_0 \in (a, b)$ là một điểm cực trị của f thì ta có $f'(x_0) = 0$.

Chứng minh. Ta chứng minh trong trường hợp x_0 là điểm cực tiểu. Trường hợp điểm cực đại tương tự.

Hàm f có đạo hàm trên (a, b) nên tại điểm x_0 nó có đạo hàm bên trái và đạo hàm bên phải, và hai đạo hàm này bằng nhau.

Ta có $f'(x_0^+) = \lim\limits_{x \to x_0^+} \dfrac{f(x) - f(x_0)}{x - x_0}$. Vì $x \to x_0^+$ nghĩa là $x > x_0$ (x tiến tới x_0 từ bên phải), và do x_0 là cực tiểu $f(x) - f(x_0) \geqslant 0$ nên phân số dưới dấu giới hạn lớn hơn 0. Suy ra $f'(x_0^+) \geqslant 0$.

Hoàn toàn tương tự ta chứng minh được $f'(x_0^-) \leqslant 0$. Và do $f'(x_0^+) = f'(x_0^-) = f'(x_0)$ nên $f'(x_0) = 0$.

Ta có điều phải chứng minh. □

Định lí 3. Định lí Rolle

Xét hàm số f liên tục trên đoạn $[a, b]$, có đạo hàm trên khoảng (a, b) và $f(a) = f(b)$. Khi đó tồn tại c thuộc (a, b) sao cho $f'(c) = 0$.

Định lí 4. Định lí Lagrange

Xét hàm số f liên tục trên đoạn $[a, b]$, có đạo hàm trên khoảng (a, b). Khi đó tồn tại c thuộc (a, b) sao cho $f'(c)(b - a) = f(b) - f(a)$.

Định nghĩa 2. Hàm lõm

Hàm số f liên tục trên khoảng $\mathbb{I}$ nếu với mọi α, β mà $\alpha + \beta = 1$ ta đều có

$$f(\alpha x + \beta y) \leqslant \alpha f(x) + \beta f(y), \quad \forall x, y \in \mathbb{I} \tag{6.1}$$

Chương 7

Quantum computing

7.1 Qubit và toán tử quantum

Trên máy tính hiện nay, đơn vị xử lý cơ bản là bit (0 hoặc 1). Trong máy tính lượng tử, đơn vị tính toán là qubit (quantum bit).

Qubit

Mỗi qubit $|\psi\rangle$ được biểu diễn dưới dạng tổ hợp tuyến tính của cơ sở gồm $|0\rangle = (1,0)$ và $|1\rangle = (0,1)$. Khi đó qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Ở đây $\alpha, \beta \in \mathbb{C}$ (tập số phức).

Tích của n qubit là các tổ hợp $|0,0,\ldots,0\rangle$, $|0,0,\ldots,1\rangle$, ..., $|1,1,\ldots,1\rangle$. Ta cũng ký hiệu $|0\rangle \otimes |1\rangle = |01\rangle$.

Ví dụ 1. Nếu $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ và $|\Psi\rangle = \gamma|0\rangle + \delta|1\rangle$ thì

$$|\psi\rangle \otimes |\Psi\rangle = (\alpha|0\rangle + \beta|1\rangle) \otimes (\gamma|0\rangle + \delta|1\rangle) = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle$$

Tiếp theo là **toán tử quantum** và tương ứng với nó là các cổng (gate) trên mạch.

Toán tử quantum tác động lên một qubit hoặc tích của nhiều qubit.

Qubit có dạng $|\psi\rangle = a|0\rangle + b|1\rangle$. Ta có thể viết hệ số dưới dạng vector cột $\begin{pmatrix} a \\ b \end{pmatrix}$. Khi đó, toán tử quantum sẽ là một ma trận 2×2 biến đổi vector trên thành vector mới $\begin{pmatrix} c \\ d \end{pmatrix}$ tương ứng với qubit $|\Psi\rangle = c|0\rangle + d|1\rangle$.

Nói cách khác, đặt toán tử quantum là ma trận $\mathcal{U} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ thì ta có

$$|\psi\rangle \to |\Psi\rangle = \mathcal{U}|\psi\rangle, \quad \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

Các toán tử quantum thường gặp

1. **Toán tử đồng nhất**. Toán tử đồng nhất identity giữ nguyên qubit đầu vào. Ma trận tương ứng là ma trận đơn vị $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

2. **Toán tử NOT**. Toán tử NOT có ma trận tương ứng là $\text{NOT} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Khi đó $\text{NOT}|\psi\rangle = b|0\rangle + a|1\rangle$ với $x \in \{0, 1\}$.

Khi qubit là $|0\rangle$ hoặc $|1\rangle$, tác dụng của toán tử NOT là phép XOR nên ta có $\text{NOT}|x\rangle = |x \oplus 1\rangle$.

3. **Toán tử Hadamard**. Đây là một toán tử đặc biệt và được quan tâm nhiều.

Ma trận của toán tử Hadamard là $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Ví dụ 2. Xét qubit $|\psi\rangle = a|0\rangle + b|1\rangle$, toán tử Hadamard tương ứng với phép nhân ma trận

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(a+b) \\ \frac{1}{\sqrt{2}}(a-b) \end{pmatrix}$$

Ta chuyển cột kết quả về lại dạng tổ hợp tuyến tính thì cổng Hadamard hoạt động trên qubit $|\psi\rangle = a|0\rangle + b|1\rangle$ cho kết quả là

$$H|\psi\rangle = H(a|0\rangle + b|1\rangle) = \frac{1}{\sqrt{2}}(a+b)|0\rangle + \frac{1}{\sqrt{2}}(a-b)|1\rangle$$

Nếu $|\psi\rangle \equiv |0\rangle$ thì tương đương với $a = 1, b = 0$. Ta có $H|0\rangle = \dfrac{|0\rangle + |1\rangle}{\sqrt{2}}$.

Nếu $|\psi\rangle \equiv |1\rangle$ thì tương đương với $a = 0, b = 1$. Ta có $H|1\rangle = \dfrac{|0\rangle - |1\rangle}{\sqrt{2}}$.

Tổng quát ta nhận thấy, với $x \in \{0, 1\}$ thì $H|x\rangle = \dfrac{|0\rangle + (-1)^x|1\rangle}{\sqrt{2}}$.

Ta thấy rằng toán tử ngược của toán tử Hadamard là chính nó.

4. **Toán tử control**. Đây là toán tử thường được dùng nhất khi tính toán trên tích của nhiều qubit.

Như đã xem xét ở trên, tích của n qubit sẽ có 2^n phần tử tương ứng các bộ $|0, 0, \ldots, 0, 0\rangle$, $|0, 0, \ldots, 0, 1\rangle$, ... Do đó toán tử control sẽ là ma trận kích thước $2^n \times 2^n$.

Gọi $\mathcal{U} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ là toán tử tác động lên một qubit (ví dụ như 3 toán tử

đã đề cập). Xét hai qubit là $|x\rangle = a|0\rangle + b|1\rangle$ và $|y\rangle = c|0\rangle + d|1\rangle$. Từ phía trên

$$|x\rangle \otimes |y\rangle = ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle$$

Khi đó toán tử control có dạng ma trận là

$$c\mathcal{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c_{11} & c_{12} \\ 0 & 0 & c_{21} & c_{22} \end{pmatrix}$$

Hay viết dưới dạng ma trận khối là $c\mathcal{U} = \begin{pmatrix} I & \mathcal{O} \\ \mathcal{O} & \mathcal{U} \end{pmatrix}$.

Ta cũng viết tích $|x\rangle \otimes |y\rangle$ dưới dạng vector cột (4 phần tử). Khi đó

$$\mathcal{U}(|x\rangle \otimes |y\rangle) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c_{11} & c_{12} \\ 0 & 0 & c_{21} & c_{22} \end{pmatrix} \cdot \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ c_{11} \cdot bc + c_{12} \cdot bd \\ c_{21} \cdot bc + c_{22} \cdot bd \end{pmatrix}$$

Hai phần tử đầu của vector kết quả không thay đổi, còn phần sau có "một phần" là $\mathcal{U}|y\rangle$. Khi viết lại kết quả dưới dạng qubit thì

$$ac|00\rangle + ad|01\rangle + (c_{11} \cdot bc + c_{12} \cdot bd)|10\rangle + (c_{21} \cdot bc + c_{22} \cdot bd)|11\rangle$$

Ta có một số nhận xét sau đây.

Nếu $|x\rangle \equiv |0\rangle$, tức là $a = 1, b = 0$ thì tích trên tương ứng với $c|00\rangle + d|01\rangle + 0|10\rangle + 0|11\rangle = |0\rangle \otimes (c|0\rangle + d|1\rangle) = |x\rangle \otimes |y\rangle$.

Nếu $|x\rangle \equiv |1\rangle$, tức là $a = 0, b = 1$ thì tích trên tương ứng với $0|00\rangle + 0|01\rangle + (c_{11}c + c_{12}d)|10\rangle + (c_{21}c + c_{22}d)|11\rangle = |1\rangle \otimes ((c_{11}c + c_{12}d)|0\rangle + (c_{21}c + c_{22}d)|1\rangle) = |1\rangle \otimes \mathcal{U}|y\rangle = |x\rangle \otimes \mathcal{U}|y\rangle$.

Tổng kết lại, với $x \in \{0, 1\}$ thì

- nếu $x = 0$ thì $|x\rangle \otimes |y\rangle \to |x\rangle \otimes |y\rangle$.
- nếu $x = 1$ thì $|x\rangle \otimes |y\rangle \to |x\rangle \otimes \mathcal{U}|y\rangle$.

Tùy vào x là 0 hay 1 mà toán tử quantum $\mathcal{U}$ sẽ bị bỏ qua hoặc xem xét. Ở đây qubit $|x\rangle$ đóng vai trò điều khiển nên đây được gọi là toán tử control.

5. **Toán tử control CNOT (Control NOT)**. Toán tử quantum CNOT có ma trận là

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} I & \mathcal{O} \\ \mathcal{O} & \text{NOT} \end{pmatrix}$$

Qubit $|x\rangle$ với $x \in \{0, 1\}$ đóng vai trò control cho qubit $|y\rangle$. Khi $x \equiv 0$ thì y giữ nguyên, hay $|y \oplus 0\rangle = |y \oplus x\rangle$. Khi $x \equiv 1$ thì áp dụng cổng NOT bên trên, khi đó y biến đổi thành $y \oplus 1 = y \oplus x$.

Một ví dụ về qubit trong NSUCRYPTO 2022

Đố vui: nếu ta có qubit là $\alpha|0\rangle + \beta|1\rangle$, hãy xây dựng mạch logic để biến đổi qubit trên thành $\alpha|000\rangle + \beta|111\rangle$.

Giải: sử dụng toán tử Hadamard và NOT. Ta có

$$NOT(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle$$

và

$$\mathcal{H}(\alpha|0\rangle + \beta|1\rangle) = \frac{1}{\sqrt{2}}(\alpha+\beta)|0\rangle + \frac{1}{\sqrt{2}}(\alpha-\beta)|1\rangle$$

Toán tử $CNOT$ được biểu diễn bởi ma trận

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Giả sử ta có hai qubit là $|\alpha\rangle = x|0\rangle + y|1\rangle$ và $|\beta\rangle = z|0\rangle + t|1\rangle$.
Khi đó $|\alpha\rangle \otimes |\beta\rangle = xz|00\rangle + xt|01\rangle + yz|10\rangle + yt|11\rangle$.
Qua toán tử $CNOT$ ta có

$$\begin{aligned} CNOT(|\alpha\rangle \otimes |\beta\rangle) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} xz \\ xt \\ yz \\ yt \end{pmatrix} = \begin{pmatrix} xz \\ xt \\ yt \\ yz \end{pmatrix} \\ &= xz|00\rangle + xt|01\rangle + yt|10\rangle + yz|11\rangle \end{aligned}$$

Trường hợp $|\alpha\rangle = |0\rangle$ thì $x = 1, y = 0$. Khi đó ta có tương đương

$$CNOT(|0\rangle \otimes |\beta\rangle) = z|00\rangle + t|01\rangle = |0\rangle \otimes (z|0\rangle + t|1\rangle) = |0\rangle \otimes |\beta\rangle.$$

Trường hợp $|\alpha\rangle = |1\rangle$ thì $x = 0, y = 1$. Khi đó ta có tương đương

$$CNOT(|1\rangle) \otimes |\beta\rangle) = t|10\rangle + z|11\rangle = |1\rangle \otimes (t|0\rangle + z|1\rangle) = |1\rangle \otimes NOT(|\beta\rangle)$$

Nói cách khác, nếu $x \in \{0, 1\}$ thì

$$CNOT(|x\rangle \otimes |\beta\rangle) = \begin{cases} |x\rangle \otimes |\beta\rangle, & \text{nếu } x = 0 \\ |x\rangle \otimes NOT(|\beta\rangle), & \text{nếu } x = 1 \end{cases}$$

Do đó các toán tử có ma trận $\begin{pmatrix} I_n & \mathcal{O} \\ \mathcal{O} & \mathcal{U} \end{pmatrix}$ được gọi là toán tử kiểm soát (controlled).

Một trường hợp riêng nữa là khi $\beta = 0$ hoặc $\beta = 1$. Khi đó, với toán tử NOT bên trên ta suy ra

$$CNOT(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |x \oplus y\rangle$$

Chương 8

Lý thuyết xác suất

8.1 Định nghĩa xác suất

Định nghĩa 1. Định nghĩa cổ điển của xác suất

Định nghĩa thống kê của xác suất nói rằng, giả sử trong một phép thử có n khả năng có thể xảy ra. Xét một biến cố A xảy ra khi thực hiện phép thử có k khả năng xảy ra. Khi đó xác suất của biến cố A ký hiệu là $P(A)$ và được tính

$$P(A) = \frac{k}{n}$$

Dễ thấy, do biến cố A là một trường hợp nhỏ trong tổng thể tất cả trường hợp khi thực hiện phép thử, do đó $0 \leq k \leq n$. Nghĩa là

$$0 \leq P(A) \leq 1$$

với mọi biến cố A bất kì.

Ví dụ 1. Xét phép thử tung hai đồng xu. Gọi A là biến cố hai đồng xu cùng mặt.

Ta ký hiệu S là đồng xu sấp, N là đồng xu ngửa. Khi đó các trường hợp có thể xảy ra của phép thử là $S-S$, $S-N$, $N-S$, $N-N$ (4 trường hợp).

Trong khi đó, các trường hợp có thể xảy ra của biến cố A là $S-S$, $N-N$ (2 trường hợp).

Kết luận: $P(A) = \frac{2}{4} = \frac{1}{2}$

Chúng ta gọi tập hợp tất cả các trường hợp khi thực hiện phép thử là **không gian mẫu** và ký hiệu là Ω. Mỗi phần tử trong không gian mẫu được gọi là **biến cố sơ cấp**. Trong ví dụ trên, $\Omega = \{S-S, S-N, N-S, N-N\}$.

Tập hợp các trường hợp có thể xảy ra của biến cố gọi là **không gian biến cố** và ký hiệu là Ω_A. Trong ví dụ trên, $\Omega_A = \{S-S, N-N\}$.

Như vậy, $P(A) = \frac{|\Omega_A|}{|\Omega|}$

Ví dụ 2. Tung hai con súc sắc cân đối và đồng chất. Tính xác suất tổng số nút của hai con súc sắc bằng 4.

Việc tung mỗi con súc sắc có 6 trường hợp. Do đó $|\Omega| = 6^2 = 36$

Gọi A là biến cố tổng số nút của hai con súc sắc bằng 4. Ta có các trường hợp là $4 = 1 + 3 = 3 + 1 = 2 + 2$ (3 trường hợp).

Như vậy $|\Omega_A| = 3$ và $P(A) = \frac{3}{36} = \frac{1}{12}$

Định nghĩa 2. Biến cố xung khắc

Hai biến cố được gọi là **xung khắc** nếu biến cố này xảy ra thì biến cố kia chắc chắn không xảy ra. Nói cách khác giao của chúng bằng rỗng.

Khi đó, nếu A và B là hai biến cố xung khắc,

$$P(A + B) = P(A) + P(B)$$

Ta còn có thể ký hiệu $P(A + B)$ là $P(A \cup B)$ (hợp hai biến cố).

Định nghĩa 3. Biến cố độc lập

Hai biến cố được gọi là **độc lập** nếu việc xảy ra của biến cố này không ảnh hưởng đến việc xảy ra của biến cố kia.

Khi đó, nếu A và B là hai biến cố độc lập thì

$$P(AB) = P(A)P(B)$$

8.2 Xác suất có điều kiện

Xét hai tập hợp A và B. Số phần tử của phép hợp hai tập hợp trong trường hợp tổng quát được tính như sau:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Tương tự, xác suất của phép cộng xác suất đối với hai biến cố có giao khác rỗng là:

$$P(A + B) = P(A) + P(B) - P(A \cap B)$$

Xét các tập hợp A_1, A_2, ..., A_n. Khi đó, số phần tử khi hợp các tập hợp này là:

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_n| = |A_1| + |A_2| + \cdots + |A_n| - \sum_{i,j} |A_i \cap A_j| \\ + \sum_{i,j,k} |A_i \cap A_j \cap A_k| + \cdots \\ = \sum_{i=1}^{n} (-1)^{i+1} \sum_{j_1, j_2, \cdots, j_i} |A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i}|\end{aligned}$$

Tương tự, ta có phép cộng xác suất:

Định lí 1. Phép cộng xác suất mở rộng

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} (-1)^{i+1} \sum_{j_1, j_2, \cdots, j_i} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i})$$

Định lí 2. Xác suất có điều kiện

Xét hai biến cố A và B. Khi đó xác suất xảy ra của biến cố B với điều kiện biến cố A xảy ra là:

$$P(B|A) = \frac{P(AB)}{P(A)} \tag{8.1}$$

Lúc này, A và B không độc lập.

Tổng quát, nếu n biến cố A_i, $i = 1, \ldots, n$ không độc lập thì:

$$\begin{aligned}P(A_1 A_2 \cdots A_n) =& P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 A_1) \cdots \\ & P(A_n|A_1 A_2 \cdots A_{n-1})\end{aligned}$$

Ví dụ 3. Xét hai câu hỏi trắc nghiệm có 4 lựa chọn. Tính xác suất học sinh trả lời đúng câu thứ hai với điều kiện câu đầu trả lời sai.

Giải. Gọi A là biến cố câu đầu tiên học sinh trả lời sai. $P(A) = \frac{3}{4}$

Gọi B là biến cố câu thứ hai học sinh trả lời đúng. $P(B) = \frac{1}{4}$.

Do A và B là hai biến cố độc lập nên $P(AB) = P(A)P(B) = \frac{3}{16}$

Như vậy, xác suất học sinh trả lời đúng câu thứ hai với điều kiện câu đầu trả lời đúng là: $P(B|A) = \frac{P(AB)}{P(A)} = \frac{3/16}{3/4} = \frac{1}{4}$

8.3 Công thức xác suất đầy đủ

Định nghĩa 4. Hệ biến cố đầy đủ

Xét phép thử có không gian mẫu là Ω. Một hệ các biến cố A_1, A_2, ..., A_n được gọi là **đầy đủ** nếu chúng thỏa các điều kiện:

- $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$
- $A_i \cap A_j = \emptyset$ với mọi $i \neq j$

Định lí 3. Công thức xác suất đầy đủ

Gọi $A_1, A_2, \ldots, A_n$ là một hệ biến cố đầy đủ. Khi đó, với biến cố B bất kì trong phép thử:

$$P(B) = P(A_1) \cdot P(B|A_1) + \cdots P(A_n) \cdot P(B|A_n) \tag{8.2}$$

Định lí 4. Công thức Bayes

Xét hệ có n biến cố đầy đủ $\{A_1, A_2, \ldots, A_n\}$.

Với biến cố B bất kì thì:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

với $1 \leq i \leq n$.

Chương 9

Biến ngẫu nhiên

9.1 Biến ngẫu nhiên

Xét phép thử với không gian mẫu Ω. Với mỗi biến cố sơ cấp $\omega \in \Omega$ ta liên kết với một số thực $\xi(\omega) \in \mathbb{R}$ thì ξ được gọi là **biến ngẫu nhiên** (BNN).

Định nghĩa 1. Biến ngẫu nhiên

Biến ngẫu nhiên ξ của một phép thử với không gian mẫu Ω là ánh xạ:

$$\xi = \xi(\omega), \quad \omega \in \Omega$$

Giá trị $\xi(\omega)$ được gọi là một giá trị của biến ngẫu nhiên ξ.

- Nếu $\xi(\Omega)$ là một tập hữu hạn $\{\xi_1, \xi_2, \ldots, \xi_n\}$ hay tập vô hạn đếm được thì ξ được gọi là **biến ngẫu nhiên rời rạc**.
- Nếu $\xi(\Omega)$ là một khoảng của $\mathbb{R}$ hay toàn bộ $\mathbb{R}$ thì ξ được gọi là **biến ngẫu nhiên liên tục**.

Định nghĩa 2. Hàm phân phối

Hàm phân phối của biến ngẫu nhiên ξ là hàm số $F(x)$, xác định bởi:

$$F(x) = P(\xi \leq x), \quad x \in \mathbb{R} \tag{9.1}$$

Ở đây ta viết gọn $P(\xi \leq x)$ từ $P(\{\omega : \xi(\omega) \leq x\})$. Tập hợp $\{\omega : \xi(\omega) \leq x\}$ có thể không thuộc một biến cố nào, do đó có thể là tập rỗng (ứng với xác suất là 0).

9.2 Tính chất của hàm phân phối

Tính chất 1. Hàm phân phối $F(x)$ không giảm trên mọi đoạn thẳng.

Chứng minh. Đặt $x_2 > x_1$. Ta thấy rằng

$$\{\xi \le x_2\} = \{\xi \le x_1\} + \{x_1 < \xi \le x_2\},$$

Do đó nếu ta lấy xác suất thì cũng có

$$P(\xi \le x_2) = P(\xi \le x_1) + P(x_1 < \xi \le x_2)$$

Xác suất luôn không âm, hay $P(x_1 < \xi \le x_2) \ge 0$, suy ra $P(\xi \le x_2) \ge P(\xi \le x_1)$, hay $F(x_2) \ge F(x_1)$. □

Tính chất 2. $\lim\limits_{x\to-\infty} F(x) = 0$.

Tính chất 3. $\lim\limits_{x\to+\infty} F(x) = 1$.

Tính chất 4. Hàm phân phối $F(x)$ liên tục phải trên toàn trục số.

Để chứng minh các tính chất 2, 3, 4 chúng ta cần các tiên đề của sự liên tục (continunity axioms) và sẽ không đề cập ở đây.

9.3 Biến ngẫu nhiên rời rạc

Cho BNN rời rạc $\xi = \xi(\omega)$, $\xi = \{a_1, a_2, \ldots, a_n, \ldots\}$. Giả sử $a_1 < a_2 < \ldots < a_n < \ldots$ với xác suất tương ứng là $P(\xi = a_i) = p_i$, $i = 1, 2, \ldots$

Ta có thể biểu diễn biến ngẫu nhiên và xác suất tương ứng của nó bằng bảng phân phối xác suất của ξ.

ξ	a_1	a_2	$\cdots$	a_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

Rõ ràng rằng $p_n \ge 0$ với mọi n. Hơn nữa

$$\sum_{n=1}^{\infty} p_n = 1$$

Không gian mẫu lúc này là hợp của các tập biến ngẫu nhiên rời rạc:

$$\Omega = \{\xi = a_1\} \cup \{\xi = a_2\} \cup \ldots$$

Các biến ngẫu nhiên xung khắc nhau (vì ξ không thể nhận hai giá trị khác nhau cùng lúc), do đó xác suất cả không gian mẫu là

$$1 = P(\Omega) = P(\xi = a_1) + P(\xi = a_2) + \ldots = p_1 + p_2 + \ldots$$

Định nghĩa 3. Phân phối nhị thức

Biến ngẫu nhiên ξ được gọi là có **phân phối nhị thức** với tham số p, n, với $p \in (0,1)$ và n là số tự nhiên, nếu ξ nhận các giá trị $0, 1, \ldots, n$ và

$$P(\xi = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n \tag{9.2}$$

Ở đây $q = 1 - p$.

Ví dụ 1. Một bài kiểm tra có 100 câu hỏi trắc nghiệm bốn đáp án. Xác suất chọn ngẫu nhiên đúng đáp án của mỗi câu hỏi thì bằng nhau và bằng $\frac{1}{4}$.

