二次筛法

我们知道很多密码系统的安全性依赖于大整数因子分解问题的困难性。比如 RSA 加密算法，其破解的核心其实就是分解其中的大数，找到其两个质因子。

对于一般的数位位数较小的情况（几十 bit 的数量级），常用的分解整数的方法是试除法。而当数位达到百 bit 数量级时，试除法因为其效率低下，明显不适合大数分解。

解决较大的整数分解问题，有椭圆曲线算法、特殊数域筛法、二次筛法等，而二次筛法是 500bit 及以下整数分解时，已知的最快算法。

看雪的工具板块中的工具 PPSIQS，比 RSATool 快很多，使用的就是二次筛法。

我自己没有找到很好的介绍二次筛法中文资料。因此作文一篇，在此简单介绍下二次筛法。

1. 二次筛法思想简述

了解二次筛法前，需要先了解下费马分解法，因为二次筛法是来源于费马分解法的。

1.1 费马分解法

在费马分解法发明之前的很长一段世间，人们只会使用试除法分解大数，而费马分解法的发明，使得分解大数效率大大提高。费马分解法的具体操作是怎样的呢？

其实很简单，以分解大数 N 为例，在费马分解法中，我们并不直接去找 N 的质因子。而是试图先找一对具有以下关系的平方数：

$$x^2 \equiv y^2 \pmod{N}$$

如果能找到以上关系的 x、y，那我们可以继续变形得到：

$$\rightarrow x^2 - y^2 \equiv 0 (\mathrm{mod}\ N)$$

$$\rightarrow (x + y) * (x - y) \equiv 0 (\mathrm{mod}\ N)$$

这相当于说明(x+y)*(x-y)能够被 N 整除，那么(x-y)中一定有与 N 共有的公约数。通过欧几里得算法，可以很快地找到(x-y)与 N 的最大公约数，自然就是 N 的分解了。

所以，在费马分解法的指导下，我们的问题转变了，我们只需要找到 $x^2 \equiv y^2 \pmod N$ 关系即可。然而，如果盲目去尝试，这样的 x,y 也是很不容易找到的……

二次筛法的发明，就是帮助我们有步骤地、有效地找到满足以上条件的 x、y 关系。

1.2 二次筛法中的二次函数

在二次筛法中，我们先构造一个二次函数：

$$Q(x) = (x + \lfloor \sqrt{N} \rfloor)^2 - N$$

其中 x 是从 0 开始的任意整数。之所以构造一个这样的二次函数，是因为它天然地满足一些我们期待的要求（可以帮助我们进行 N 分解），理由如下：

因为 $Q(x) = (x + \lfloor \sqrt{N} \rfloor)^2 - N$

可以推导出：

$$Q(x) - (x + \lfloor \sqrt{N} \rfloor)^2 = -N$$

这说明$Q(x) - (x + \lfloor \sqrt{N} \rfloor)^2$可以被 N 整除，用同余语言描述，就是：

$$Q(x) \equiv \left(x + \lfloor \sqrt{N} \rfloor\right)^2 \ (\mathrm{mod}\ N)$$

以上式子的右边，已经是一个平方数了，所以，我们的问题进一步简化和明确，只需要 Q(x)是一个（模 N 的）平方数，即可用费马分解法分解。

1.3 二次筛法中的因子基(Factor Bae)

如何比较高效地寻找 Q(x)，确保它是一个平方数呢？联想到所有的整数都可以分解成质数的幂的乘积形式。所谓平方数，其实就是所有质因子的幂是偶数的的情况（如 $400=2^4 5^2$）。

在二次筛法中，我们并不试图去直接找满足平方数要求的 Q(x）,而是先找一系列能够被某个质数集合完全分解的数，再由这些数，拼凑出 Q(x)来。

而“某个质数集合”，称为“因子基”（Factor Base）。能够被因子基中的质数完全分解的数，我们称为其是“光滑的”（Smoothness）。

比如，我们选在$\{2, 5\}$作为因子基，则 400 相对于这个因子基就是光滑的，而 30 相对于这个因子基就是不光滑的。

1.4 利用 $Q(x_i)$构造 Q(x)

如上文所述，我们并不直接获取 Q(x)，而是先确定“因子基”，再确定一系列的光滑的 $Q(x_i)$，当收集到足够的 $Q(x_i)$后，满足最终要求的 Q(x)可以被构造出来。

这是因为，假定有一系列光滑的 $Q(x_i)$：

记作 T = {Q（x_1），Q（x_2），Q（x_3）...Q（x_i）}，则它们的乘积一定满足：

$$Q(x_1) * Q(x_2) * \ldots . Q(x_j) \equiv \left((x_1 + \lfloor\sqrt{N}\rfloor) * (x_2 + \lfloor\sqrt{N}\rfloor) * \ldots * (x_j + \lfloor\sqrt{N}\rfloor)\right)^2 \ (\mathrm{mod}\ N)$$

这时，右边已经符合平方数，而左边，因为 $Q(x_i)$都是光滑的，我们只要选择适合的 $Q(x_i)$相乘，确保乘积结果的**所有质因子的幂是偶数**即可，这个过程本质上是解一个线性方程的过程。

