

浅谈数形结合思想在信息学竞赛中的应用

安徽省芜湖一中 周源

目录

目录	1
摘要	2
关键字	2
引子	3
以形助数	3
[例一]Raney 引理的证明	3
[题意简述]	3
[分析]	3
目标图形化	3
小结	4
[例二]最大平均值问题(USACO 2003 March Open)	4
[题意简述]	4
[分析]	5
目标图形化	5
构造下凸折线	5
维护下凸折线	6
最后的优化: 利用图形的单调性	7
小结	7
以数助形	7
[例三]画室(POI oi V Stage I)	8
[题意简述]	8
[分析]	8
目标数值化	9
动态规划解题	9
小结	10
总结	10
附录	11
关于 2003 年上海市选拔赛题 Sequence	11
[题意简述]	11
[分析]	11
论文附件	12
参考文献	12

摘要

数与形是数学中两个最古老而又最基本的对象，数形结合又是一种重要的数学思想。

本文主要以当今信息学奥赛中几道试题为例，从以形助数和以数助形两个侧重点讨论了数形结合思想在信息学竞赛解题中广阔的应用前景。

最后文章分析指出数形结合思想的两个重要特性并由此提出“数形结合”重在有机的结合，希望对同学们在实际比赛中灵活的运用数形结合思想有一些帮助。

关键字

信息学竞赛 数学思想 数形结合思想

以数助形 以形助数

辩证矛盾 多元性 个体差异性

思维、编程、时间、空间复杂度

引子

数与形是数学中两个最古老而又最基本的对象，数形结合又是一种重要的数学思想。

在当今信息学竞赛中，某些纷繁复杂的试题背后，往往蕴含着丰富的几何背景，而计算几何类问题却又需要借助计算机强大的实数运算能力。正如华罗庚先生所说的“数形结合千般好”，在算法和程序设计中，巧妙地运用数形结合思想，可以顺利的破解问题，化难为易，找到问题的解题思路。

数形结合思想常包括以形助数、以数助形两个方面。

以形助数

正如前文所述，一些试题中繁杂的代数关系身后往往隐藏着丰富的几何背景，而借助背景图形的性质，可以使那些原本复杂的数量关系和抽象的概念，显得直观，从而找到设计算法的捷径。

[例一]Raney 引理的证明

[题意简述]

设整数序列 $A = \{A_i, i=1, 2, \dots, N\}$ ，且部分和 $S_k = A_1 + \dots + A_k$ ，序列中所有的数字的和 $S_N = 1$ 。

证明：在 A 的 N 个循环表示¹中，有且仅有一个序列 B ，满足 B 的任意部分和 S_i 均大于零。

[分析]

先来看一个例子，若有序列 $A = \langle 1, 4, -5, 3, -2, 0 \rangle$ ，其 6 个循环表示为

1. $\langle 1, 4, -5, 3, -2, 0 \rangle$
2. $\langle 4, -5, 3, -2, 0, 1 \rangle$
3. $\langle -5, 3, -2, 0, 1, 4 \rangle$
4. $\langle 3, -2, 0, 1, 4, -5 \rangle$
5. $\langle -2, 0, 1, 4, -5, 3 \rangle$
6. $\langle 0, 1, 4, -5, 3, -2 \rangle$

其中只有第 4 个序列，部分和为 3, 1, 1, 2, 6, 1，满足成为序列 B 的条件。

若要用一般的代数或是组合方法来证明这个有趣的结论，似乎无从下手，但若想到了用“形”来帮忙，问题就简单多了。

目标图形化

周期性的推广 A 序列，得到一个无穷序列，便于观察其循环表示，得到：

¹ 先设一个序列是环状的，则从其任意一个字符处断开以后形成的非环序列即为该序列的一个循环表示。

$$\langle A_1, A_2, \dots, A_N, A_1, A_2, \dots, A_N, \dots \rangle$$

同时计算这个序列的部分和 S_i ，因为这个序列是周期性的，因此对于所有的 $k > 0$ ，均有 $S_{k+N} = S_k + 1$ 。如果做出这个函数的图像，则可以说函数有一个“平均斜率”为 $\frac{1}{N}$ ：每沿横轴正方向走 N 个单位，函数值就增加 1。于是如下图所示，

可以用两条斜率为 $\frac{1}{N}$ 的直线“夹住”函数包含的所有点：

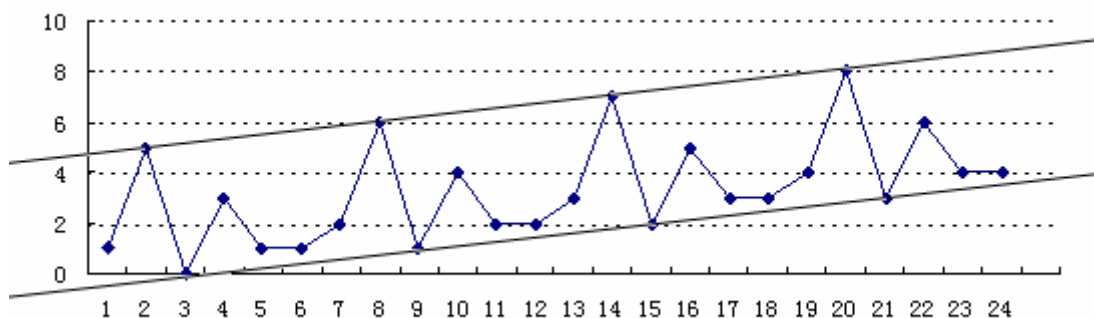


图 1 无穷序列的部分和函数图像

图示中 $N=6$ ，且使用了上文举的例子。注意较低的那条直线，在每连续的 N 个单位长度中，它与函数图像有且仅有一个交点，这是因为斜率是 $\frac{1}{N}$ 的直线在每 N 个单位长度中最多到达一次整数点。这个交点是在这以后的 N 个点中的最低值，因此由此处的后一个位置导出的循环表示的所有部分和均为正数。而同时每连续 N 个单位长度仅有一个交点也证明了解的唯一性。

小结

一个简单的几何论证就证明了著名的 Raney 引理，其简练是其他方法不能企及的。

Raney 引理有很广泛的应用，Catalan 数以及扩展 Catalan 数的组合公式就可以用该引理轻松解决。比如今年上海市选拔赛第二天比赛中的序列(Sequence)以及 OIBH 练习赛中的项链，使用 Raney 引理都是最简单的方法之一。²

用几何图形辅助思考，不只停留在组合计数这一类中，更渗透在算法设计和优化的每一个分支中，近年来流行的“斜率优化”法是另一个很好的例子。

[例二]最大平均值问题(USACO 2003 March Open)