Ở đây xác suất chọn ngẫu nhiên đúng đáp án của một câu hỏi bất kì là $p = \frac{1}{4}$, và số lượng câu hỏi là $n = 100$.

Gọi ξ là biến ngẫu nhiên số câu hỏi trả lời đúng. Khi đó ξ nhận các giá trị $0, 1, \ldots, 100$.

Do đó bài toán này có phân phối nhị nhức và

$$P(\xi = k) = C_{100}^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{100-k}$$

Định nghĩa 4. Phân phối Poisson

Biến ngẫu nhiên ξ được gọi là có **phân phối Poisson** với tham số λ, nếu ξ nhận các giá trị $0, 1, \ldots, n$ và

$$P(\xi = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots, n \tag{9.3}$$

Tham số λ thể hiện số lần trung bình mà một sự kiện xảy ra trong một khoảng thời gian nhất định. Khi đó, nếu một biến ngẫu nhiên có số lần xuất hiện trung bình của một sự kiện trong thời gian t thì nó có phân phối Poisson với tham số λt, với λ là số lần trung bình trong một đơn vị thời gian.

9.4 Biến ngẫu nhiên liên tục

Định nghĩa 5. Biến ngẫu nhiên liên tục

Biến ngẫu nhiên ξ được gọi là **liên tục**, nếu nó nhận giá trị tại mọi điểm thuộc một đoạn liên tục nào đó trên trục số, và tồn tại một hàm số không âm $p(x)$ sao cho với mọi đoạn $[a, b]$ (hữu hạn hoặc vô hạn) ta có

$$P(a \leq \xi \leq b) = \int_a^b p(x)\,dx \tag{9.4}$$

Hàm $p(x)$ được gọi là **hàm mật độ** của biến ngẫu nhiên ξ.

Tương tự biến ngẫu nhiên rời rạc, $p(x) \geq 0$ với mọi $x \in \mathbb{R}$ và khi hai cận là vô cực thì biến ngẫu nhiên bao quát toàn bộ không gian mẫu. Nghĩa là

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

Từ định nghĩa của hàm phân phối $F(x) = P(\xi \leq x)$ ta có hai tính chất của hàm mật độ:

1. $F(x) = \int_{-\infty}^{x} p(x)\,dx$
2. $p(x) = F'(x)$

Tính chất thứ nhất là từ định nghĩa hàm phân phối. Tính chất thứ hai suy ra từ việc cận trên của tích phân là hữu hạn.

Hàm mật độ của X là

$$f(x) = \begin{cases} p_i & \text{khi } x = x_i, \\ 0 & \text{khi } x \neq x_i, \forall i \end{cases}$$

Nhận xét 1

Ta có các lưu ý sau:

- $p_i \geq 0$, $\sum p_i = 1$, $i = 1, 2, \ldots$
- $P(a < X \leq b) = \sum_{a < x_i \leq b} p_i$

9.5 Hàm mật độ của biến ngẫu nhiên liên tục

Định nghĩa 6

Hàm số $f : \mathbb{R} \mapsto \mathbb{R}$ được gọi là **hàm mật độ** của biến ngẫu nhiên liên tục X nếu:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \forall a, b \in \mathbb{R}$$

Nhận xét 2

Với mọi $x \in \mathbb{R}$, $f(x) \geq 0$ và $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

Ý nghĩa hình học. Xác suất của biến ngẫu nhiên X nhận giá trị trong $[a, b]$ bằng diện tích hình thang cong giới hạn bởi $x = a$, $x = b$, $y = f(x)$ và Ox.

Chương 10

Hình học giải tích

10.1 Theo dòng lịch sử

Hình học xuất hiện từ thời xa xưa, xuất phát từ những nhu cầu thực tế nhất của con người là đo đạc để phân chia đất đai, xây dựng, canh tác, ... Từ đó con người đã có nhận thức rất sớm về quan hệ song song và vuông góc giữa hai đường thẳng.

Một cách hình ảnh (mà thật ra hình học là môn học về hình ảnh) thì hai đường thẳng song song không cắt nhau dù có kéo dài chúng ra vô tận. Các đường thẳng song song luôn có nhiều điều thú vị, cả ở mặt phẳng Euclid lẫn trong không gian. Đầu tiên phải kể đến định lý mang tên triết gia vĩ đại của Hy Lạp: Thales.

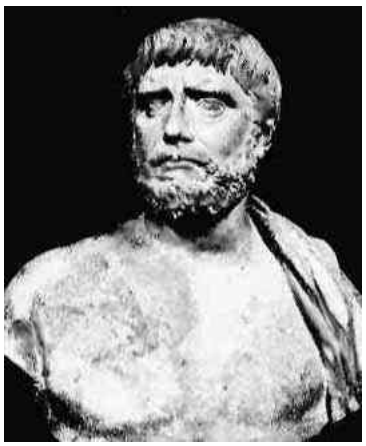
Thales của Miletus

Thales của Miletus được cho rằng sinh vào khoảng năm 624 Trước Công nguyên (TCN) và mất năm 547 TCN tại Miletus (Thổ Nhĩ Kì ngày nay)[1].

Ông được xem là nhà triết học đầu tiên khi không cố gắng giải thích tự nhiên bằng thần thoại hay các thế lực siêu nhiên như trước. Trường phái triết học do ông sáng lập, trường phái Milet, cho rằng mọi vật có nguồn gốc từ nước. Nhà triết học nổi tiếng Aristotle đánh giá rằng Thales là người sáng lập ra *triết học duy vật sơ khai*.

Trong toán học, Thales được biết tới với định lý mang tên ông về các đường song song. Định lý Thales được phát biểu như sau:

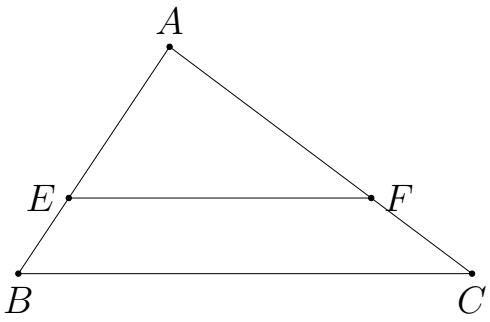
[1]https://mathshistory.st-andrews.ac.uk/Biographies/Thales/



Hình 10.1: Thales của Miletus

Định lí 1. Định lý Thales

Trong một tam giác, đường thẳng song song với một cạnh chắn trên hai cạnh còn lại các đoạn thẳng tương ứng tỉ lệ.



Hình 10.2: Định lý Thales trên mặt phẳng

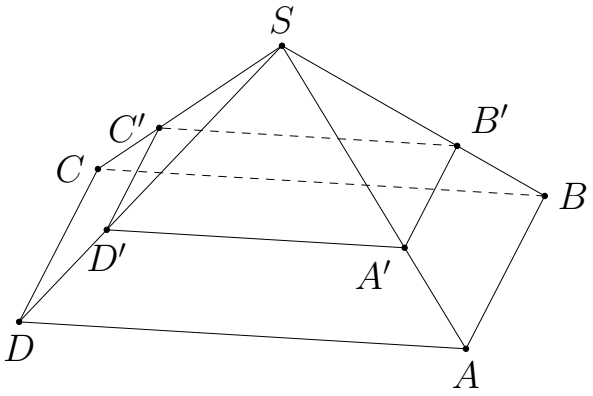
Theo định lý Thales, nếu EF song song với BC thì ta có $\dfrac{AE}{AB} = \dfrac{AF}{AC} = \dfrac{EF}{BC}$ (hình 10.2).

Không dừng lại ở mặt phẳng, khi mở rộng lên không gian định lý Thales cũng cho chúng ta một kết quả quan trọng khi nói tới các mặt phẳng song song nhau.

Định lí 2. Định lý Thales trong không gian

Trong khối chóp, mặt phẳng song song mặt đáy chắn các cạnh nối từ đỉnh hình chóp tới các đỉnh của mặt phẳng đáy các đoạn thẳng tương ứng tỉ lệ.

Theo định lý Thales, nếu mặt phẳng $(ABCD)$ song song với mặt phẳng $(A'B'C'D')$ thì $\dfrac{SA}{SA'} = \dfrac{SB}{SB'} = \dfrac{SC}{SC'} = \dfrac{SD}{SD'}$ (hình 10.3).



Hình 10.3: Định lý Thales trong không gian

Pythagoras của Samos

Khi nhắc tới vuông góc, chúng ta thường nhớ tới định lý ngày nào được học ở thời học sinh: định lý Pythagoras. Định lý này nói về quan hệ giữa độ dài các cạnh trong một tam giác vuông. Định lý tuy đơn giản nhưng có ý nghĩa rất quan trọng trong đời sống và khoa học của con người suốt chiều dài lịch sử. Đây cũng là tiền đề cho định lý mang tính lịch sử của nhân loại: định lý cuối cùng của Fermat.

Hình 10.4: Pythagoras của Samos

Pythagoras của Samos cũng là nhà triết học Hy Lạp cổ, được cho rằng sinh vào khoảng năm 570 TCN và mất năm 490 TCN[2]. Ông được học tập từ nhà triết học Thales và cũng có nhiều đóng góp cho sự phát triển của toán học, thiên văn học và âm nhạc. Tuy nhiên khác với thầy mình, trường phái triết học của ông cho rằng những con số là nguồn gốc của vạn vật và sử dụng những con số để giải thích những hiện tượng khoa học. Từ đây, các lý thuyết về âm nhạc được ra đời, cụ thể là các mối liên hệ về tần số với sự rung của dây nhạc cụ.

Ông là một trong những người hiếm hoi cho phép cả phụ nữ đi học ở lớp của mình vào thời ấy. Điều đó giúp phổ biến toán học nói riêng và kiến thức

[2]https://mathshistory.st-andrews.ac.uk/Biographies/Pythagoras/

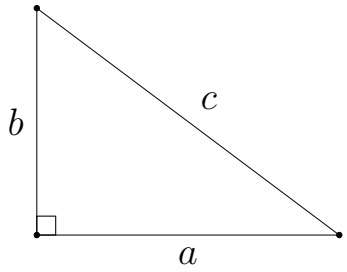
nói chung tới nhiều tầng lớp nhân dân. Tuy nhiên ông cũng có một hội kín rất thú vị. Như đã nói ở trên, trường phái triết học Pythagoras cố gắng giải thích nguồn gốc vạn vật bằng những con số. Điều này đã dẫn họ tới những khám phá động trời vào thời ấy.

Một trong những khám phá đó là về sự tồn tại của số vô tỉ dựa vào định lý mang tên ông. Lịch sử đã chỉ ra rằng trước Pythagoras, người Babylon và Ai Cập đã tìm ra rất nhiều bộ số nguyên (a, b, c) thỏa mãn $a^2 + b^2 = c^2$ là độ dài ba cạnh tam giác vuông. Định lý Pythagoras mà ngày nay chúng ta biết được phát biểu rằng:

Định lí 3. Định lý Pythagoras

Trong một tam giác vuông, bình phương độ dài cạnh huyền bằng tổng bình phương độ dài hai cạnh góc vuông.

Như vậy nếu gọi độ dài cạnh huyền là c, độ dài hai cạnh góc vuông lần lượt là a và b thì $a^2 + b^2 = c^2$ (hình 10.5).



Hình 10.5: Định lý Pythagoras

Nếu $a = b = 1$ thì sao? Khi đó bình phương độ dài cạnh huyền $c^2 = 2$. Tuy nhiên không thể tìm ra một số hữu tỉ nào để bình phương lên là 2 cả. Phát hiện này là một chấn động đối với thời Pythagoras và ông yêu cầu tất cả thành viên trong hội phải giữ kín bí mật về sự phát hiện này. Tuy nhiên thông tin vẫn lọt ra ngoài và truyền thuyết kể rằng ông đã xử tội chết cho thành viên của hội không tuân thủ.

Pythagoras đã đưa một khái niệm cực kì quan trọng trong toán học, gọi là *chứng minh* (proof). Để chứng minh một mệnh đề là đúng, chúng ta cần các mệnh đề (thường đơn giản hơn) đúng trước đó. Bằng các phép suy luận thích hợp dựa trên các mệnh đề đúng trước đó, chúng ta có thể kết luận rằng mệnh đề cần chứng minh là đúng. Phép chứng minh có thể gọi là *xương sống* của toán học, vì nếu không có một phép chứng minh đúng đắn thì một mệnh đề không thể được xác định được là có đúng hay không. Trong trường hợp của Fermat, khi ông đưa ra định lý Fermat nhưng không kèm chứng minh (vì lề sách quá chật nên không viết lời giải được) thì chúng ta không thể biết định lý Fermat có đúng hay không (?).

Nếu việc suy luận dựa trên các mệnh đề, hoặc định lý, đã đúng trước đó, thì phải có một lúc nào đó việc này dừng lại. Chúng ta không thể suy ngược tới vô hạn lần được. Do đó chúng ta cần những mệnh đề luôn đúng nhưng tính đúng đắn của nó được kiểm nghiệm trong thực tiễn. Chúng được gọi là *tiên đề* (axiom). Nhân vật tiếp theo được đề cập tới sẽ dẫn chúng ta tới hệ thống tiên đề làm nền tảng cho hình học.

Euclid của Alexandria

Đúng vậy, Euclid là người đặt nền móng cho hình học với bộ sách nổi tiếng *Elements* của mình. Trong bộ sách này đề cập tới những tiên đề, định lý làm nền tảng cho bộ môn hình học và vẫn còn ý nghĩa cho tới tận ngày nay. Những gì viết trong đó không quá xa lạ với những gì được giảng dạy trong nhà trường.

Hình 10.6: Euclid của Alexandria

Euclid của Alexandria sinh vào khoảng năm 325 TCN và mất vào khoảng năm 265 TCN[3]. Thông tin về ông không có nhiều. Nhưng chỉ mỗi bộ sách *Elements* cũng đủ để người đời sau cho rằng ông là người có ảnh hưởng nhất trong 2000 năm lịch sử phát triển của toán học.

Năm tiên đề cơ bản của hình học được ông phát biểu trong bộ *Elements* được phát biểu như sau:

1. Qua hai điểm bất kì luôn vẽ được một đường thẳng
2. Đường thẳng có thể kéo dài vô hạn về cả hai phía
3. Ta có thể xác định một đường tròn bằng tâm và bán kính của nó
4. Mọi góc vuông đều bằng nhau
5. Nếu một đường thẳng cắt hai đường thẳng khiến tổng hai góc trong cùng phía nhỏ hơn hai vuông thì hai đường thẳng đó chắc chắn sẽ cắt nhau tại một điểm nào đó

Tiên đề số 5 là rắc rối và phức tạp nhất. Nó không thực sự tự nhiên và có nhiều sự vướng mắc. Đây chính là tiên đề cho sự ra đời của hình học phi-Euclid

[3]https://mathshistory.st-andrews.ac.uk/Biographies/Euclid/

hơn 1500 năm sau.

Bộ *Elements* của Euclid bao gồm 13 quyển. Trong đó đề cập tới rất nhiều vấn đề của hình học, từ những phần tử đơn giản nhất cấu tạo nên hình học là điểm, đoạn thẳng, đường thẳng, tới những hình học lớn hơn như hình chữ nhật, hình tròn, đa giác, mặt phẳng. Thậm chí ông cũng đã có những dấu chân ở hình học không gian như hình chóp, hình cầu, hình nón ([1], [3]).

10.2 Phương pháp tọa độ trong mặt phẳng

Cuộc cách mạng trong hình học xảy ra khi nhà toán học lãng tử René Descartes phát minh ra hệ tọa độ và từ đó mọi đối tượng hình học có thể được biểu diễn bởi các phương pháp đại số như phương trình, đẳng thức.

Danh mục thuật ngữ và ký hiệu

Đầu tiên chúng ta thống nhất các thuật ngữ cũng như ký hiệu được sử dụng kể từ đây.

Điểm là đơn vị cơ bản của hình học. Bất kì đối tượng hình học nào cũng là một *tập hợp điểm*. Điểm được ký hiệu bởi chữ in hoa, ví dụ như A, B_1, B_2.

Đường thẳng đi qua hai điểm phân biệt cho trước. Đường thẳng có thể kéo dài vô hạn về hai phía. Đường thẳng được ký hiệu bởi chữ in thường hoặc chữ Hy Lạp trong ngoặc đơn, ví dụ như (d), (Δ).

Đoạn thẳng chỉ phần đường thẳng nằm giữa hai điểm.

Nửa đường thẳng chỉ phần đường thẳng nằm một phía của một điểm trên đường thẳng và chỉ kéo dài vô hạn về phía đó.

Vector là đoạn thẳng có hướng. Với điểm đầu là A và điểm cuối là B thì vector từ A tới B được ký hiệu là $\overrightarrow{AB}$. Để chỉ một vector không cần biết điểm đầu và điểm cuối ta dùng chữ thường in đậm, ví dụ như $\boldsymbol{a}$.

Góc giữa hai vector $\overrightarrow{OA}$ và $\overrightarrow{OB}$ là góc $\angle AOB$ và ký hiệu là $(\overrightarrow{OA}, \overrightarrow{OB})$.

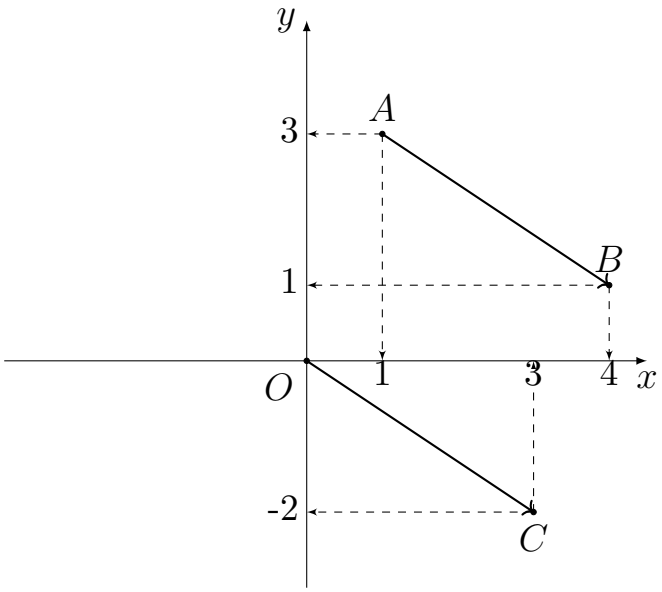
Tương tự đối với vector $\boldsymbol{a}$ và $\boldsymbol{b}$ thì góc giữa chúng ký hiệu là $(\boldsymbol{a}, \boldsymbol{b})$.

Vector trong mặt phẳng

Trong hệ tọa độ Oxy với tâm O và hai trục Ox (trục hoành) và Oy (trục tung) vuông góc nhau, đặt $O = (0, 0)$ là tọa độ của tâm O.

Tiếp theo, mọi điểm trong mặt phẳng Euclid đi liền với cặp số (x, y) chỉ tọa độ của điểm đó. Ví dụ $A = (1, 3)$, $B = (4, 1)$.

Tọa độ của điểm cũng là tọa độ của vector từ O tới điểm đó. Với hình 10.7 thì $\overrightarrow{OA} = (1, 3)$ và $\overrightarrow{OB} = (4, 1)$. Tọa độ của vector $\overrightarrow{AB}$ khi đó sẽ là



Hình 10.7: Tọa độ của điểm trong mặt phẳng

$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = (4,1) - (1,3) = (3,-2)$. Cũng theo hình 10.7 thì ta thấy $\overrightarrow{AB} = \overrightarrow{OC} = (3,-2)$.

Như vậy, nếu ta có hai điểm $A = (x_A, y_A)$ và $B = (x_B, y_B)$ thì vector $\overrightarrow{AB}$ là

$$\overrightarrow{AB} = (x_B - x_A, y_B - y_A) \tag{10.1}$$

Tích vô hướng của hai vector $\boldsymbol{a} = (x_1, y_1)$ và $\boldsymbol{b} = (x_2, y_2)$ được định nghĩa là

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = x_1 x_2 + y_1 y_2 \tag{10.2}$$

Ta cũng có thể ký hiệu tích vô hướng là $\boldsymbol{a} \cdot \boldsymbol{b}$.

Ta ký hiệu $\|\boldsymbol{a}\|$ là độ dài (chuẩn Euclid, Euclid norm) của vector $\boldsymbol{a}$. Trong hệ tọa độ Descartes vuông góc, theo định lý Pythagoras, độ dài của vector là độ dài cạnh huyền tam giác vuông (hình 10.7). Như vậy, độ dài đoạn thẳng AB với $A = (x_A, y_A)$ và $B = (x_B, y_B)$ là

$$AB = \|\overrightarrow{AB}\| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \tag{10.3}$$

Khi đó cosin góc giữa hai vector $\boldsymbol{a}$ và $\boldsymbol{b}$ là

$$\cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|} = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \cdot \sqrt{x_2^2 + y_2^2}} \tag{10.4}$$

Nếu góc giữa hai vector bằng 90 độ thì hai vector được gọi là vuông góc nhau. Khi đó tích vô hướng $\boldsymbol{a} \cdot \boldsymbol{b} = 0$.

Phương trình đường thẳng trong mặt phẳng

Theo tiên đề Euclid, một đường thẳng được xác định khi biết hai điểm phân biệt thuộc đường thẳng đó. Trong hệ tọa độ, chúng ta có hai cách tìm phương trình đường thẳng.

Bằng vector pháp tuyến. Vector pháp tuyến của đường thẳng là vector vuông góc với mọi vector có phương là đường thẳng đó. Giả sử $\boldsymbol{v} = (a, b)$ là vector pháp tuyến của đường thẳng đi qua điểm $M_0 = (x_0, y_0)$. Khi đó đường thẳng đi qua qua M_0 nhận $\boldsymbol{v}$ làm vector pháp tuyến là *tập hợp điểm* $M = (x, y)$ trên mặt phẳng sao cho $\boldsymbol{v} \cdot \overrightarrow{M_0M} = 0$. Điều này tương đương với

$$\boldsymbol{v} \cdot \overrightarrow{M_0M} = a \cdot (x - x_0) + b \cdot (y - y_0) = 0 \tag{10.5}$$

Bằng vector chỉ phương. Vector chỉ phương của đường thẳng là vector có phương song song với đường thẳng đó. Giả sử $\boldsymbol{v}' = (a', b')$ là vector chỉ phương của đường thẳng đi qua điểm $M_0 = (x_0, y_0)$. Khi đó đường thẳng đi qua M_0 nhận $\boldsymbol{v}'$ làm vector chỉ phương là *tập hợp điểm* $M = (x, y)$ trên mặt phẳng sao cho $\boldsymbol{v}' \parallel \overrightarrow{M_0M}$. Điều này tương đương với

$$\boldsymbol{v}' \parallel \overrightarrow{M_0M} \Leftrightarrow \frac{x - x_0}{a'} = \frac{y - y_0}{b'} \tag{10.6}$$

1. Cả hai cách biểu diễn khi khai triển ra đều có dạng $ax + by + c = 0$ với c là hằng số. Đây được gọi là dạng tổng quát của phương trình đường thẳng.
2. Cách viết $\dfrac{x - x_0}{a'} = \dfrac{y - y_0}{b'}$ được gọi là dạng chính tắc của phương trình đường thẳng.
3. Dạng chính tắc của phương trình đường thẳng còn có một tác dụng đặc biệt khác

$$\frac{x - x_0}{a'} = \frac{y - y_0}{b'} = t$$

với $t \in \mathbb{R}$. Khi đó tọa độ $M = (x, y)$ có thể được biểu diễn dưới dạng

$$\begin{cases} x = x_0 + a't \\ y = y_0 + b't \end{cases}, \quad t \in \mathbb{R} \tag{10.7}$$

Đây được gọi là phương trình dạng tham số.

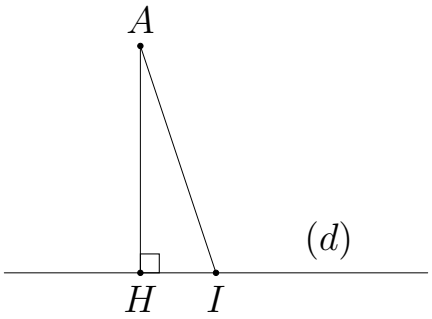
Chúng ta chú ý rằng nếu đường thẳng song song với một trong hai trục tọa độ thì vector chỉ phương của nó sẽ cùng phương với vector đơn vị $(1, 0)$ hoặc $(0, 1)$. Do đó không thể viết dưới dạng chính tắc được (không thể chia cho 0) nhưng có thể viết dưới dạng tổng quát hoặc dạng tham số.

