揭开神秘面纱

——2017 年集训队一道测试题解法探秘

冯跃峰

2017 年中国国家集训队第 3 次测试中, 瞿振华提供了一道非常有趣的组合问题, 给出的两个解答也很精彩. 尤其是付云浩还给出一个非常简洁明了的解答, 巧妙至极, 读之令人拍案叫绝 (见《走向 IMO (2017)》p118).

我们自然会问, 怎么会想到如此精妙绝伦的解答? 愚禁不住好奇心驱使, 对其解答的思维过程妄加揣测, 写成如下的文字, 作为对原解答的学习和膜拜.

题目. 设 X 是一个 100 元集, 求具有下列性质的最小自然数 n: 对于任意由 X 的子集构成的长为 n 的序列 $A_1, A_2, \cdots, A_n$, 都存在 $1 \le i < j < k \le n$, 满足 $A_i \subseteq A_j \subseteq A_k$, 或者 $A_k \subseteq A_j \subseteq A_i$.

(2017 年中国国家集训队测试题第 3 轮)

【题感】 从条件看, 任何长为 n 的集合序列 $A_1, A_2, \cdots, A_n$, 都存在 $1 \le i < j < k \le n$, 满足 $A_i \subseteq A_j \subseteq A_k$, 或者$A_k \subseteq A_j \subseteq A_i$. 这个条件将被反复运用, 它是“任意型”的, 但不能“赋值”(适当取长为 n 的集合序列), 因为 n 并不知道; 也无法叠合, 所以只能反面思考. 注意到其性质的表述却比较啰嗦, 可引入新定义:

如果长为 n 的序列 $A_1, A_2, \cdots, A_n$, 存在 $1 \le i < j < k \le n$, 满足$A_i \subseteq A_j \subseteq A_k$, 或者$A_k \subseteq A_j \subseteq A_i$, 则称 (A_i, A_j, A_k) 是该序列的一个长为 3 的同序链.

这样, 问题变为: X **的任意长为** n **的子集列都含有长为 3 的同序链, 求** n **的最小值.**

从目标看, 求 n 的最小值, 包括“不等式 $n \ge \cdots$”与“等式 $n = \cdots$时合乎要求”两个方面.

一般地说, 不等式的建立有 4 个最常用的方法:

收稿日期: **2018-1-25.**

(i) 建立算法: 得到控制式 $f(n) \geq 0$;

(ii) 穷举否定: 证明 $n \neq 1, n \neq 2, \cdots, n \neq r-1$, 得到 $n \geq r$;

(iii) 反面思考: 找最大的常数 n', 使不满足限定条件, 由此推出 $n \geq n'+1$.

(iv) 从等号突破: 找到一个满足限定条件的 n_0, 然后证明 $n \geq n_0$.

对于本题, 由于题目的条件不提供算法, 从而方法 (i) 不行; 而方法 (ii) 则太繁, 比如 $n = 1, 2, \cdots, 100, \cdots$ 都不合乎要求. 同样理由, 方法 (iv) 也不行, 所以我们考虑方法 (iii).

【反面思考】取尽可能大的 n, 使题目限定条件不满足, 即 X 存在长为 n 的子集列不含长为 3 的同序链.

这自然想到找一个充分条件: 考虑怎样的集合列一定不含同序链.

【充分条件】集合列不含同序链的一个充分条件是: 任何两个集合互不包含. 进而, 两个不同集合互不包含的一个充分条件是它们的容量相等. 于是, 取若干个容量相等的互异集合排成一个序列 U, 则 U 不含同序链.

为使序列 U 尽可能长, 可取 U 中各子集的容量为 $|X|/2 = 50$. 此时, 序列 U 的长度为 C_{100}^{50}.

【局部扩展】由于允许所构造的集合列包含长为 2 的同序链 (包括含相同集合), 于是可在序列 U 的后面再接一个序列 U, 得到集合列 $M = (U, U)$. 至此, 集合列 M 的长度达到了$\mathrm{C}_{100}^{50} + \mathrm{C}_{100}^{50}$.

能否继续扩展? 显然, 集合列 M 的左右两边都不能再排任何集合. 比如 M 后面排一个集合 A, 当 $|A| < 50$ 时, U 中一定有一个集合 B 包含 A; 当 $|A| > 50$ 时, U 中一定有一个集合 B 被 A 包含, 于是 M 中有同序链 BBA.

是不是 M 已达到最大? 非也! ——在两个集合列 U 之间还可以"插入"容量不为 50 的集合. 为使序列 M 尽可能长, 可插入一个"各集合容量都为 51"的子集列 V, 得到集合列 $M = (U, V, U)$.

至此, 集合列 M 的长度达到了 $\mathrm{C}_{100}^{50} + \mathrm{C}_{100}^{51} + \mathrm{C}_{100}^{50} = \mathrm{C}_{101}^{51} + \mathrm{C}_{100}^{50}$.

能否让 V 也出现两次? 不能! 否则有形如 AAB 的同序链. 但注意到长为 4 的大小交替排列不含长为 3 的同序链, 所以在集合列 U, V 之间还可以"插入"容量小于 50 的集合. 为使构造的序列尽可能长, 可插入一个"各集合容量都为 49"的子集列 W, 得到集合列 $M = (U, W, V, U)$.

