案例报告

Blockly 平方数分解

盛国楷

物理科学与技术学院　　320170936321

1. 程序功能和问题背景

程序的功能是将输入的正整数 N 分解为 L 个平方数，问题的解往往不唯一，但我们总是以所有可能的值中最小的 L 来分解 N，并且 L 不会大于 4。如：$43 = 3^2+3^2+5^2$，$533 = 2^2 + 23^2 = 7^2 + 22^2$ 等等。数学背景方面，有如下定理作为支撑。

Lagrange 定理. *每个正整数都是四个平方数之和，其中允许有* 0^2 *项存在。*

程序在设计中并没有用到上面的定理，该定理只是确保问题总是有解和 L 总可以不大于 4。此外，以下不加证明的应用了两条初等数论中的小结论，它们只起到判定作用，计算时用不到它们。

2.Blockly 程序示意与整体思路

整体的设计思路非常朴素直接，主要是依靠嵌套的循环来求各个平方数，加上靠数论中的结论来确定嵌套循环的层数。问题中的 L 被限制在 1,2,3,4 这四个可能的数之间。

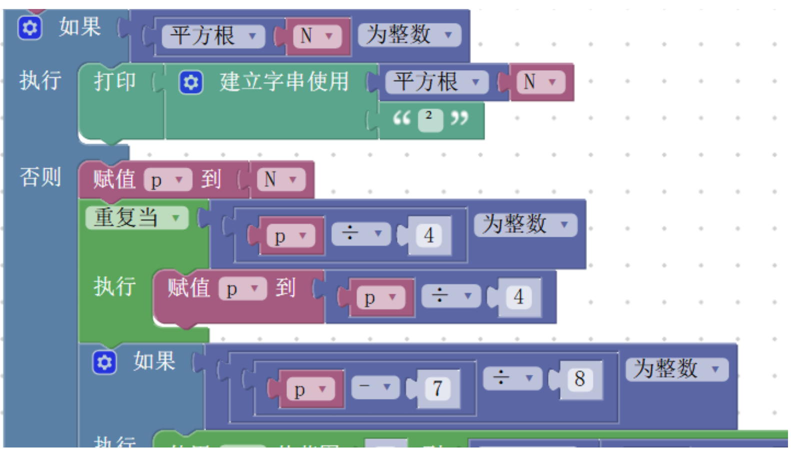

图 1: 第一部分如图所示

对上一页的图片作出解释。首先检验 N 是否为形如 $3^2, 7^2$ 这类完全平方数，即 L 是否为 1。如果不是，检验 N 是否可以表示为 $(8a+7)\cdot 4^b$ 的形式。

推论 1. *$N \neq (8a+7)\cdot 4^b$ 是 N 可以用三个平方数的和来表示的充要条件，其中 a 和 b 皆为非负整数.*

然后依据该推论我们可以确定 L 是否为 4，即确定 N 是否最少也需要 4 个平方数的和来表示。比如令 a, b 都等于零，得到数字 7，而 7 可以且仅可以被表示成 $1^2+1^2+1^2+2^2$ 的形式。如果 N 不为 $(8a+7)\cdot 4^b$ 的形式，则用 1 个平方数，2 个平方数，或者 3 个平方数都有可能能把 N 表示出来。

