第 6 章：数论基础
Number Theory

赵俊舟

西安交通大学网安学院
junzhou.zhao@xjtu.edu.cn

2025 年 12 月 20 日

目录

目录

数论简介

- 数论主要研究整数集合 $\mathbb{Z}$ 的性质，尤其关注正整数集合 $\mathbb{Z}^+$

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} \qquad \mathbb{Z}^+ = \{1, 2, 3, \ldots\}$$

- 按数的可分性，一个正整数总能分为以下三类：
  - 单位元：1
  - 素数：$2, 3, 5, 7, 11, 13, 17, 19, \ldots$
  - 合数：$4, 6, 8, 9, 10, 12, 14, 15, \ldots$
- 一个正整数 $p > 1$ 是素数当且仅当它只有因子 1 和 p。
- 素数是数论的核心，因为所有整数 $n > 1$ 都可以进行素因子分解：

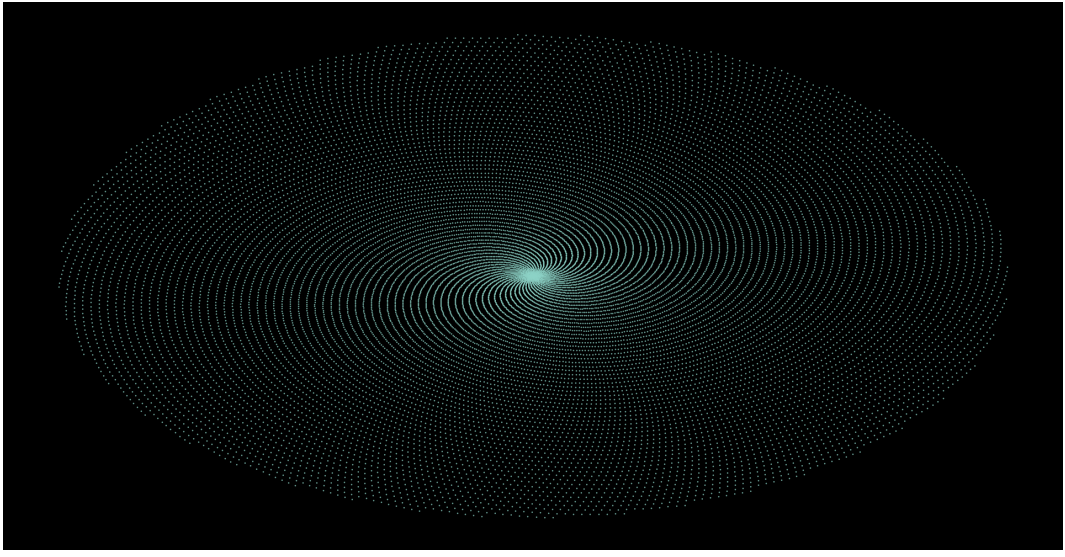
$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

其中 $p_1 < p_2 < \cdots < p_k$ 为素数，$\alpha_1, \alpha_2, \ldots, \alpha_k$ 为正整数。

关于素数的研究问题

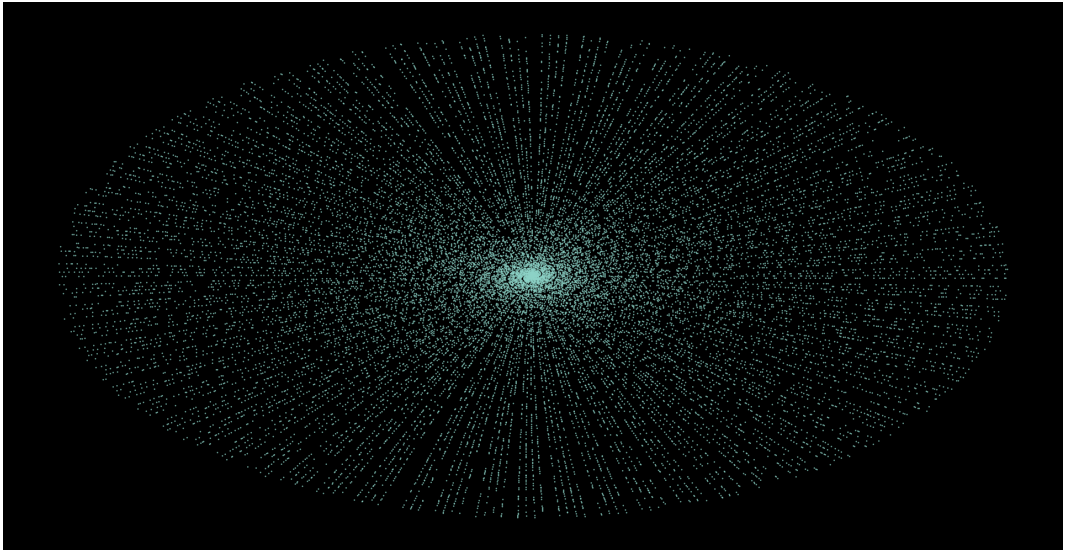
- 数论已经被研究了 2000 多年，关于素数仍有许多问题悬而未决。
- Q1: 素数的分布情况
  做出关键贡献的学者：欧几里得（300BC）、黎曼（1859）、阿达马和普桑（1896）等
- Q2: 孪生素数的分布情况
  做出关键贡献的学者：哥德巴赫（1742）、陈景润（1966）、张益唐（2013）等
- Q3: 等差素数列的分布情况
  做出关键贡献的学者：陶哲轩（2007）等
- 数论中的问题通常很容易表述，但这些问题往往很难解决。