至此, M 无法再扩充, 其长度为 $\mathrm{C}_{101}^{51} + \mathrm{C}_{100}^{50} + \mathrm{C}_{100}^{49} = \mathrm{C}_{101}^{51} + \mathrm{C}_{101}^{50} = \mathrm{C}_{102}^{51}$.

由于 M 中不存在长为 3 的同序链, 所以 $n \leq \mathrm{C}_{102}^{51}$ 不合乎要求, 故 $n \geq \mathrm{C}_{102}^{51} + 1$.

下面证明 $n=\mathrm{C}_{102}^{51}+1$ 合乎要求.

注意条件的反面: 存在集合列“不含长为 3 的同序链”更便于推理, 宜进行反面思考.

【反面思考】反设 $n={C_{102}}^{51}+1$ 不合乎要求, 则必定存在长为 n 的集合序列 $A_1,A_2,\cdots,A_n$, 其中不含长为 3 的同序链.

【结构联想】注意到集合序列“不存在长为 3 的同序链”的一个充分条件是各个集合互不包含, 这似乎与斯佩拉定理相关, 而数据 C_{102}^{51} 恰好是斯佩拉定理中的“最值常数”: 102 元集合中互不包含的子集的最多个数, 它指引我们这样探索: 能否由 X 的不含长为 3 的同序链的子集列 $(A_1,A_2,\cdots,A_n)$, 对应某个 102 元集合的一个互不包含子集列 $(B_1,B_2,\cdots,B_n)$.

若然, 则由斯佩拉定理, 必有 $n\le\mathrm{C}_{102}^{51}$, 与 $n=\mathrm{C}_{102}^{51}+1$ 矛盾.

【反面思考】如何使子集 $B_1,B_2,\cdots,B_n$ 互不包含? 采用反证法的推理模式: 假设 $B_1,B_2,\cdots,B_n$ 中存在两个集合 $B_i,B_j\ (1\le i<j\le n)$, 其中一个包含另一个, 则有以下两种情形: $B_i\subseteq B_j\ (1\le i<j\le n)$ 或 $B_i\subseteq B_j\ (1\le j<i\le n)$. 要导出矛盾, 只需使其具有如下性质:

当 $B_i\subseteq B_j\ (1\le i<j\le n)$ 时, 必定有 $A_i\subseteq A_j\ (1\le i<j\le n)$, 且存在 $j<k\le n$, 使得 $A_j\subseteq A_k$, 这样便有 $A_i\subseteq A_j\subseteq A_k\ (1\le i<j<k\le n)$, 产生矛盾;

当 $B_i\subseteq B_j\ (1\le j<i\le n)$ 时, 必定有 $A_i\subseteq A_j\ (1\le j<i\le n)$, 且存在 $1\le k<j$, 使得 $A_j\subseteq A_k$, 这样便有 $A_i\subseteq A_j\subseteq A_k\ (1\le k<j<i\le n)$, 产生矛盾.

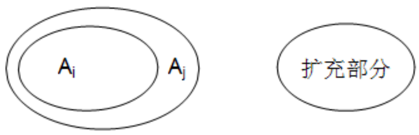
现在考虑如何构造 $B_1,B_2,\cdots,B_n$ 使其具有上述性质, 取一个代表元来研究: 考虑如何构造 $B_t\ (1\le t\le n)$.

由于要满足多个条件, 可先构造满足部分条件的拟对象.

【拟对象逼近】(i) 首先, B_t 是某个 102 元集合的子集, 这个 102 元集显然为 $X\cup\{x,y\}$, 其中 x,y 是不属于 X 的任意两个不同元素.

(ii) 其次, 考虑如何定义 $B_t\ (1\le t\le n)$, 方能使其满足“$B_i\subseteq B_j\ (1\le j<i\le n)$ 时, 必定有 $A_i\subseteq A_j\ (1\le j<i\le n)$”.

【找充分条件】一个充分条件是: B_t 是由 A_t 扩充的, 如下图:



不管 A_i,A_j 扩充的元素是否相同, 只要 $B_i\subseteq B_j$, 就一定有 $A_i\subseteq A_j$.

再结合 (i), 扩充的元素必定属于 $\{x,y\}$, 从而扩充方式有 4 种可能:

$B_t=A_t; B_t=A_t\cup\{x\}; B_t=A_t\cup\{y\}; B_t=A_t\cup\{x,y\}$.

究竟 A_t 用哪种扩充方式, 应依据目标要求来确定.

(iii) 最后考虑, 如何定义 B_t, 才能保证"存在 $j<k\le n$, 使得 $A_j\subseteq A_k$". 它可理解为: 在序列 $A_1,A_2,\cdots,A_n$ 位于 A_j 的右边存在包含 A_j 的集合 A_k.

为叙述问题方便, 我们先给出如下定义:

如果 $A_i\subseteq A_j$ $(1\le i<j\le n)$, 则称 A_i 被 A_j "右包含", A_j 为右包含 A_i 的集合; 如果 $A_i\subseteq A_j$ $(1\le j<i\le n)$, 则称 A_i 被 A_j "左包含", A_j 为左包含 A_i 的集合.