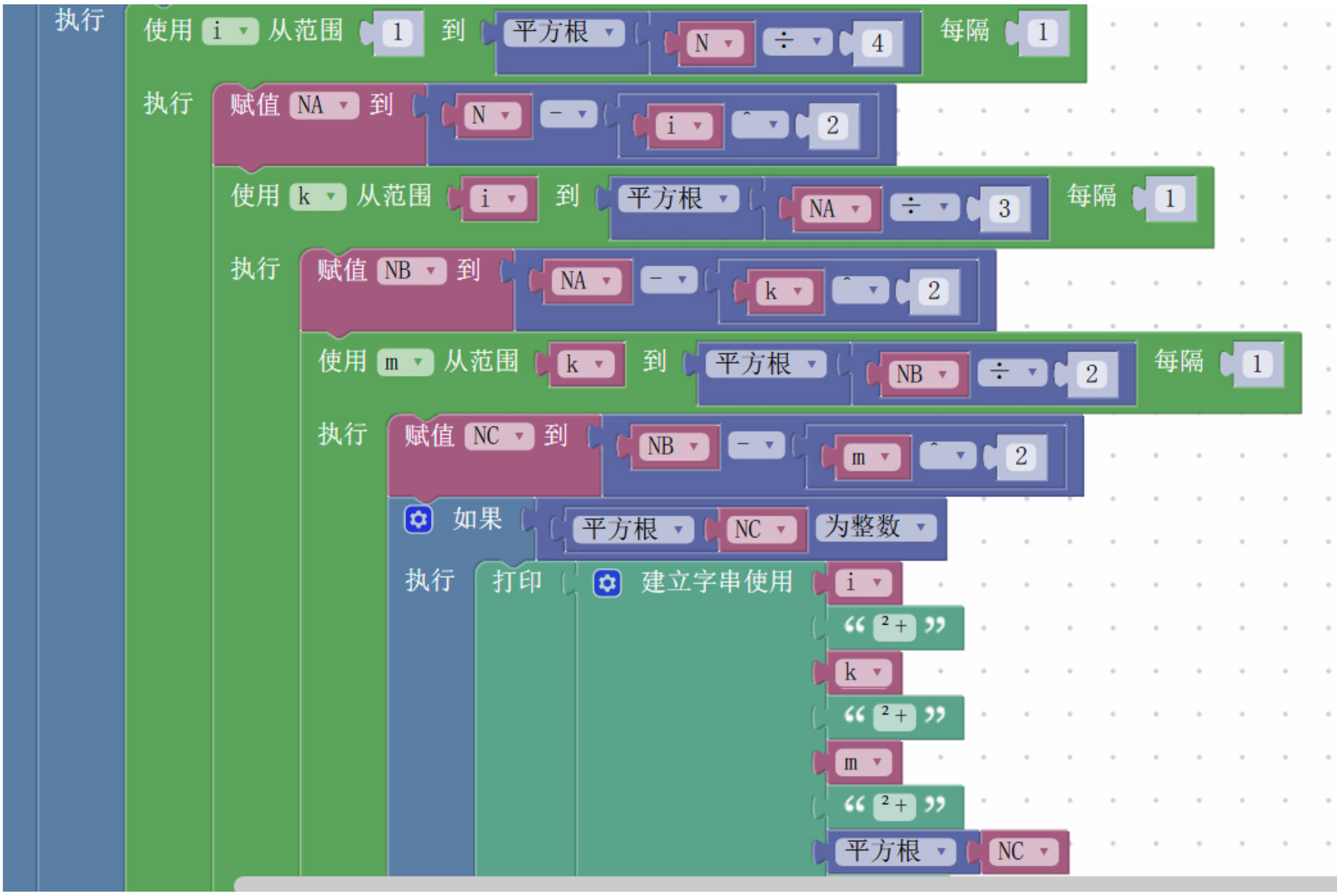

图 2: 用来计算四平方和的嵌套循环部分如图所示

可以看到，使用多个循环相嵌，L 越小，则需要加入的嵌套层数就越少，像 $L=1$ 时连循环都不需要，用一个检验就可以了；而 $L=4$ 时要额外引入循环变量 i, k, m。之前已经检验了 L 是否为 1 和 4，那么只需要检验一下 L 是否为 2 或者 3 就能对每个 N 进行合适的计算。如果对不同的 N 都使用上图所示循环，不但会使计算冗余，还会产生麻烦的结果。但是实际上不好操作。即使事实上有以下的推论：

推论 2. *将 N 分解为 $N = p_1p_2p_3...p_rM^2$，其中 $p_1, p_2, ...p_r$ 是互不相同的素因子，则 N 可表示成两个平方数之和的充要条件是每个 p_i 或为 2 或满足 $p_i \equiv 1(mod 4)$。*