2.算法简述

1. 得到待分解整数 N。
2. 选取因子基$S = \{p_1 ，p_2 ， \dots\}$，其中，$p_n(n \geq 2)$ ，p_n需要满足：

   p_n 是素数，且$N \ mod \ p_n$ 是二次剩余的，也就是满足表达式：

   $\left(\frac{\mathrm{N}}{p_n}\right)$ (N 对 p_n的勒让德符号) $= 1$。
3. 计算出一系列的 Q（ x_i ），i ∈(0 , m)。
4. 通过筛法找到对于因子基 S 是光滑的所有 Q（ x_j ）。
5. 根据 Q（ x_j ）构造指数矩阵 S。
6. 尝试找到矩阵 R，满足 R * S ≡ [0,0,0...0] (mod 2)。
   a) 如果找不到矩阵 R，回到第 3 步并扩大 i 的取值范围。

3.**实例**

我们来简单看一下利用二次筛法分解 $\mathrm{N} = 15347$。

3.1 收集 $Q(x_i)$

1. 首先选择因子基$S = \{2 ，17 ，23 ，29 ，31\}$（ 3,5,7,11,13,19 在 S 中被忽略，因为对于这些素数来说，$\left(\frac{\mathrm{n}}{\mathrm{p}}\right) = -1$ ）。
2. 选取 Q（ x_i ），i ∈(0 , 100) 。

$$T_1 = \{Q(x_1)，Q(x_2)，Q(x_3)，Q(x_4)，Q(x_5)，Q(x_6) \ldots\ldots Q(x_{99})\}$$

$$= \{29，278，529，782，1037，1294，\ \ldots\ldots 34382\}$$

3. 利用筛法进行筛选：

a) T_1中的每个数一直除以 S 里的第一个数：2，直到不能整除为止，得到T_2。

$$T_2 = \{29，139，529，391，1037，647，\ \ldots\ldots 17191\}$$

b) T_2中的每个数一直除以 S 里的第二个数：17，直到不能整除为止，得到T_3。

$$T_3 = \{29，139，529，23，61，647，\ \ldots\ldots 17191\}$$

c) 以此类推，最后得到T_n：

$$T_n = \{1，139，1，1，61，647，\ \ldots\ldots 17191\}$$

在 T_n 中会出现若干个 1，很容易我们就能理解，所有值为 1 对应的 Q（x_i）就是能够被因子基完全平方分解的，也是我们需要的值，我们将这些值做成一张表格，如下表：

x_i	$x_i + \lfloor\sqrt{N}\rfloor$	Q（x_i）	Q（x_i）的因子基分解	指数矩阵 mod 2
1	124	29	$2^0 * 17^0 * 23^0 * 29^1 * 31^0$	$(0，0，0，1，0)$
3	126	529	$2^0 * 17^0 * 23^2 * 29^0 * 31^0$	$(0，0，0，0，0)$
4	127	782	$2^1 * 17^1 * 23^1 * 29^0 * 31^0$	$(1，1，1，0，0)$
55	178	16337	$2^0 * 17^1 * 23^0 * 29^0 * 31^2$	$(0，1，0，0，0)$

72	195	22678	$2^1 * 17^1 * 23^1 * 29^1 * 31^0$	$(1，1，1，1，0)$

3.2 处理 Q(x_i)

通过上表我们不难看出很明显 Q(3)直接满足条件。那么我们就有：

因为：$Q(3) = \left(3 + \left\lfloor\sqrt{15347}\right\rfloor\right)^2 - 15347$

$= 126^2 - 15347 = 23^2$

所以：$126^2 \equiv 23^2 \pmod{15347}$

$\gcd(126 + 23，15347) = 149$，$\gcd(126 - 23，15347) = 103$。最后我们就可以求出 15347 的两个质因子 149 和 103。

一般运气没那么好，并不能直接得到 Q(x)，而是构造来的，比如，我们剔除掉 Q（3），用剩余几个 Q（x_i）来构造。剩余几个 Q（x_i）的指数矩阵 S：

$$S = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

我们想要找到矩阵 R：

$$R * \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \equiv [0 \quad 0 \quad 0 \quad 0 \quad 0] \pmod 2$$

解以上线性方程得到：

$R = [1 \quad 1 \quad 0 \quad 1]$

则我们知道：

因为：$\mathrm{Q(x)} = Q(1) * Q(4) * Q(72)$

$= 29 * 782 * 22678$

$= 22678^2$

$$\left((1 + \lfloor\sqrt{15347}\rfloor) * (4 + \lfloor\sqrt{15347}\rfloor) * (72 + \lfloor\sqrt{15347}\rfloor)\right)^2$$

$= 124^2 * 127^2 * 195^2 = 3070860^2$

所以：$22678^2 \equiv 3070860^2 \ (\mathrm{mod}\ 15347)$。

$\gcd(3070860 + 22678，15347) = 149$

$\gcd(3070860 - 22678，\ 15347) = 103$

显然 $149 * 103 = 15347$。我们也成功分解了整数 15347。

4.程序

看雪工具板块的 PPSIQS 使用的就是二次筛法，效率已经很高，不过没有将分解 n 封装成接口供调用。

我正在按照 PPSIQS 的算法，重写一个分解 N 的接口。