[题意简述]

读入一系列正数， a_1, a_2, \dots, a_N ，以及一个数 F 。定义 $ave(i, j) = \frac{a_i + \dots + a_j}{j - i + 1}$ ， $i \leq j$ 。

² 用 Raney 引理解答 Sequence 的过程，详见附录。

求 $\text{Max}\{\text{ave}(a, b), 1 \leq a, b \leq N, \text{且 } a \leq b - F + 1\}$, 即求一段长度大于等于 F 且平均值最大的子串。

范围: $F \leq N \leq 10^5$ 。

[分析]

简单的枚举算法可以这样描述: 每次枚举一对满足条件的 (a, b) , 即 $a \leq b - F + 1$, 检查 $\text{ave}(a, b)$, 并更新当前最大值。

然而这题中 N 很大, N^2 的枚举算法显然不能使用, 但是能不能优化一下这个效率不高的算法呢? 答案是肯定的。

目标图形化

首先一定会设序列 a_i 的部分和: $S_i = a_1 + a_2 + \dots + a_i$, 特别的定义 $S_0 = 0$ 。

这样可以很简洁的表示出目标函数 $\text{ave}(i, j) = \frac{S_j - S_{i-1}}{j - (i-1)}$!

如果将 S 函数绘在平面直角坐标系内, 这就是过点 S_j 和点 S_{i-1} 直线的斜率!

于是问题转化为: 平面上已知 $N+1$ 个点, $P_i(i, S_i)$, $0 \leq i \leq N$, 求横向距离大于等于 F 的任意两点连线的最大斜率。

构造下凸折线

有序化一下, 规定对 $i < j$, 只检查 P_j 向 P_i 的连线, 对 P_i 不检查与 P_j 的连线。也就是说对任意一点, 仅检查该点与在其前方的点的斜率。于是我们定义点 P_i 的检查集合为

$$G_i = \{P_j, 0 \leq j \leq i - F\}$$

特别的, 当 $i < F$ 时, G_i 为空集。

其明确的物理意义为: 在平方级算法中, 若要检查 $\text{ave}(a, b)$, 那么一定有 $P_a \in G_b$; 因此平方级的算法也可以这样描述, 首先依次枚举 P_b 点, 再枚举 $P_a \in G_b$, 同时检查 $k(P_a P_b)$ 。

若将 P_i 和 G_i 同时列出, 则不妨称 P_i 为检查点, G_i 中的元素都是 P_i 的被检查点。

当我们考察一个点 P_i 时, 朴素的平方级算法依次选取 G_i 中的每一个被检查点 p , 考察直线 pP_i 的斜率。但仔细观察, 若集合内存在三个点 P_i, P_j, P_k , 且 $i < j < k$, 三个点形成如下图所示的关系, 即 P_j 点在直线 $P_i P_k$ 的上凸部分: $k(P_i, P_j) > k(P_j, P_k)$, 就很容易可以证明 P_j 点是多余的。

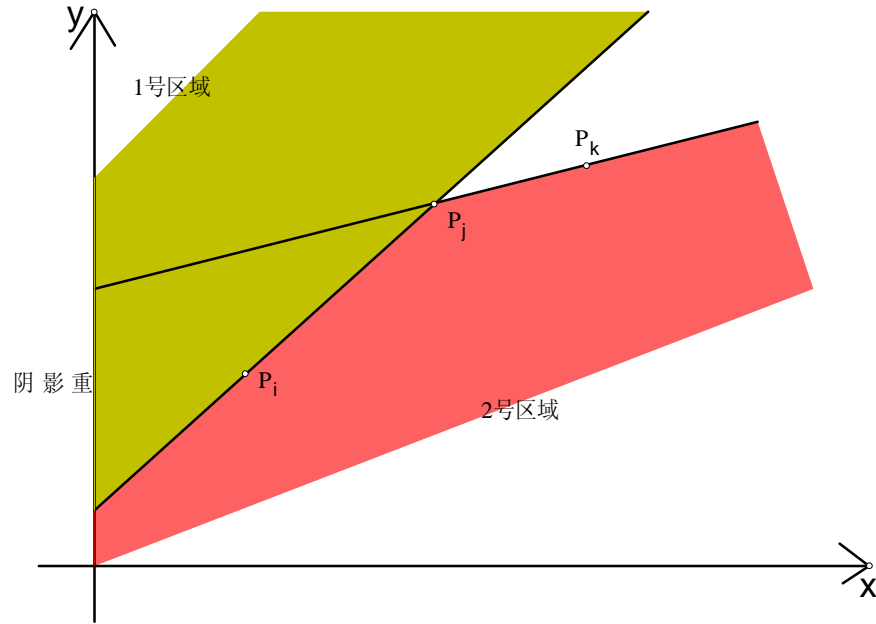


图 2

若 $k(P_t, P_j) > k(P_t, P_i)$, 那么可以看出, P_t 点一定要在直线 P_iP_j 的上方, 即阴影所示的 1 号区域。同理若 $k(P_t, P_j) > k(P_t, P_k)$, 那么 P_t 点一定要在直线 P_jP_k 的下方, 即阴影所示的 2 号区域。

综合上述两种情况, 若 P_iP_j 的斜率同时大于 P_iP_t 和 P_jP_t 的, P_t 点一定要落在两阴影的重叠部分, 但这部分显然不满足开始时 $t > j$ 的假设。于是, P_t 落在任何一个合法的位置时, P_iP_j 的斜率要么小于 P_iP_t , 要么小于 P_jP_t , 即不可能成为最大值, 因此 P_j 点多余, 完全可以从检查集中删去。

这个结论告诉我们, 任何一个点 P_t 的检查集中, 不可能存在一个对最优结果有贡献的上凸点, 因此我们可以删去每一个上凸点, 剩下的则是一个下凸折线。最后需要在这个下凸折线上找一点与 P_t 点构成的直线斜率最大——显然这条直线是在与折线相切时斜率最大, 如图所示。