由"$A_j\subseteq A_k(1\le j<k\le n)$"的意义: A_j 存在"右包含"集合, 想到这样定义 B_t:

当 A_t 存在"右包含"集合时, 令 $B_t=A_t\cup\{x\}$ (并假定此时不存在左包含).

进而想到: 当 A_t 存在"左包含"集合时, 令 $B_t=A_t\cup\{y\}$ (假定此时不存在右包含);

当 A_t 既存在"左包含"集合, 又存在"右包含"集合时, 令 $B_t=A_t\cup\{x,y\}$;

当 A_t 既不存在"左包含"集合, 又不存在"右包含"集合时, 令 $B_t=A_t$.

【验证】(i) 若 $B_i\subseteq B_j(1\le i<j\le n)$, 此时显然有 $A_i\subseteq A_j(1\le i<j\le n)$. 这表明, A_i 存在" 右包含"集合 (不否定 A_i 同时存在"左包含"集合), 由 B_i 的定义,

$$B_i=A_i\cup\{x\}, \text{或者 } B_i=A_i\cup\{x,y\}.$$

由此可见, $x\in B_i$. 又 $B_i\subseteq B_j$, 所以 $x\in B_j$.

进而由 B_j 的定义, A_j 存在" 右包含"集合, 即存在 $j<k\le n$, 使 $A_j\subseteq A_k$, 于是 $A_i\subseteq A_j\subseteq A_k$ $(1\le i<j<k\le n)$, 矛盾.

(ii) 若 $B_j\subseteq B_i$ $(1\le i<j\le n)$, 此时显然有 $A_j\subseteq A_i$ $(1\le i<j\le n)$. 这表明, A_j 存在"左包含"集合 (不否定 A_j 同时存在"右包含"集合), 由 B_j 的定义,

$$B_j=A_j\cup\{y\}, \text{或者 } B_j=A_j\cup\{x,y\}.$$

由此可见, $y\in B_j$. 又 $B_j\subseteq B_i$, 所以 $y\in B_i$.

进而由 B_i 的定义, A_i 存在"左包含"集合, 即存在 $1\le k<i\le n$, 使 $A_i\subseteq A_k$, 于是 $A_j\subseteq A_i\subseteq A_k$ $(1\le k<i<j\le n)$, 矛盾.

【新写】对 X 的子集序列 $A_1,A_2,\cdots,A_n$, 称满足题目条件的 3 个集合 A_i,A_j,A_k 为长为 3 的同序链. 如果 $A_i\subseteq A_j$ 且 $1\le i<j\le n$ $(1\le j<i\le n)$,

则称 A_j 为右(左)包含 A_i 的集合.

将 X 的所有 50 (49, 51) 元集排成一个序列 $U(V,W)$, 构造集合列 $M=(U,V,W,U)$. 则 M 的长度为 $\mathrm{C}_{100}^{50}+\mathrm{C}_{100}^{49}+\mathrm{C}_{100}^{51}+\mathrm{C}_{100}^{50}=\mathrm{C}_{102}^{51}$.

由于 M 中不存在长为 3 的同序链, 所以 $n\geq \mathrm{C}_{102}^{51}+1$.

当 $n=\mathrm{C}_{102}^{51}+1$ 时, 反设存在长为 n 的不含长为 3 的同序链的集合序列 $A_1,A_2,\cdots,A_n$, 任取一个 102 元集: $X\cup\{x,y\}$. 对每一个 $1\leq t\leq n$, 当 A_t 在序列中只存在"右(左)包含"集合时, 令 $B_t=A_t\cup\{x\}$ $(A_t\cup\{y\})$; 当 A_t 的"左、右包含"集合都存在时, 令 $B_t=A_t\cup\{x,y\}$; 当 A_t 的"左、右包含"集合都不存在时, 令$B_t=A_t$.

若存在 $B_i\subseteq B_j$ $(1\leq i<j\leq n)$, 此时显然有 $A_i\subseteq A_j$ $(1\leq i<j\leq n)$. 这表明, A_i 存在"右包含"集合, 由 B_i 的定义,

$$B_i=A_i\cup\{x\},\text{或者}B_i=A_i\cup\{x,y\}.$$

由此可见, $x\in B_i$. 又 $B_i\subseteq B_j$, 所以 $x\in B_j$.

进而由 B_j的定义, A_j 存在"右包含"集合, 即存在 $j<k\leq n$, 使 $A_j\subseteq A_k$, 于是 $A_i\subseteq A_j\subseteq A_k$ $(1\leq i<j<k\leq n)$, 矛盾. 若存在 $B_j\subseteq B_i$ $(1\leq i<j\leq n)$ 同样矛盾.

所以 $B_1,B_2,\cdots,B_n$ 互不包含, 由斯佩拉定理, 有 $n\leq \mathrm{C}_{102}^{51}$, 与 $n=\mathrm{C}_{102}^{51}+1$ 矛盾.

综上所述, n 的最小值为 $\mathrm{C}_{102}^{51}+1$.