不好操作的原因是涉及到质因数分解，除了一个用来判定“是否为质数”的模块，我并没有在 Blockly 中找到其他合适的模块。以我现有的能力，没有其他更方便的命令或者不自己做模块的话要把 N 分解成互不相同的素因子需要用很多的模块来拼装，也就是更长的代码。如果有更方便的关于质数的模块，或者关于储存平方数的矩阵数组一类的模块的话就可以简洁很多。不得已我增加了额外的计算量来实现原来的想法。

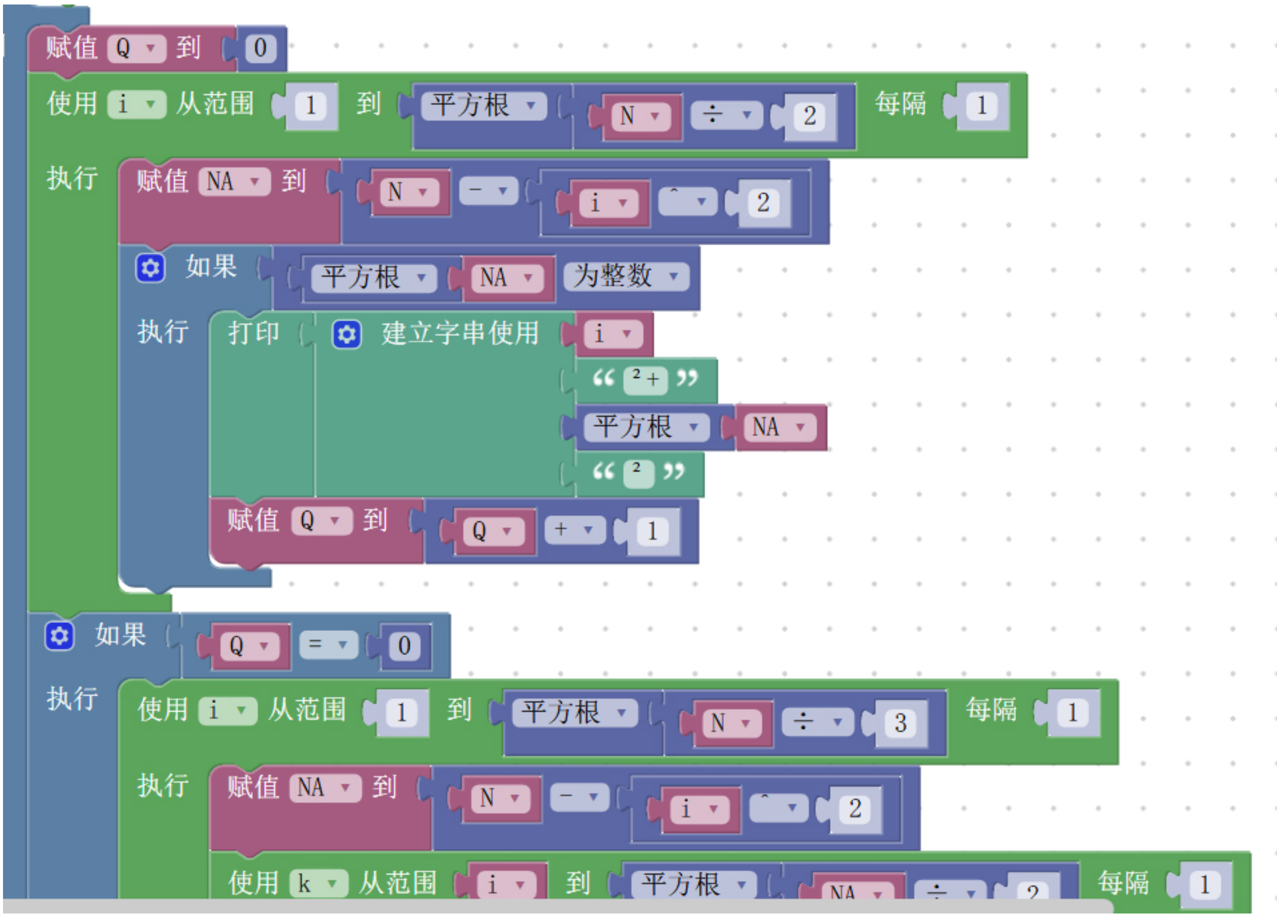

图 3: 如图所示，依据 Q 的值来决定是否进行额外的计算

顺便提一下，我不太会用 Blockly，用它的”改变变量 Q 从 0”模块程序跑不出来想要的结果，我就省懒用赋值 $Q = Q + 1$ 来代替了。

3. 重要的细节问题

课堂上演示过一个输入一个正整数 n，判定其是否为质数的案例，里面的循环是从 1 到 $\sqrt{n}$。这里在如果也类似的设置起始值与终止值的话，会导致如下两个后果。

1. 会造成结果混乱。以 $x^2 + y^2 = 10$ 为例子，从 0 到 $\sqrt{10}$ 的循环会给出”$1^2 + 3^2$”和”$3^2 + 1^2$”两个结果，显然这两个结果是同解，只是对换了一下 x 和 y 的顺序。当计算四个平方和时这个问题就更严重了，一个解共会有 $4! = 24$ 个结果。

2. 能看出来，第一个后果本质上是出现了冗余的计算。