Q1: 素数的分布（前 2 万个整数的分布）



$$x \to (x\cos(x), x\sin(x))$$

Q1: 素数的分布（前 2 万个素数的分布情况）



$$x \to (x\cos(x), x\sin(x))$$

Q1: 素数的分布（固定区间长度中的素数）

2	101	211	307	401	503	601	701	809	907	1009	1103	1201	1301	1409	1511	1601	1709	1801	1901
3	103	223	311	409	509	607	709	811	911	1013	1109	1213	1303	1423	1523	1607	1721	1811	1907
5	107	227	313	419	521	613	719	821	919	1019	1117	1217	1307	1427	1531	1609	1723	1823	1913
7	109	229	317	421	523	617	727	823	929	1021	1123	1223	1319	1429	1543	1613	1733	1831	1931
11	113	233	331	431	541	619	733	827	937	1031	1129	1229	1321	1433	1549	1619	1741	1847	1933
13	127	239	337	433	547	631	739	829	941	1033	1151	1231	1327	1439	1553	1621	1747	1861	1949
17	131	241	347	439	557	641	743	839	947	1039	1153	1237	1361	1447	1559	1627	1753	1867	1951
19	137	251	349	443	563	643	751	853	953	1049	1163	1249	1367	1451	1567	1637	1759	1871	1973
23	139	257	353	449	569	647	757	857	967	1051	1171	1259	1373	1453	1571	1657	1777	1873	1979
29	149	263	359	457	571	653	761	859	971	1061	1181	1277	1381	1459	1579	1663	1783	1877	1987
31	151	269	367	461	577	659	769	863	977	1063	1187	1279	1399	1471	1583	1667	1787	1879	1993
37	157	271	373	463	587	661	773	877	983	1069	1193	1283		1481	1597	1669	1789	1889	1997
41	163	277	379	467	593	673	787	881	991	1087		1289		1483		1693			1999
43	167	281	383	479	599	677	797	883	997	1091		1291		1487		1697			
47	173	283	389	487		683		887		1093		1297		1489		1699			
53	179	293	397	491		691				1097				1493					
59	181			499										1499					
61	191																		
67	193																		
71	197																		
73	199																		
79																			
83																			
89																			
97																			

Q1: 素数的分布

- 欧几里得在其著作《几何原本》中证明有无穷多个素数。
- 最小的素数是 2，目前发现的最大素数是 $2^{136,279,841}-1$（发现于 2024 年 10 月 21 日），共 $41,024,320$ 位数，比上一个发现的最大素数（2018 年）多 $1,600$ 位。
- 用 $\pi(x)$ 表示不超过 x 的素数个数，欧几里得定理其实说明

$$\pi(x) \to \infty \quad \text{as} \quad x \to \infty$$

- 更准确的结论是素数定理，由阿达马（1896）等人证明

$$\pi(x) \sim \frac{x}{\ln x}$$

- 可以近似认为，在区间 $[1, x]$ 碰到一个素数的概率为 $1/\ln x$；或者说，在 x 附近，每 $\ln x$ 个整数中有一个素数。

Q1: 素数的分布

- 如果黎曼猜想为真，那么素数定理可以进一步精确为

$$\pi(x) = \int_2^x \frac{\mathrm{d}t}{\ln t} + O(xe^{-c\sqrt{\ln x}})$$

- 黎曼猜想是复分析中的一个著名猜想：复平面上 ζ 函数

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s = \sigma + it, \sigma, t \in \mathbb{R}$$

  满足 $0 < Re(s) < 1$ 的所有非平凡零点都位于 $Re(s) = 1/2$ 上，即满足 $\zeta(\rho) = 0$ 的点 ρ 具有形式 $\rho = 1/2 + it$。

- 黎曼猜想是克雷数学研究所于 2000 年提出的 7 个千禧年大奖难题之一，每个难题奖金 100 万美元。

Q2: 孪生素数的分布

- 孪生素数指相差为 2 的素数对，例如 $(3,5),(5,7),(11,13)$ 等。目前发现的最大孪生素数为（发现于 2016 年）：

$$2,996,863,034,895 \times 2^{1,290,000} \pm 1$$

- 用 $\pi_2(x)$ 表示不超过 x 的孪生素数数量，孪生素数猜想说明

$$\pi_2(x) \to \infty \quad \text{as} \quad x \to \infty$$

- 如果素数分布服从独立同分布，那么

$$\pi_2(x) \sim \frac{x}{(\ln x)^2}$$

- 素数分布显然不独立，哈代和利特尔伍德猜想：

$$\pi_2(x) = 2\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2}\int_2^x \frac{\mathrm{d}t}{(\ln t)^2} \approx 1.320323632\int_2^x \frac{\mathrm{d}t}{(\ln t)^2}$$

Q2: 孪生素数的分布

- 陈景润利用筛法证明：有无穷多个整数对 $(p, p+2)$ 其中 p 为素数，$p+2$ 是不超过 2 个素数的乘积（1966–1973）。
- 张益唐证明：间距小于 7000 万的素数对有无穷多个（2013）。
- 仅仅过了几个月，素数对之差被缩小为 246。
- 差是 2 的素数对为孪生素数对，差是 4 的素数对为表亲素数对，差是 6 的素数对为性感素数对……。
- 波里尼亚克猜想（也称为弱孪生素数猜想，1849）：存在无穷多个素数对 $(p, p+2k), k = 1, 2, 3, \ldots$。

Q3: 等差素数列的分布

- 等差素数列是如下形式的素数数列

$$p, p+d, p+2d, \ldots, p+kd$$

其中 p 是首项，d 是公差，$p+kd$ 是尾项。例如 $(3,5,7)$，$(5,11,17,23,29)$。

- 目前发现的最长等差素数列为（记为 AP27，2009）

$$224,584,605,939,537,911 + 81292139 \cdot 23\# \cdot k \quad k = 0, \ldots, 26$$

其中 $23\#$ 是不超过 23 的素数的乘积。

- 格林-陶定理（2004）：存在任意长的等差素数列。
- 陶哲轩等人于 2006 年获菲尔兹奖，等同于数学诺贝尔奖。
- 目前仍不清楚怎样去发现任意长等差素数列，也不清楚等差连续素数列的存在情况。

关于素数的参考资料

- Great Internet Mersenne Prime Search:
  `https://www.mersenne.org`
- The largest known simultaneous primes:
  `http://primerecords.dk/simultprime.htm`
- 目前发现的最长等差素数列:
  `http://primerecords.dk/aprecords.htm`
- 目前发现的最长等差连续素数列:
  `http://primerecords.dk/cpap.htm`

目录

因子和带余除法

定义 (因子)

如果 $a = mb$，其中 a, b, m 为整数，且 $b \neq 0$，则称 b 能整除 a，或 a 除以 b 余数为 0，记 $b|a$，称 b 是 a 的一个因子。

- 如果 $a|1$，则 $a = \pm 1$
- 如果 $a|b$，且 $b|a$，则 $a = \pm b$
- 任何 $b \neq 0$ 能整除 0
- 如果 $a|b$，且 $a|c$，则对任何整数 X 和 Y 有 $a|(Xb + Yc)$

定义 (带余除法)