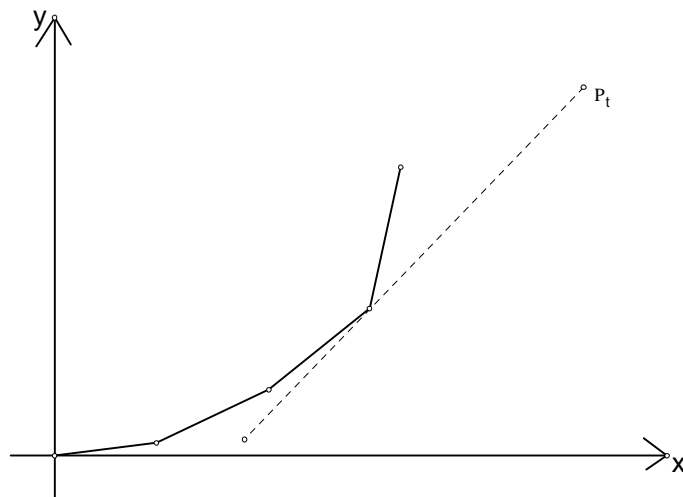


图 3

维护下凸折线

这一小节中, 我们的目标是: 用尽可能少的时间得到每一个检查点的下凸折

线。

算法首先从 P_F 开始执行：它是检查集合非空的最左边的一个点，集合内仅有一个元素 P_0 ，而这显然满足下凸折线的要求，接着向右不停的检查新的点： P_{F+1} , P_{F+2} , ..., P_N 。

检查的过程中，维护这个下凸折线：每检查一个新的点 P_i ，就可以向折线最右端加入一个新的点 P_{i-F} ，同时新点的加入可能会导致折线右端的一些点变成上凸点，我们用一个类似于构造凸包的过程依次删去这些上凸点，从而保证折线的下凸性。由于每个点仅被加入和删除一次，所以每次维护下凸折线的平摊复杂度为 $O(1)$ ，即我们用 $O(N)$ 的时间得到了每个检查集合的下凸折线。

最后的优化：利用图形的单调性

最后一个问题就是如何求过 P_i 点，且与折线相切的直线了。一种直接的方法就是二分，每次查找的复杂度是 $O(\log_2 N)$ 。但是从图形的性质上很容易得到另一种更简便更迅速的方法：由于折线上过每一个点切线的斜率都是一定的³，而且根据下凸函数斜率的单调性，如果在检查点 P_i 时找到了折线上的已知一个切点 A ，那么 A 以前的所有点都可以删除了：过这些点的切线斜率一定小于已知最优解，不会做出更大的贡献了。

于是另外保留一个指针不回溯的向后移动以寻找切线斜率即可，平摊复杂度为 $O(1)$ 。

至此，此题算法时空复杂度均为 $O(N)$ ，得到了圆满的解决。

小结

回顾本题的解题过程，一开始就确立了以平面几何为思考工具的正确路线，很快就发现了检查集合中对最优解有贡献的点构成一个下凸函数这个重要结论，之后借助计算几何中求凸包的方法维护一个下凸折线，最后还利用下凸函数斜率的单调性发现了找切线简单方法。题解围绕平面几何这个中心，以斜率为主线，整个解题过程一气呵成，又避免了令人头晕的代数式变换，堪称以形助数的经典例题。

顺便提一下：这种方法在加速决策过程，很多动态规划算法都可以运用本题“斜率优化”的方法提高算法效率。如 IOI 2002 的 batch 和 BOI 2003 的 euro 等。至于这类题目的共同特点，还是很值得研究的，但不在本文讨论范围内，因而不再讨论，但欢迎有兴趣的同学以后和我交流。

以数助形

古希腊的毕达哥拉斯认为“万物皆数”，的确，数是反映事物本质特征的最好方法之一。数学发展史上，正是在解析几何创立之后，人们才对各种繁杂的曲线有了更深入的了解。如今信息时代中，计算机处理各类事物，最终无不是归结于二进制数的基本运算，数的重要性可见一斑。

在当今信息学竞赛中，一些试题给出的描述中图形极为复杂，容易使选手陷入“迷魂阵”，在这种情况下，以数助形，一举抓住其本质特征，不失为解题的一种好方法。

³ 由于折线没有连续性，因此更准确的应该说，过每一个点切线斜率的范围都一定的。

[例三]画室(P01 oi V Stage I)

[题意简述]

定义尺寸为 0 的方阵为一个 1×1 的矩阵,在其唯一的一个方格中有一个小孔。
对于 $i > 0$, 递归的定义尺寸为 i 的方阵如下图所示:

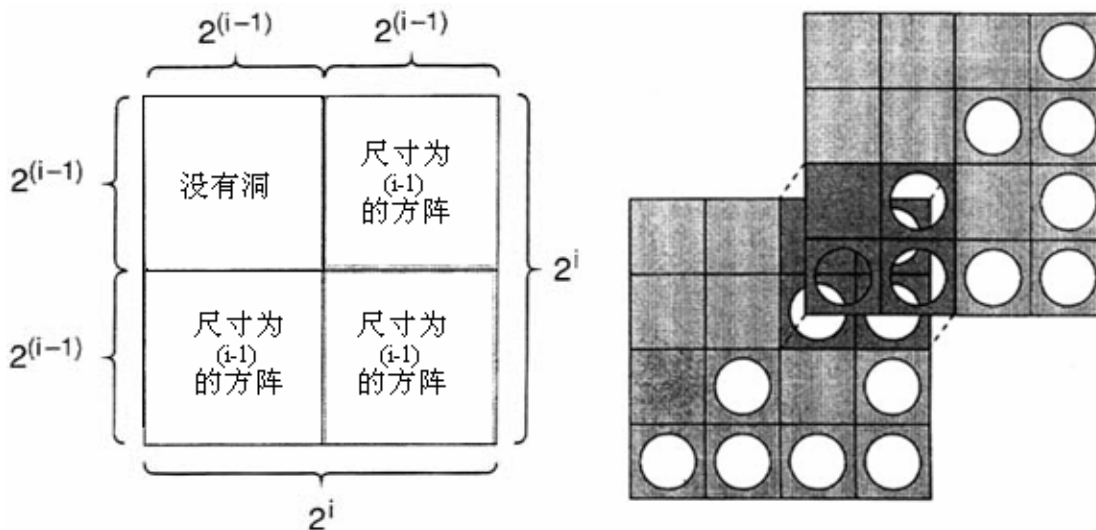


图 4

给定方阵的尺寸 N , 以及另外两个参数 X 和 Y 。准备两个尺寸为 N 的方阵,一个叠放在另一个上面,再将上面的方阵向右移动 X 列,同时向上移动 Y 行。

如此操作之后,求两个方阵有多少个公共的孔。

如右上图,尺寸为 2 的方阵,向右平移 2 列,向上平移 2 行。则两个方阵有 3 个公共小孔。

范围: $N \leq 100$ 。

[分析]