为了解决问题，考虑在不同的求解过程中设置不同的循环的范围。

二平方和求解

只需要把从 1 到 $\sqrt{N}$ 改成从 1 到 $\sqrt{N/2}$。考察一个圆，它的方程是 $x^2+y^2=N$，直线 $y=x$ 与圆交于点 $(\sqrt{N/2},\sqrt{N/2})$，以此点为界限将第一象限内的圆对称的分为两部分。二平方和问题的解就是圆弧落在网格上的点。所以只需要考察 x 从 0 变化到 $\sqrt{N/2}$ 的部分就能检验完毕所有可能的点且不会有重复。

三平方和求解

基于同样的考虑，考察球面 $x^2+y^2+z^2=N$ 和空间平面 $x=y,y=z,z=x$，某个解 $x^2+y^2+z^2$ 对应着空间中 $3!=6$ 个点，而前面提到的三个平面恰好将第一卦限内的球面分为对称的 6 个部分，只需要任选其中一个区域检验而不是全部计算一遍。

简洁起见，选取图中左下角与 XY 平面相交的区域，赋予 i 为高度 z 的几何意义，那么 z 的范围就是从 0 至 $\sqrt{N/3}$，每一层高度 i 的圆的半径的平方为 $NA=N-i^2$，再赋予 k 为横坐标 x 的几何意义，由 $x=y,z=x$，$x^2+y^2+z^2=N$ 以及循环中的 $z=i$ 这些方程可以算出对于固定高度 i 上的 x, 即 k,，其变化范围从 i 到 $\sqrt{NA/2}$。显然问题中的 N 和 NA 在这里的几何意义分别是球的半径的平方和高度 i 上的圆半径的平方。