给定任意正整数 a 和 b，用 a 除以 b，得到商 q 和余数 r，即

$$a = qb + r$$

其中 $0 \leq r < b$，记 $q = \lfloor a/b \rfloor$。

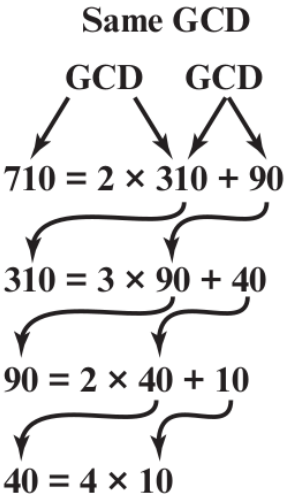
最大公约数和欧几里得算法

- 欧几里得算法是数论中的一个最基本技巧，可以求两个正整数的最大公约数。
- 对任意整数 a, b，且 $a \geq b > 0$，有

$$\gcd(a, b) = \gcd(b, a \bmod b)$$

即辗转相除。

```
Euclid(a, b){
  if (b=0) then
    return a
  else
    return Euclid(b, a mod b)
}
```



欧几里得算法的原理

- 假设要求整数 a 和 b 的最大公因子，不妨令 $a \geq b > 0$.
- b 除 a 可以表示为 $a = qb + r$，其中 $0 \leq r < b$ 为余数。
- 如果 $r = 0$，则 $\gcd(a, b) = b$。
- 如果 $r \neq 0$，考虑 $\gcd(a, b)$ 和 $\gcd(b, r)$ 之间的关系：
  - 令 $d = \gcd(a, b)$。因为 $d|a$ 且 $d|b$，所以 $d|(a - qb)$，即 $d|r$。也就是说，d 是 b, r 的公因子。那么 $d \leq \gcd(b, r)$。
  - 令 $c = \gcd(b, r)$。因为 $c|b$ 且 $c|r$，所以 $c|(qb + r)$，即 $c|a$。也就是说，c 是 a, b 的公因子。因为 a, b 的最大公因子是 d，所以 $c = \gcd(b, r) \leq d$。
- 所以 $\gcd(a, b) = \gcd(b, r)$，即求 a 和 b 的最大公因子可以转化为求 b 和 r 的最大公因子。

扩展欧几里得算法

- 给定两个整数 a 和 b，扩展欧几里得算法可以得到两个整数 x 和 y，满足 $ax + by = \gcd(a, b)$。
- 利用欧几里得算法，并且假设每步 i 都可得到 x_i 和 y_i 满足 $r_i = ax_i + by_i$。有以下关系式：

$$\begin{aligned}
a &= q_1 b + r_1 & r_1 &= ax_1 + by_1 \\
b &= q_2 r_1 + r_2 & r_2 &= ax_2 + by_2 \\
r_1 &= q_3 r_2 + r_3 & r_3 &= ax_3 + by_3 \\
&\vdots & &\vdots \\
r_{n-2} &= q_n r_{n-1} + r_n & r_n &= ax_n + by_n \\
r_{n-1} &= q_{n+1} r_n + 0
\end{aligned}$$

- 从而得到 $d = r_n = ax_n + by_n = ax + by$，即 $x = x_n, y = y_n$。

几个性质

性质

对于正整数 a, b，存在整数 X, Y，使 $Xa + Yb = \gcd(a, b)$，并且 $\gcd(a, b)$ 是能够表示成这种形式的最小正整数。

- 定义集合 $I \triangleq \{\hat{X}a + \hat{Y}b | \hat{X}, \hat{Y} \in \mathbb{Z}\}$。由于 $a, b \in I$，所以 $I \neq \emptyset$
- 令 d 为集合 I 中的最小正整数，如果 $d = \gcd(a, b)$，则 d 可以写成 $d = Xa + Yb \in I$。
- 要证 $d = \gcd(a, b)$，需证 $d|a, d|b$ 且 d 最大。其实 d 可以整除 I 中所有元素，任取元素 $c \in I$ 并且写成 $c = X'a + Y'b$。
- 用带余除法，得 $c = qd + r$，$0 \leq r < d$，故 $r = c - qd = X'a + Y'b - q(Xa + Yb) = (X' - qX)a + (Y' - qY)b \in I$
- 如果 $r \neq 0$，那么 $r < d$，这与 d 是 I 中最小正整数矛盾，故 $r = 0$。所以 d 能同时整除 a 和 b，是 a, b 的公约数。

几个性质

- 如果存在 $d' > d$，且 $d'|a, d'|b$，所以 $d'|(Xa + Yb)$。
- 由于 $d = Xa + Yb$，所以 $d'|d$。因为 $d' > d$，产生矛盾。

性质

如果 $c|ab$ 且 $\gcd(a, c) = 1$，则 $c|b$。当 p 是素数时，如果 $p|ab$，则 $p|a$ 或 $p|b$。

$c|ab \Rightarrow ab = \gamma c$。$\gcd(a, c) = 1 \Rightarrow Xa + Yc = 1$。两边同乘以 b

$$b = Xab + Ybc = X\gamma c + Yc = (X\gamma + Y)c$$

性质

如果 $a|c, b|c$ 且 $\gcd(a, b) = 1$，那么 $ab|c$。

由 $c = Aa, c = Bb, Xa + Yb = 1$，得

$$c = Xac + Ybc = XaBb + YbAa = ab(XB + YA)$$

模运算和同余

定义 (模运算)

如果 a 是整数，n 是正整数，定义 a 除以 n 所得余数为 a 模 n，记为 $a \bmod n$。对于任意整数 a，有

$$a = \lfloor a/n \rfloor \times n + (a \bmod n)$$

例如，$11 \bmod 7 = 4$，$-11 \bmod 7 = 3$.

定义 (同余)

如果 $a \bmod n = b \bmod n$，则称整数 a 和 b 是模 n 同余，表示为 $a \equiv b \pmod{n}$ 或 $a \equiv_n b$。

例如，$73 \equiv 4 \pmod{23}$，$21 \equiv -9 \pmod{10}$，$13 \equiv 8 \pmod{5}$

同余的性质

性质

- $n|(a-b) \Leftrightarrow a \equiv b \pmod{n}$
- **对称性**：$a \equiv b \pmod{n} \Leftrightarrow b \equiv a \pmod{n}$
- **传递性**：$a \equiv b \pmod{n}$ 且 $b \equiv c \pmod{n} \Rightarrow a \equiv c \pmod{n}$

证明.