Khoảng cách giữa điểm và đường thẳng

Nhắc lại một chút kiến thức cơ sở. **Khoảng cách** từ một điểm A nằm ngoài đường thẳng (d) là độ dài đoạn thẳng AH với $H \in (d)$ sao cho AH nhỏ nhất

(hình 10.8).

Khi đó H được gọi là **hình chiếu** của A lên đường thẳng (d) và AH là **khoảng cách** từ A tới (d). Do AH là đoạn thẳng có độ dài ngắn nhất, điều này xảy ra khi $AH \perp (d)$.



Hình 10.8: Hình chiếu và khoảng cách tới đường thẳng

Như vậy, để tìm hình chiếu của điểm A lên đường thẳng (d), ta dựng đường thẳng đi qua điểm A và vuông góc với (d).

Giả sử phương trình đường thẳng (d) với vector pháp tuyến $\boldsymbol{v} = (a, b)$ là $(d) : ax + by + c = 0$.

Gọi (d') là đường thẳng đi qua $A = (x_0, y_0)$ và vuông góc với d. Do $\boldsymbol{v}$ là vector pháp tuyến của (d) nên $\boldsymbol{v}$ là vector chỉ phương của (d'). Khi đó phương trình dạng tham số của (d') là

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \end{cases}, t \in \mathbb{R}$$

Gọi H là hình chiếu của A lên (d). Khi đó H là giao điểm của (d) và (d'). Vì $H \in (d')$ nên tọa độ của H có dạng $(x_0 + at, y_0 + bt)$ với t nào đó thuộc $\mathbb{R}$. Chúng ta sẽ đi tìm t này.

Vì $H \in (d)$ nên ta thay tọa độ của H vừa tìm được vào phương trình của (d) thu được

$$a(x_0 + at) + b(y_0 + bt) + c = 0 \Leftrightarrow t = -\frac{ax_0 + by_0 + c}{a^2 + b^2}$$

Như vậy là ta đã tìm được t từ đó xác định được tọa độ của H.

Từ đây ta tính được khoảng cách từ A tới (d) hay nói cách khác là độ dài đường AH. Ta có $A = (x_0, y_0)$ và $H = (x_0 + at, y_0 + bt)$ nên $\overrightarrow{AH} = (at, bt)$. Suy ra

$$\begin{aligned} AH =& \|\overrightarrow{AH}\| = \sqrt{(at)^2 + (bt)^2} = |t|\sqrt{a^2 + b^2} \\ =& \Big| -\frac{ax_0 + by_0 + c}{a^2 + b^2} \Big| \cdot \sqrt{a^2 + b^2} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \end{aligned}$$

10.3 Đạo hàm

Phép tính vi tích phân đã được con người nghiên cứu từ lâu. Câu chuyện về ai là người phát minh ra phép tính vi tích phân: Newton hay Leibniz, được coi là một trong những vụ tranh cãi đáng xấu hổ nhất lịch sử toán học. Nhưng họ cũng đã để lại một mảnh đất màu mỡ cho toán học về sau.

Cơ học và sự ra đời của đạo hàm

Trường phái Newton sử dụng đạo hàm như công cụ khảo sát vận tốc từ quãng đường. Ở bậc trung học chúng ta biết rằng *vận tốc trung bình* bằng quãng đường chia thời gian. Tuy nhiên điều đó chỉ đúng cho *chuyển động thẳng đều*. Nếu quãng đường là một hàm số phụ thuộc thời gian (quãng đường là $s(t)$ với t là thời gian) thì điều đó không đúng nữa.

Do quãng đường phụ thuộc thời gian nên có thể là vận tốc cũng phụ thuộc thời gian? Hợp lí đấy. Nhưng với mỗi một giá trị thời gian t cho ta một vị trí $s(t)$ trên trục số, còn vận tốc thì không thể phụ thuộc một giá trị thời gian được. Rõ ràng vật phải di chuyển một quãng đường từ thời gian t_0 tới t_1 thì mới có vận tốc trên quãng đường đó chứ?

Cách tiếp cận ở đây là, chúng ta cho sự thay đổi thời gian, tức hiệu $\Delta t = t_1 - t_0$, rất nhỏ. Khi đó vật đi từ $s(t_0)$ tới $s(t_1)$, vậy là chúng ta có thể tính vận tốc với công thức $v = \dfrac{s(t_1) - s(t_0)}{t_1 - t_0}$. Do Δt rất nhỏ, hay *tiến về 0*, thì vận tốc gần như xảy ra vào đúng một thời điểm. Do đó vận tốc lúc này được gọi là *vận tốc tức thời*. Đó cũng chính là ý nghĩa cơ học và sự ra đời của đạo hàm theo trường phái Newton.

Định nghĩa đạo hàm

Xét hàm số $f(x)$ liên tục trên khoảng (a, b) có chứa điểm x_0. Đạo hàm của $f(x)$ tại x_0 được định nghĩa là giới hạn

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{10.8}$$

Lưu ý rằng nếu giới hạn trên không phải là giới hạn hữu hạn (không tồn tại hoặc tiến tới vô cực) thì hàm số không có đạo hàm tại điểm x_0.

Ví dụ, để tính đạo hàm của hàm số $f(x) = x^3 + 2x^2 - 4$ tại $x_0 = 4$, ta khai

triển

$$\begin{aligned}\frac{f(x)-f(x_0)}{x-x_0} =& \frac{f(x)-f(4)}{x-4}\\ =& \frac{x^3+2x^2-4-(4^3+2\cdot 4^2-4)}{x-4}\\ =& \frac{(x^3-4^3)+2(x^2-4^2)}{x-4}\\ =& \frac{(x-4)(x^2+4x+16)+2(x-4)(x+4)}{x-4}\\ =& x^2+4x+16+2(x+4)\end{aligned}$$

Cho x tiến tới 4 thì ta có đạo hàm tại $x=4$

$$\begin{aligned}f'(4) =& \lim_{x\to 4}\frac{f(x)-f(4)}{x-4}\\ =& \lim_{x\to 4}(x^2+4x+16+2(x+4))\\ =& 4^2+4\cdot 4+16+2\cdot(4+4)=64\end{aligned}$$

Trong định nghĩa ở 10.8, nếu ta đặt $\Delta x = x - x_0$ và $\Delta y = y - y_0 = f(x) - f(x_0)$, ta gọi Δx là *số gia* của biến x, tương tự Δy là *số gia* của biến y.

Trong định nghĩa, x tiến tới x_0 tương đương với Δx tiến tới 0. Chuyển vế x_0 ta có $x = x_0 + \Delta x$ và từ đó $f(x) = f(x_0 + \Delta x)$. Định nghĩa đạo hàm ở trên có thể được viết lại

$$f'(x_0) = \lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \tag{10.9}$$

Nếu hàm số có đạo hàm tại mọi điểm trên khoảng (a, b) thì ta nói hàm số khả vi trên khoảng đó.

Ví dụ đối với hàm số $f(x) = x^3 + 2x^2 - 4$ như trên. Với mọi $x_0 \in \mathbb{R}$ ta có

$$\begin{aligned}f'(x_0) =& \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}\\ =& \lim_{x\to x_0}\frac{x^3+2x^2-4-(x_0^3+2x_0^2-4)}{x-x_0}\\ =& \lim_{x\to x_0}\frac{(x^3-x_0^3)+2(x^2-x_0^2)}{x-x_0}\\ =& \lim_{x\to x_0}(x^2+xx_0+x_0^2)+2(x+x_0)\\ =& x_0^2+x_0\cdot x_0+x_0^2+2(x_0+x_0)=3x_0^2+4x_0\end{aligned}$$

Ta thấy rằng giới hạn trên luôn tồn tại với mọi $x_0 \in \mathbb{R}$ nên thay x_0 thành x ta có đạo hàm $f'(x) = 3x^2 + 4x$ của $f(x)$ trên $\mathbb{R}$.

Vi phân

Trong cách ký hiệu

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

ta thay Δy thành dy và Δx thành dx thì vi phân được định nghĩa là

$$f'(x) = \frac{dy}{dx} \Leftrightarrow dy = f'(x)\,dx \tag{10.10}$$

Cách ký hiệu vi phân có ý nghĩa là vế trái là vi phân theo biến y và vế phải là vi phân theo biến x. Do $y = f(x)$ nên khi vi phân hai vế sẽ cho ra $dy = f'(x)\,dx$ (vế trái là đa thức bậc 1 biến y).

Ví dụ phương trình $y^2 = x^3 + 4x - 7$ thì khi vi phân hai vế ta có

$$(y^2)'\,dy = (x^3 + 4x - 7)\,dx \Leftrightarrow 2y\,dy = (3x^2 + 4)\,dx$$

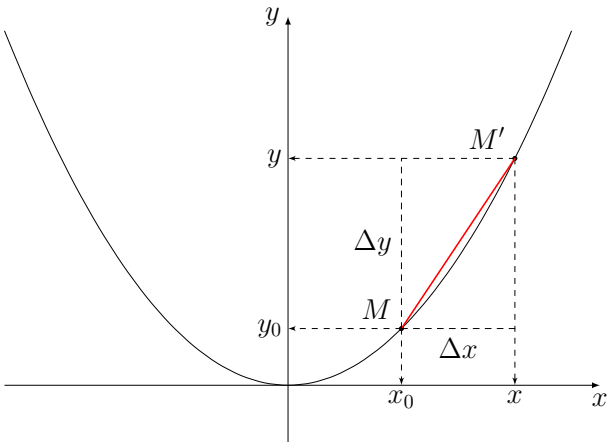
Ý nghĩa hình học của đạo hàm

Xét hàm số $y = f(x)$ liên tục trên khoảng (a, b) chứa điểm x_0.

Gọi $M' = (x, y)$ là một điểm thuộc hàm số $y = f(x)$. Khi đó đạo hàm của $f(x)$ tại x_0 là giới hạn

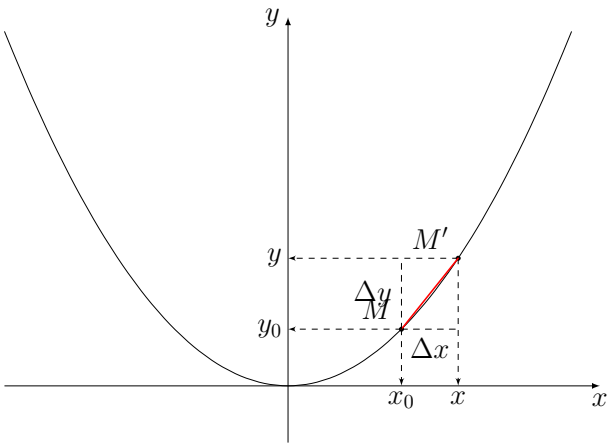
$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Xét hình 10.9, tỉ số $\Delta y/\Delta x$ là tangent của góc hợp bởi trục hoành Ox và đường thẳng MM'.



Hình 10.9: Hệ số góc (trường hợp 1)

Tiếp theo, xét hình 10.10, ta thấy đường thẳng MM' ngày càng tiến sát lại với đường cong. Như vậy, khi Δx tiến tới 0 thì đường thẳng MM' cắt đường



Hình 10.10: Hệ số góc (trường hợp 2)

cong tại hai điểm càng sát nhau. Đến khi hai điểm đó trùng nhau, đường thẳng MM' chỉ đi qua đúng một điểm thuộc đường cong và khi đó MM' trở thành tiếp tuyến của đường cong tại điểm $M = (x_0, y_0)$.

Khi đó $f'(x_0)$ là tangent của góc hợp bởi MM' và trục hoành Ox, hay nói cách khác là *hệ số góc* của đường tiếp tuyến. Thêm nữa $f'(x_0) = \dfrac{\Delta y}{\Delta x} = \dfrac{y - y_0}{x - x_0}$ nên phương trình đường tiếp tuyến đi qua $M = (x_0, y_0)$ là

$$y = f'(x_0)(x - x_0) + y_0 \tag{10.11}$$

10.4 Tích phân

Tích phân là khái niệm quan trọng trong giải tích. Sau đây sẽ trình bày cách tính tích phân theo tổng Riemann.

Tích phân và phân chia diện tích

Xét phương trình của một đường cong $y = f(x) > 0$ trên đoạn $[a, b]$.

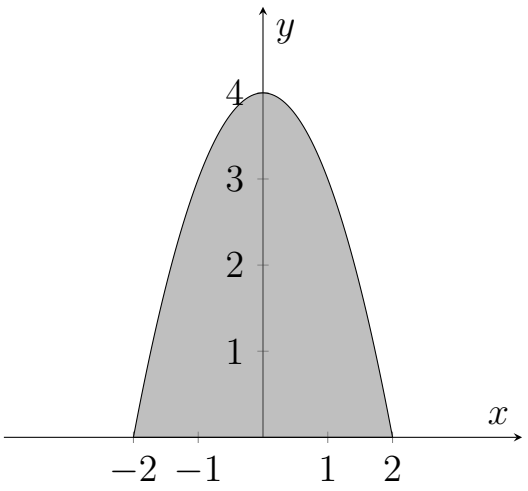
Theo định nghĩa, tích phân từ a tới b là diện tích phần hình phẳng giới hạn bởi đường cong $y = f(x)$, trục hoành Ox và hai trục đứng $x = a$, $x = b$.

Ở hình 10.11, diện tích phần tô màu xám là tích phân từ -2 tới 2 của hàm số $f(x) = -x^2 + 4$.

Chúng ta có thể tính diện tích hình chữ nhật, hình thang, hình vuông. Vậy có cách nào để tính diện tích một hình giới hạn bởi các đường cong bất kì không? Có đấy. Chúng ta sẽ tính xấp xỉ bằng tổng diện tích các hình chữ nhật.

Ví dụ với hàm số $f(x) = -x^2 + 4$ ở trên, ta chia đoạn $[a, b]$ thành n phần bằng nhau

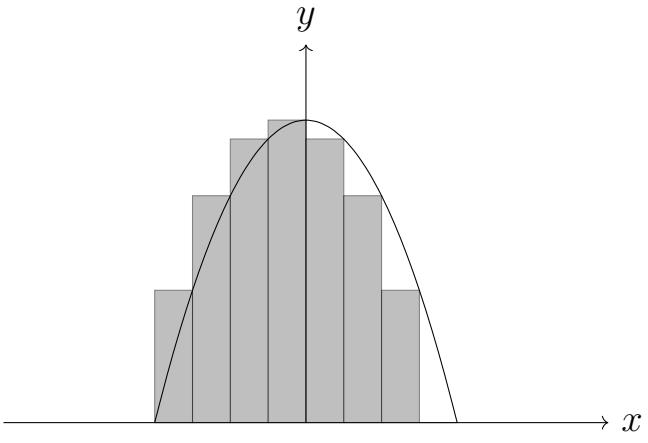
$$a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$$



Hình 10.11: Tích phân từ -2 tới 2 của $f(x) = -x^2 + 4$

Trong đó $x_{i+1} - x_i$ cố định và bằng $\dfrac{b-a}{n}$.

Đối với hình 10.12 ta xấp xỉ bằng 7 hình chữ nhật. Đối với hình 10.13 ta xấp xỉ bằng 15 hình chữ nhật. Đối với hình 10.14 ta xấp xỉ bằng 31 hình chữ nhật.

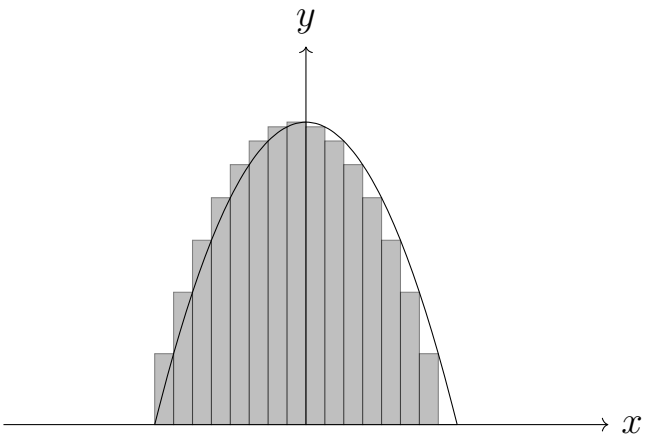


Hình 10.12: Xấp xỉ diện tích bởi 7 hình chữ nhật

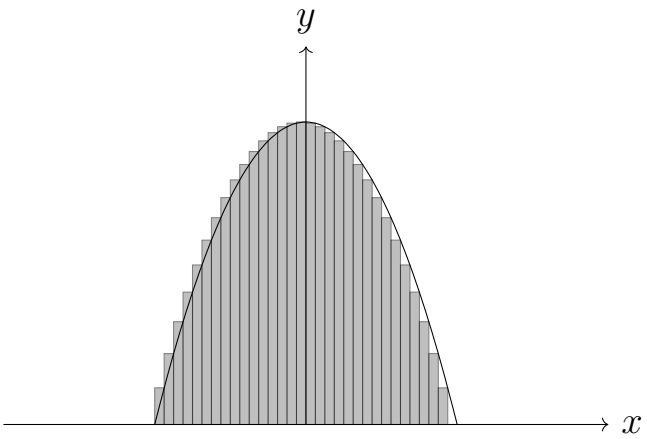
Càng dùng nhiều hình chữ nhật, tổng diện tích của chúng càng gần với diện tích cần tìm, hay là tích phân cần tìm.

Ở ba hình trên, mỗi hình chữ nhật trong đó có chiều rộng bằng nhau là $\dfrac{b-a}{n}$ với n là số đoạn. Chiều dài là $f(x_i)$ với $x_i = a + \dfrac{b-a}{n}i$, $i = 1, 2, \ldots, n$ (biên sau).

Cụ thể hơn, hình chữ nhật từ x_{i-1} tới x_i sẽ có chiều dài là $f(x_i)$ và chiều rộng là $\dfrac{b-a}{n}$.



Hình 10.13: Xấp xỉ diện tích bởi 15 hình chữ nhật



Hình 10.14: Xấp xỉ diện tích bởi 31 hình chữ nhật

Khi đó, tổng diện tích của các hình chữ nhật là

$$\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i) = \sum_{i=1}^{n}\frac{b-a}{n}f(x_i) \tag{10.12}$$

Khi số lượng hình chữ nhật tăng lên tới vô hạn thì tổng diện tích sẽ tiến tới diện tích chính xác của hình cần tìm, hay nói cách khác là tích phân. Do đó kết quả sẽ là

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n}\frac{b-a}{n}f(x_i), \quad x_i = a + \frac{b-a}{n}i \tag{10.13}$$

Ví dụ tính tích phân qua tổng Riemann

Ví dụ, tính tích phân từ -2 tới 2 của hàm số $f(x) = -x^2 + 4$ ở trên. Ta có $b = 2$ và $a = -2$ nên

$$\begin{aligned}\frac{b-a}{n} f(x_i) =& \frac{4}{n}\Big(-\Big(-2 + \frac{4}{n}i\Big)^2 + 4\Big) \\ =& \frac{4}{n}\Big(-4 + \frac{16}{n}i - \frac{16}{n^2}i^2 + 4\Big) \\ =& \frac{64}{n}\Big(\frac{i}{n} - \frac{i^2}{n^2}\Big)\end{aligned}$$

Tính tổng i từ 1 tới n ta có $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Tính tổng i^2 từ 1 tới n ta có $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

Suy ra

$$\begin{aligned}\sum_{i=1}^{n} \frac{64}{n}\Big(\frac{i}{n} - \frac{i^2}{n^2}\Big) =& \frac{64}{n^2}\sum_{i=1}^{n} i - \frac{64}{n^3}\sum_{i=1}^{n} i^2 \\ =& -\frac{64}{n^2}\cdot\frac{n(n+1)}{2} - \frac{64}{n^3}\cdot\frac{n(n+1)(2n+1)}{6}\end{aligned}$$

Khi n tiến tới vô cực thì biểu thức trên tiến tới $\dfrac{64}{2} - \dfrac{64\cdot 2}{6} = \dfrac{32}{3}$. Đây chính là giá trị của tích phân $\displaystyle\int_{-2}^{2} (-x^2 + 4)\, dx$.

Chương 11

Hình học affine

11.1 Không gian affine

Không gian affine

Định nghĩa 1. Không gian affine

Cho $\mathcal{V}$ là một không gian vector trên trường $\mathbb{F}$, và $\mathcal{A}$ là một tập khác rỗng mà các phần tử của nó gọi là **điểm**. Giả sử có ánh xạ φ

$$\begin{aligned}\varphi : \mathcal{A} \times \mathcal{A} &\to \mathcal{V} \\ (M, N) &\to \varphi(M, N)\end{aligned}$$

thỏa mãn hai điều kiện sau

1. Với mọi điểm $M \in \mathcal{A}$ và vector $\overrightarrow{v} \in \mathcal{V}$ có duy nhất một điểm $N \in \mathcal{A}$ sao cho $\varphi(M, N) = \overrightarrow{v}$;
2. Với ba điểm M, N, P bất kì ta luôn có

$$\varphi(M, N) + \varphi(N, P) = \varphi(M, P).$$

Ta nói $\mathcal{A}$ là một **không gian affine**.

Tên gọi đầy đủ: **$\mathcal{A}$ là không gian affine trên trường $\mathbb{F}$ liên kết với không gian vector $\mathcal{V}$ bởi ánh xạ liên kết φ.**

Khi đó, $\mathcal{V}$ được gọi là **không gian vector liên kết** với (hay **không gian nền**) của $\mathcal{A}$ và được ký hiệu là $\overrightarrow{\mathcal{A}}$.

φ được gọi là **ánh xạ liên kết**. Ta ký hiệu $\varphi(M, N) = \overrightarrow{MN}$ từ đây về sau. Khi đó hai điều kiện trên được viết lại là

1. Với mọi $M \in \mathcal{A}$, với mọi $\overrightarrow{v} \in \mathcal{V}$, tồn tại duy nhất $N \in \mathcal{A}$ sao cho $\overrightarrow{MN} = \overrightarrow{v}$
2. Với mọi M, N, P thuộc $\mathcal{A}$, $\overrightarrow{MN} + \overrightarrow{NP} = \overrightarrow{MP}$

Biểu thức ở điều kiện 2 còn được gọi là **hệ thức Chales**.

Nếu $\mathbb{F} = \mathbb{R}$ thì ta gọi là không gian affine thực.

Nếu $\mathbb{F} = \mathbb{C}$ thì ta gọi là không gian affine phức.

Nếu muốn nhấn mạnh trường $\mathbb{F}$ ta nói là $\mathbb{F}$-không gian affine.

Ta ký hiệu một không gian affine là bộ $(\mathcal{A}, \overrightarrow{\mathcal{A}}, \varphi)$. Ta cũng có thể ghi tắt là $\mathcal{A}(\mathbb{F})$ hoặc chỉ là $\mathcal{A}$.

Nếu $\overrightarrow{\mathcal{A}}$ là không gian vector n chiều thì ta nói $\mathcal{A}$ là không gian affine n chiều và ký hiệu là $\mathcal{A}^n$. Như vậy

$$\dim \mathcal{A} = \dim \overrightarrow{\mathcal{A}}$$

Ví dụ 1. Xét tập hợp các điểm trong không gian $\mathbb{R}^3$ học ở THTP. Khi đó $\mathcal{A} = \mathbb{R}^3$ là tập hợp các điểm, $\overrightarrow{\mathcal{A}}$ là tập hợp các vector trong $\mathbb{R}^3$. Một vector từ điểm A tới điểm B (theo nghĩa hình học) là một đoạn thẳng có hướng nối từ A tới B.

Lưu ý. Ở THPT chúng ta học rằng tọa độ của một điểm M cũng chính là tọa độ của vector $\overrightarrow{OM}$. Tuy nhiên điều đó không phải lúc nào cũng đúng. Ở các phần sau sẽ giải thích lý do tại sao.

Tính chất của không gian affine

Với mọi M, N, Q thuộc $\mathcal{A}$ ta có

1. $\overrightarrow{MN} = \overrightarrow{0}$ khi và chỉ khi $M \equiv N$
2. $\overrightarrow{MN} = -\overrightarrow{NM}$
3. $\overrightarrow{MN} = \overrightarrow{PQ}$ khi và chỉ khi $\overrightarrow{MP} = \overrightarrow{NQ}$
4. $\overrightarrow{MN} = \overrightarrow{PN} - \overrightarrow{PM}$

Chứng minh. Để chứng minh các tính chất trên ta sử dụng hai điều kiện trong định nghĩa không gian affine (đặc biệt là hệ thức Chales).