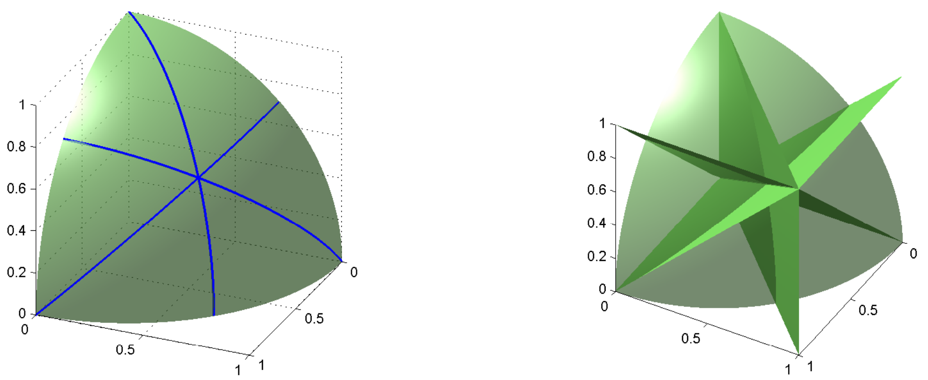

图 4: 对称的六个平面区域

四平方和求解

尽管四维的球体已经无法直观的想象其几何形状，还是可以用求解二、三平方和的方法来计算。从对称性考虑，选取方程组

$$\begin{cases} x^2+y^2+z^2+w^2=N, \\ w=z, \\ x=y \quad or \quad z=x, \end{cases} \tag{0.1}$$

利用此方程组可以确定 $4! = 24$ 个对称空间区域的边界。设 $NA = N - i^2, NB = NA - k^2, NC = NB - m^2$, 则各循环变量变化范围为

$$\begin{cases} 1 \leq w = i \leq \sqrt{N/4}, \\ i \leq z = k \leq \sqrt{NA/3}, \\ k \leq x = m \leq \sqrt{NB/2}, \end{cases} \tag{0.2}$$

4. 对运算规模的估算比较和由其引发的困惑

估算

为方便起见，忽略掉计算机测试、初始化等等的运算时间，只取每次赋值和随后的开根号这一过程的运算时间为主部，称这样的一段运算时间为“1 个单位运算时间”。很显然当 N 足够大时，无论循环变量的初始值与终止值如何设置，运算规模的主部 T 总是满足 $T(L=2)$[1] $\sim \sqrt{N}$，$T(L=3) \sim N$，$T(L=4) \sim N\sqrt{N}$ 的关系，每多一层嵌套循环，表达式就再多乘一个 $\sqrt{N}$，只是表达式前边的系数不同。记 T 为设置循环变量从 1 变化到 $\sqrt{N}$ 的程序的运算规模主部，t 为修改后的程序的运算规模主部。

下面通过计算表达式前的系数来更精确的估计计算规模，再比较一下 t 和 T 的大小。设 $N = R^2$，进行如下运算[2]并默认 N 是个大数：

$$t(2) = \sqrt{N}$$

$$T\ 2\ = \sqrt{N/2}$$

$$\frac{t(2)}{T(2)} \approx 71\%$$

$$t(3) = \sum_{i=1}^{R} \sqrt{R^2 - i^2} = R^2 \int_0^1 \sqrt{1 - x^2} = 0.79R^2 = 0.79N$$

$$T(3) = \sum_{i=1}^{\frac{\sqrt{3}}{3}R} \sqrt{\frac{R^2 - i^2}{2}} - i = R^2 \int_0^{\frac{\sqrt{3}}{3}} \sqrt{\frac{1 - x^2}{2}} - x = 0.22R^2 = 0.22N$$

$$\frac{t(3)}{T(3)} \approx 28\%$$

$$t(4) = \sum_{i=1}^{R} \sum_{k=1}^{\sqrt{R^2 - i^2}} \sqrt{R^2 - i^2 - k^2} = \iint_{D_1} \sqrt{R^2 - i^2 - k^2} \mathrm{d}x\mathrm{d}y = 0.52R^3 = 0.52N\sqrt{N}$$

$$T(4) = \sum_{i=1}^{\frac{R}{2}} \sum_{k=i}^{\sqrt{\frac{R^2 - i^2}{3}}} \sqrt{\frac{R^2 - i^2 - k^2}{3}} = \iint_{D_2} \sqrt{\frac{R^2 - x^2 - y^2}{2}} - x \ \mathrm{d}x\mathrm{d}y = 0.52R^3 = 0.04R^3 = 0.04N\sqrt{N}$$

$$\frac{t(4)}{T(4)} \approx 7.7\%$$

[1] 这里的“$\sim$ ”是“正比于”的意思，T 以刚刚规定的”单位运算时间“为单位

[2] 运算中出现的平面区域 D_1 为圆 $x^2 + y^2 \leq R^2$ 在第一象限内的那一部分；平面区域 D_2 为直线 $y = 0, y = x$, 和椭圆 $3x^2 + y^2 = R^2$ 在第一象限所围的部分。