- $(\Rightarrow)$ 如果 $n|(a-b)$，则有 $(a-b)=kn$，k 为某个整数，所以 $a=b+kn$。故 $a \bmod n=(b+kn) \bmod n=b \bmod n$。
- $(\Leftarrow)$ 如果 $a \equiv b \pmod{n}$，那么 $a=k_1n+r, b=k_2n+r$，进而 $n|(a-b)$。

模算术运算

性质 (模运算的分配率)

运算 $\circ \in \{+, \times\}$，则

$$(a \circ b) \bmod n = [(a \bmod n) \circ (b \bmod n)] \bmod n$$

性质 (模运算的加性和乘性)

如果 $a \equiv b \pmod{n}$ 且 $c \equiv d \pmod{n}$，则

$$(a \circ c) \equiv (b \circ d) \pmod{n}$$

- $n|(a-b) \wedge n|(c-d) \Rightarrow n|(a-b+c-d) \Rightarrow n|[(a+c)-(b+d)] \Rightarrow (a+c) \equiv (b+d) \pmod{n}$
- $n|(a-b) \wedge n|(c-d) \Rightarrow n|[c(a-b)+b(c-d)] \Rightarrow n|(ac-bd) \Rightarrow ac \equiv bd \pmod{n}$

模 n 乘法逆元

定义 (模 n 乘法逆元)

对于整数 b，如果存在整数 c 使 $bc \equiv 1 \pmod{n}$，则称 c 为 b 的模 n 乘法逆元，记为 $c = b^{-1} \bmod n$。

除法转化为乘法运算：$a/b \bmod n \triangleq ab^{-1} \bmod n$

×	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

模 7 乘法

×	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

模 8 乘法

模 n 乘法逆元存在的条件

定理

令 $b \geq 1, n \geq 1$ 为整数，则 b 存在模 n 乘法逆元的充要条件是 $\gcd(b, n) = 1$。

证明.

- 如果 b 存在模 n 乘法逆元，记 $c = b^{-1} \bmod n$。
- 从而 $bc \equiv 1 \pmod{n}$，则 $bc - 1 = \gamma n$，或 $bc - \gamma n = 1$。
- 因为 $\gcd(b, n)$ 是能写成上述形式的最小正整数，所以只能 $\gcd(b, n) = 1$。
- 如果 $\gcd(b, n) = 1$，则存在 X, Y 使 $Xb + Yn = 1$。
- 两边模 n，得到 $Xb \bmod n = 1$，即 $X = b^{-1} \bmod n$。

目录

群（Groups）

定义 (群，Groups)

定义了一个二元运算 $\cdot$ 的集合，记作 $\{\mathbb{G}, \cdot\}$，且满足下列公理：

(A1) **封闭性**：如果 a 和 b 都属于 $\mathbb{G}$，则 $a \cdot b$ 也属于 $\mathbb{G}$；

(A2) **结合律**：对于 $\mathbb{G}$ 中任意元素 a, b, c，都有 $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ 成立；

(A3) **单位元**：$\mathbb{G}$ 中存在一个元素 e，对于 $\mathbb{G}$ 中任意元素 a，都有 $a \cdot e = e \cdot a = a$ 成立；

(A4) **逆元**：对于 $\mathbb{G}$ 中任意元素 a，$\mathbb{G}$ 中都存在一个元素 a'，使得 $a \cdot a' = a' \cdot a = e$ 成立。

注：当群中的运算符是加法时，习惯上记它的单位元为 0，a 的逆元是 $-a$，并且减法用以下的规则定义：$a - b = a + (-b)$

群（Groups）

- 如果群的元素是有限个，则称为有限群；否则称为无限群。
- 有限群中元素的个数称为有限群的阶。

定义 (交换群，阿贝尔群，Abelian Groups)

还满足以下条件的群称为交换群：

(A5) 交换律 Commutative：对于 $\mathbb{G}$ 中任意的元素 a, b，都有 $a \cdot b = b \cdot a$ 成立。

例 (加法群 $\mathbb{Z}_N$)

定义集合 $\mathbb{Z}_N \triangleq \{0, \ldots, N-1\}$，$N > 1$，以及模 N 加法运算

$$a + b \triangleq (a + b) \bmod N$$

则 $(\mathbb{Z}_N, +)$ 构成群，同时也是交换群，记为群 $\mathbb{Z}_N$。

群中的指数运算

- 对于加法群，将群中的元素 g 累加 m 次，记为
$$mg = m \cdot g \triangleq \underbrace{g + \cdots + g}_{\text{累加 } m \text{ 次}}$$
- 习惯的运算规则依旧成立，例如 $mg + m'g = (m + m')g$，$m(m'g) = (mm')g$，$1 \cdot g = g$。
- 对于乘法群，将群中的元素 g 连乘 m 次，记为
$$g^m \triangleq \underbrace{g \cdots g}_{\text{连乘 } m \text{ 次}}$$
- 习惯的运算规则依旧成立，例如 $g^m \cdot g^{m'} = g^{m+m'}$，$(g^m)^{m'} = g^{mm'}$，$g^1 = g$，$g^{-m} \triangleq (g^{-1})^m = (g^m)^{-1}$。

群的几个性质

性质

令 $\mathbb{G}$ 为有限群，阶为 $m = |\mathbb{G}|$。对于群中任意元素 $g \in \mathbb{G}$，有 $g^m = 1$。

证明.

- 以交换群为例进行证明，该结论其实对有限群都成立。
- 对于任意 $g_i \neq g_j$，则 $gg_i \neq gg_j$，否则两边同乘 g^{-1} 得 $g_i = g_j$。
- 进而

$$g_1 g_2 \cdots g_m = (gg_1)(gg_2)\cdots(gg_m) = g^m(g_1 \cdot g_2 \cdots g_m)$$

- 所以 $g^m = 1$。

群的几个性质

推论

令 $\mathbb{G}$ 为有限群，阶为 $m = |\mathbb{G}| > 1$。对于群中任意元素 $g \in \mathbb{G}$ 和整数 x，有 $g^x = g^{x \bmod m}$。

说明进行指数运算时，可以先对指数进行模运算。

证明.

令 $x = qm + r$，则

$$g^x = g^{qm+r} = g^r = g^{x \bmod m}$$

群的几个性质

推论

令 $\mathbb{G}$ 为有限群，阶为 $m = |\mathbb{G}| > 1$。令 e 为正整数，定义函数 $f_e : \mathbb{G} \mapsto \mathbb{G}$ 为 $f_e(g) = g^e$。当 $\gcd(e, m) = 1$ 时，f_e 为群 $\mathbb{G}$ 上的一个置换映射（即双射）。令 $d = e^{-1} \bmod m$，则 f_d 为 f_e 的反函数。

证明.