直接分析两个方阵相交后的情况是可行的,我曾经看过一些集训队前辈的解题报告,都是这么分析的,但是方法很繁,思考量很大。

下图是某解题报告中的一个说明附图,报告中先标出两个方阵的相交区域,再分情况讨论。显然可以看出,直接从“形”来分析本题,路子是很坎坷的。

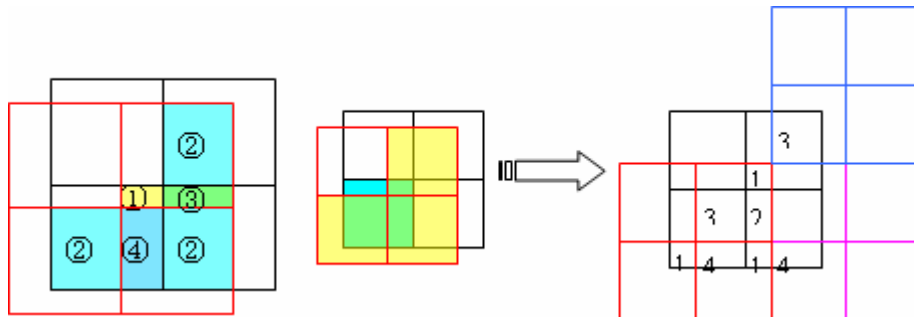


图 5

目标数值化

我们不如换至和“形”相对的另一面“数”来思考，按照下图所示的 x, y 方向为每行每列从 0 开始编号，最大至 2^N-1 ，于是每一个方格都有唯一的坐标 (x, y) 。

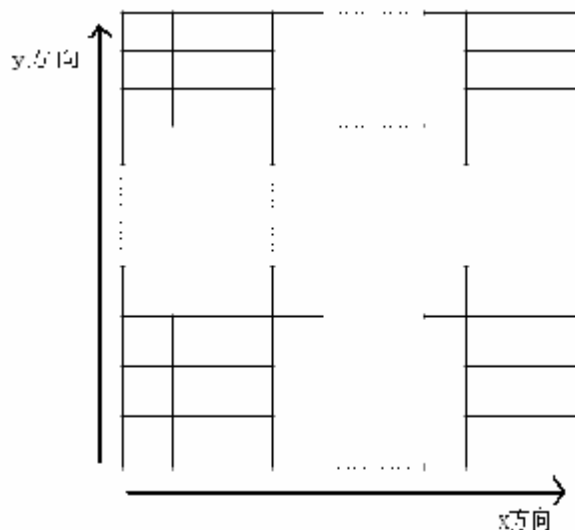


图 6

下面来研究一下在什么条件下，一个方格 $P(x, y)$ 内有小孔。由于方阵是二分递归定义的，于是我们很自然联想到将 x 和 y 化为二进制。设 x 和 y 的二进制表示分别为：

$$a_1a_2a_3\dots a_N \text{ 和 } b_1b_2b_3\dots b_N$$

来看两个数的第 1 位， a_1 和 b_1 ，如下图，它们一共有 4 种取值方法，其分布分别对应着递归定义中的左上、左下、右上、右下四块区域。显然当 $a_1=0$ 且 $b_1=1$ 时无论以后各位取什么数， P 点内都不会有小孔：因为其已经落在了左上无孔区。否则可以同理讨论两个数的第 2 位，第 3 位……

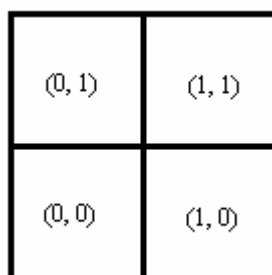


图 7 示意 (a_1, b_1) 的取值分布情况

得到的结论是，当且仅当不存在 $1 \leq i \leq N$ ，满足 $a_i=0$ 且 $b_i=1$ 时，方格 P 内有小孔。不妨称这个为方格的**有孔性质**。

动态规划解题

后面的问题就非常简单了，题目要找的无非是这样的有序数对 (x, y) 的个数： $0 \leq x, y, x+X, y+Y \leq 2^N-1$ ，且 $(x, y), (x+X, y+Y)$ 的二进制表示都满足有孔性质称这个为方格的**有公共孔性质**。

我们可以采用动态规划的方法：首先将 X, Y 也都转化成二进制形式：

$$p_1p_2p_3\dots p_N \text{ 和 } q_1q_2q_3\dots q_N$$

以位数为阶段，通过记录进位情况保证无后效性： $f(i, k1, k2)$ 表示第 i 位至第

N 位部分满足有公共孔性质的有序对总数，且要满足这一部分有序对的坐标和对应部分的 X, Y 相加进位分别是 k_1 和 k_2 ：显然 $0 \leq k_1, k_2 \leq 1$ 。

动态规划的状态转移是非常简单的，但描述比较复杂：每一次转移需要约 $(2^2)^2=16$ 次运算，因此不再赘述，有兴趣的读者可以查看附件中的程序。

最后说明一下，题目所要的答案就是 $f(1, 0, 0)$ 。

算法若不算上高精度，时间复杂度为 $O(N)$ ，若使用循环数组，空间上仅需要常数个高精度数组⁴，而且实现程序也极为简单，包括高精度也不过 100 多行。对比从“形”上得出的算法，“数”的优越性是不言而喻的。

小结

回顾解题过程，当分析发现两方阵相交情况较复杂，不宜讨论时，我们决定避开“形”的正面冲突，而从“数”这方面下手，很快便取得了令人满意的效果：方格的有孔性质和有公共孔性质使题目的要求显得简单了许多。到此就可以套用经典的动态规划算法了。

可以说本题是一个较好的例子，但类似以数助形的例题似乎比较罕见。事实上，正如前文所述，一般的计算机都是以数为基础的，同学们在写各类程序的时候，最终还是要归结到“数”来实现，对数的重要作用多少有些熟视无睹了。

而上例又可以看出，如果试题加以适当的“误导”，选手们背离“数”的捷径，南辕北辙也不是没有可能的。因此，在遇到如同上例的题目时，面对多元化的复杂图形，化形归数，往往是抓住题目要害的好方法。

总结

数与形是现实世界中客观事物的抽象和反映，是数学的基石，也是信息学竞赛命题涉及的两个主要方面。数形结合是一种古老的数学思想，新兴的信息学奥林匹克竞赛又赋予她新的活力。

上文举了三个实例，大体上来说，都巧妙的运用了数形结合思想。但从细节上分析，它们之间仍略有差异。

其一，三者从两个不同的侧重点阐述了数形结合思想的内涵，即以形助数和以数助形。但在实际问题中，数和形决没有明确的界限，数形结合思想也并不仅仅局限于文中提出的两个方面。更多的情况下，数与形互相促进、互相包含，在一定条件下互相转化，可以用“数形互助”一词来形容。