分析与困惑

可以看到 $\frac{t(L)}{T(L)}$ 的值随 L 变大而变小，且总有 $t < T$, 所以差不多可以说对循环变量做出调整确实使计算规模减小了一部分，但其实这个结论是不严谨的。原因在于上面在计算时近似每次赋值和随后的开根号这一过程所占用运算时间都是相等的，并以此作为单位。但是很显然大数和小数所占用的时间是有差异的。这个差异有多大呢？会不会大到甚至改变估算结果的数量级呢？我无从知晓。我并没有学过计算机的专业课，并没有学过正规的编程语言，只会用一点粗浅的 MATLAB。百度上没找到精细的解释，也没有其他人能询问。我在这个问题上考虑了很长时间，没找到令我满意的答案。

又想到了另一个方面，如果这个差异造成的影响不大，循环的初始值终止值还能继续优化修改，如 $L = 2$ 时将 i 从 0 到 $\sqrt{\frac{N}{2}}$ 改为从 $\sqrt{\frac{N}{2}}$ 到 $\sqrt{N}$ 的范围[3]，这样可以再降低一半计算规模；$L = 3$ 时当然也可以把 k 从 i 到 $\sqrt{\frac{N-i^2}{2}}$ 改成从 $\sqrt{\frac{N-i^2}{2}}$ 到 $\sqrt{N-2i^2}$ 的范围。[4]

此后可以以此类推。但我不觉得这个差异的影响会这么小，但具体多大我也不知道。我想这其中需要一些计算机基础知识来做更深的分析。

5. 感想

这个报告里我用在案例本身上的时间并不多，在其他方面倒是用了不少时间。主要是我想借这次难得的机会训练一下平常没什么机会用的 MATLAB 和 LaTeX，尤其是后者，学了就没用到过。

Blockly 上手确实简单，可以使我这种计算机基础几乎为 0 的人也能实现一些简单的想法。不过美中不足的是 Blockly 在使用上有点小瑕疵，作报告的过程中发现了几个疑似有 BUG 的模块，从模块转成 XML 代码的过程中间也会出错，转出来的 XML 代码我再转成模块发现不一样了。

MARLAB 是最简单的部分，主要是用来画图，这个功能平常不怎么用。发现真要用时书上现成的命令是不够用而且很死板的，要自己去写函数和去论坛上学习别人的代码才能应对。不知道是不是我用的不对，感觉 MATLAB 画出来的图有种迷之锯齿感，试了很多都这样，最后随便挑了俩缩小一下比例放上来了。

LaTeX 这个是整个过程中最难的部分，要用时发现和重新学一遍几乎没啥区别，宏包指令什么的都是要翻书才能用。操作比 Word 难很多，不过打出来的数学公式确实是名不虚传的漂亮啊，而且越用越上手，我本来没想估算运算规模的，只是想多用用不同的数学组件临时加的。

这个过程中我意识到有必要学习一些计算机方面的基础知识，掌握一门真正意义上的编程语言越来越重要。下学期的信息学院也有开设相关的任选课，可以考虑选修一下。

---

[3] 易知 $\sqrt{\frac{N}{2}} > \sqrt{N-i^2} - \sqrt{\frac{N}{2}}$

[4] 由柯西不等式知 $(1^2+1^2)(N-2i^2+i^2) \geq (i+\sqrt{N-2i^2})^2$, 即 $\sqrt{\frac{N-i^2}{2}} - i \geq \sqrt{N-2i^2} - \sqrt{\frac{N-i^2}{2}}$

没电脑真不方便啊，周围人的笔记本都借了一圈了，后来图书馆的机房临近期末也关闭了，现在打下这一段的我是坐在网吧里的。周围的人除了一个玩天龙八部的都在吃鸡，没想到绝地求生火到这个程度。下学期一定要买一台笔记本电脑。