如果一个函数存在反函数，说明这个函数是双射，因此只需证明 f_d 是 f_e 的反函数即可。对于任意 $g \in \mathbb{G}$，有

$$f_d(f_e(g)) = f_d(g^e) = g^{ed} = g^{ed \bmod m} = g$$

乘法群 $\mathbb{Z}_N^*$

- 定义在 $\mathbb{Z}_N = \{0, \ldots, N-1\}$ 上的模 N 加法构成加法群。
- 注意 $\mathbb{Z}_N$ 里的元素都存在加法逆元，但不一定都存在乘法逆元。
- 为了保证元素存在乘法逆元，定义集合

$$\mathbb{Z}_N^* \triangleq \{b \in \mathbb{Z}_N | \gcd(b, N) = 1\}$$

及乘法运算 $ab \triangleq ab \bmod N$。

性质

定义在集合 $\mathbb{Z}_N^*$ 上的模 N 乘法运算构成乘法群，同时也是交换群，记为群 $\mathbb{Z}_N^*$。

欧拉函数（Euler's Totient Function）

- 记群 $\mathbb{Z}_N^*$ 的阶为 $\phi(N) \triangleq |\mathbb{Z}_N^*|$，称为欧拉函数。
- 当 $N = p$ 为素数时，$\phi(p) = p - 1$。

性质

p, q 是素数且 $p \neq q$，则 $\phi(pq) = \phi(p)\phi(q) = (p-1)(q-1)$。

- 考虑集合 $\{1, \ldots, pq-1\}$，其中不与 pq 互素的数构成的集合为 $\{p, 2p, \ldots, (q-1)p\}$ 和 $\{q, 2q, \ldots, (p-1)q\}$。
- 这两个集合无交集：假设存在 $1 \leq i \leq q-1$ 和 $1 \leq j \leq p-1$，满足 $ip = jq$，两边模 p 得 $jq \bmod p = 0$，因为 p, q 为素数，故 $jq \bmod p \neq 0$，所以这两个集合不可能有交集。
- 两个集合共有 $p - 1 + q - 1$ 个整数，所以

$$\phi(n) = (pq-1) - (p-1+q-1) = (p-1)(q-1) = \phi(p)\phi(q)$$

欧拉函数（Euler's Totient Function）

性质

p 是素数，则 $\phi(p^k) = p^{k-1}(p-1)$。

性质

正整数 n 的素因子分解为 $n = p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}$，则

$$\phi(n) = \prod_{i=1}^{t} p_i^{k_i - 1}(p_i - 1)$$

- 对一个正整数 n 进行素因子分解是很困难的事情，因此目前尚不存在通用的计算 $\phi(n)$ 的高效算法。
- 数论中可以证明，计算一个正整数 n 的欧拉函数 $\phi(n)$ 等同于对 n 进行素因子分解。

欧拉定理和费马定理

推论 (欧拉定理和费马定理)

对于任意整数 $N > 1$ 及 $a \in \mathbb{Z}_N^*$，则

$$a^{\phi(N)} \equiv 1 \pmod{N}$$

特别的，当 $N = p$ 为素数时，对于 $a \in \mathbb{Z}_p$，有

$$a^{p-1} \equiv 1 \pmod{p}$$

推论

令 $N > 1$，e 为正整数，定义函数 $f_e : \mathbb{Z}_N^* \mapsto \mathbb{Z}_N^*$ 为 $f_e(x) = x^e \bmod N$。当 $\gcd(e, \phi(N)) = 1$ 时，f_e 为群 $\mathbb{Z}_N^*$ 上的一个置换。令 $d = e^{-1} \bmod \phi(N)$，则 f_d 为 f_e 的反函数。

目录

素性测试

- 密码学中常常需要寻找大素数。
- 传统的方法是用试除法，即依次除小于该数平方根的所有整数，这种方法只对较小的数有用。
- 可以采用基于素数特性的统计素性测试方法：
  - 其中所有的素数都满足素数特性。
  - 但是有一些被称为伪素数的合数也满足素数特性。
- 也可使用一种较慢的确定性素性测试方法 AKS。

奇整数的表示

$n \geq 3$ 的奇整数可表示为 $n - 1 = 2^k q$，其中 $k > 0$，q 是奇数。

素数的两个性质

性质 (性质一)

若 p 是素数，a 是小于 p 的正整数，则 $a^2 \bmod p = 1$ 当且仅当 $a \bmod p = 1$ 或 $a \bmod p = p - 1$。

证明.

- $\Rightarrow$: 由 $a^2 \bmod p = 1$ 知 $p|(a^2 - 1)$ 即 $p|(a+1)(a-1)$。由于 p 是素数，故只能是 $p|(a+1)$ 或 $p|(a-1)$，得 $a \bmod p = 1$ 或 $a \bmod p = -1 \bmod p = p - 1$。
- $\Leftarrow$: 当 $a \bmod p = 1$ 或 $a \bmod p = p - 1$ 时，有 $p|(a-1)$ 或 $p|(a+1)$，所以 $p|(a+1)(a-1)$，即 $p|(a^2 - 1)$，从而得到 $a^2 \bmod p = 1$。

某些合数也可能成立。$p = 4$，当 $a = 1$ 或 3 时，$a^2 \bmod p = 1$

素数的两个性质

性质 (性质二)

设 p 是大于 2 的素数，有 $p-1=2^k q$，$k>0$，q 是奇数。设 a 是小于 p 的整数，则以下两个结论必然有一个成立：

- $a^q \bmod p = 1$。
- 在整数 $a^q \bmod p, a^{2q} \bmod p, \ldots, a^{2^{k-1}q} \bmod p$ 中存在一个数为 $p-1$。即存在 $0 \le j \le k-1$，满足 $a^{2^j q} \bmod p = p-1$。

例

- $p=29$ 为素数，$29-1=2^2 \times 7$。
- 取 $a=2$，则 $a^q \bmod p = 2^7 \bmod 29 = 12$，$a^{2q} \bmod p = 2^{14} \bmod 29 = 28$，满足第二个结论，故该性质对素数 29 成立。

素数的两个性质

证明.