这，体现了数形的**辩证矛盾关系**和数形结合思想的**多元性**。

其二，用“形”来解例题二，似乎是唯一的出路，但在例一和例三中，并不是仅仅能用文中提到的方法解题，其他精彩解法我也略知一二。但相比而言，巧妙的使用数形结合思想会大大降低思考和编程复杂度，为我们在短短的竞赛时间中迅速解题开辟了一条便捷的道路。

需要指出的是，不同的人有不同的知识结构，比赛经验等，他们对某一算法难度系数的感觉也是不同的。因而对同一题而言，不同的人可能会选择不同的数形之路解题。这，体现了数形结合思想的**个体差异性**。

而本文提出的三个例子，都是选择了大多数人能够接受的算法，却并不能说是每位读者心目中最简单的算法。但醉翁之意不在酒，几个小例子仅作抛砖引玉，重点在于探讨如何在信息学竞赛中运用数形结合思想。

⁴ 直接用“形”的方法做出的程序，空间复杂度是 $O(N)$ 的，而且程序很长，详见附录。

在信息学竞赛中运用数形结合思想，就是在处理问题时，斟酌问题的具体情形，善于抓住问题的主要矛盾，使数量关系的问题借助于几何图形直观而形象化，或者使图形问题借助于数量关系而本质化。

数形结合，重在“结合”二字。灵活的运用数形结合思想，需要重视思想的个体差异性，根据各人的现有知识水平和思维方式，有机的将抽象的数学、计算机语言与直观的图形结合起来，将抽象思维与形象思维结合起来，实现抽象概念与具体形象的联系和转化，更快更好更简单的解决实际问题。

附录

关于 2003 年上海市选拔赛题 Sequence

[题意简述]

一个序列 $\{A_i, i=0, 1, 2, \dots, 3N\}$ 由 $3N+1$ 项组成，每一项要么为 1，要么为 -2。定义部分和 $S_K=A_0+A_1+\dots+A_K$ ，求所有满足性质 P 的序列 A 的数目，性质 P 为： $S_{3N}=1$ 且对于所有的 $K=0, 1, 2, \dots, 3N-1, 3N$ ，有 $S_K>0$ 。即所有项的和为 1，且所有部分和为正。

例如 $N=2$ 的时候，共有 3 组这样的序列：

1, 1, 1, -2, 1, 1, -2,

1, 1, 1, 1, -2, 1, -2,

1, 1, 1, 1, 1, -2, -2。

范围： $N \leq 1000$ 。

[分析]

[引理]任一序列 A，它的任何一种循环表示都不与自身相同。

[证明]若相同，根据循环串的性质，其必定可以分成 $d>1$ 个完全相同部分。设每部分和为 s ，显然有 $s*d=1$ ，而 $d>1$ ，则 s 一定不是整数，这与序列中所有项都是整数矛盾。

因此，A 的任意循环表示都不等于 A。Q.E.D.

[定理]满足性质 P 的序列个数为 $\frac{1}{3N+1} C_{3N+1}^N$ 。

[证明]列出所有的 A 序列，一共有 C_{3N+1}^N 个。

根据其循环表示分类，由于[引理]的成立，每一类中一定都有 $3N+1$ 个序列，即一共 $\frac{1}{3N+1} C_{3N+1}^N$ 类。又因为 Raney 引理成立，所以每一类中有且只有一个序列满足性质 P。

即满足性质 P 的序列总数为 $\frac{1}{3N+1} C_{3N+1}^N$ 。Q.E.D.

同样的方法，可以推出 Catalan 数的公式，这里不再赘述。

论文附件

POI oi V Stage I Painter's Studio 一题，集训队前辈的解题报告：



画室解题报告.doc

特别感谢湖南长郡中学的金恺提供这份报告。

POI oi V Stage I Painter's Studio 一题，数形结合思想算法的程序：



Mal.pas

参考文献

CONCRETE MATHEMATICS by Ronald L. Graham & Donald E. Knuth & Oren Patashnik

USA Computing Olympiad : <http://ace.delos.com/usacogate>

Polish Olympiad in Informatics : <http://www.oi.edu.pl/>

上海市 NOI'2003 选拔赛 (SHTSC 2003) 试题

数形结合思想在数学教学中的妙用 from 教育教学论文网

伸展树的基本操作与应用

安徽省芜湖一中 杨思雨

目录

【关键字】	2
【摘要】	2
【引言】	2
【伸展树的基本操作】	2
伸展操作 $Splay(x,S)$	3
伸展树的基本操作.....	4
时间复杂度分析.....	5
【伸展树的应用】	7
【总结】	8
【参考书目】	9
【附录】	9

【关键字】

伸展树 基本操作 应用

【摘要】

本文主要介绍了伸展树的基本操作以及其在解题中的应用。全文可以分为以下四个部分。

第一部分引言，主要说明了二叉查找树在信息学竞赛中的重要地位，并且指出二叉查找树在某些情况下时间复杂度较高，进而引出了在时间复杂度上更为优秀的伸展树。

第二部分介绍了伸展树的基本操作。并给出了对伸展树时间复杂度的分析和证明，指出伸展树的各种基本操作的平摊复杂度均为 $O(\log n)$ ，说明伸展树是一种较平衡的二叉查找树。

第三部分通过一个例子介绍了伸展树在解题中的应用，并将它与其它树状数据结构进行了对比。

第四部分指出了伸展树的优点，总结全文。

【引言】

二叉查找树（Binary Search Tree）能够支持多种动态集合操作。因此，在信息学竞赛中，二叉排序树起着非常重要的作用，它可以被用来表示有序集合、建立索引或优先队列等。

作用于二叉查找树上的基本操作的时间是与树的高度成正比的。对一个含 n 各节点的完全二叉树，这些操作的最坏情况运行时间为 $O(\log n)$ 。但如果树是含 n 个节点的线性链，则这些操作的最坏情况运行时间为 $O(n)$ 。而有些二叉查找树的变形，其基本操作在最坏情况下性能依然很好，比如红黑树、AVL 树等等。

本文将要介绍的伸展树（Splay Tree），也是对二叉查找树的一种改进，虽然它并不能保证树一直是“平衡”的，但对于伸展树的一系列操作，我们可以证明其每一步操作的平摊复杂度都是 $O(\log n)$ 。所以从某种意义上说，伸展树也是一种平衡的二叉查找树。而在各种树状数据结构中，伸展树的空间要求与编程复杂度也都是很优秀的。