1. Nếu $M \equiv N$ thì $\overrightarrow{MM} = \overrightarrow{MN} + \overrightarrow{NM} = \overrightarrow{MM} + \overrightarrow{MM}$. Suy ra $\overrightarrow{MM} = \overrightarrow{0}$ hay $\overrightarrow{MN} = \overrightarrow{0}$. Từ đây, nếu $\overrightarrow{MN} = \overrightarrow{0}$ thì theo điều kiện 1 trong định nghĩa, tồn tại duy nhất điểm N thỏa $\overrightarrow{MN} = \overrightarrow{0}$. Điều này tương đương với $M \equiv N$.

2. Từ hệ thức Chales ta có

$$\overrightarrow{0} = \overrightarrow{MM} = \overrightarrow{MN} + \overrightarrow{NM} \Leftrightarrow \overrightarrow{MN} = -\overrightarrow{NM}$$

3. $\overrightarrow{MN} = \overrightarrow{MP} + \overrightarrow{PN}$ và $\overrightarrow{PQ} = \overrightarrow{PN} + \overrightarrow{NQ}$ nên $\overrightarrow{MP} + \overrightarrow{PN} = \overrightarrow{PN} + \overrightarrow{NQ}$, hay $\overrightarrow{MP} = \overrightarrow{NQ}$.

4. $\overrightarrow{PM} + \overrightarrow{MN} = \overrightarrow{PN}$, chuyển vế $\overrightarrow{PM}$ ta có điều phải chứng minh. □

Phẳng

Ở THPT ta có điểm tương đương 0-phẳng, đường thẳng tương đương 1-phẳng, mặt phẳng tương đương 2-phẳng.

Trong mặt phẳng Oxy, một đường thẳng được xác định khi biết một điểm thuộc nó và một vector chỉ phương $\overrightarrow{v} \neq \overrightarrow{0}$. Khi đó đường thẳng đi qua P nhận $\overrightarrow{v}$ làm vector chỉ phương là tập hợp các điểm $M \in \mathbb{R}^2$ sao cho $\overrightarrow{PM}$ cùng phương $\overrightarrow{v}$. Nói cách khác

$$d = \{M \in \mathbb{R}^2 : \overrightarrow{PM} = a\overrightarrow{v},\ a \in \mathbb{R}\}$$

Trong không gian $Oxyz$, tương tự một đường thẳng xác định khi biết một điểm thuộc nó và một vector chỉ phương $\overrightarrow{v}$ tương ứng

$$d = \{M \in \mathbb{R}^3 : \overrightarrow{PM} = a\overrightarrow{v},\ a \in \mathbb{R}\}$$

Một mặt phẳng trong $\mathbb{R}^3$ xác định khi biết một điểm thuộc nó và một cặp vector chỉ phương $\overrightarrow{u}$, $\overrightarrow{v}$ của nó

$$\alpha = \{M \in \mathbb{R}^3 : \overrightarrow{PM} = a\overrightarrow{u} + b\overrightarrow{v},\ a, b \in \mathbb{R}\}$$

Trong hình học affine ta mở rộng các khái niệm phẳng trên.

Định nghĩa 2. Phẳng

Cho không gian affine $(\mathcal{A}, \overrightarrow{\mathcal{A}}, \varphi)$, P là một điểm thuộc $\mathcal{A}$ và $\overrightarrow{\alpha}$ là một không gian vector con của $\overrightarrow{\mathcal{A}}$. Khi đó tập hợp

$$\alpha = \{M \in \mathcal{A} : \overrightarrow{PM} \in \alpha\}$$

được gọi là **phẳng** đi qua P với (không gian chỉ) phương $\overrightarrow{\alpha}$.

Nếu $\dim \overrightarrow{\alpha} = m$ thì ta nói α là một **phẳng m chiều** hay một m-phẳng và viết $\dim \alpha = m$. Như vậy

$$\dim \alpha = \dim \overrightarrow{\alpha}$$

Theo cách gọi thông thường, 1-phẳng là đường thẳng, 2-phẳng là mặt phẳng. **Siêu phẳng** là tên gọi của phẳng có đối chiều 1, tức là nếu số chiều của không gian là n thì số chiều của siêu phẳng là $n - 1$.

Nhận xét 1

1. Nếu α là phẳng đi qua P thì $P \in \alpha$ và với mọi M, N thuộc α ta có $\overrightarrow{MN} = \overrightarrow{PN} - \overrightarrow{PM}$ cũng thuộc α;
2. 0-phẳng là tập chỉ gồm một điểm. Do đó ta có thể xem một điểm là một 0-phẳng;
3. Điểm P trong định nghĩa không có vai trò quan trọng gì. Mọi điểm P trong α đều có ý nghĩa như nhau;
4. Giả sử α là phẳng đi qua P với phương $\overrightarrow{\alpha}$, β là phẳng đi qua Q với phương $\overrightarrow{\beta}$. Khi đó $\alpha \subset \beta$ khi và chỉ khi $P \in \beta$ và $\overrightarrow{\alpha} \subset \overrightarrow{\beta}$. Suy ra $\alpha \equiv \beta$ khi $P \in \beta$ (hay $Q \in \alpha$) và $\overrightarrow{\alpha} \equiv \overrightarrow{\beta}$;
5. Nếu α là phẳng với phương $\overrightarrow{\alpha}$ thì α được gọi là không gian affine liên kết với $\overrightarrow{\alpha}$ bởi ánh xạ liên kết

$$\varphi_{\alpha\times\alpha} : \alpha \times \alpha \to \overrightarrow{\alpha}$$

Vì vậy ta có thể xem phẳng là không gian affine con.

Để xác định đường thẳng ta chỉ cần biết một vector chỉ phương là đủ. Để xác định mặt phẳng ta chỉ cần biết hai vector không song song của mặt phẳng đó là đủ. Tổng quát, để xác định phương $\overrightarrow{\alpha}$ của m-phẳng α ta chỉ cần biết một cơ sở là đủ.

Từ định nghĩa của không gian vector (tập sinh) ta thấy rằng một m-phẳng chỉ có một không gian chỉ phương duy nhất, nhưng có thể có nhiều cơ sở khác nhau.

Độc lập affine và phụ thuộc affine

Định nghĩa 3

Hệ $m+1$ điểm $\{A_0, A_1, \ldots, A_m\}$ $(m \geq 1)$ của không gian affine $\mathcal{A}$ được gọi là **độc lập affine** nếu hệ m vector

$$\{\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \ldots, \overrightarrow{A_0A_m}\}$$

của $\overrightarrow{\mathcal{A}}$ là một hệ vector độc lập tuyến tính.

Hệ điểm không độc lập tuyến tính được gọi là **phụ thuộc affine**.

Chúng ta có một số lưu ý từ định nghĩa.

1. Tập chỉ gồm một điểm A_0 bất kì được quy ước là luôn độc lập;

2. Trong định nghĩa trên điểm A_0 bình đẳng như các điểm khác vì nếu hệ

$$\{\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \ldots, \overrightarrow{A_0A_m}\}$$

độc lập affine thì hệ

$$\{\overrightarrow{A_iA_0}, \ldots, \overrightarrow{A_iA_{i-1}}, \overrightarrow{A_iA_{i+1}}, \ldots, \overrightarrow{A_iA_m}\}$$

cũng độc lập affine;

3. Hệ $\{A_0, \ldots, A_m\}$ phụ thuộc affine thì hệ

$$\{\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_m}\}$$

phụ thuộc affine;

4. Hệ con của một hệ độc lập thì độc lập, nhưng hệ con của một hệ phụ thuộc chưa chắc phụ thuộc.

Ta sẽ chứng minh lưu ý thứ hai.

Chứng minh. Ta xét tổ hợp tuyến tính

$$\lambda_1\overrightarrow{A_0A_1} + \lambda_2\overrightarrow{A_0A_2} + \ldots + \lambda_m\overrightarrow{A_0A_m}$$

Do hệ vector độc lập tuyến tính nên $\lambda_1 = \ldots = \lambda_m = 0$. Khi đó ta khai triển vế trái

$$\begin{aligned}
&\lambda_1\overrightarrow{A_0A_1} + \lambda_2\overrightarrow{A_0A_2} + \ldots + \lambda_m\overrightarrow{A_0A_m}\\
=&\lambda_1(\overrightarrow{A_iA_1} - \overrightarrow{A_iA_0}) + \lambda_2(\overrightarrow{A_iA_2} - \overrightarrow{A_iA_0}) + \ldots\\
&+ \lambda_{i-1}(\overrightarrow{A_iA_{i-1}} - \overrightarrow{A_iA_0}) - \lambda_i\overrightarrow{A_iA_0} + \lambda_{i+1}(\overrightarrow{A_iA_{i+1}} - \overrightarrow{A_iA_0})\\
&+ \lambda_m(\overrightarrow{A_iA_m} - \overrightarrow{A_iA_0})\\
=&\lambda_1\overrightarrow{A_iA_1} + \lambda_2\overrightarrow{A_iA_2} + \ldots + \lambda_{i-1}\overrightarrow{A_iA_{i-1}} + \lambda_{i+1}\overrightarrow{A_iA_{i+1}} + \ldots\\
&+ \lambda_m\overrightarrow{A_iA_m} - (\lambda_1 + \lambda_2 + \ldots + \lambda_m)\overrightarrow{A_iA_0} = \overrightarrow{0}
\end{aligned}$$

Do $\lambda_1 = \ldots = \lambda_m = 0$ nên tổ hợp tuyến tính ứng với các vector $\overrightarrow{A_iA_j}$ $(j \neq i)$ độc lập tuyến tính và ta có điều phải chứng minh. □

Định lí 1

Trong không gian affine n chiều $\mathcal{A}^n$, với $0 < m \leq n+1$, luôn tồn tại các hệ m điểm độc lập. Mọi hệ gồm hơn $n+1$ điểm đều phụ thuộc.

11.2 Giao của các phẳng. Bao affine

Cho $\{\alpha_i : i \in I\}$ là một họ không rỗng các phẳng trong không gian affine $\mathcal{A}$.

Định lí 2

Nếu $\bigcap_{i\in I} \alpha_i \neq \emptyset$ thì $\bigcap_{i\in I} \alpha_i$ là một phẳng có phương $\bigcap_{i\in I} \overrightarrow{\alpha_i}$.

Chứng minh. Vì $\bigcap_{i\in I} \alpha_i \neq \emptyset$ nên tồn tại $P \in \bigcap_{i\in I} \alpha_i$, hay $P \in \alpha_i$ với $i \in I$.

Nếu $M \in \bigcap_{i\in I}$ thì $M \in \alpha_i$ với $i \in I$. Suy ra $\overrightarrow{PM} \in \alpha_i$. Do đó

$$\bigcap_{i\in I} \alpha_i = \{M \in \mathcal{A} : \overrightarrow{PM} \in \bigcap_{i\in I} \overrightarrow{\alpha_i}\}$$

Điều này nghĩa là $\bigcap_{i\in I} \alpha_i$ là phẳng đi qua P với không gian chỉ phương là $\bigcap_{i\in I} \overrightarrow{\alpha_i}$. □

Định nghĩa 4. Phẳng giao

Phẳng $\bigcap_{i\in I} \alpha_i$ trong định lý trên được gọi là **phẳng giao** của các phẳng α_i.

Từ định nghĩa trên ta thấy rằng $\bigcap_{i\in I} \alpha_i$ là phẳng lớn nhất (theo quan hệ bao hàm) chứa trong tất cả các phẳng α_i, $i \in I$.

Định nghĩa 5. Bao affine

Cho X là một tập con khác rỗng của không gian affine $\mathcal{A}$. Khi đó giao của mọi phẳng chứa X trong $\mathcal{A}$ sẽ là một phẳng, gọi là **bao affine** của X, ký hiệu là $\langle X \rangle$.

Như vậy, bao affine $\langle X \rangle$, theo quan hệ bao hàm, của tập X là phẳng bé nhất chứa X.

Tương tự phép giao và hợp của hai tập hợp, chúng ta có phép giao các phẳng ở trên và phép tổng của các phẳng sẽ đề cập sau đây.

Định nghĩa 6. Phẳng tổng

Cho $\{\alpha_i : i \in I\}$ là một họ không rỗng các phẳng. Bao affine của tập hợp $\bigcup_{i\in I}\alpha_i$ được gọi là **phẳng tổng** (hay **tổng**) của các phẳng α_i, ký hiệu là $\sum_{i\in I}\alpha_i$.

Như vậy, phẳng tổng là phẳng bé nhất chứa tất cả các phẳng α_i, $i \in I$.

Ta có nhận xét sau. Nếu X là một hệ hữu hạn điểm $X = \{P_0, P_1, \ldots, P_m\}$ thì tổng $P_0 + P_1 + \ldots + P_m$ (ta xem các P_i là các 0-phẳng) là phẳng có số chiều bé nhất đi qua các điểm này. Hơn nữa

$$\dim(P_0 + P_1 + \ldots + P_m) = \operatorname{rank}\{\overrightarrow{P_0P_1}, \overrightarrow{P_0P_2}, \ldots, \overrightarrow{P_0P_m}\}$$

Do đó hệ điểm $\{P_0, P_1, \ldots, P_m\}$ độc lập thì $\dim(P_0 + P_1 + \ldots + P_m) = m$.

Chứng minh. Đặt $I = P_0 + P_1 + \ldots + P_m$ là phẳng tổng của hệ điểm

$$\{P_0, P_1, \ldots, P_m\}$$

Khi đó I đi qua các điểm P_0, P_1, ..., P_m.

Đặt α_i là phẳng đi qua P_0 và P_i, $i = 1, 2, \ldots, m$. Khi đó α_i có phương là $\overrightarrow{P_0P_i}$. Tổng I chính là tổng các phẳng $\alpha_1 + \alpha_2 + \ldots + \alpha_m$, và $\overrightarrow{P_0P_1}$, $\overrightarrow{P_0P_2}$, ..., $\overrightarrow{P_0P_m}$ là các vector chỉ phương của nó. Như vậy nếu $\overrightarrow{I}$ là không gian chỉ phương của I thì nó gồm các vector độc lập tuyến tính $\overrightarrow{P_0P_{i_1}}$, $\overrightarrow{P_0P_{i_2}}$, ..., $\overrightarrow{P_0P_{i_k}}$. Khi đó $\dim I = \dim \overrightarrow{I} = k = \operatorname{rank}\{\overrightarrow{P_0P_1}, \overrightarrow{P_0P_2}, \ldots, \overrightarrow{P_0P_m}\}$. Từ đây ta có điều phải chứng minh. □

Định lí 3

Cho α và β là hai phẳng. Nếu $\alpha \cap \beta \neq \emptyset$ thì với mọi $P \in \alpha$ và với mọi $Q \in \beta$ ta có $\overrightarrow{PQ} = \overrightarrow{\alpha} + \overrightarrow{\beta}$.

Ngược lại nếu có điểm $P \in \alpha$ và $Q \in \beta$ sao cho $\overrightarrow{PQ} = \overrightarrow{\alpha} + \overrightarrow{\beta}$ thì $\alpha \cap \beta \neq \emptyset$.

Chứng minh. Giả sử $\alpha \cap \beta \neq \emptyset$. Khi đó tồn tại điểm $M \in \alpha \cap \beta$, suy ra $M \in \alpha$ và $M \in \beta$. Với mọi $P \in \alpha$ và với mọi $Q \in \beta$ thì $\overrightarrow{PM} \in \overrightarrow{\alpha}$ và $\overrightarrow{MQ} \in \overrightarrow{\beta}$. Từ đó $\overrightarrow{PQ} = \overrightarrow{PM} + \overrightarrow{MQ} = \overrightarrow{\alpha} + \overrightarrow{\beta}$.

Đảo lại, giả sử ta có điểm $P \in \alpha$ và điểm $Q \in \beta$ sao cho $\overrightarrow{PQ} = \overrightarrow{\alpha} + \overrightarrow{\beta}$. Khi đó tồn tại hai vector $\overrightarrow{u}$ và $\overrightarrow{v}$ sao cho $\overrightarrow{PQ} = \overrightarrow{u} + \overrightarrow{v}$ với $\overrightarrow{u} \in \overrightarrow{\alpha}$ và $\overrightarrow{v} \in \overrightarrow{\beta}$. Theo định nghĩa của không gian affine thì với điểm $P \in \alpha$, tồn tại duy nhất điểm $M \in \alpha$ sao cho $\overrightarrow{PM} = \overrightarrow{u}$. Tương tự với điểm $Q \in \beta$ tồn tại duy nhất điểm $N \in \beta$ sao cho $\overrightarrow{QN} = \overrightarrow{v}$. Suy ra $\overrightarrow{PQ} = \overrightarrow{u} + \overrightarrow{v} = \overrightarrow{PM} - \overrightarrow{QN}$. Chuyển vế $\overrightarrow{QN}$ ta có $\overrightarrow{PM} = \overrightarrow{u} = \overrightarrow{PQ} + \overrightarrow{QN} = \overrightarrow{PN}$. Điều này chỉ xảy ra khi $M \equiv N$, hay nói cách khác M và N thuộc $\alpha \cap \beta$. Như vậy $\alpha \cap \beta \neq \emptyset$. □

Định lí 4

Giả sử α và β là hai phẳng với phương lần lượt là $\overrightarrow{\alpha}$ và $\overrightarrow{\beta}$. Khi đó

1. Nếu $\alpha \cap \beta \neq \emptyset$ thì

$$\dim(\alpha + \beta) = \dim(\alpha) + \dim(\beta) - \dim(\alpha \cap \beta)$$

2. Nếu $\alpha \cap \beta = \emptyset$ thì

$$\dim(\alpha + \beta) = \dim(\alpha) + \dim(\beta) - \dim(\overrightarrow{\alpha} \cap \overrightarrow{\beta}) + 1$$

Chứng minh. 1. Nếu $\alpha \cap \beta \neq \emptyset$ thì theo định lý 11.2 ta có $\alpha \cap \beta$ là một phẳng có phương $\overrightarrow{\alpha} \cap \overrightarrow{\beta}$. Lấy $P \in \alpha \cap \beta$ và gọi γ là phẳng đi qua P với phương $\overrightarrow{\gamma} = \overrightarrow{\alpha} + \overrightarrow{\beta}$. Ta có $\alpha \subset \gamma$ và $\beta \subset \gamma$. Ngoài ra nếu có phẳng γ' chứa α và β thì $P \in \gamma'$ và phương của γ' phải chứa $\overrightarrow{\alpha}$ và $\overrightarrow{\beta}$. Nói cách khác $\gamma \subset \gamma'$. Vậy γ là phẳng bé nhất chứa α và β, tức là $\gamma = \alpha + \beta$. Do đó

$$\begin{aligned}\dim(\alpha + \beta) &= \dim \gamma = \dim \overrightarrow{\gamma} = \dim(\alpha + \beta) \\ &= \dim \overrightarrow{\alpha} + \dim \overrightarrow{\beta} - \dim(\overrightarrow{\alpha} \cap \overrightarrow{\beta}) \\ &= \dim \alpha + \dim \beta - \dim(\overrightarrow{\alpha} \cap \overrightarrow{\beta})\end{aligned}$$

2. Nếu $\alpha \cap \beta = \emptyset$ thì theo định lý 11.2, nếu ta lấy $P \in \alpha$ và $Q \in \beta$ thì $\overrightarrow{PQ} \notin \overrightarrow{\alpha} + \overrightarrow{\beta}$. Gọi $\overrightarrow{\gamma}$ là không gian con một chiều sinh bởi $\overrightarrow{PQ}$, ta có $(\overrightarrow{\alpha} + \overrightarrow{\beta}) \cap \overrightarrow{\gamma} = \{\overrightarrow{0}\}$ (các không gian vector không có vector nào chung ngoài $\overrightarrow{0}$). Gọi η là phẳng đi qua P có không gian chỉ phương là $\overrightarrow{\alpha} + \overrightarrow{\beta} + \overrightarrow{\gamma}$ thì ta có $\alpha \subset \eta$ và $\beta \subset \eta$. Suy ra $\alpha + \beta \subset \eta$.

3. η' là một phẳng chứa α và β thì $P \in \eta'$ và phương $\overrightarrow{\eta'}$ của η' phải chứa $\overrightarrow{\alpha}$, $\overrightarrow{\beta}$ và $\overrightarrow{\gamma}$. Từ đó $\eta \subset \eta'$. Suy ra η là phẳng bé nhất chứa α và β, hay $\eta = \alpha + \beta$.

Do $\dim((\overrightarrow{\alpha}+\overrightarrow{\beta})\cap\overrightarrow{\gamma})=0$ nên

$$\begin{aligned}\dim(\alpha+\beta)&=\dim\eta=\dim(\overrightarrow{\alpha}+\overrightarrow{\beta}+\overrightarrow{\gamma})\\&=\dim\overrightarrow{\alpha}+\dim\overrightarrow{\beta}+\dim\overrightarrow{\gamma}-\dim(\overrightarrow{\alpha}\cap\overrightarrow{\beta})\\&=\dim\alpha+\dim\beta+1-\dim(\overrightarrow{\alpha}\cap\overrightarrow{\beta})\end{aligned}$$

□

Vị trí tương đối

Định nghĩa 7. Cắt nhau, chéo nhau, song song

Hai phẳng α và β được gọi là **cắt nhau cấp** r nếu $\alpha\cap\beta$ là một r-phẳng.

Chúng được gọi là **chéo nhau cấp** r nếu $\alpha\cap\beta=\emptyset$ và $\dim(\overrightarrow{\alpha}\cap\overrightarrow{\beta})=r$.

Chúng được gọi là **song song** (với nhau) nếu $\overrightarrow{\alpha}\subset\overrightarrow{\beta}$ hoặc $\overrightarrow{\beta}\subset\overrightarrow{\alpha}$.

Theo định lý về dim bên trên, trong $\mathbb{R}^3$ không tồn tại hai mặt phẳng chéo nhau cấp 0 hoặc 1.

Định lí 5

Cho hai phẳng song song α và β. Nếu $\alpha\cap\beta\neq\emptyset$ thì $\alpha\subset\beta$ hoặc $\beta\subset\alpha$.

Chứng minh. Do α và β có điểm chung nên $\alpha\cap\beta$ là một phẳng có phương $\overrightarrow{\alpha}\cap\overrightarrow{\beta}$. Theo định nghĩa về sự song song, α song song β dẫn tới $\overrightarrow{\alpha}\subset\overrightarrow{\beta}$ hoặc $\overrightarrow{\beta}\subset\overrightarrow{\alpha}$. Nếu $\overrightarrow{\alpha}\subset\overrightarrow{\beta}$ thì $\overrightarrow{\alpha}\cap\overrightarrow{\beta}=\overrightarrow{\alpha}$. Suy ra $\alpha\cap\beta=\alpha$ hay $\alpha\subset\beta$. Trường hợp $\overrightarrow{\beta}\subset\overrightarrow{\alpha}$ tương tự. □

Định lí 6

Qua một điểm A có một và chỉ một m-phẳng song song với m-phẳng α đã cho.

Chứng minh. Gọi β là m-phẳng đi qua A với phương là $\overrightarrow{\alpha}$. Khi đó β là phẳng m chiều song song với α. Nếu β' cũng là m-phẳng đi qua A và song song với α thì $\overrightarrow{\beta'}=\overrightarrow{\beta}=\overrightarrow{\alpha}$. Do β và β' có điểm chung nên theo định lý 11.2 ta có $\beta\equiv\beta'$ □

Định lí 7

Trong không gian affine n chiều $\mathcal{A}^n$ cho một siêu phẳng α và một m-phẳng β $(1 \leq m \leq n-1)$. Khi đó α và β hoặc song song hoặc cắt nhau theo một $(m-1)$-phẳng.

Mục tiêu và tọa độ affine

Định nghĩa 8. Mục tiêu affine

Cho $\mathcal{A}^n$ là một không gian affine n chiều. Hệ $\{O, \overrightarrow{e_1}, \overrightarrow{e_2}, \ldots, \overrightarrow{e_n}\}$ gồm một điểm $O \in \mathcal{A}^n$ và một cơ sở $\{\overrightarrow{e_1}, \overrightarrow{e_2}, \ldots, \overrightarrow{e_n}\}$ của $\overrightarrow{\mathcal{A}^n}$ được gọi là **mục tiêu affine** (hay **mục tiêu**) của $\mathcal{A}^n$.

Điểm O được gọi là **gốc**, vector $\overrightarrow{e_i}$ được gọi là **vector cơ sở thứ** i, $i = 1, 2, \ldots, n$.