- 因为 p 是素数，则由费马定理可知 $a^{p-1} \equiv 1 \pmod{p}$。由于 $p-1 = 2^k q$，则 $a^{2^k q} \bmod p = 1$。
- 观察下述数列：

  $$a^q \bmod p, a^{2q} \bmod p, \ldots, a^{2^{k-1}q} \bmod p, a^{2^k q} \bmod p = 1$$

  这个数列最后一个数为 1，而且每个数为前一个数的平方。
- 最后一个数为 1，那么前一个数只能为 1 或 $p-1$。如果倒数第二个数为 1，则它前一个数只能为 1 或 $p-1$；依次类推。
- 所以，这个数列要么全是 1，即第一个数为 1；要么数列中某个数为 $p-1$，从这个数之后全为 1。

Miller-Rabin 素性测试

- 若 n 为素数，$n-1=2^k q$，$a \in \{1,\ldots,n-1\}$，那么数列
$$a^q \bmod n, a^{2q} \bmod n, \ldots, a^{2^{k-1}q} \bmod n$$
要么第一个数为 1，要么数列中某个数为 $n-1$。
- 如果不满足上一条，那么 n 必为合数。
- 注意，如果上述条件满足，也不一定推出 n 一定为素数。

例

- $n = 2047 = 23 \times 89$，则 $n-1 = 2 \times 1023$。
- 计算 $2^{1023} \bmod 2047 = 1$，所以虽然 $n = 2047$ 满足条件，但不是素数。

Miller-Rabin 素性测试

算法 1: PrimeTest(n)

输入: 奇整数 n

输出: n 是不是素数

找出整数 k, q，其中 $k > 0$，q 是奇数，使 $n - 1 = 2^k q$。

随机选取整数 $a \in \{1, \ldots, n-1\}$。

if $a^q \bmod n = 1$ **then**

    return 可能是素数。

for $j = 0$ to $k - 1$ **do**

    if $a^{2^j q} \bmod n = n - 1$ **then**

        return 可能是素数。

return 是合数。

Miller-Rabin 素性测试

- 如果返回 ‘‘合数’’，则这个数必为合数。否则可能为素数。
- 有结论：给定一个非素奇数 n 和一个随机整数 $a, 1 < a < n-1$，程序 PrimeTest 误报的概率小于 $1/4$（即当程序返回 “n 可能是素数” 时，误报的概率小于 $1/4$）。
- 因此，如果选择 t 个不同 a 进行测试，则它们都能通过测试并产生误报的概率小于 $(1/4)^t$。
- 对随机选取的 a，重复调用 PrimeTest(n)，如果某时刻 PrimeTest 返回 ‘‘合数’’，则 n 一定不是素数。
- 若 PrimeTest 连续 t 次返回 ‘‘可能是素数’’，当 t 足够大时，可以相信 n 是素数。

Miller-Rabin 素性测试举例

例 (考虑素数 $n = 29$)

- $n - 1 = 28 = 2^2 \times 7 = 2^k q$
- 选取 $a = 2$
  - $a^q \bmod n = 12$，$a^{2q} \bmod n = 28$，返回 “有可能是素数”
- 选取 $a = 10$
  - $a^q \bmod n = 17$，$a^{2q} \bmod n = 28$，返回 “有可能是素数”

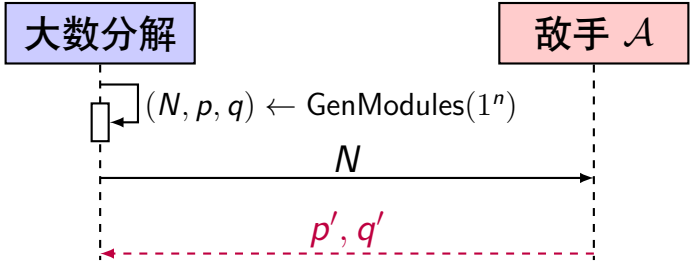
例 (考虑合数 $n = 13 \times 17 = 221$)

- $n - 1 = 220 = 2^2 \times 55 = 2^k q$
- 选取 $a = 5$
  - $a^q \bmod n = 112$
  - $a^{2q} \bmod n = 168$。返回 “合数”

目录

大数分解困难性假设

- 考虑一个 PPT 算法 GenModules，输入 1^n，输出 (N, p, q)，其中 $N = pq$，p, q 以可忽略概率 $\mathrm{negl}(n)$ 为两个 n 比特的素数。



- 若 $(p', q') = (p, q)$，则敌手成功，记 $\mathrm{Factor}_{\mathcal{A},\mathrm{GenModulus}}(n) = 1$
- 如果对于任意 PPT 敌手 $\mathcal{A}$，有

$$\Pr[\mathrm{Factor}_{\mathcal{A},\mathrm{GenModulus}}(n) = 1] \leq \mathrm{negl}(n)$$

称基于 GenModulus 的大数分解问题是困难的。

RSA 问题困难性假设

- 大数分解问题已经研究了几百年，未找到高效解法，但大数分解问题难以直接应用于密码设计。
- 1978 年，Rivest, Shamir, Adleman 提出一个与大数分解问题相关的问题，称为 RSA 问题。
- 考虑一个 PPT 算法 GenRSA，输入 1^n，输出 (N, e, d)。

算法 2: GenRSA(1^n)

输入: 安全参数 1^n

输出: N, e, d

1 $(N, p, q) \leftarrow \text{GenModulus}(1^n)$

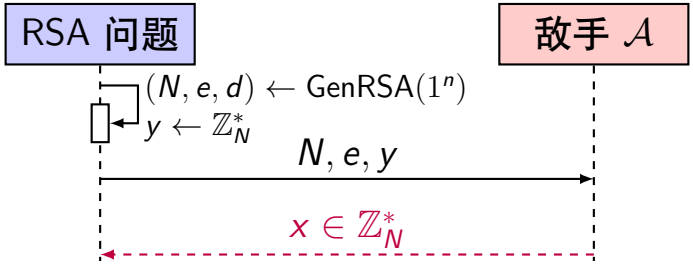
2 $\phi(N) = (p-1)(q-1)$

3 选择 $e > 1$ 且 $\gcd(e, \phi(N)) = 1$

4 计算 $d = e^{-1} \bmod \phi(N)$

5 **return** N, e, d

RSA 问题困难性假设



- 若 $x^e \equiv y \pmod{N}$，则敌手成功，记 $\mathsf{RSAinv}_{\mathcal{A},\mathsf{GenRSA}}(n) = 1$
- 如果对于任意 PPT 敌手 $\mathcal{A}$，有
$$\Pr[\mathsf{RSAinv}_{\mathcal{A},\mathsf{GenRSA}}(n) = 1] \leq \mathsf{negl}(n)$$
称基于 GenRSA 的 RSA 问题是困难的。
- 若 d、$\phi(N)$ 或 N 的分解未知，则 RSA 问题难以求解，这是 RSA 公钥算法安全性的基础。