【伸展树的基本操作】

伸展树是二叉查找树的一种改进，与二叉查找树一样，伸展树也具有有序性。即伸展树中的每一个节点 x 都满足：该节点左子树中的每一个元素都小于 x ，而其右子树中的每一个元素都大于 x 。与普通二叉查找树不同的是，伸展树可以自我调整，这就要依靠伸展操作 $Splay(x,S)$ 。

伸展操作 $\text{Splay}(x,S)$

伸展操作 $\text{Splay}(x,S)$ 是在保持伸展树有序性的前提下，通过一系列旋转将伸展树 S 中的元素 x 调整至树的根部。在调整的过程中，要分以下三种情况分别处理：

情况一：节点 x 的父节点 y 是根节点。这时，如果 x 是 y 的左孩子，我们进行一次 Zig（右旋）操作；如果 x 是 y 的右孩子，则我们进行一次 Zag（左旋）操作。经过旋转， x 成为二叉查找树 S 的根节点，调整结束。如图 1 所示

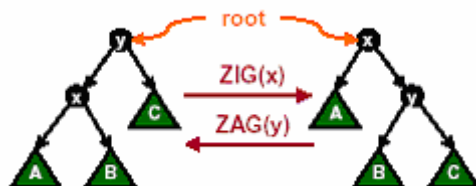


图 1

情况二：节点 x 的父节点 y 不是根节点， y 的父节点为 z ，且 x 与 y 同时是各自父节点的左孩子或者同时是各自父节点的右孩子。这时，我们进行一次 Zig-Zig 操作或者 Zag-Zag 操作。如图 2 所示

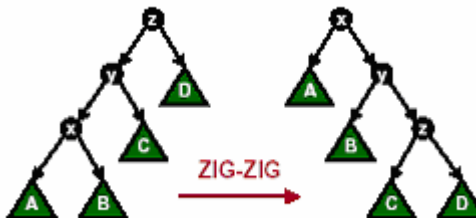


图 2

情况三：节点 x 的父节点 y 不是根节点， y 的父节点为 z ， x 与 y 中一个是其父节点的左孩子而另一个是其父节点的右孩子。这时，我们进行一次 Zig-Zag 操作或者 Zag-Zig 操作。如图 3 所示



图 3

如图 4 所示，执行 $\text{Splay}(1,S)$ ，我们将元素 1 调整到了伸展树 S 的根部。再执行 $\text{Splay}(2,S)$ ，如图 5 所示，我们从直观上可以看出在经过调整后，伸展树比原来“平衡”了许多。而伸展操作的过程并不复杂，只需要根据情况进行旋转就

可以了，而三种旋转都是由基本得左旋和右旋组成的，实现较为简单。

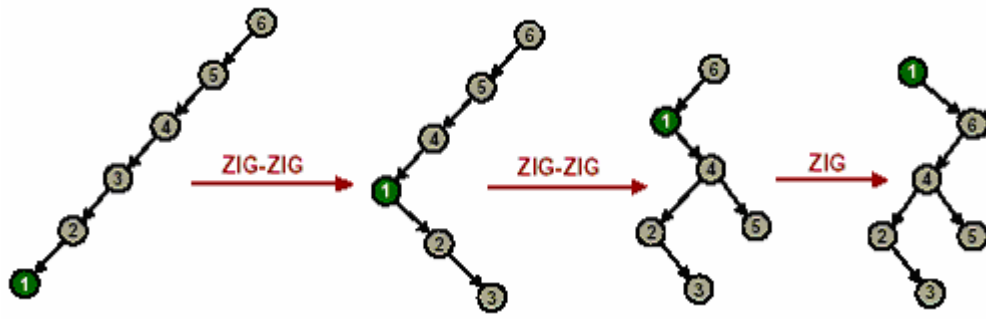


图 4 Splay(1,S)

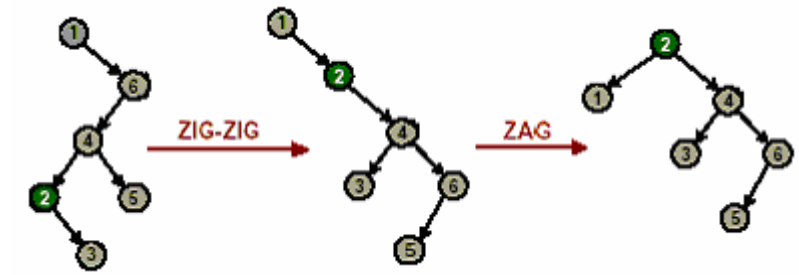


图 5 Splay(2,S)

伸展树的基本操作

利用 Splay 操作，我们可以在伸展树 S 上进行如下运算：

(1)Find(x,S): 判断元素 x 是否在伸展树 S 表示的有序集中。

首先，与在二叉查找树中的查找操作一样，在伸展树中查找元素 x 。如果 x 在树中，则再执行 Splay(x,S)调整伸展树。

(2)Insert(x,S): 将元素 x 插入伸展树 S 表示的有序集中。

首先，也与处理普通的二叉查找树一样，将 x 插入到伸展树 S 中的相应位置上，再执行 Splay(x,S)。

(3>Delete(x,S): 将元素 x 从伸展树 S 所表示的有序集中删除。

首先，用在二叉查找树中查找元素的方法找到 x 的位置。如果 x 没有孩子或只有一个孩子，那么直接将 x 删去，并通过 Splay 操作，将 x 节点的父节点调整到伸展树的根节点处。否则，则向下查找 x 的后继 y ，用 y 替代 x 的位置，最后执行 Splay(y,S)，将 y 调整为伸展树的根。

(4)Join($S1,S2$): 将两个伸展树 $S1$ 与 $S2$ 合并成为一个伸展树。其中 $S1$ 的所有元素都小于 $S2$ 的所有元素。

首先，我们找到伸展树 $S1$ 中最大的一个元素 x ，再通过 Splay($x,S1$)将 x 调整到伸展树 $S1$ 的根。然后再将 $S2$ 作为 x 节点的右子树。这样，就得到了新的伸展树 S 。如图 6 所示

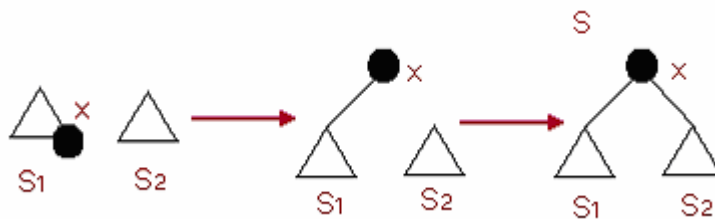


图 6

(5)Split(x,S): 以 x 为界, 将伸展树 S 分离为两棵伸展树 S1 和 S2, 其中 S1 中所有元素都小于 x, S2 中的所有元素都大于 x。