Giả sử $\{O, \overrightarrow{e_i}\}$ là một mục tiêu của không gian affine $\mathcal{A}^n$. Khi đó với mọi $M \in \mathcal{A}^n$, vector $\overrightarrow{OM} \in \overrightarrow{\mathcal{A}^n}$ nên ta có biểu diễn tuyến tính của $\overrightarrow{OM}$ qua các cơ sở $\{\overrightarrow{e_i}\}$

$$\overrightarrow{OM} = \sum_{i=1}^{n} x_i \overrightarrow{e_i}$$

Nhắc lại đại số tuyến tính, lúc này vector $\overrightarrow{OM}$ có tọa độ $(x_1, x_2, \ldots, x_n)$ đối với cơ sở $\{\overrightarrow{e_i}\}$, $x_i \in \mathbb{F}$, $i = 1, 2, \ldots, n$.

Khi đó bộ $(x_1, x_2, \ldots, x_n)$ được gọi là *tọa độ* của M trong mục tiêu $\{O, \overrightarrow{e_i}\}$ và x_i được gọi là tọa độ thứ i. Ta ký hiệu tọa độ của M là $M(x_i)$ hoặc (x_i).

Giả sử M có tọa độ (x_i) và N có tọa độ (y_i) đối với mục tiêu $\{\overrightarrow{e_i}\}$. Ta có $\overrightarrow{MN} = \overrightarrow{ON} - \overrightarrow{OM} = (y_i - x_i)$. Như vậy $(y_i - x_i)$ là tọa độ của vector $\overrightarrow{MN}$ trong mục tiêu $\{O, \overrightarrow{e_i}\}$.

Nhận xét 2

1. Giả sử trên $\mathcal{A}^n$ đã chọn được mục tiêu cố định $\{O, \overrightarrow{e_i}\}$. Xét ánh xạ

$$\begin{aligned} \varphi : \mathcal{A} &\to \mathbb{F}^n \\ M &\to (x_i) \end{aligned}$$

với (x_i) là tọa độ của M. Khi đó φ là song ánh và mỗi điểm được đồng nhất với một phần tử của $\mathbb{F}^n$. Nói cách khác đối tượng hình học được đồng nhất với đối tượng đại số.

2. Xét mục tiêu affine $\{O, \overrightarrow{e_i}\}$ của $\mathcal{A}^n$ và gọi $E_i \in \mathcal{A}$ là các điểm sao cho $\overrightarrow{OE_i} = \overrightarrow{e_i}$. Khi đó hệ điểm $\{O, E_1, E_2, \ldots, E_n\}$ độc lập affine. Ngược lại, một hệ gồm $n+1$ điểm $\{O, E_1, E_2, \ldots, E_n\}$ độc lập affine xác định một mục tiêu affine $\{O, \overrightarrow{e_i}\}$ với $\overrightarrow{e_i} = \overrightarrow{OE_i}$. Nếu ta chọn $O = (0, \ldots, 0)$ và $E_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ với số 1 ở vị trí i thì đây được gọi là cơ sở chính tắc.

3. Siêu phẳng đi qua n điểm độc lập O, E_1, E_2, ..., E_{i-1}, E_{i+1}, ..., E_n được gọi là **siêu phẳng tọa độ thứ** i. Dễ thấy M thuộc siêu phẳng tọa độ thứ i khi và chỉ khi $x_i = 0$ với x_i là tọa độ thứ i của M.

Chương 12

Machine Learning

12.1 Các thuật toán cơ sở

Linear Regression

Giả sử ta có N điểm dữ liệu đầu vào $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ với $\boldsymbol{x}_i \in \mathbb{R}^d$. Ứng với từng điểm dữ liệu đầu vào $\boldsymbol{x}_i$ ta có một đầu ra y_i. Nghĩa là ta có N cặp dữ liệu $(\boldsymbol{x}_i, y_i)$.

Mục tiêu là xây dựng hàm số $\hat{y} = f(x_1, x_2, \ldots, x_d)$ sao cho tổng sai số của y_i và $\hat{y}_i$ là nhỏ nhất, tức là

$$\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2 \to \min$$

Để hàm số đạt giá trị nhỏ nhất (hoặc lớn nhất) ta tìm cực trị của hàm số và khảo sát. Tuy nhiên không phải hàm số nào cũng đạo hàm được. Một cách tiếp cận đơn giản là sử dụng hàm tuyến tính, dễ xây dựng và luôn khả vi. Ta đặt

$$\hat{y} = f(x_1, x_2, \ldots, x_d) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

Lúc này, hàm mất mát ở trên có dạng

$$\mathcal{L} = \sum_{i=1}^{N} \|y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}\|^2$$

Bình phương chuẩn Euclid chính là bình phương của vector. Do đó dưới dấu tổng là các hàm số bình phương. Khi đạo hàm riêng theo w_j ta có

$$\frac{\partial \mathcal{L}}{\partial w_j} = \sum_{i=1}^{N} 2x_{ij}(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}))$$

với $1 \leq j \leq d$. Với $j = 0$ có chút khác biệt, $\dfrac{\partial \mathcal{L}}{\partial w_0} = \sum_{i=1}^{N} 2(y_i - (w_0 + w_1 x_{i1} + \ldots + w_d x_{id}))$.

Ta cho các đạo hàm riêng $\dfrac{\partial \mathcal{L}}{\partial w_j}$ bằng 0 thì được

$$\begin{aligned}
&\sum_{i=1}^{N} x_{ij}(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}) = \sum_{i=1}^{N} x_{ij} y_i \\
\Leftrightarrow\ & w_0 \sum_{i=1}^{N} x_{ij} + w_1 \sum_{i=1}^{N} x_{ij} x_{i1} + w_2 \sum_{i=1}^{N} x_{ij} x_{i2} \\
&\qquad + \cdots + w_d \sum_{i=1}^{N} x_{ij} x_{id} = \sum_{i=1}^{N} x_{ij} y_i
\end{aligned}$$

Bây giờ chúng ta cần biểu diễn các dấu tổng lại thành dạng đại số (ma trận, vector) vì chúng sẽ được sử dụng để nhân với vector $\boldsymbol{w} = (w_0, w_1, \ldots, w_d)$.

Ta có $\sum_{i=1}^{N} x_{ij} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{pmatrix}$.

Ta cũng có $\sum_{i=1}^{N} x_{ij} x_{i1} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} \end{pmatrix} \cdot \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{pmatrix}$.

Cứ tương tự như vậy, ta xếp các dấu sigma thành dạng cột thì tương đương với

$$\begin{pmatrix} * & \sum_{i=1}^{N} x_{ij} & * \\ * & \sum_{i=1}^{N} x_{ij} x_{i1} & * \\ \vdots & \vdots & \vdots \\ * & \sum_{i=1}^{N} x_{ij} x_{id} & * \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ \cdots & \cdots & \ddots & \cdots \\ x_{1d} & x_{2d} & \cdots & x_{Nd} \end{pmatrix} \cdot \begin{pmatrix} * & x_{1j} & * \\ * & x_{2j} & * \\ \vdots & \vdots & \vdots \\ * & x_{Nj} & * \end{pmatrix}$$

Ghép các cột theo thứ tự j từ 0 tới d ta có

$$\begin{aligned}\begin{pmatrix} w_0 & w_1 & \cdots & w_d \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ \cdots & \cdots & \ddots & \cdots \\ x_{1d} & x_{2d} & \cdots & x_{Nd} \end{pmatrix} \\ \times \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & x_{N1} & \cdots & x_{Nd} \end{pmatrix} \\ = \begin{pmatrix} y_1 & y_2 & \cdots & y_N \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & x_{N1} & \cdots & x_{Nd} \end{pmatrix}\end{aligned}$$

Hay nói cách khác, nếu ta đặt $\boldsymbol{w} = (w_0, w_1, \ldots, w_d)$ là ma trận hàng, $\boldsymbol{X}$ là ma trận có các hàng là các input, thì phương trình trên được viết lại là $\boldsymbol{wX}^T\boldsymbol{X} = \boldsymbol{yX}$.

Nếu đặt $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$ và $\boldsymbol{b} = \boldsymbol{yX}$ thì đây là hệ phương trình theo các ẩn $w_0, w_1, \ldots, w_d$. Tuy nhiên không phải lúc nào $\boldsymbol{A}$ cũng khả nghịch nên chúng ta sẽ sử dụng một khái niệm gọi là *giả nghịch đảo* để tìm nghiệm cho hệ phương trình.

Ký hiệu $\boldsymbol{A}^\dagger$ là giả nghịch đảo của ma trận $\boldsymbol{A}$. Khi đó nghiệm của hệ phương trình là $\boldsymbol{w} = \boldsymbol{bA}^\dagger$.

K-Means clustering

Một công việc thường được quan tâm là phân loại một nhóm các đối tượng thành những nhóm nhỏ hơn theo những tiêu chí nhất định.

Tương tự như phần trước, chúng ta có N điểm dữ liệu $\boldsymbol{x}_i$ thuộc $\mathbb{R}^d$. Ta muốn phân cụm các vector này vào những cluster (cụm) sao cho chúng gần nhau nhất (về mặt khoảng cách Euclid).

Giả sử ta muốn phân N điểm dữ liệu trên vào $K < N$ cluster. Ta cần tìm các điểm $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_K$ là tâm của các cụm, sao cho tổng khoảng cách từ các điểm $\boldsymbol{x}_i$ tới tâm cluster mà nó được phân vào là nhỏ nhất. Nghĩa là ứng với center $\boldsymbol{m}_1$ ta cần tìm các điểm $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_t}$ sao cho $\sum_{j=1}^{t} \|\boldsymbol{x}_{i_j} - \boldsymbol{m}_1\|^2$ nhỏ nhất. Tương tự cho các tâm khác.

Nhưng câu chuyện phức tạp ở đây là, tâm nằm ở đâu để có thể bao quát các điểm? Tâm được chọn phải có tính tổng quát, và việc phân các điểm vào cluster tương ứng với tâm thực hiện như thế nào?

Một kỹ thuật thường được sử dụng là *one-hot*. Với mỗi điểm dữ liệu $\boldsymbol{x}_i$ ta thêm một label $\boldsymbol{y}_i = (y_{i1}, \cdots y_{iK})$. Điểm $\boldsymbol{x}_i$ sẽ thuộc cluster j khi $y_{ij} = 1$, không

thuộc thì bằng 0. Như vậy chỉ có đúng một phần tử của $\boldsymbol{y}_i$ bằng 1, còn lại bằng 0. Như vậy ràng buộc của $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iK})$ là $y_{ij} \in \{0, 1\}$ và $\sum_{j=1}^{K} y_{ij} = 1$.

Khi đó, ta mong muốn phân các điểm $\boldsymbol{x}_i$ vào cluster $\boldsymbol{m}_k$ để khoảng cách tới tâm $\boldsymbol{m}_k$ là ngắn nhất, hay $\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 \to \min$. Thêm nữa, với cách ký hiệu y_{ij} như trên, biểu thức tương đương với

$$\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 = y_{ik}\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 = \sum_{j=1}^{K} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

vì điểm $\boldsymbol{x}_i$ sẽ thuộc cluster $\boldsymbol{m}_k$ nào đó với $1 \leq k \leq K$.

Sai số cho toàn bộ dữ liệu lúc này sẽ là

$$\mathcal{L}(\boldsymbol{Y}, \boldsymbol{M}) = \sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

Ta cần tối ưu $\boldsymbol{Y}$ và $\boldsymbol{M}$. Việc tối ưu hai ma trận cùng lúc là rất khó thậm chí bất khả thi. Do đó chúng ta có một cách tiếp cận khác là luân phiên cố định một bên và tối ưu bên còn lại. Từ đó công việc được chia làm hai bước.

Bước 1. Cố định $\boldsymbol{M}$, tìm $\boldsymbol{Y}$.

Giả sử ta đã biết các center $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_K$. Lúc này ta cần phân các điểm $\boldsymbol{x}_i$ vào cluster gần nó nhất. Dễ thấy rằng center gần nó nhất sẽ có khoảng cách Euclid ngắn nhất. Do đó ta tìm j sao cho $\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$ đạt nhỏ nhất. Không cần thiết phải tính căn bậc hai để giảm độ phức tạp.

Bước 2. Cố định $\boldsymbol{Y}$, tìm $\boldsymbol{M}$.

Khi đã biết $\boldsymbol{Y}$ tức là ta đã biết điểm nào được phân vào cluster nào. Khi đó ta cần tìm tâm cho từng cluster. Gọi $l(\boldsymbol{m}_j)$ là hàm tổng bình phương khoảng cách các điểm trong cluster tới tâm $\boldsymbol{m}_j$. Nghĩa là

$$l(\boldsymbol{m}_j) = \sum_{i=1}^{N} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

Mục tiêu của chúng ta là tối ưu tâm $\boldsymbol{m}_j$. Do đó ta đạo hàm theo vector $\boldsymbol{m}_j$ thu được $\dfrac{\partial l(\boldsymbol{m}_j}{\partial \boldsymbol{m}_j} = \sum_{i=1}^{N} 2y_{ij}(\boldsymbol{x}_i - \boldsymbol{m}_j)$. Cho đạo hàm bằng 0 và biến đổi ta có

$$\begin{aligned}
&2\sum_{i=1}^{N} y_{ij}(\boldsymbol{x}_i - \boldsymbol{m}_j) = 0 \\
\Leftrightarrow\ &\boldsymbol{m}_j \sum_{i=1}^{N} y_{ij} = \sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i \\
\Leftrightarrow\ &\boldsymbol{m}_j = \frac{\sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i}{\sum_{i=1}^{N} y_{ij}}
\end{aligned}$$

Để ý rằng, $\sum_{i=1}^{N} y_{ij}$ là số lượng điểm trong cluster, và $\sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i$ là tổng các điểm trong cluster. Như vậy $\boldsymbol{m}_j$ là trung bình cộng các điểm trong cluster j.

Algorithm 3 Thuật toán K-Means clustering

Require: Dữ liệu $\boldsymbol{X}$ (có N điểm dữ liệu) và số cluster K
Ensure: Các center $\boldsymbol{M}$ và label $\boldsymbol{y}$ cho mỗi điểm dữ liệu
1. Chọn K điểm bất kì làm các cluster ban đầu.
2. Phân mỗi điểm dữ liệu vào cluster gần nó nhất (cố định M, tìm Y).
3. Nếu việc phân dữ liệu vào các cluster ở bước 2 không thay đổi so với trước đó thì dừng thuật toán.
4. Cập nhật center mới cho mỗi cluster bằng cách lấy trung bình cộng các điểm trong cluster (cố định Y, tìm M).
5. Quay lại bước 2.

Gradient Descent

Trong nhiều trường hợp chúng ta thường không thể tìm nghiệm của phương trình đạo hàm để từ đó tìm các cực trị địa phương. Một phương pháp hiệu quả là gradient descent.

Hàm một biến

Giả sử x^* là local extremum (cực trị địa phương) của hàm số $f(x)$. Khi đó chúng ta xây dựng dãy số $\{x_n\}$ hội tụ về x^*. Ý tưởng thực hiện là dựa trên nhận xét, nếu x_n nằm bên phải x^* thì x_{n+1} nằm giữa x^* và x_n. Ta đã biết nếu x^* là một điểm cực trị thì $f'(x) > 0$ với $x > x^*$ mà x_n đi từ bên phải sang bên trái (ngược chiều Ox nên mang dấu âm). Từ đó chúng ta có công thức chung sau

$$x_{n+1} = x_n - \eta f'(x_n)$$

Trong đó η là một số dương nhỏ, gọi là *learning rate* (tốc độ học).

Ta chọn x_0 là một điểm bất kì. Tuy nhiên việc chọn x_0 cũng có thể ảnh hưởng đến tốc độ hội tụ.

Ví dụ với hàm số $f(x) = x^2 + 5\sin x$. Ta có đạo hàm là $f'(x) = 2x + 5\cos x$. Việc giải phương trình đạo hàm bằng 0 là điều không dễ dàng. Do đó gradient descent tỏ ra hiệu quả trong trường hợp này.

Chọn $\eta = 0.1$ và $x_0 = 5$. Sau đó chọn $\eta = 0.1$ và $x_0 = -5$. Ta thấy trường hợp sau tốn ít vòng lặp hơn do $x_0 = -5$ gần điểm cực trị hơn (≈ -1.11).

Hàm nhiều biến

Lúc này đầu vào của hàm số là một vector $\boldsymbol{x}$. Đặt $\nabla f(\boldsymbol{x})$ là đạo hàm của hàm f theo vector $\boldsymbol{x}$. Tương tự, ta xây dựng dãy vector $\{\boldsymbol{x}_n\}$ hội tụ về cực trị $\boldsymbol{x}^*$. Công thức lúc này là

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \eta \cdot \nabla f(\boldsymbol{x}_n)$$

Ta đã biết đạo hàm của hàm số theo vector cũng là vector cùng cỡ. Do đó giả sử $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$ thì đạo hàm của nó là

$$\nabla f(\boldsymbol{x}) = \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\Big)$$

Với ví dụ là bài toán Linear Regression, lúc này hàm mất mát là

$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N}\|y_i - \boldsymbol{x}_i\boldsymbol{w}^T\|^2 = \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}^T\|^2$$

Đạo hàm của hàm mất mát là

$$\nabla\mathcal{L} = \frac{1}{N}(\boldsymbol{w}\boldsymbol{X}^T - \boldsymbol{y})\boldsymbol{X}$$

Lúc này, với vector khởi đầu $\boldsymbol{w}_0$ chúng ta xây dựng dãy $\{\boldsymbol{w}_n\}$ tới khi nhận được $\boldsymbol{w}_n/d < \varepsilon$, với d là độ dài vector $\boldsymbol{w}$.

Perception Learning Algorithm

Một trong những nhiệm vụ quan trọng nhất của ML là phân loại (tiếng Anh - classification).

Perception là thuật toán phân loại cho trường hợp đơn giản nhất khi có hai lớp. Nếu ta có các điểm dữ liệu cho trước trong không gian d chiều, ta muốn tìm một siêu phẳng ($(d-1)$-phẳng) chia các điểm dữ liệu đó thành hai phần. Sau đó khi có một điểm dữ liệu mới ta chỉ cần bỏ nó vào bên này hoặc bên kia của siêu phẳng.

Trong dạng này, mỗi điểm dữ liệu được biểu diễn ở dạng cột của ma trận. Giả sử các điểm dữ liệu là $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, với $\boldsymbol{x}_i \in \mathbb{R}^d$, thì ma trận dữ liệu là $\boldsymbol{X} = \begin{pmatrix}\boldsymbol{x}_1^T & \boldsymbol{x}_2^T & \cdots & \boldsymbol{x}_N^T\end{pmatrix}$. Ta gọi nhãn tương ứng với N điểm dữ liệu trên là vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ với $y_i = 1$ nếu $\boldsymbol{x}_i$ thuộc class xanh, và $y_i = -1$ nếu $\boldsymbol{x}_i$ thuộc class đỏ.

Một siêu phẳng có phương trình là

$$f_{\boldsymbol{w}}(\boldsymbol{x}) = w_0 + w_1x_1 + \ldots + w_dx_d = \boldsymbol{w}\cdot\boldsymbol{x}^T$$

Một điểm thuộc nửa không gian (tạm gọi là *bên này*) đối với siêu phẳng thì $f_{\boldsymbol{w}}(\boldsymbol{x}) < 0$, nếu thuộc nửa bên kia thì $f_{\boldsymbol{w}}(\boldsymbol{x}) > 0$, nếu nằm trên phẳng thì bằng 0.

Gọi label($\boldsymbol{x}$) là nhãn của điểm $\boldsymbol{x}$. Khi đó điểm $\boldsymbol{x}$ thuộc một trong hai bên của phẳng nên label($\boldsymbol{x}$) = sgn($\boldsymbol{w} \cdot \boldsymbol{x}^T$) với sgn là hàm dấu. Ta quy ước sgn(0) = 1.

Khi một điểm bị phân loại sai class thì ta nói điểm đó bị **misclassified**. Ý tưởng của thuật toán là làm giảm thiểu số lượng điểm bị misclassified qua nhiều lần lặp. Đặt

$$J_1(\boldsymbol{w}) = \sum_{\boldsymbol{x}_i \in \mathcal{M}} (-y_i \cdot \text{sgn}(\boldsymbol{w} \cdot \boldsymbol{w}_i^T))$$

trong đó $\mathcal{M}$ là tập các điểm bị misclassified (tập này sẽ thay đổi theo $\boldsymbol{w}$.

Nếu $\boldsymbol{x}_i$ bị misclassified thì y_i và sgn($\boldsymbol{w} \cdot \boldsymbol{x}_i^T$) ngược dấu nhau. Nói cách khác, $-y_i \cdot \text{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}_i^T) = 1$. Từ đó $J_1(\boldsymbol{w})$ là hàm đếm số lượng điểm bị misclassified. Ta thấy rằng $J_1(\boldsymbol{w}) \geq 0$ nên ta cần tối ưu để hàm này đạt giá trị nhỏ nhất bằng 0. Khi đó không điểm nào bị misclassified.

Tuy nhiên có một vấn đề. Hàm $J_1(\boldsymbol{w})$ là hàm rời rạc (hàm sgn) nên rất khó tối ưu vì không thể tính đạo hàm. Do đó chúng ta cần một cách tiếp cận khác, một hàm mất mát khác tốt hơn.

Nếu ta bỏ đi hàm sgn thì có hàm

$$J(\boldsymbol{w}) = \sum_{\boldsymbol{x}_i \in \mathcal{M}} (-y_i \cdot \boldsymbol{w} \cdot \boldsymbol{x}^T)$$

Nhận xét. Một điểm bị misclassified nằm càng xa biên giới (siêu phẳng) thì giá trị $\boldsymbol{w} \cdot \boldsymbol{x}_i^T$ càng lớn, tức là hàm J đi ra xa so với giá trị nhỏ nhất. Hàm J cũng đạt min ở 0 nên ta cũng có thể dùng hàm này để loại bỏ các điểm bị misclassified.

Lúc này hàm $J(\boldsymbol{x})$ khả vi nên ta có thể dùng GD hoặc SGD để tìm nghiệm cho bài toán.

Nếu xét tại một điểm thì

$$J(\boldsymbol{w}, \boldsymbol{x}_i, y_i) = -y_i \cdot \boldsymbol{w} \cdot \boldsymbol{x}_i^T \Rightarrow \frac{\partial J}{\partial \boldsymbol{w}} = -y_i \boldsymbol{x}_i$$

Khi đó quy tắc để cập nhật là $\boldsymbol{w} = \boldsymbol{w} + \eta \cdot y_i \cdot \boldsymbol{x}_i$ với η là learning rate (thường chọn bằng 1). Nói cách khác ta đang xây dựng dãy $\{\boldsymbol{w}_n\}$ hội tụ lại nghiệm bài toán với công thức $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \eta \cdot y_i \cdot \boldsymbol{x}_i$.

Thuật toán PLA có thể được mô tả như sau:

1. Chọn ngẫu nhiên vector $\boldsymbol{w}$ với w_i xấp xỉ 0.
2. Duyệt ngẫu nhiên qua các $\boldsymbol{x}_i$:

- Nếu $\boldsymbol{x}_i$ được phân lớp đúng, tức sgn($\boldsymbol{w} \cdot \boldsymbol{x}_i^T$) = y_i thì ta không cần làm gì.
- Nếu $\boldsymbol{x}_i$ bị misclassified, ta cập nhật $\boldsymbol{w}$ theo công thức $\boldsymbol{w} = \boldsymbol{w} + \eta \cdot y_i \cdot \boldsymbol{x}$.

3. Kiểm tra xem có bao nhiêu điểm bị misclassified. Nếu không còn điểm nào thì ta dừng thuật toán, ngược lại thì quay lại bước 2.

Phụ lục A

NSUCRYPTO 2023

Problem 1. Affine cipher

Đây là bài 1 của round 2 và được giải bởi bạn Chương.

Đề bài

Ta xét bảng chữ cái A, ..., Z, α, β, γ có 29 chữ cái. Ta đánh số A, ..., Z từ 0 tới 25, và α, β, γ là 26, 27, 28.

Ta sử dụng cryptosystem mã hóa từng khối 2 ký tự, gọi là bigram. Với x và y là hai ký tự của bigram, thì plaintext sẽ là $P = 29x + y$.

Mã hóa sử dụng biến đổi affine (giống hệ mã affine) là $C = aP+b \pmod{841}$.

Khi phân tích một đoạn văn bản dài, người ta phát hiện ra rằng các bigram sau xuất hiện nhiều nhất "$\beta\gamma$", "UM" và "LC". Đồng thời, trong tiếng Anh thì các bigram "TH", "HE" và "IN" cũng xuất hiện nhiều nhất.