目录

目录

循环群及其性质

- 令 $\mathbb{G}$ 为有限群，阶为 m，对于任意元素 g，考虑序列
$$\langle g \rangle \triangleq \{g^0, g^1, \ldots\}$$
- 由群的性质知 $g^m = 1$，令 $i \leq m$ 为最小正整数使 $g^i = 1$，则
$$\langle g \rangle \triangleq \{g^0, \ldots, g^{i-1}\}$$
- 称 $\langle g \rangle$ 为 $\mathbb{G}$ 的子群，g 为群 $\langle g \rangle$ 的生成元，元素 g 的阶为 i。

性质

令 $\mathbb{G}$ 为有限群，$g \in \mathbb{G}$ 为阶为 i 的元素，则
$$g^x = g^y \Leftrightarrow x \equiv y \pmod{i}$$

- 如果 $x \equiv y \pmod{i}$，则 $g^x = g^{x \bmod i} = g^{y \bmod i} = g^y$
- 如果 $g^x = g^y$，则 $1 = g^{x-y} = g^{(x-y) \bmod i}$。因为 $(x-y) \bmod i < i$ 且 i 是 g 的阶，则只能 $(x-y) \bmod i = 0$。

循环群及其性质

性质

令 $\mathbb{G}$ 为有限群，阶为 m，元素 g 的阶为 i，则 $i|m$。

$$g^m = 1 = g^0 \Rightarrow m \equiv 0 \pmod{i} \Rightarrow i|m$$

推论

令 $\mathbb{G}$ 为素数阶 p 的有限群，则 $\mathbb{G}$ 为循环群，并且除单位元外，其他元素都是生成元。

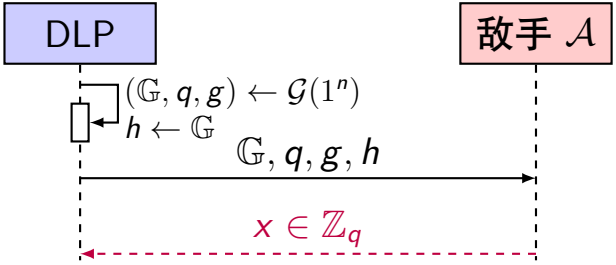
定理

令 p 为素数，则 $\mathbb{Z}_p^*$ 为循环群，并且阶为 $p-1$。

目录

离散对数问题

- 考虑阶为 q 生成元为 g 的循环群 $\mathbb{G} = \{g^0, \ldots, g^{q-1}\}$。
- 对于任意 $h \in \mathbb{G}$，存在 $x \in \mathbb{Z}_q$，使 $g^x = h$，称 x 为 h 以 g 为底的离散对数，记为 $x \triangleq \log_g h$。



当 $g^x = h$ 时，称敌手成功，记 $\mathsf{DLog}_{\mathcal{A},\mathcal{G}}(n) = 1$

- 如果对于任意 PPT 敌手 $\mathcal{A}$，有

$$\Pr[\mathsf{DLog}_{\mathcal{A},\mathcal{G}}(n) = 1] \leq \mathsf{negl}(n)$$

称基于 $\mathcal{G}$ 的离散对数问题是困难的。

Diffie-Hellman 问题

- Diffie-Hellman 问题与离散对数问题密切相关，但目前尚不能证明两者等价。
- 给定循环群 $\mathbb{G}$ 及其生成元 g。

Computational Diffie-Hellman (CDH) 问题

对于群中任意两个元素 $h_1, h_2 \in \mathbb{G}$，计算

$$DH_g(h_1, h_2) = g^{\log_g h_1 \cdot \log_g h_2}.$$

- 若离散对数问题可解，则 CDH 问题也可解。
- 首先计算 $x_1 = \log_g h_1$，然后得到 $DH_g(h_1, h_2) = h_2^{x_1}$。

Diffie-Hellman 问题

Decisional Diffie-Hellman (DDH) 问题

对于随机采样的整数 $x, y, z \in \mathbb{Z}_q$，如果对于任意 PPT 算法 $\mathcal{A}$，都有

$$\left|\Pr[\mathcal{A}(\mathbb{G}, q, g, g^x, g^y, g^z) = 1] - \Pr[\mathcal{A}(\mathbb{G}, q, g, g^x, g^y, g^{xy}) = 1]\right| \leq \mathsf{negl}(n)$$

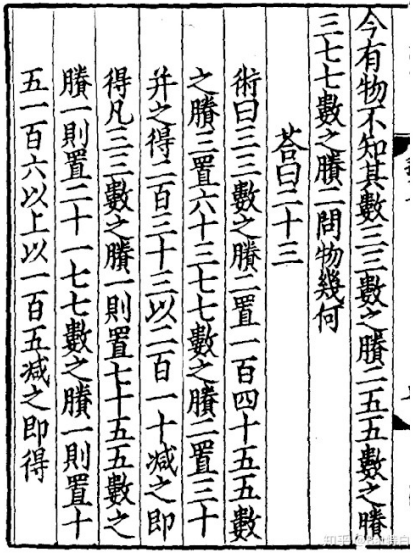
则称 DDH 问题是困难的。

- 对于群中的随机元素 h_1, h_2，任何 PPT 敌手都无法区分 $DH_g(h_1, h_2)$ 和群中的随机元素。
- 如果 CDH 问题可解，则 DDH 问题可解，但反过来不一定成立。

目录

中国余数定理 Chinese Remainder Theorem (CRT)

- 最早见于中国南北朝时期的数学著作《孙子算经》中的“物不知其数”问题，也称“孙子定理”。
- 中国余数定理说明某一范围内的整数可通过它对两两互素的整数取模所得的余数来重构。


今有物不知其數三三數之賸二五五數之賸二三七七數之賸二問物幾何
荅曰二十三
術曰三三數之賸二置一百四十五五數之賸三置六十三七七數之賸二置三十并之得二百三十三以二百一十減之即得凡三三數之賸一則置七十五五數之賸一則置二十一七七數之賸一則置十五一百六以上以一百五減之即得

例 (如何由余数重构整数)

