Barrett Reduction

Barrett Reduction是一种在计算模运算（求余数）时使用的优化算法，可以避免昂贵的除法运算。它的基本思想相当直接。

当要计算 $a \bmod N$（a 模 N）的余数 r 时，常规做法是首先计算商 $q = a \div N$，然后计算余数 $r = a - q \times N$。

然而，计算 q 时需要一次除法，而除法对 CPU 的运算开销比较大。Barrett Reduction 通过引入一个预计算的常数来避免这个除法运算。

算法核心思想

1. 选择一个足够大的 $R = 2^k$，使得 $R > N$。
2. 预计算 $\mu = \lfloor R/N \rfloor$。（$\lfloor \rfloor$ 表示向下取整）
3. 当需要计算 $a \bmod N$ 时，用以下步骤代替：
    - 计算 $q = \lfloor (a \times \mu)/R \rfloor$
    - 计算 $r = a - q \times N$
    - 如果 $r \geq N$，则 $r = r - N$

这里的关键优化是，由于 R 是 2 的幂，$(a \times \mu)/R$ 可以通过简单的位移操作实现：$q = (a \times \mu) \gg k$。

对于一个固定的有限域，N 通常是固定的，那么 μ 也是固定的，可以预先计算并存储。

误差分析

这种方法可能会导致 q 略小于实际的商，因为 μ 是 R/N 的下界（由于向下取整）。这就是为什么我们需要在计算 $r = a - q \times N$ 后，判断 r 是否小于 N，如果大于或等于 N 则需要再减去 N。

关于最差情况下需要减去 N 的次数：

- 误差的上界可以被证明是 2。也就是说，计算出的 r 最多比实际的余数大 2N。
- 因此，最多需要减去 N 两次就可以得到正确的余数。

算法优势

1. 避免了昂贵的除法运算，替换为乘法和位移操作。
2. 对于固定的模数 N，μ 可以预计算，进一步提高效率。
3. 即使在最坏情况下，也只需要进行有限次（最多两次）的额外减法。

计算μ的误差范围

我们设 q' 为实际的 a/N 值，q 为经过 Barrett Reduction 的值 $q' = \lfloor a/N \rfloor$，$q = \lfloor (a \times \mu)/R \rfloor$

1. $R = 2^k$，其中 k 是一个整数，且 $R > N$。
2. $\mu = \lfloor R/N \rfloor$

3. 假设 $a < R$（这是一个重要的假设，在实际应用中通常成立）

上界：$q \leq q'$

我们知道 $\mu = \lfloor R/N \rfloor$，所以 $\mu \leq R/N$。

$$\begin{aligned} q &= \lfloor (a \times \mu)/R \rfloor \\ &\leq \lfloor (a \times (R/N))/R \rfloor \quad \text{（因为}\mu \leq R/N\text{）} \\ &= \lfloor a/N \rfloor \\ &= q' \end{aligned} \tag{1}$$

所以 $q \leq q'$

下界：$q > q' - 2$

首先，我们知道 $R/N - 1 < \mu \leq R/N$ （因为 μ 是 R/N 的向下取整） 所以：

$$\begin{aligned} q &= \lfloor (a \times \mu)/R \rfloor \\ &> \lfloor (a \times (R/N - 1))/R \rfloor \quad \text{（因为}\mu > R/N - 1\text{）} \\ &= \lfloor a/N - a/R \rfloor \end{aligned} \tag{2}$$

接下来，我们可以写出： $q > a/N - a/R - 1$ （因为对任何 x，$\lfloor x \rfloor > x - 1$）

现在，利用假设 $a < R$，我们有 $a/R < 1$，所以：

$$\begin{aligned} q &> a/N - 1 - 1 \\ &= a/N - 2 \end{aligned} \tag{3}$$

但是 $q' = \lfloor a/N \rfloor$，所以 $a/N - 1 < q' \leq a/N$ 将这个不等式代入上面的式子： $q > (q' - 1) - 2 = q' - 3$

因为 q 是整数，所以 $q \geq q' - 2$

结合之前得到的 $q \leq q'$，我们可以得出 $q' - 2 < q \leq q'$