首先执行 Find(x,S), 将元素 x 调整为伸展树的根节点, 则 x 的左子树就是 S1, 而右子树为 S2。如图 7 所示

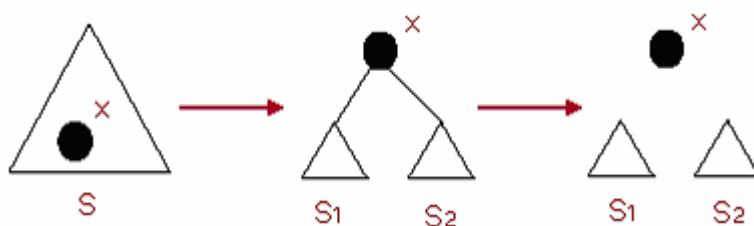


图 7

除了上面介绍的五种基本操作, 伸展树还支持求最大值、求最小值、求前趋、求后继等多种操作, 这些基本操作也都是建立在伸展操作的基础上的。

时间复杂度分析

由以上这些操作的实现过程可以看出, 它们的时间效率完全取决于 Splay 操作的时间复杂度。下面, 我们就用会计方法来分析 Splay 操作的平摊复杂度。

首先, 我们定义一些符号: $S(x)$ 表示以节点 x 为根的子树。|S| 表示伸展树 S 的节点个数。令 $\mu(S) = \lceil \log|S| \rceil$, $\mu(x) = \mu(S(x))$ 。如图 8 所示

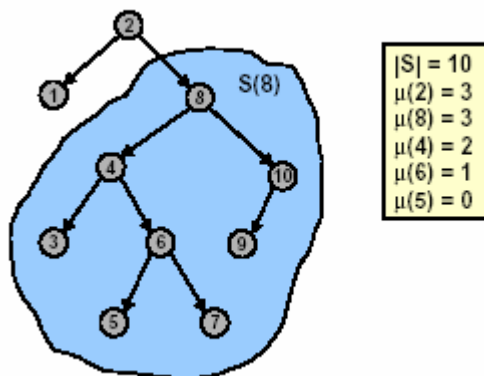


图 8

我们用 1 元钱表示单位代价（这里我们将对于某个点访问和旋转看作一个单位时间的代价）。定义**伸展树不变量**：在任意时刻，伸展树中的任意节点 x 都至少有 $\mu(x)$ 元的存款。

在 **Splay** 调整过程中，费用将会用在以下两个方面：

(1) 为使用的时间付费。也就是每一次单位时间的操作，我们要支付 1 元钱。

(2) 当伸展树的形状调整时，我们需要加入一些钱或者重新分配原来树中每个节点的存款，以保持不变量继续成立。

下面我们给出关于 **Splay** 操作花费的定理：

定理：在每一次 **Splay(x,S)** 操作中，调整树的结构与保持伸展树不变量的总花费不超过 $3\mu(S)+1$ 。

证明：用 $\mu(x)$ 和 $\mu'(x)$ 分别表示在进行一次 Zig、Zig-Zig 或 Zig-Zag 操作前后节点 x 处的存款。

下面我们分三种情况分析旋转操作的花费：

情况一：如图 9 所示

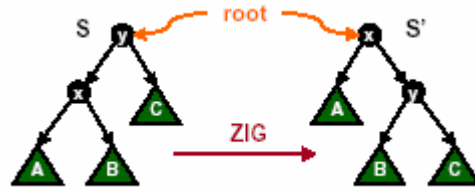


图 9

我们进行 Zig 或者 Zag 操作时，为了保持伸展树不变量继续成立，我们需要花费：

$$\begin{aligned} \mu'(x) + \mu'(y) - \mu(x) - \mu(y) &= \mu'(y) - \mu(x) \\ &\leq \mu'(x) - \mu(x) \\ &\leq 3(\mu'(x) - \mu(x)) \\ &= 3(\mu(S) - \mu(x)) \end{aligned}$$

此外我们花费另外 1 元钱用来支付访问、旋转的基本操作。因此，一次 Zig 或 Zag 操作的花费至多为 $3(\mu(S) - \mu(x))$ 。

情况二：如图 10 所示

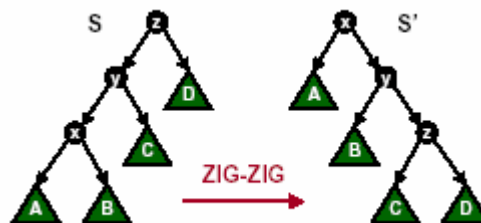


图 10

我们进行 Zig-Zig 操作时，为了保持伸展树不变量，我们需要花费：

$$\begin{aligned} \mu'(x) + \mu'(y) + \mu'(z) - \mu(x) - \mu(y) - \mu(z) &= \mu'(y) + \mu'(z) - \mu(x) - \mu(y) \\ &= (\mu'(y) - \mu(x)) + (\mu'(z) - \mu(y)) \end{aligned}$$

$$\begin{aligned} &\leq (\mu'(x) - \mu(x)) + (\mu'(x) - \mu(x)) \\ &= 2(\mu'(x) - \mu(x)) \end{aligned}$$

与上种情况一样，我们也需要花费另外的 1 元钱来支付单位时间的操作。

当 $\mu'(x) < \mu(x)$ 时，显然 $2(\mu'(x) - \mu(x)) + 1 \leq 3(\mu'(x) - \mu(x))$ 。也就是进行 Zig-Zig 操作的花费不超过 $3(\mu'(x) - \mu(x))$ 。

当 $\mu'(x) = \mu(x)$ 时，我们可以证明 $\mu'(x) + \mu'(y) + \mu'(z) < \mu(x) + \mu(y) + \mu(z)$ ，也就是说我们不需要任何花费保持伸展树不变量，并且可以得到退回来的钱，用其中的 1 元支付访问、旋转等操作的费用。为了证明这一点，我们假设 $\mu'(x) + \mu'(y) + \mu'(z) > \mu(x) + \mu(y) + \mu(z)$ 。

联系图 9，我们有 $\mu(x) = \mu'(x) = \mu(z)$ 。那么，显然 $\mu(x) = \mu(y) = \mu(z)$ 。于是，可以得出 $\mu(x) = \mu'(z) = \mu(z)$ 。令 $a = 1 + |A| + |B|$ ， $b = 1 + |C| + |D|$ ，那么就有

$$\lceil \log a \rceil = \lceil \log b \rceil = \lceil \log(a+b+1) \rceil. \quad ①$$