Q. Có thể giải mã "KEUDCR" mà không cần khóa hay không? Còn key thì sao?

Giải

Theo thống kê các bigram xuất hiện nhiều nhất trong ciphertext và trong plaintext sẽ khớp nhau. Do đó có thể thấy "TH" mã hóa thành "$\beta\gamma$" và "HE" mã hóa thành "LC". Như vậy ta có hệ phương trình

$$\begin{aligned} 812 &= a \cdot 558 + b \pmod{841} \\ 321 &= a \cdot 207 + b \pmod{841} \end{aligned}$$

Giải hệ ta có $a = 15, b = 10$. Đây là key.

Từ đây chúng ta có thể giải mã thành **CRYPTO** là plaintext ban đầu. Bài này ăn trọn 4/4 điểm.

Problem 2. Simple ideas for primes

Đề bài

Chúng ta xem xét một số dãy số bao gồm các số nguyên tố.

• *Số Fermat*, $F_k = 2^{2^k} + 1$, với k bắt đầu từ 0. Ta có các số F_0, F_1, F_2, F_3, F_4 là các số nguyên tố, còn F_5 thì không phải.

• *Số Mersenne*, $M_k = 2^k - 1$. Ta có M_2, M_3, M_5, M_7 là các số nguyên tố, trong khi M_{11} là hợp số. Các số nguyên tố Mersenne là các số dạng $2^k - 1$ với k là số nguyên tố.

• Dãy số $31, 331, 3\,331, 33\,331, 333\,331, 3\,333\,331, 33\,333\,331$ là các số nguyên tố được xây dựng theo quy tắc trên, nhưng số $333\,333\,331$ là hợp số chia hết cho 17.

Ta nói dãy Fermat trên có *sequence primality parameter* là 5, dãy Mersenne bằng 4, dãy cuối cùng bằng 7.

Q. Xây dựng một dãy bao gồm các số nguyên tố như vậy. Điều kiện quan trọng ở đây là các số hạng được xác định bởi chỉ số của dãy, không phụ thuộc vào các số trước nó.

Giải

Bắt đầu với dãy Euler

$$f(n) = n^2 + n + 41$$

Đây là dãy các số nguyên tố với $n = 0, 1, \ldots, 39$ và $f(40)$ là hợp số. Như vậy đây là dãy nguyên tố độ dài 40.

Và tất nhiên, dãy "ai cũng biết" thì chỉ được 2 điểm thôi 😳.

Sau khi tham khảo những thí sinh khác thì có một số cách xây dựng nhằm cải tiến điều này, tham khảo từ [1].

Nếu ta chuyển dãy trên thành

$$g(n) = f(n - 40) = (n - 40)^2 + (n - 40) + 41 = n^2 - 79n + 1601$$

thì thu được dãy số nguyên tố với độ dài 80. Các nhà toán học thế kỉ 20 đã chứng minh được rằng, nếu $p(x)$ là một đa thức sinh ra dãy số nguyên tố với $0 \leqslant x \leqslant n$ thì đa thức $p(n - x)$ cũng vậy.

Trong bảng, dãy nguyên tố có độ dài lớn nhất là 56 được biểu diễn bởi đa thức

$$\frac{1}{4}\left(n^5 - 133n^4 + 6729n^3 - 158379n^2 + 1720294n - 6823316\right)$$

[1] https://mathworld.wolfram.com/Prime-GeneratingPolynomial.html

Problem 3. Mixed hashes

Đề bài

Alice và Bob trao đổi các thông điệp mã hóa. Họ dùng thuật toán mã hóa khối PRESENT với key 80-bit và ECB mode. Ở đây, thông tin được lưu dạng ảnh .ppm.

Header của file .ppm gồm 3 dòng theo dạng **P6\nX\nY\n255**. Trong đó X và Y là kích thước của ảnh theo chiều ngang và dọc.

Để đảm bảo an toàn, header sẽ được loại bỏ trước khi encrypt. Để có thể khôi phục header, hash của header sẽ được gửi đi thay vì header. Khi đó 3 phần của header sẽ được ngăn cách bởi dấu cách (space) thay vì newline như trên. Nghĩa là "P6 X Y 255".

Bob chuẩn bị 8 ảnh (trong file đính kèm) mà không có header. Bob encrypt 8 ảnh đó với cùng một key theo thuật toán PRESENT và ECB mode. Bob cũng gửi hash của 8 headers đi kèm. Tuy nhiên các hash đã bị trộn lẫn với nhau. Liệu chúng ta có thể khôi phục thông điệp mà Bob muốn gửi Alice?

Giải

Bài này là bài 3 ở round 1 và round 2. Trong thời gian 2 round mình đều giải ra (round 2 chi tiết hơn và trình bày đẹp hơn :v).

Đề cho một file mẫu là mikky.ppm. Khi phân tích file này mình thấy rằng, nếu gọi w và h là độ rộng và độ cao của ảnh (lấy từ header) thì độ dài file không có header là $3 \cdot w \cdot h$.

Sau khi encrypt bằng thuật toán mã hóa khối với ECB mode, độ dài sẽ là $3 \cdot w \cdot h + pd$, trong đó pd là padding. Theo thuật toán **PRESENT** thì $0 \leqslant pd \leqslant 8$.

Với dự đoán rằng $w \approx h$, mình lấy căn bậc hai của độ dài các filel đề cho, và đưa ra dự đoán $w, h \in [400, 600]$. Nếu sai thì mình tăng độ rộng khoảng này thôi.

Tiếp theo, bruteforce w và h trong khoảng này, cho tới khi hash "P6 x y 255" xuất hiện trong số các hash trên, và

$$0 \leqslant \text{len}(ciphertext) - 3 * w * h \leqslant 8$$

thì mình lấy w và h này. Thế là mình có header.

Do cả 8 file được encrypt bởi cùng một key **PRESENT**, và key có 80 bit tương ứng 10 bytes, hay 10 ký tự, nhìn đề mình nhận thấy có chuỗi **P6 X Y 255** là hợp lý. Như vậy key cho PRESENT là chuỗi **P6 X Y 255**.

Cuối cùng, mình giải mã lần lượt từng file với key trên, ghép header tương ứng vào, như vậy là mình giải mã được tất cả file rồi.

Vậy thông điệp gốc là ”♡Loveyou”. Bài này được 5/6 điểm vì không nộp code tính toán header, mất điểm vì chủ quan.

Problem 5. Primes

Đây là bài 5 của round 2 và được giải bởi người đồng đội Uyên.

Đề bài

Marcus chọn hai số nguyên tố lớn p và q rồi tính $n = p \cdot q$ và $m = p + q$. Sau đó số $n \cdot m$ được sử dụng trong cryptosystem.

Khi test Marcus thấy các số p và q cho tích $n \cdot m$ kết thúc bởi 2023. Điều đó khả thi không?

Giải

Do p và q là các số nguyên tố lớn nên chúng lẻ. Suy ra $m = p + q$ là số chẵn, nên tích $n \cdot m$ cũng là số chẵn.

Số chẵn kết thúc bởi 0, 2, 4, 6, 8 nên không thể kết thúc bởi 3. Do đó điều này không thể xảy ra.

Problem 6. An aggregated signature

Bài này không biết làm 🥹.

Đề bài

Trong một tổ chức quốc tế lớn, gọi là **NSUCRYPTO association**, mọi người quyết định tổ chức một tờ báo thông tin (news journal) trong lĩnh vực mật mã. Tổ chức muốn rằng, tin tức chỉ được công bố khi đã được kiểm duyệt bởi một nhóm lớn các nhà mật mã. Vì vậy, 10 000 chuyên gia mật mã đã được mời tới làm biên tập cho tờ báo.

Quy định công bố như sau. Tin tức được công bố khi nó được ký bởi tất cả các thành viên biên tập. Tuy nhiên các nhà mật mã không rảnh để thu thập 10 000 chữ ký cá nhân (gặp mình thì mình cũng không muốn 🥲). Do đó mọi người muốn một chữ ký postquantum dùng chung mà không thể chia nhỏ ra các chữ ký cá nhân.

Yêu cầu là cần xây dựng mô hình chữ ký thỏa mãn các yêu cầu sau:

- kích thước chữ ký không quá lớn, có thể hơn vài kilobytes;

• kích thước public key (để kiểm tra chữ ký) là nhỏ. Kích thước của key nên là cố định (hoặc gần cố định) kể cả khi số lượng chuyên gia tăng lên, ví dụ 20 000;
• việc kiểm tra chữ ký tốn không quá 2 phút;
• chữ ký có thể chống lại các tấn công trên máy tính lượng tử.

Problem 7. A unique coding

Bài này khi nhìn đề thì "có vẻ" câu hỏi Q2 là trường hợp nhỏ hơn của Q1. Mình giải Q2 (không chắc đúng hoàn toàn) nên lời giải sau đây áp dụng cho cả Q1 và Q2.

Đề bài

Xét binary error-correcting code $\mathcal{C}$ với độ dài n. Code này chỉ đơn giản là tập con của $\mathbb{F}_2^n$ thôi và ta truyền một phần tử của code này qua các kênh truyền.

Khi đi qua các kênh truyền các bit có thể bị sai, dẫn tới bị đảo bit. Khi nhận được vector $\boldsymbol{y} \in \mathbb{F}_2^n$, ta sẽ decode thành một phần tử thuộc $\mathcal{C}$ mà có khoảng cách gần $\boldsymbol{y}$ nhất, nói cách khác là Hamming weight ngắn nhất.

Xét cơ chế decode maximal-likelihood. Giả sử ta nhận được $\boldsymbol{y} \in \mathbb{F}_2^n$, ta muốn xét các trường hợp lỗi xảy ra ít nhất, gọi là $d_{\boldsymbol{y}}$, nghĩa là

$$d_{\boldsymbol{y}} = \min_{\boldsymbol{x} \in \mathcal{C}} wt(\boldsymbol{x}, \boldsymbol{y})$$

Tiếp theo, đặt $\mathcal{D}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathcal{C} : wt(\boldsymbol{x}, \boldsymbol{y}) = d_{\boldsymbol{y}}\}$. Cuối cùng ta decode $\boldsymbol{y}$ thành phần tử $\boldsymbol{x}$ nào đó trong $\mathcal{D}(\boldsymbol{y})$.

Chúng ta quan tâm tới các trường hợp code $\mathcal{C}$ khiến $|\mathcal{D}(\boldsymbol{y})| = 1$ với mọi $\boldsymbol{y} \in \mathbb{F}_2^n$. Nói cách khác khi nhận được $\boldsymbol{y}$ bất kì của $\mathbb{F}_2^n$ ta có thể decode thành một dạng duy nhất.

Q1. Code $\mathcal{C}$ như nào thì thỏa mãn tính chất này?

Q2. Code $\mathcal{C}$ như nào thỏa mãn tính chất này và là không gian tuyến tính con của $\mathbb{F}_2^n$?

Giải

Đầu tiên mình có nhận xét (khá rõ ràng) sau đây:

Nhận xét 1

Với mọi n, code $\mathcal{C} \equiv \mathbb{F}_2^n$ thỏa mãn tính chất trên.

Chúng ta có thể thấy rằng với mọi $\boldsymbol{y} \in \mathbb{F}_2^n$ nhận được thì sẽ decode thành chính nó trong $\mathcal{C}$.

Tiếp theo, ta xét nửa trên của $\mathbb{F}_2^n$. Trong hệ thập phân thì đó là các số từ 0 tới $2^{n-1}-1$ và có dạng

$$\boldsymbol{x} = (0, x_1, x_2, \ldots, x_{n-1})$$

Nói cách khác, code $\mathcal{C}$ là tập hợp

$$\mathcal{C} = \{\boldsymbol{x} = (0, x_1, x_2, \ldots, x_{n-1}) : x_i \in \mathbb{F}_2\}$$

Code $\mathcal{C}$ này thỏa mãn tính chất trên và mình sẽ chứng minh ngay sau đây.

Chứng minh. Giả sử ta nhận được $\boldsymbol{y} \in \mathbb{F}_2^n$. Ta có hai trường hợp:

• nếu $\boldsymbol{y} = (0, y_1, y_2, \ldots, y_{n-1})$, hay nói cách khác biểu diễn thập phân của $\boldsymbol{y}$ là từ 0 tới $2^{n-1}-1$, thì $\boldsymbol{y}$ được decode thành chính nó trong $\mathcal{C}$. Khi này $d_{\boldsymbol{y}} = 0$ nhỏ nhất và không có vector nào khác cho Hamming weight bằng 0 trừ chính nó.

• nếu $\boldsymbol{y} = (1, y_1, y_2, \ldots, y_{n-1})$, hay nói cách khác biểu diễn thập phân của $\boldsymbol{y}$ là từ 2^{n-1} tới 2^n-1, thì $\boldsymbol{y}$ được decode thành $\boldsymbol{x} = (0, y_1, y_2, \ldots, y_{n-1})$ trong $\mathcal{C}$. Khi này $d_{\boldsymbol{y}} = 1$ nhỏ nhất vì khác mỗi bit đầu tiên và cũng không có vector nào khác cho Hamming weight bằng 1.

□

Tiếp theo, mình viết các vector trong $\mathcal{C}$ thành các hàng của 1 ma trận $2^{n-1} \times n$. Gọi A là ma trận hoán vị các cột của ma trận $2^{n-1} \times n$ đó. Khi đó A là ma trận có tính chất: trên mỗi hàng và trên mỗi cột có đúng một phần tử (bằng 1) và ma trận A khả nghịch. Ví dụ, với $n = 4$, ma trận để hoán vị cột 2 với cột 4 là $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$.

Khi đó, nếu mình nhân ma trận $2^{n-1} \times n$ của code $\mathcal{C}$ với bất kì ma trận A nào như vậy thì code $\mathcal{C}'$ nhận được cũng thỏa mãn tính chất trên.

Ví dụ 1. Với $n = 4$ thì code $\mathcal{C}$ gồm các vector

$$\mathcal{C} = \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111\}$$

Với ma trận A hoán vị cột 2 và 4 như trên ta có

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} = \mathcal{C}'$$

Bây giờ mình sẽ chứng minh rằng với mọi ma trận A hoán vị các cột như vậy thì code $\mathcal{C}'$ cũng thỏa mãn tính chất.

Chứng minh. Đặt

$$\mathcal{C} = \{(0, x_1, x_2, \ldots, x_{n-1}), x_i \in \mathbb{F}_2\}$$

Gọi A là ma trận hoán vị cột kích thước $n \times n$. Khi đó ánh xạ

$$A : \mathbb{F}_2^n \to \mathbb{F}_2^n, \quad \boldsymbol{y} \to \boldsymbol{y} \cdot A$$

là song ánh do A là ma trận khả nghịch. Khi đó xét code

$$\mathcal{C}' = \{\boldsymbol{x} \cdot A : \boldsymbol{x} \in \mathcal{C}\}$$

Mình vẫn có hai trường hợp.

Trường hợp 1. Với $\boldsymbol{y} = (0, y_1, y_2, \ldots, y_{n-1}) \in \mathbb{F}_2^n$ từ 0 tới $2^{n-1} - 1$ như trên. Xét $\boldsymbol{y}' = \boldsymbol{y} \cdot A$.

Khi đó, với $\boldsymbol{x} = (0, y_1, y_2, \ldots, y_{n-1}) \in \mathcal{C}$, ta có $\boldsymbol{x}' = \boldsymbol{x} \cdot A \in \mathcal{C}'$. Từ đây suy ra

$$wt(\boldsymbol{x}' \oplus \boldsymbol{y}') = wt((\boldsymbol{x} \cdot A) \oplus (\boldsymbol{y} \cdot A)) = wt((\boldsymbol{x} \oplus \boldsymbol{y}) \cdot A) = wt(\mathbf{0} \cdot A) = 0$$

Ở đây $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{F}_2^n$.

Nói cách khác $d_{\boldsymbol{y}'} = 0$ và có duy nhất một vector $\boldsymbol{x}'$ được định nghĩa như trên thỏa mãn. Do đó $|\mathcal{D}(\boldsymbol{y}')| = 1$.

Trường hợp 2. Với $\boldsymbol{y} = (1, y_1, y_2, \ldots, y_{n-1}) \in \mathbb{F}_2^n$ từ 2^{n-1} tới $2^n - 1$ như trên. Ta cũng xét $\boldsymbol{y}' = \boldsymbol{y} \cdot A$.

Khi đó, với $\boldsymbol{x} = (0, y_1, y_2, \ldots, y_{n-1}) \in \mathcal{C}$, ta cũng có $\boldsymbol{x}' = \boldsymbol{x}\cdot \in \mathcal{C}'$. Từ đây ta có

$$wt(\boldsymbol{x}' \oplus \boldsymbol{y}') = wt((\boldsymbol{x} \cdot A) \oplus (\boldsymbol{y} \cdot A)) = wt((\boldsymbol{x} \oplus \boldsymbol{y}) \cdot A) = wt((1, 0, 0, \ldots, 0) \cdot A) = 1$$

Ở phép nhân vector $(1, 0, \ldots, 0)$ với ma trận A, vì ma trận A chỉ có duy nhất một cột có dạng $(1, 0, \ldots, 0)^T$ nên kết quả phép nhân là một vector có đúng một phần tử 1, còn lại là 0.

Nói cách khác $d_{\boldsymbol{y}'} = 1$ và có duy nhất một vector $\boldsymbol{x}'$ được định nghĩa như trên thỏa mãn. Do đó $|\mathcal{D}(\boldsymbol{y}')| = 1$.

Như vậy ta đã chứng minh xong. □

Hoàn toàn tương tự, khi code $\mathcal{C}$ là các vector bắt đầu với hai số 0 thì ta lần lượt xét $\boldsymbol{y}$ trong các khoảng $[0, 2^{n-2}-1]$, $[2^{n-2}, 2^{n-1}-1]$, $[2^{n-1}, 2^{n-1}+2^{n-2}-1]$, $[2^{n-1}+2^{n-2}-1, 2^n-1]$. Nghĩa là

$$\mathcal{C} = \{\boldsymbol{x} = (0, 0, x_1, x_2, \ldots, x_{n-2} : x_i \in \mathbb{F}_2^n)\}$$

Khi đó ta xét các vector $\boldsymbol{y}$ có dạng:

- $\boldsymbol{y} = (0, 0, y_1, y_2, \ldots, y_{n-2})$
- $\boldsymbol{y} = (0, 1, y_1, y_2, \ldots, y_{n-2})$
- $\boldsymbol{y} = (1, 0, y_1, y_2, \ldots, y_{n-2})$
- $\boldsymbol{y} = (1, 1, y_1, y_2, \ldots, y_{n-2})$

Theo quy nạp thì code C bắt đầu với i số 0 đều đúng, $0 \leqslant i \leqslant n$. Nghĩa là

$$\mathcal{C} = \{\boldsymbol{x} = (0, \ldots, 0, x_1, x_2, \ldots, x_{n-i}) : x_i \in \mathbb{F}_2\}$$

Sau đó chúng ta lại áp dụng phép nhân với ma trận hoán vị cột A như bên trên thì các code $\mathcal{C}'$ cũng thỏa mãn.

Vấn đề ở đây là, những code $\mathcal{C}$ như vậy là không gian vector sinh bởi i vector $(0 \leqslant i \leqslant n)$ trong các vector sau:

$$\begin{aligned}
\boldsymbol{v}_1 &= (1, 0, 0, \ldots, 0, 0) \\
\boldsymbol{v}_2 &= (0, 1, 0, \ldots, 0, 0) \\
\boldsymbol{v}_3 &= (0, 0, 1, \ldots, 0, 0) \\
\ldots &= \ldots \\
\boldsymbol{v}_n &= (0, 0, 0, \ldots, 0, 1)
\end{aligned}$$

Số cách chọn i vector từ n vector là

$$C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$$

cách. Nói cách khác có 2^n code $\mathcal{C}$ thỏa tính chất đề bài và là không gian tuyến tính của $\mathbb{F}_2^n$.

Ví dụ 2. Với $n = 3$ thì các code sau thỏa mãn tính chất

$$\mathcal{C}_1 = \{000\},$$
$$\mathcal{C}_2 = \{000, 001\},$$
$$\mathcal{C}_3 = \{000, 010\},$$
$$\mathcal{C}_4 = \{000, 100\},$$
$$\mathcal{C}_5 = \{000, 001, 010, 011\},$$
$$\mathcal{C}_6 = \{000, 001, 100, 101\},$$
$$\mathcal{C}_7 = \{000, 010, 100, 110\},$$
$$\mathcal{C}_8 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

Bình luận

Đối với Q1 có thể thấy rằng bất cứ code nào chỉ chứa đúng một vector sẽ thỏa mãn điều kiện. Lý do là vì bất cứ $\boldsymbol{y}$ nào được gửi tới cũng sẽ decode ra vector đó.

Bài này mình được 6/12 điểm vì đưa ra cách xây dựng tốt, trình bày đẹp.

Problem 8. Algebraic cryptanalysis

Bài này là bài 7 ở round 1 và là bài 8 ở round 2. Bài này mình giải khá qua loa ở round 1 và được giải đầy đủ, rõ ràng hơn bởi người đồng đội vip pro Chương ở round 2.

Đề bài

Bob muốn xây dựng stream cipher **BOB-0.1**.

Bob sử dụng một binary key độ dài 8 là $K = (k_1, \ldots, k_8)$. Sau đó anh ấy sinh ra dãy nhị phân β theo quy tắc:

- $\beta_n = k_n$ khi $n = 1, 2, \ldots, 8$
- $\beta_n = \beta_{n-1} \oplus \beta_{n-8}$ khi $n \geqslant 9$

Sau đó Bob sinh dãy nhị phân γ dùng trong phép XOR với plaintext. Dãy γ được tạo bởi quy tắc $\gamma_n = \beta_n \cdot \beta_{n+2} \oplus \beta_{n+7}$ với $n \geqslant 1$.

Alice chặn được 8 bit của γ sau khi để lỡ 1200 bit. Các bit đó là '00100001'. Liệu Alice có thể tìm lại được key K ban đầu không?

Giải

Độ dài K là 8 bit, nếu chúng ta brutefore $K = (k_1, \ldots, k_8)$ rồi sinh ra 1208 bit γ theo quy tắc trên và so sánh xem $\gamma_{1201}, \ldots, \gamma_{1208}$ nào khớp với 8 bit trên

thì ta có thể biết được K ban đầu là gì.

Và, bất ngờ chưa, có tới hai trường hợp K thỏa mãn :v :v

Bây giờ thì chúng ta cần xem xem tại sao lại có hai trường hợp thỏa mãn.

Cùng nhau khai triển $\beta_{n+1}, \ldots, \beta_{n+8}$ theo $(\beta_{n-7}, \ldots, \beta_n)$ nào.

- $\beta_{n+1} = \beta_n \oplus \beta_{n-7}$
- $\beta_{n+2} = \beta_{n+1} \oplus \beta_{n-6} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6}$
- $\beta_{n+3} = \beta_{n+2} \oplus \beta_{n-5} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5}$
- $\beta_{n+4} = \beta_{n+3} \oplus \beta_{n-4} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5} \oplus \beta_{n-4}$
- $\beta_{n+5} = \beta_{n+4} \oplus \beta_{n-3} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5} \oplus \beta_{n-4} \oplus \beta_{n-3}$
- $\beta_{n+6} = \beta_{n+5} \oplus \beta_{n-2} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5} \oplus \beta_{n-4} \oplus \beta_{n-3} \oplus \beta_{n-2}$
- $\beta_{n+7} = \beta_{n+6} \oplus \beta_{n-1} = \beta_n \oplus \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5} \oplus \beta_{n-4} \oplus \beta_{n-3} \oplus \beta_{n-2} \oplus \beta_{n-1}$
- $\beta_{n+8} = \beta_{n+7} \oplus \beta_n = \beta_{n-7} \oplus \beta_{n-6} \oplus \beta_{n-5} \oplus \beta_{n-4} \oplus \beta_{n-3} \oplus \beta_{n-2} \oplus \beta_{n-1}$

Nếu viết ở dạng phép nhân ma trận modulo 2 ta có

$$\begin{pmatrix} \beta_{n+1} \\ \beta_{n+2} \\ \beta_{n+3} \\ \beta_{n+4} \\ \beta_{n+5} \\ \beta_{n+6} \\ \beta_{n+7} \\ \beta_{n+8} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_{n-7} \\ \beta_{n-6} \\ \beta_{n-5} \\ \beta_{n-4} \\ \beta_{n-3} \\ \beta_{n-2} \\ \beta_{n-1} \\ \beta_n \end{pmatrix}$$

Ma trận to to kia là ma trận khả nghịch. Do đó, nếu chúng ta có các $\beta_{n+1}, \ldots, \beta_{n+8}$ thì chúng ta có thể tìm ngược lại $\beta_{n-7}, \ldots, \beta_n$. Tiếp tục quá trình này cuối cùng ta có thể tìm lại $(\beta_1, \ldots, \beta_8) = K$.