- $\mathbb{Z}_{10}$ 中的数可通过它们对 2 和 5 取模所得的两个余数来重构。
- 假设数 x 的余数 $r_2 = 0$ 且 $r_5 = 3$;
- 则 x 是 $\mathbb{Z}_{10}$ 中的偶数且被 5 除余 3，唯一解 $x = 8$。

CRT 的几种表述形式

令 $n_1, \ldots, n_k$ 两两互素，$N = \prod_{i=1}^{k} n_i$，则以下两种表述等价：

表述一

$\mathbb{Z}_N$ 中的任一整数 $a \in \mathbb{Z}_N$ 都对应一个 k 元组 $(x_1, \ldots, x_k)$，其中 $x_i = a \bmod n_i$，$i = 1, \ldots, k$。

表述二

一元线性同余方程组

$$\begin{cases} x \equiv x_1 \pmod{n_1} \\ x \equiv x_2 \pmod{n_2} \\ \cdots \\ x \equiv x_k \pmod{n_k} \end{cases}$$

在 $\mathbb{Z}_N$ 中有一个公共解 x。

CRT 的作用

- 模数 N 很大时，模 N 的运算可以转换为模较小的数 n_i 上的运算，事先需分解 $N = n_1 \times \cdots \times n_k$。
- $\mathbb{Z}_N$ 中的算术运算可以转换为 k 元组上的算术运算。若

$$A \leftrightarrow (a_1, \ldots, a_k)$$
$$B \leftrightarrow (b_1, \ldots, b_k)$$

则

$$(A + B) \bmod N \leftrightarrow ((a_1 + b_1) \bmod n_1, \ldots, (a_k + b_k) \bmod n_k)$$
$$(A \times B) \bmod N \leftrightarrow ((a_1 \times b_1) \bmod n_1, \ldots, (a_k \times b_k) \bmod n_k)$$

因为

$$(A \cdot B) \bmod n_i = (A \bmod n_i \cdot B \bmod n_i) \bmod n_i = (a_i \cdot b_i) \bmod n_i$$

其中 $\cdot \in \{+, \times\}$。

CRT 的证明

当 $k = 2$ 时

已知

$$\begin{cases} x \equiv x_1 \pmod{n_1} \\ x \equiv x_2 \pmod{n_2} \end{cases}$$

其中 n_1 和 n_2 互素且 $N = n_1 n_2$，求 $x \in \mathbb{Z}_N$。

- 由扩展欧几里得算法可得整数 m_1, m_2 且 $m_1 n_1 + m_2 n_2 = 1$
- 一元线性同余方程组的解为 $x = (x_1 m_2 n_2 + x_2 m_1 n_1) \bmod N$
- 因为

$$x \bmod n_1 = x_1 m_2 n_2 \bmod n_1 = x_1$$
$$x \bmod n_2 = x_2 m_1 n_1 \bmod n_2 = x_2$$

CRT 的证明

推论

如果

$$\begin{cases} x \equiv y \pmod{p} \\ x \equiv y \pmod{q} \end{cases}$$

其中 p 和 q 互素，那么

$$x \equiv y \pmod{pq}$$

CRT 的证明

当 $k = 3$ 时

已知

$$\begin{cases} x \equiv x_1 & (\mathrm{mod}\ n_1) \\ x \equiv x_2 & (\mathrm{mod}\ n_2) \\ x \equiv x_3 & (\mathrm{mod}\ n_3) \end{cases}$$

其中 n_1, n_2, n_3 两两互素且 $N = n_1 n_2 n_3$，求 $x \in \mathbb{Z}_N$。

- $k > 2$ 时的情况可以归约为 $k = 2$ 时的情况。
- 由前两个等式可以确定 $x \equiv x_{12} \pmod{n_1 n_2}$。
- 再与第三个等式可以确定 $x \equiv x_{123} \pmod{n_1 n_2 n_3}$。

CRT 的证明

完整证明.

- $\forall i$， N/n_i **与** n_i **互素** $\Rightarrow \exists y_i, (N/n_i)y_i \bmod n_i = 1$。
- $\forall i \neq j$， N/n_i **有因子** $n_j \Rightarrow (N/n_i)y_i \bmod n_j = 0$。
- **令**

$$x \triangleq \sum_{i=1}^{k} \frac{N}{n_i} y_i x_i \bmod N$$

因为

$$x \bmod n_j = \frac{N}{n_j} y_j x_j \bmod n_j = x_j$$

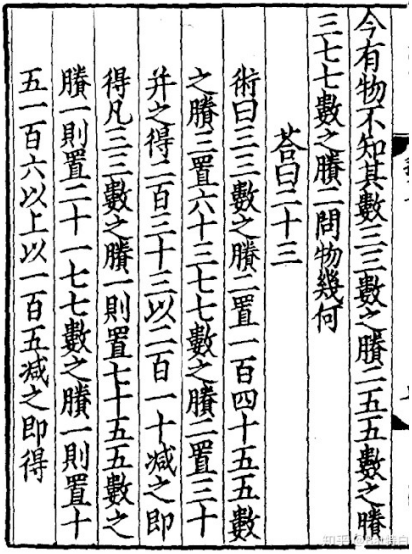
所以 x **是** $x \bmod n_j = x_j, j = 1, \ldots, k$ **的公共解**。

“物不知其数”问题求解

$x \bmod 3 = 2$, $x \bmod 5 = 3$,
$x \bmod 7 = 2$
$n_1 = 3, n_2 = 5, n_3 = 7$
$x_1 = 2, x_2 = 3, x_3 = 2$
$N = 3 \times 5 \times 7 = 105$

- 求 $y_i = (N/n_i)^{-1} \bmod n_i$，得 $y_1 = 2, y_2 = 1, y_3 = 3$
- 代入前面得到的公式中

$$x = \sum_{i=1}^{3} \frac{N}{n_i} y_i x_i \bmod N = 23$$


今有物不知其數三三數之賸二五五數之賸二三七七數之賸二問物幾何
荅曰二十三
術曰三三數之賸二置一百四十五五數之賸三置六十三七七數之賸二置三十并之得二百三十三以二百一十減之即得凡三三數之賸一則置七十五五數之賸一則置二十一七七數之賸一則置十五一百六以上以一百五減之即得

“物不知其数”问题求解

《孙子歌诀》

三人同行七十希，
五树梅花廿一支，
七子团圆正半月，
除百零五便得知。

明朝数学家程大位《算法统宗》

小结

1. 数论简介
2. 模算术
3. 群
4. 素性测试
5. 大数分解与 RSA 问题
6. 循环群与离散对数
7. 中国余数定理