我们不妨设 $b \geq a$ ，则有

$$\begin{aligned} \lceil \log(a+b+1) \rceil &\geq \lceil \log(2a) \rceil \\ &= 1 + \lceil \log a \rceil \\ &> \lceil \log a \rceil \end{aligned} \quad ②$$

①与②矛盾，所以我们可以得到 $\mu'(x) = \mu(x)$ 时，Zig-Zig 操作不需要任何花费，显然也不超过 $3(\mu'(x) - \mu(x))$ 。

情况三：与情况二类似，我们可以证明，每次 Zig-Zag 操作的花费也不超过 $3(\mu'(x) - \mu(x))$ 。

以上三种情况说明，Zig 操作花费最多为 $3(\mu(S) - \mu(x)) + 1$ ，Zig-Zig 或 Zig-Zag 操作最多花费 $3(\mu'(x) - \mu(x))$ 。那么将旋转操作的花费依次累加，则一次 Splay(x,S) 操作的费用就不会超过 $3\mu(S) + 1$ 。也就是说对于伸展树的各种以 Splay 操作为基础的基本操作的平摊复杂度，都是 $O(\log n)$ 。所以说，伸展树是一种时间效率非常优秀的数据结构。

【伸展树的应用】

伸展树作为一种时间效率很高、空间要求不大的数据结构，在解题中有很大的用武之地。下面就通过一个例子说明伸展树在解题中的应用。

例：营业额统计 Turnover (湖南省队 2002 年选拔赛)

题目大意

Tiger 最近被公司升任为营业部经理，他上任后接受公司交给的第一项任务便是统计并分析公司成立以来的营业情况。Tiger 拿出了公司的账本，账本上记录了公司成立以来每天的营业额。分析营业情况是一项相当复杂的工作。由于节假日，大减价或者是其他情况的时候，营业额会出现一定的波动，当然一定的波动是能够接受的，但是在某些时候营业额突变得很高或是很低，这就证明公司此时的经营状况出现了问题。经济管理学上定义了一种**最小波动值**来衡量这种情况：

该天的最小波动值 = $\min \{ | \text{该天以前某一天的营业额} - \text{该天的营业额} | \}$

当最小波动值越大时，就说明营业情况越不稳定。而分析整个公司的从成立

到现在营业情况是否稳定，只需要把每一天的最小波动值加起来就可以了。你的任务就是编写一个程序帮助 Tiger 来计算这一个值。

注：第一天的最小波动值为第一天的营业额。

数据范围：天数 $n \leq 32767$ ，每天的营业额 $a_i \leq 1,000,000$ 。最后结果 $T \leq 2^{31}$ 。

初步分析

题目的意思非常明确，关键是要每次读入一个数，并且在前面输入的数中找到一个与该数相差最小的一个。

我们很容易想到 $O(n^2)$ 的算法：每次读入一个数，再将前面输入的数一次查找一遍，求出与当前数的最小差值，记入总结果 T 。但由于本题中 n 很大，这样的算法是不可能在规定时间内出解的。而如果使用线段树记录已经读入的数，就需要记下一个 $2M$ 的大数组，这在当时比赛使用 TurboPascal 7.0 编程的情况下是不可能实现的。而前文提到的红黑树与平衡二叉树虽然在时间效率、空间复杂度上都比较优秀，但过高的编程复杂度却让人望而却步。于是我们想到了伸展树算法。

算法描述

进一步分析本题，解题中，涉及到对于有序集的三种操作：插入、求前趋、求后继。而对于这三种操作，伸展树的时间复杂度都非常优秀，于是我们设计了如下算法：

开始时，树 S 为空，总和 T 为零。每次读入一个数 p ，执行 $\text{Insert}(p,S)$ ，将 p 插入伸展树 S 。这时， p 也被调整到伸展树的根节点。这时，求出 p 点左子树中的最右点和右子树中的最左点，这两个点分别是有序集中 p 的前趋和后继。然后求得最小差值，加入最后结果 T 。

解题小结

由于对于伸展树的基本操作的平摊复杂度都是 $O(\log n)$ 的，所以整个算法的时间复杂度是 $O(n \log n)$ ，可以在时限内出解。而空间上，可以用数组模拟指针存储树状结构，这样所用内存不超过 $400K$ ，在 TP 中使用动态内存就可以了。编程复杂度方面，伸展树算法非常简单，程序并不复杂。虽然伸展树算法并不是本题唯一的算法，但它与其他常用的数据结构相比还是有很多优势的。下面的表格就反映了在解决这一题时各个算法的复杂度。从中可以看出伸展树在各方面都是优秀的，这样的算法很适合在竞赛中使用。

	顺序查找	线段树	AVL 树	伸展树
时间复杂度	$O(n^2)$	$O(n \log a)$	$O(n \log n)$	$O(n \log n)$
空间复杂度	$O(n)$	$O(a)$	$O(n)$	$O(n)$
编程复杂度	很简单	较简单	较复杂	较简单

【总结】

由上面的分析介绍，我们可以发现伸展树有以下几个优点：

(1) 时间复杂度低，伸展树的各种基本操作的平摊复杂度都是 $O(\log n)$ 的。在树状数据结构中，无疑是非常优秀的。

(2)空间要求不高。与红黑树需要记录每个节点的颜色、AVL 树需要记录平衡因子不同，伸展树不需要记录任何信息以保持树的平衡。

(3)算法简单，编程容易。伸展树的基本操作都是以 Splay 操作为基础的，而 Splay 操作中只需根据当前节点的位置进行旋转操作即可。

虽然伸展树算法与 AVL 树在时间复杂度上相差不多，甚至有时候会比 AVL 树慢一些，但伸展树的编程复杂度大大低于 AVL 树。在竞赛中，使用伸展树在编程和调试中都更有优势。

在信息学竞赛中，不能只一味的追求算法有很高的时间效率，而需要在时间复杂度、空间复杂度、编程复杂度三者之间找到一个“平衡点”，合理的选择算法。这也需要我们在平时对各种算法反复琢磨，深入研究，在竞赛中才能够游刃有余的应用。

【参考书目】

- [1]傅清祥，王晓东.《算法与数据结构》.电子工业出版社.1998.01
- [2]严蔚敏，吴伟民.《数据结构(第二版)》.清华大学出版社.1992.06
- [3]《Introduction to Algorithms,Second Edition》.2001

【附录】

- (一) 伸展操作和其他各种基本操作，其实现参见：[Splay Tree.doc](#)
- (二)文中提到的伸展树的基本操作，具体过程可参照动画：[Splay Tree.htm](#)
- (三)针对文中例题，作者用伸展树算法编写了程序：[Turnover.pas](#)