Tiếp theo, cũng tương tự, chúng ta biểu diễn dãy γ theo β..

- $\gamma_{n+1} = \beta_{n+1} \cdot \beta_{n+3} \oplus \beta_{n+8}$
- $\gamma_{n+2} = \beta_{n+2} \cdot \beta_{n+4} \oplus \beta_{n+1} \oplus \beta_{n+8}$
- $\gamma_{n+3} = \beta_{n+3} \cdot \beta_{n+5} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+8}$
- $\gamma_{n+4} = \beta_{n+4} \cdot \beta_{n+6} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+8}$
- $\gamma_{n+5} = \beta_{n+5} \cdot \beta_{n+7} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+8}$
- $\gamma_{n+6} = \beta_{n+6} \cdot \beta_{n+8} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \oplus \beta_{n+8}$
- $\gamma_{n+7} = \beta_{n+7} \cdot \beta_{n+1} \oplus \beta_{n+7} \cdot \beta_{n+8} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \oplus \beta_{n+6} \oplus \beta_{n+8}$
- $\gamma_{n+8} = \beta_{n+8} \cdot \beta_{n+1} \oplus \beta_{n+8} \cdot \beta_{n+2} \oplus \beta_{n+1} \oplus \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \oplus \beta_{n+6} \oplus \beta_{n+7}$

Trường hợp 1. $\beta_{n+1} = 0$. Khi đó từ γ_{1201} tới γ_{1208} tương đương với hệ phương

trình

$$
\begin{aligned}
0 &= \beta_{n+8} \\
0 &= \beta_{n+2} \cdot \beta_{n+4} \\
1 &= \beta_{n+3} \cdot \beta_{n+5} \oplus \beta_{n+2} \\
0 &= \beta_{n+4} \cdot \beta_{n+6} \oplus \beta_{n+2} \oplus \beta_{n+3} \\
0 &= \beta_{n+5} \cdot \beta_{n+7} \oplus \beta_{n+2} \oplus \beta_{n+3} \cdot \beta_{n+4} \\
0 &= \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \\
0 &= \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \oplus \beta_{n+6} \\
1 &= \beta_{n+2} \oplus \beta_{n+3} \oplus \beta_{n+4} \oplus \beta_{n+5} \oplus \beta_{n+6} \oplus \beta_{n+7}
\end{aligned}
$$

Hệ phương trình trên có nghiệm duy nhất $(\beta_i) = (0, 1, 1, 0, 0, 0, 1, 0)$.

Trường hợp 2. $\beta_{n+1} = 1$. Tương tự hệ phương trình cũng có nghiệm duy nhất $(\beta_i) = (1, 1, 1, 0, 0, 0, 0, 1)$.

Như vậy ta có hai nghiệm thỏa mãn chuỗi 8 bit $\gamma_{1201}, \ldots, \gamma_{1208}$. Do không có điều kiện nào thêm, ta không thể xác định đâu là khóa trong hai trường hợp trên.

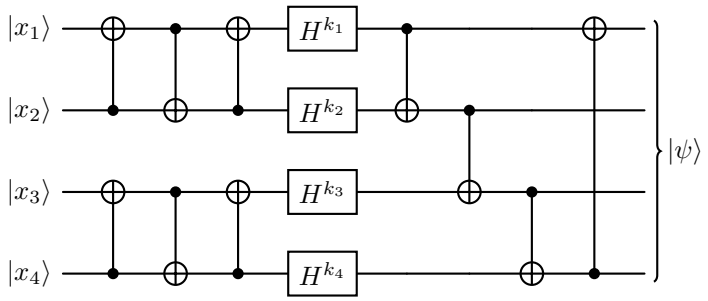
Bài này bạn Chương được 4/4 điểm. Good job nigga 😂.

Problem 10. Quantum encryption

Đây là bài 8 của round 1 và bài 10 của round 2. Bài này sai gần bước cuối mới cay 😭.

Đề bài

Bob tạo một thuật toán mã hóa encrypt 4 bit (x_1, x_2, x_3, x_4) bằng key cũng 4 bit (k_1, k_2, k_3, k_4) với mạch sau:



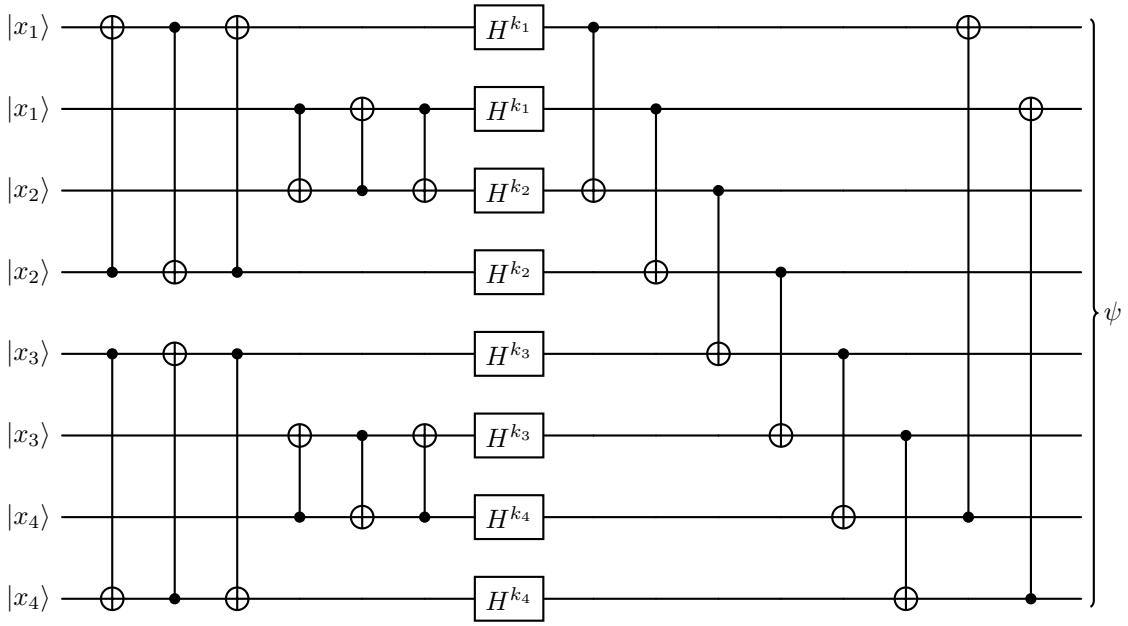
Plaintext 4 bit (x_1, x_2, x_3, x_4) được biểu diễn ở dạng 4-qubit "plainstate" $|x1, x2, x3, x4\rangle$. Quantum state này là input cho mạch ở dạng qubit đơn đi qua các cổng.

Ở đây hai loại cổng được sửa dụng là CNOT và Hadamard.

Ký hiệu H^b với $b \in \{0, 1\}$ có nghĩa là, nếu $b = 0$ thì cổng đồng nhất I được sử dụng (không thay đổi), còn nếu $b = 1$ thì cổng Hadamard sẽ được sử dụng.

Kết quả sau khi qua mạch là "cipherstate" $|\psi\rangle$.

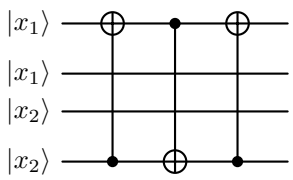
Bob có nhiệm vụ tăng số qubit đầu vào lên nhằm giảm các sai số khi tính toán và truyền dữ liệu trên kênh quantum. Do đó Bob biến đổi thành mạch sau:



Alice nói rằng cô ấy có thể tìm lại được key nếu biết N amplitude của kết quả $|\psi\rangle$. Do có 8 qubits ở kết quả nên số lượng amplitude tối đa là $2^8 = 256$, nói cách khác $N \leqslant 256$. Vậy Alice cần ít nhất bao nhiêu amplitude là đủ để tìm lại key?

Giải

Đầu tiên xét 4 dây trên, 4 dây dưới tương tự.



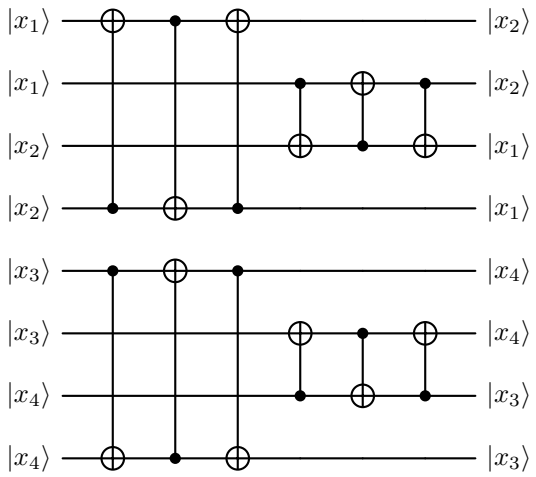
Chúng ta xét dây 1 và 4 của mạch (tương tự cho dây 2 và 3). Áp dụng cổng CNOT liên tiếp 3 lần ta có

$$|x_1\rangle \otimes |x_2\rangle \to |x_1 \oplus x_2\rangle \otimes |x_2\rangle \to |x_1 \oplus x_2\rangle \otimes |x_1\rangle \to |x_2\rangle \otimes |x_1\rangle$$

Nói cách khác là đảo bit 😆.

(a) Dây 1 và dây 4 (b) Dây 2 và dây 3

Tương tự cho các cặp dây (5, 8) và (6, 7). Do đó khi tới trước các cổng Hadamard thì thứ tự các qubit từ trên xuống dưới là hình A.2.



Hình A.2: Qubits trước Hadamard

Mạch ở dây 1 và 2 đều có dạng $|x_2\rangle$ đi qua H^{k_1} nên sau khi qua cổng mình đặt $|z_2\rangle = H^{k_1}|x_2\rangle$.

Tương tự, $|z_1\rangle = H^{k_2}|x_1\rangle$ cho dây 3 và 4, $|z_4\rangle = H^{k_3}|x_4\rangle$ cho dây 5 và 6, $|z_3\rangle = H^{k_4}|x_3\rangle$ cho dây 7 và 8.

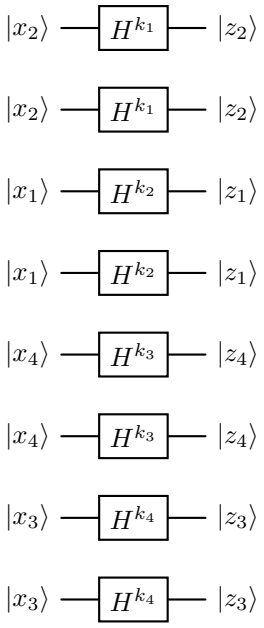
Mạch sau khi đi qua Hadamard có dạng A.3

Ở đây chúng ta có một lưu ý nhỏ có thể giúp ích trong việc giới hạn số lượng amplitude theo đề bài. Nếu $k_1 = 0$ thì $|z_2\rangle = |x_2\rangle$. Nếu $k_1 = 1$ thì $|z_2\rangle = \dfrac{|0\rangle + (-1)^{x_2}|1\rangle}{\sqrt{2}}$. Như vậy, hệ số trước $|0\rangle$ của $|z_2\rangle$ có thể là $0, 1, \dfrac{1}{\sqrt{2}}$ đều không âm.

Bây giờ chúng ta quay lại toán tử CNOT. Ma trận tương ứng của toán tử CNOT là $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. Kết quả sau khi thực hiện toán tử CNOT là hệ số trước $|00\rangle$ và $|01\rangle$ giữ nguyên, còn hệ số trước $|10\rangle$ và $|11\rangle$ đổi chỗ cho nhau.

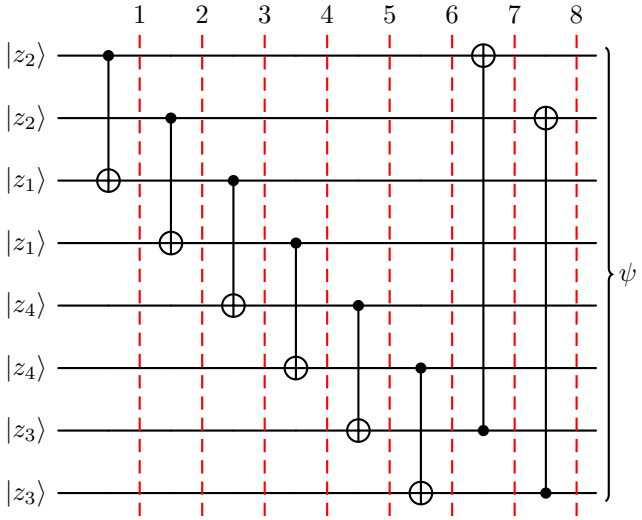
Đối với 3 qubit, mình **dự đoán** tương tự.

Ở cổng CNOT đầu tiên, dây 1 control dây 3. Nếu mình chỉ xét 3 dây đầu thì tích các qubit gồm $|000\rangle$, $|001\rangle$, $|010\rangle$, $|011\rangle$, $|100\rangle$, $|101\rangle$, $|110\rangle$, $|111\rangle$.



Hình A.3: Qubits sau Hadamard

Áp dụng "chiến thuật" tương tự, mình chỉ quan tâm vị trí 1 và 3. Nghĩa là hệ số của $|0x0\rangle$ và $|0x1\rangle$ giữ nguyên, còn hệ số trước $|1x0\rangle$ và $|1x1\rangle$ đổi chỗ cho nhau, với $x \in \{0, 1\}$. Nói cách khác, 8 hệ số trước amplitude chỉ thay đổi vị trí chứ không nhiều hơn hay ít đi, hay tập hợp hệ số giữ nguyên.



Như vậy, giả sử $|z_2\rangle = a|0\rangle + b|1\rangle$, $|z_1\rangle = c|0\rangle + |1\rangle$, $|z_4\rangle = e|0\rangle + f|1\rangle$, $|z_3\rangle = g|0\rangle + h|1\rangle$. Khi đó kết quả cipherstate là

$$|\psi\rangle = |z_2\rangle \otimes |z_2\rangle \otimes |z_1\rangle \otimes |z_1\rangle \otimes |z_4\rangle \otimes |z_4\rangle \otimes |z_3\rangle \otimes |z_3\rangle$$

Xét $|z_2\rangle \otimes |z_2\rangle = a^2|00\rangle + ab|01\rangle + ab|10\rangle + b^2|11\rangle$. Ở đây có 3 hệ số khác nhau là (a^2, ab, b^2). Với lưu ý bên trên $a \geqslant 0$ nên từ a^2 tính được a. Từ a, ta cần thêm ab để xác định b.

Như vậy mình cần $2^4 = 16$ hệ số để tìm lại các key ban đầu.

Bình luận

Thế éo nào mình lại nhầm khúc cuối mà lấy cả a^2, ab và b^2 nên kết quả ra $3^4 = 81$. Tất nhiên là **SAI BÉT** nên chỉ được 2/8 😭.

Problem 11. AntCipher

Bài này là bài số 2 ở round 1 và là bài số 11 ở round 2. Lúc thi round 1 mình không biết giải, còn ở round 2 thì mình đã giải theo cách như sau.

Đề bài

Đặt

$$f = (x_1 \vee x_2 \vee x_9) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_9) \wedge (\neg x_1 \vee x_2 \neg x_9) \wedge (x_1 \vee \neg x_2 \vee x_9) \wedge$$
$$(x_1 \vee x_2 \vee x_3) \wedge (\neg x_9 \vee \neg x_{10} \vee \neg x_3) \wedge (x_1 \vee \neg x_2 \vee x_4) \wedge (\neg x_9 \vee x_{10} \vee \neg x_4) \wedge$$
$$(\neg x_1 \vee x_2 \vee x_5) \wedge (x_9 \vee \neg x_{10} \vee \neg x_5) \wedge (\neg x_1 \vee \neg x_2 \vee x_6) \wedge (x_9 \vee x_{10} \vee \neg x_6) \wedge$$
$$(x_1 \vee x_2 \vee x_3 \vee x_4 \vee \neg x_7) \wedge (x_2 \vee x_3 \vee x_4 \vee \neg x_7 \vee \neg x_8)$$

Hàm f gồm 10 biến được viết dưới dạng CNF (conjunctive normal form). Thuật toán mã hóa dựa trên hàm f biến đổi hai bit plaintext (x_1, x_2) thành hai bit ciphertext (x_9, x_{10}) khi giá trị hàm $f = True$. Hàm f này có 10 biến $x_1, x_2, \ldots, x_{10}$ và 46 literals, là các hạng tử trong biểu diễn CNF của hàm. Ví dụ với dấu ngoặc thứ hai có 3 literals là $\neg x_1$, $\neg x_2$ và $\neg x_9$.

Q. Vì các giới hạn tính toán nên chúng ta chỉ có thể sử dụng tối đa 16 biến với 20 literals. Nhắc lại rằng hàm f ở trên có 10 biến và 46 literals. Hãy tìm cách biểu diễn tương đương của thuật toán mã hóa trên với giới hạn đã cho.

Giải

Khi mình code hàm để tính giá trị hàm f và xem xét những vector

$$\boldsymbol{x} = (x_1, \ldots, x_{10})$$

mà $f = True$, mình nhận thấy rằng:

- nếu $(x_1, x_2) = (0, 0)$ thì $(x_9, x_{10}) = (1, 0)$
- nếu $(x_1, x_2) = (0, 1)$ thì $(x_9, x_{10}) = (1, 1)$
- nếu $(x_1, x_2) = (1, 0)$ thì $(x_9, x_{10}) = (0, 0)$
- nếu $(x_1, x_2) = (1, 1)$ thì $(x_9, x_{10}) = (0, 1)$

Mình nhận ra rằng các biến $x_3, x_4, \ldots, x_7, x_8$ hoàn toàn không tác động lên việc mã hóa từ (x_1, x_2) thành (x_9, x_{10}) (ít nhất là ở những chỗ $f = True$:v).

Như vậy bài toán được rút gọn thành hàm boolean 4 biến x_1, x_2, x_9 và x_{10}. Ở đó $f(0010) = f(0111) = f(1000) = f(1101) = 1$. Các vector còn lại thì $f = 0$. Ở dưới là bảng chân trị A.1.

x_1	x_2	x_9	x_{10}	f
0	0	0	0	0
0	0	0	1	0
0	0	1	0	1
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	0
1	1	0	0	0
1	1	0	1	1
1	1	1	0	0
1	1	1	1	0

Bảng A.1: Bảng chân trị hàm f

Từ bảng chân trị trên, sử dụng phương pháp bìa Karnaugh mình rút gọn được thành

$$\begin{aligned} f(x_1, x_2, x_9, x_{10}) =& (\neg x_1 \vee \neg x_9) \wedge (x_1 \vee x_9) \wedge \\ & (\neg x_1 \vee \neg x_2 \vee x_{10}) \wedge (x_1 \vee x_2 \vee \neg x_{10}) \wedge \\ & (\neg x_1 \vee x_2 \vee \neg x_{10}) \wedge (x_1 \vee \neg x_2 \vee x_{10}) \end{aligned}$$

CNF này có 4 biến và 16 literals, thỏa mãn yêu cầu đề bài và ăn trọn 6/6 điểm 😇.

Phụ lục B

Ôn thi

B.1 Ôn thi ngày 20/11/2023

Toán tử tuyến tính

Toán tử tuyến tính là một ánh xạ

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

Nếu A là một ma trận cỡ $m \times n$ thì đây là một ánh xạ tuyến tính với phép nhân ma trận với vector $A \cdot \boldsymbol{x} = \boldsymbol{y}$.

Ở đây $\boldsymbol{x} \in \mathbb{R}^n$ và $\boldsymbol{y} \in \mathbb{R}^m$.

Định nghĩa 1. Hạt nhân

Hạt nhân của ánh xạ tuyến tính A là tập hợp nghiệm của hệ thuần nhất và được ký hiệu là $\ker(A)$. Nói cách khác

$$\ker(A) = \{\boldsymbol{x} \in \mathbb{R}^n : \ A \cdot \boldsymbol{x} = \boldsymbol{0}\} \tag{B.1}$$

Định nghĩa 2. Ảnh

Ảnh của ánh xạ tuyến tính A là tập hợp tất cả giá trị có thể của phép nhân ma trận và được ký hiệu là $\mathrm{im}(A)$. Nói cách khác

$$\mathrm{im}(A) = \{A \cdot \boldsymbol{x} : \ \boldsymbol{x} \in \mathbb{R}^n\} \tag{B.2}$$

Tính chất đối với ánh xạ $A : \mathbb{R}^n \to \mathbb{R}^m$ là $\dim(\ker A) + \dim(\mathrm{im}\, A) = n$.

Trị riêng và vector riêng

Định nghĩa 3. Trị riêng, vector riêng

Xét hệ phương trình tuyến tính thuần nhất biểu diễn bởi phép nhân

$$A \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x}$$

Giá trị λ khiến phương trình có nghiệm không tầm thường được gọi là **trị riêng** (eigenvalue) của ánh xạ tuyến tính.

Vector $\boldsymbol{x}$ là cơ sở của không gian vector nghiệm khi đó được gọi là **vector riêng** (eigenvector) ứng với trị riêng λ.

Lưu ý rằng có thể có nhiều vector riêng tương ứng với một trị riêng.

Để tìm trị riêng ta giải phương trình đặc trưng $\det(A - \lambda I) = 0$ và tìm tất cả nghiệm thực λ của phương trình.

Sau đó ta thế từng λ vào hệ $A\boldsymbol{x} = \lambda\boldsymbol{x}$ và tìm cơ sở của không gian nghiệm. Các vector trong cơ sở là vector riêng tương ứng với λ đó.

Một số tính chất của trị riêng và vector riêng (giả sử rằng đối với ma trận A cỡ $n \times n$ thì phương trình đặc trưng có đầy đủ n nghiệm thực).

1. $\operatorname{tr} A = \lambda_1 + \lambda_2 + \ldots + \lambda_n$
2. $\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$

Tính chất liên quan đến rank và trace:

1. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$
2. $\operatorname{rank}(AB) \leqslant \min(\operatorname{rank}(A), \operatorname{rank}(B))$

Bài tập

Bài 1. Cho vector cột $\boldsymbol{v} \in \mathbb{R}^n$. Đặt $A = \boldsymbol{v} \cdot \boldsymbol{v}^T$. Tìm spa A.

Các cột của A có dạng $\boldsymbol{v} \cdot v_1$, $\boldsymbol{v} \cdot v_2$, ..., $\boldsymbol{v} \cdot v_n$. Như vậy các cột đều tỉ lệ với cột đầu nên $\operatorname{rank} A = 1$.

Suy ra $\dim \ker A = n - 1$ và do đó $\lambda = 0$ là nghiệm bậc $n - 1$ trong phương trình đặc trưng.

Như vậy phương trình đặc trưng còn một nghiệm $\lambda \neq 0$.

Do $(\boldsymbol{v} \cdot \boldsymbol{v}^T)\boldsymbol{x} = \lambda\boldsymbol{x} \Leftrightarrow \boldsymbol{v}(\boldsymbol{v}^T \cdot \boldsymbol{x}) = \lambda\boldsymbol{x}$.

Đặt $\boldsymbol{v}^T \cdot \boldsymbol{x} = \alpha$ thì $\alpha\boldsymbol{v} = \lambda\boldsymbol{x}$. Suy ra $\boldsymbol{x} = \boldsymbol{v}$ và do đó $\alpha = \lambda = \|\boldsymbol{v}\|^2$.

Vậy spa $A = \{\|\boldsymbol{v}\|^2, 0, 0, \ldots, 0\}$.

Bài 3. Cho ma trận $A_{3\times3}$. Biết rằng $\operatorname{tr} A = \operatorname{tr} A^{-1} = 0$ và $\det A = 1$. Chứng minh rằng $A^3 = I$.

Phương trình đặc trưng có dạng $P_3(\lambda) = -\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$.

Theo tính chất trên thì $a_2 = \sum \lambda = \operatorname{tr} A = 0$.

Do λ là trị riêng nên $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Do A khả nghịch nên $\dfrac{1}{\lambda}\boldsymbol{x} = A^{-1}\boldsymbol{x}$.

Nghĩa là $\dfrac{1}{\lambda}$ là trị riêng của ma trận A^{-1}. Suy ra $\dfrac{1}{\lambda_1} + \dfrac{1}{\lambda_2} + \dfrac{1}{\lambda_3} = \operatorname{tr} A^{-1} = 0$.

Từ đó suy ra $\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 = 0$.

Cuối cùng $\det A = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 = 1$.

Vậy phương trình đặc trưng là $P_3(\lambda) = -\lambda^3 + 1$. Theo định lý Cayley-Hamilton thì $P_3(A) = -A^3 + I = 0$, hay $A^3 = I$.

Bài 4. Cho ma trận $A_{n\times n}$, $A_{ij} \geqslant 0$. Giả sử ma trận có đủ n trị riêng thực. Chứng minh rằng $\lambda_1^k + \lambda_2^k + \ldots + \lambda_n^k \geqslant 0$ với mọi $k \in \mathbb{N}$.

Ta thấy rằng với $k = 1$ thì $\lambda_1 + \ldots + \lambda_n = \operatorname{tr}(A) \geqslant 0$.

Vì λ_i là thỏa phương trình $A\boldsymbol{x} = \lambda_i\boldsymbol{x}$ nên nhân hai vế cho A ta có $A \cdot A\boldsymbol{x} = A \cdot \lambda_i\boldsymbol{x}$. Tương đương với $A^2\boldsymbol{x} = \lambda_i(A\boldsymbol{x}) = \lambda_i^2\boldsymbol{x}$.

Nói cách khác, λ_i^2 là trị riêng của ma trận A^2. Thực hiện tương tự ta có λ_i^k là trị riêng của ma trận A^k.

Do đó $\lambda_1^k + \ldots + \lambda_n^k = \operatorname{tr}(A^k) \geqslant 0$.

Bài 5. Cho ma trận A khả nghịch. X là ma trận sao cho $AX + XA = 0$. Chứng minh rằng $\operatorname{tr} X = 0$.

Nhân bên trái hai vế cho A^{-1} ta có $X + A^{-1}XA = 0$. Ta biết rằng $A^{-1}XA$ là ma trận tương đương ma trận X nên $\operatorname{tr}(A^{-1}XA) = \operatorname{tr} X$.

Suy ra $\operatorname{tr} X + \operatorname{tr} X = \operatorname{tr} 0 = 0$. Từ đây có $\operatorname{tr} X = 0$.

Phụ lục C

RUDN Olympiad 2023

Lần đầu tiên mình được tham dự thi toán đồng đội theo hình thức MathBoy (trận chiến toán).

Trong cách thi này, mỗi đội có 3 vị trí: người thuyết trình (докладчик), người phản biện (оппонент) và người giám sát (наблюдатель).

Ở mỗi vòng sẽ có 3 đội thi với nhau. Mỗi đội sẽ có 1 vị trí tương ứng với 3 vị trí trên. Sau đây là ví dụ

	Đội 1	Đội 2	Đội 3
Vòng 1	О	Д	Н
Vòng 2	Н	О	Д
Vòng 3	Д	Н	О

Ở mỗi vòng, đội đóng vai trò người thuyết trình lên bảng ghi bài giải trong thời gian cho phép và thuyết trình về bài giải của đội mình. Đội phản biện có nhiệm vụ phản biện bài thuyết trình đó. Đội giám sát, dựa trên bài thuyết trình cũng như phản biện mà ghi chép lại các lỗi, chỗ khó hiểu, ... và trình lên cho giám khảo.

Ngoài ra, đội thuyết trình trước đó phải trình bài giải viết tay cho giám khảo chấm trước khi lên thuyết trình.

Ở đây có rất nhiều câu chuyện hack não đã xảy ra. Lúc mình thi vòng 1, câu hỏi quá khó nên đội thuyết trình chỉ viết được một ít. Đồng nghĩa việc đội phản biện cũng như đội giám sát ... thất nghiệp, không có gì để nói.

Đối với vòng 2, trận chiến cân bằng hơn, đội mình làm việc giám sát. Dựa trên bài giải của đội thuyết trình, chúng mình thấy những trường hợp chưa được xét tới và có thể bị sai, do đó cả ba đội đều có điểm (đội thuyết trình có nhiều điểm nhất vì các bạn giải hơn 1 nửa rồi).

Đối với vòng 3, đội mình thuyết trình. Đội mình clear bài đó nên giành điểm tuyệt đối cho phần thuyết trình. Tuy nhiên các bạn phản biện cũng không vừa, vẫn cố gắng bắt một số lỗi do trình bày quá cô đọng. Kết quả là đội mình

(thuyết trình) full điểm cho vòng 3, đội phản biện được 3 điểm.

Phụ lục D

Đường đoản thời

Lời nói đầu

Động lực để tác giả viết bài này là sau khi đọc về sự ra đời phép tính vi tích phân cùng vụ tranh cãi đáng xấu hổ trong lịch sử toán học giữa Newton và Leibniz, cùng với bài toán của Johann Bernoulli.

Bài nghiên cứu này được tham khảo nhiều nguồn (từ Miguel A. Lerma[1] và Lê Quang Ánh[2]) và là tài liệu học tập cá nhân. Tác giả hy vọng rằng bài nghiên cứu nhỏ này sẽ giúp ích được cho các bạn học sinh, sinh viên đam mê toán và vật lý (mặc dù tác giả không phải dân lý hihi).

Bối cảnh lịch sử

Thế kỷ 17 đã chứng kiến một drama có thể gọi là đáng xấu hổ nhất lịch sử toán học. Hai nhà toán học có ảnh hưởng rất lớn lại vướng vào một vụ kiện tụng và tranh cãi khó coi để xem ai là người phát minh ra phép tính vi tích phân. Vâng, chúng ta đang nói đến Newton và Leibniz. Vào thời điểm đó có một nhà toán học xuất sắc thuộc một dòng họ cũng gồm rất nhiều nhân vật xuất sắc đã đưa ra một bài toán đố cho các nhà toán học trên thế giới. Bài toán đó đã chứng minh được ưu thế vượt trội trong phương pháp vi tích phân của Leibniz.

Nhà toán học xuất sắc đó là Johann Bernoulli, thuộc dòng họ Bernoulli nổi tiếng. Bài toán đó được phát biểu như sau:

Cho hai điểm A và B nằm trong mặt phẳng thẳng đứng P (A cao hơn B). Hãy xác định đường nối hai điểm A và B và nằm trong mặt phẳng P sao cho một điểm chỉ chịu trọng lực chạy từ A đến B trong thời gian ngắn nhất.

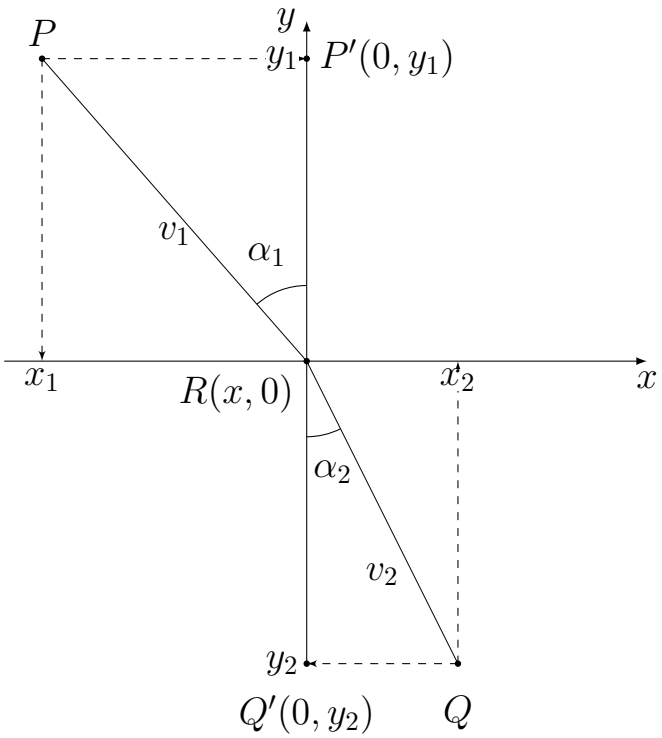
[1] *A simple derivation of the equation for the brachistochrone curve*, https://sites.math.northwestern.edu/~mlerma/papers-and-preprints/brachistochrone.pdf

[2] *Gia đình Bernoulli: một dòng họ Toán học*, trang 7, https://rosetta.vn/lequanganh/gia-dinh-bernoulli-mot-dong-ho-toan-hoc/

Chúng ta đã biết rằng đường đi ngắn nhất giữa hai điểm là đoạn thẳng nối hai điểm đó. Tuy nhiên trong bài toán của Johann Bernoulli thì đại lượng ngắn nhất cần tìm không phải khoảng cách giữa hai điểm mà là thời gian di chuyển giữa hai điểm. Mục tiêu cần làm ở bài toán này là xác định đường đi (hay quỹ đạo) thời gian ngắn nhất đó. Do đó bài toán này được gọi là bài toán *đường đoản thời* (brachistochrone curve).

Để giải bài toán này chúng ta cần một định luật cũng về thời gian ngắn nhất. Đó là nguyên lý thời gian ngắn nhất của Fermat và một hệ quả của nó là định luật Snell-Descartes.

Định luật Snell-Descartes



Hình D.1: Định luật Snell-Descartes

Nguyên lý thời gian ngắn nhất của Fermat phát biểu rằng

Khi ánh sáng truyền từ môi trường này sang môi trường khác thì nó luôn truyền đi theo đường nhanh nhất.

Hệ quả của nguyên lý của Fermat là định luật Snell-Descartes mà chúng ta thường thấy ở chương trình vật lý ở phổ thông dưới dạng định luật khúc xạ ánh sáng

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{v_1}{v_2} \tag{D.1}$$

với α_1 và α_2 lần lượt là góc hợp bởi tia vào và tia ra với pháp tuyến tại điểm tới, v_1 và v_2 là vận tốc truyền trong môi trường ở nửa trên và nửa dưới Ox (xem

hình D.1).

Để chứng minh định luật trên, ta thấy rằng v_1 là vận tốc khi di chuyển từ điểm P tới điểm R nên thời gian t_1 đi từ điểm P tới R là

$$t_1 = \frac{\|\overrightarrow{PR}\|}{v_1} = \frac{\sqrt{(x-x_1)^2 + y_1^2}}{v_1} \tag{D.2}$$

Lưu ý rằng tia sáng không truyền tới gốc tọa độ $O(0,0)$ mà truyền tới một điểm $R(x,0)$ là vì điểm bắt đầu là $P(x_1,y_1)$ và ánh sáng truyền đi theo đường nhanh nhất (theo nguyên lý Fermat) nên không có gì đảm bảo rằng nó sẽ truyền tới $O(0,0)$.

Tương tự, thời gian t_2 đi từ điểm R tới Q là

$$t_2 = \frac{\|\overrightarrow{RQ}\|}{v_2} = \frac{\sqrt{(x-x_2)^2 + y_2^2}}{v_2} \tag{D.3}$$

Kết hợp hai phương trình của t_1 và t_2 lại thì tổng thời gian di chuyển từ P tới Q biểu diễn theo x là

$$T(x) = t_1 + t_2 = \frac{\sqrt{(x-x_1)^2 + y_1^2}}{v_1} + \frac{\sqrt{(x-x_2)^2 + y_2^2}}{v_2} \tag{D.4}$$

Đạo hàm theo x ta có

$$T'(x) = \frac{x-x_1}{v_1\sqrt{(x-x_1)^2 + y_1^2}} + \frac{x-x_2}{v_2\sqrt{(x-x_2)^2 + y_2^2}} \tag{D.5}$$

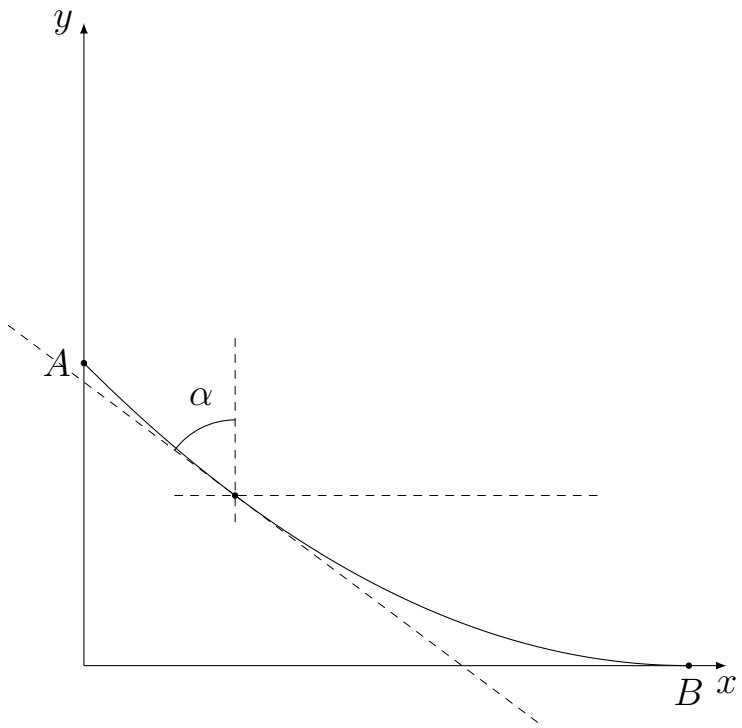
Để ý rằng $x > x_1$ nên $x - x_1 = \|\overrightarrow{PP'}\|$. Tương tự $x_2 - x = \|\overrightarrow{QQ'}\|$. Để tìm cực trị ta cho đạo hàm bằng 0 rồi tính đạo hàm cấp 2. Ta có

$$T'(x) = 0 \Leftrightarrow \frac{\|\overrightarrow{PP'}\|}{v_1\|\overrightarrow{PR}\|} - \frac{\|\overrightarrow{QQ'}\|}{v_2\|\overrightarrow{RQ}\|} = 0 \Leftrightarrow \frac{\sin\alpha_1}{v_1} - \frac{\sin\alpha_2}{v_2} = 0$$

Như vậy $\dfrac{\sin\alpha_1}{v_1} = \dfrac{\sin\alpha_2}{v_2}$. Đạo hàm cấp 2 tương ứng là

$$T''(x) = \frac{y_1^2}{v_1((x-x_1)^2 + y_1^2)} + \frac{y_2^2}{v_2((x-x_2)^2 + y_2^2)} > 0$$

Do đó x thỏa $T'(x) = 0$ ở trên là cực tiểu và định luật Snell-Descartes được chứng minh.



Hình D.2: Đường cycloid

Đường cong Cycloid

Đáp án cho bài toán mà Johann Bernoulli đặt ra là đường cong Cycloid. Sau đây sẽ trình bày cách giải bài toán của Johann Bernoulli.

Phương của vận tốc tức thời tại một điểm khi một vật đi theo một quỹ đạo đường cong là tiếp tuyến với đường cong tại điểm đó. Khi đó góc α trong định luật Snell-Descartes sẽ có liên hệ với hệ số góc của tiếp tuyến với đường cong. Nói rõ hơn, góc hợp bởi tiếp tuyến và trục Ox là $\alpha + \dfrac{\pi}{2}$ và hệ số góc của tiếp tuyến là $\tan\left(\alpha + \dfrac{\pi}{2}\right) = \dfrac{dy}{dx}$ (hình D.2).

Ta có $\tan\left(\alpha + \dfrac{\pi}{2}\right) = -\cot\alpha$ và $1 + \cot^2\alpha = \dfrac{1}{\sin^2\alpha}$ nên

$$\frac{1}{\sin^2\alpha} = 1 + \cot^2\alpha = 1 + \left(\frac{dy}{dx}\right)^2 \tag{D.6}$$

Giả sử tọa độ của A là (x_0, y_0). Khi một điểm di chuyển từ A tới B, gọi (x, y) là tọa độ của điểm đó trên đường cong. Theo định luật bảo toàn cơ năng thì

$$mgy_0 = \frac{1}{2}mv^2 + mgy$$

với v là vận tốc tức thời tại điểm (x, y) và mgy là thế năng tại điểm đó. Như vậy ta có

$$v^2 = 2g(-y + y_0) \tag{D.7}$$

Theo định luật Snell-Descartes thì $\dfrac{v}{\sin\alpha}$ là một hằng số khi nằm trong cùng môi trường. Do đó tồn tại số r cố định sao cho $\dfrac{v^2}{\sin^2\alpha} = 4gr$. Từ hai biểu thức của v^2 và $\dfrac{1}{\sin^2\alpha}$ ở trên ta có

$$\frac{v^2}{\sin^2\alpha} = 2g(-y+y_0)\left(1+\left(\frac{dy}{dx}\right)^2\right) = 4gr \tag{D.8}$$

Suy ra

$$\left(\frac{dy}{dx}\right)^2 = \frac{2r}{y_0-y} - 1 \tag{D.9}$$

Tới đây ta thấy rằng bậc của dy và dx ở vế trái là giống nhau, trong khi vế phải chỉ có y mà không có x. Do đó "bắt chước" cách đổi biến của đường tròn, đặt

$$\begin{cases} x = a\theta + b\cos\theta \\ y = c + d\sin\theta \end{cases}$$

với a, b, c, d là các số thực cần tìm, θ là góc hợp bởi Oy và đoạn thẳng nối tâm O và điểm trên đường cong (theo góc α). Lưu ý rằng khi thay $\theta = 0$ và $\theta = \pi/2$ vào hai phương trình trên ta phải thu được hai điểm trên hai trục tọa độ.

Lấy vi phân hai phương trình trên ta có

$$\begin{cases} dx = a - b\sin\theta\, d\theta \\ dy = d\cos\theta\, d\theta \end{cases}$$

Thay vào phương trình D.9 ta được

$$\frac{d^2\cos^2\theta}{(a-b\sin\theta)^2} = \frac{2r}{y_0-c-d\sin\theta} - 1 = \frac{2r-y_0+c+d\sin\theta}{y_0-c-d\sin\theta}$$

Do $\cos^2\theta = 1-\sin^2\theta = (1-\sin\theta)(1+\sin\theta)$ nên ta muốn chọn a và b có thể rút gọn được cho tử số.

Trường hợp 1. $a = b$, ta thu được

$$\frac{d^2(1+\sin\theta)}{a^2(1-\sin\theta)} = \frac{(2r-y_0+c)+d\sin\theta}{(y_0-c)-d\sin\theta}$$

Ta sẽ muốn đồng nhất hệ số tự do và hệ số trước $\sin\theta$ để dễ tính toán sau này. Do đó một cách chọn đơn giản là $2r-y_0+c = d$ và $y_0-c = d$. Suy ra $r = d$. Thu gọn phương trình ta được

$$\frac{d^2(1+\sin\theta)}{a^2(1-\sin\theta)} = \frac{1+\sin\theta}{1-\sin\theta}$$

Như vậy $a^2 = d^2$ nên $a = d$ hoặc $a = -d$. Ta xét trường hợp $a = d$, trường hợp $a = -d$ cũng cho kết quả tương tự (không thỏa mãn).

Ta có $a = b = d = r$ và $c = y_0 - d = y_0 - r$. Phương trình đường cong trong tọa độ cực sẽ là

$$\begin{cases} x = r(\theta + \cos\theta) \\ y = (y_0 - r) + r\sin\theta \end{cases}$$

Với $\theta = 0$ thì $(x, y) = (r, y_0 - r$. Với $\theta = \pi/2$ thì $(x, y) = (\pi r/2, y_0)$.

Tới đây chúng ta có thể thêm bớt một hằng số để "kéo" các tọa độ về trục.

Ta đưa tọa độ khi $\theta = 0$ về Oy thì $x' = x - r$. Tương tự tọa độ khi $\theta = \pi/2$ sẽ về Ox nên $y' = y - y_0$. Như vậy tọa độ (mới) cho hai trường hợp θ là $(0, -r)$ và $(\pi r/2 - 1, 0)$ nhưng vì r là số dương (bán kính) nên $(0, -r)$ nằm dưới trục Ox, không phù hợp với hình vẽ.

Trường hợp 2. $a = -b$, ta thu được

$$\frac{d^2(1 - \sin\theta)}{a^2(1 + \sin\theta)} = \frac{(2r - y_0 + c) + d\sin\theta}{(y_0 - c) - d\sin\theta}$$

Tương tự, để đồng nhất và rút gọn hệ số cho hợp với bên vế trái ta chọn $2r - y_0 + c = -d$ và $y_0 - c = -d$. Suy ra $d = -r$. Thu gọn phương trình ta được

$$\frac{d^2(1 - \sin\theta)}{a^2(1 + \sin\theta)} = \frac{1 - \sin\theta}{1 + \sin\theta}$$

Như vậy $a^2 = d^2$ nên $a = d$ hoặc $a = -d$ Ta xét trường hợp $a = -d$.

Khi đó $a = -b = -d = r$ và $c = y_0 + d = y_0 - r$. Phương trình đường cong trong tọa độ cực sẽ là

$$\begin{cases} x = r(\theta - \cos\theta) \\ y = (y_0 - r) - r\sin\theta \end{cases}$$

Với $\theta = 0$ thì $(x, y) = (-r, y_0)$. Với $\theta = \pi/2$ thì $(x, y) = (r(\pi/2 - 1), y_0 - 2r)$.

Tới đây ta cũng thêm bớt một hằng số vào hoành độ và tung độ để "kéo" các tọa độ về trục.

Ta đưa tọa độ khi $\theta = 0$ về Oy thì $x' = x + r$. Tương tự ta đưa tọa độ khi $\theta = \pi/2$ về Ox thì $y' = y - y_0 + 2r$. Khi đó tọa độ (mới) là $(0, 2r)$ và $(\pi r/2, 0)$. Điều này phù hợp với yêu cầu bài toán và tương đương với phương trình trong tọa độ cực

$$\begin{cases} x = r(\theta - \cos\theta) + r = r(1 + \theta - \cos\theta) \\ y = (y_0 - r) - r\sin\theta - (y_0 - 2r) = r(1 - \sin\theta) \end{cases} , 0 \le \theta \le \frac{\pi}{2} \qquad \text{(D.10)}$$

Đây chính là kết quả cần tìm. Thêm nữa vị trí ban đầu của vật là $(0, y_0)$ và tọa độ theo phương trình là $(0, 2r)$ nên suy ra $y_0 = 2r$.

Phương trình phụ thuộc thời gian

Trong phương trình đường cong có sự tham gia của bán kính r cố định và góc quét θ. Chúng ta cần mối liên hệ giữa các phương trình theo thời gian.

Nhắc lại, vận tốc tức thời tại một điểm có phương trùng với tiếp tuyến với đường cong tại điểm đó. Do đó $v = \dfrac{\sqrt{(dy)^2 + (dx)^2}}{dt}$ xác định vận tốc tức thời với quãng đường là $(dy)^2 + (dx)^2$ là bình phương khoảng cách trong mặt phẳng. Từ đây suy ra

$$v^2 = \Big(\frac{dy}{dt}\Big)^2 + \Big(\frac{dx}{dt}\Big)^2 = r^2\cos^2\theta\Big(\frac{d\theta}{dt}\Big)^2 + r^2(1+\sin\theta)^2\Big(\frac{d\theta}{dt}\Big)^2$$
$$= 2r^2(1+\sin\theta)\Big(\frac{d\theta}{dt}\Big)^2$$

Từ bên trên và $y_0 = 2r$ ta có

$$v^2 = 2g(y_0 - y) = 2g(y_0 - r + r\sin\theta) = 2gr(1+\sin\theta)$$

Suy ra

$$2r^2(1+\sin\theta)\Big(\frac{d\theta}{dt}\Big)^2 = 2gr(1+\sin\theta)$$

Hay

$$\Big(\frac{d\theta}{dt}\Big)^2 = \frac{g}{r} \Rightarrow \frac{d\theta}{dt} = \sqrt{\frac{g}{r}} = \text{const} \tag{D.11}$$

Như vậy $\theta = \sqrt{\dfrac{g}{r}}t = \omega t$. Ở đây t là thời gian tính từ lúc bắt đầu thả vật từ điểm A. Cuối cùng phương trình phụ thuộc thời gian của đường cong Cycloid là

$$\begin{cases} x = r(1 + \omega t - \cos\omega t) \\ y = r(1 - \sin\omega t) \end{cases} \tag{D.12}$$

Trong đó, r là bán kính cố định (bằng nửa độ cao ban đầu y_0 của vật), $\omega = \frac{g}{r}$ là tần số góc, y_0 là độ cao ban đầu của vật (tung độ điểm A).

Tài liệu tham khảo

[1] Euclid. *Euclid's Elements of Geometry*. Revised and corrected. Richard Fitzpatrick. ISBN: 978-0-6151-7984-1.

[2] Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. *An Introduction to Mathematical Cryptography*. 2014. DOI: 10.1007/978-1-4939-1711-2.

[3] John Casey, Euclid. *The First Six Books of the Elements of Euclid*. 2007.