

素数等差数列的最小参数计算：基于 2^{32} 范围的系统性搜索

曾飞戈
(独立研究者, 中国)

摘要

素数等差数列是数论的核心研究对象。尽管格林-陶定理断言了任意长素数等差数列的存在性，但构造具有极小参数（最小首项或最小公差）的具体实例仍是重大的计算挑战。本文提出一种系统性的二阶段搜索算法，在 2^{32} （约 4.29×10^9 ）范围内的约 2.03 亿个素数中展开大规模计算，旨在发现特定长度 k 的、参数极小的素数等差数列。主要工作如下：（1）在统一的 2^{32} 范围内，完成了对长度 $k=2$ 至 16 的素数等差数列最小公差解与最小首项解的系统性搜索，获得了一系列最优解；（2）基于详实的计算日志，揭示了算法耗时、最小公差倍数与项数 k 的数学性质（尤其是素数性 k 值）之间的内在关联；（3）通过数据集的范围对比，初步探讨了最小首项与搜索范围的关系。本研究获得的一系列最优解，为素数分布的结构特性提供了新的具体证据，并为后续建立更精确的理论模型奠定了实证基础。

关键词： 素数等差数列；最小首项；最小公差；最优解；计算数论；搜索算法；格林-陶定理；基础公差

1. 引言

素数等差数列是指由素数构成的等差数列，其形式为 $P, P+d, P+2d, \dots, P+(k-1)d$ ，其中 P 为首项， d 为公差， k 为项数。对素数等差数列的研究历史

悠久。1770 年，拉格朗日证明了，在公差为 2 的三素数等差数列中，必有一项能被 3 整除，唯一的例外是 (3,5,7)。这一结论揭示了素数分布的内在约束，也预示着寻找长素数等差数列的困难性。1923 年，哈代与李特尔伍德在其素数 k 元组猜想中，预言了任意长度素数等差数列存在的可能性。2004 年，格林与陶哲轩[1]里程碑式地证明了格林-陶定理：素数序列中包含任意长度的等差数列。这从理论上彻底解决了存在性问题。

然而，格林-陶定理的证明是非构造性的，它并未提供寻找这些长等差数列的具体方法。因此，一个巨大的计算挑战便凸显出来：如何有效地构造出这些序列，并优化其参数——特别是寻找具有最小首项或最小公差的实例。这已成为计算数论领域一个重要的研究方向。在有界间隔方向上，张益唐[7]于 2013 年证明了存在无穷多对间距有界的素数，这为素数分布的精细结构提供了重要进展。在中国，王元[8]在筛法及其应用方面做出了奠基性工作，潘承洞与潘承彪[9]的素数定理初等证明工作也为素数分布研究提供了经典基础。上述工作为素数分布的整体结构提供了理论基础，而本文则从计算角度切入，关注特定范围内最小参数的精确确定。

与寻找最长序列的“广度”追求不同，另一条研究路线专注于“深度”挖掘，即在特定长度 k 下，寻找首项或公差尽可能小的素数等差数列。此类小参数研究需要对特定范围内的素数进行精细、密集地搜索。本文沿袭这一路线，致力于在固定的大范围内，进行一次统一、系统的最小参数穷举搜索，以提供一份干净、可复现的基准数据集。

此前，作者在素数等差数列最小参数计算方向上进行了初步探索。本文是在此基础上的进一步发展和系统化。此外，作者在文献[10]中系统阐述了素数等差数列的构造性猜想体系，本文所使用的“基础公差”概念即源自该工作。

本文的主要工作：

1. **系统性计算：** 在固定的整数范围 $N = 2^{32} = 4,294,967,296$ （即 $\pi(N) = 203,280,221$ ）内，对 $k = 2$ 至 16 的素数等差数列，同时完成了最小公差解与最小首项解的完备搜索。

2. **规律揭示：** 基于精确的计算日志，定量分析了搜索难度与项数 k 的数学性质之间的关系，为“素数项数导致搜索空间跃升”这一理论推断提供了实证支持。

3. **数据奠基：** 所获得的完整数据集及搜索程序均已存档，可供独立核验，并为后续关于最小参数分布规律的理论研究提供了可靠的实证基础。

2. 定义与理论背景

定义 2.1（基础公差）

设 P_n 为第 n 个素数。对于项数 k 的素数等差数列，设 P_n 为大于等于 k 的最小素数（即满足 $P_{n-1} < k \leq P_n$ ）。其公差 d 必须是前 $n-1$ 个素数乘积的倍数。该乘积称为基础公差，记为：

$$D_0 = \prod_{i=1}^{n-1} P_i = 2 \times 3 \times 5 \times \dots \times P_{n-1}$$

（注：该乘积遍历所有小于 k 的素数，不包括 k 本身。）

示例： 当 $k=10$ 时，大于等于 10 的最小素数是 $P_5 = 11$ ，基础公差 $D_0 = 2 \times 3 \times 5 \times 7 = 210$ 。当 $k=16$ 时，大于等于 16 的最小素数是 $P_7 = 17$ ，基础公差 $D_0 = 2 \times 3 \times 5 \times 7 \times 11 \times 13 = 30030$ 。

该概念与数论中的 primorial 函数 $(k-1)\#$ 在数学上等价（即不超过 $k-1$ 的所有素数之积）。作者在素数 AP 最小参数搜索中对该概念进行了系统化应用，将其作为搜索算法的理论基础。

定义 2.2（验证范围）

本文的所有搜索和验证均在整数区间 $[1, N]$ 内进行，其中 $N = 2^{32} = 4,294,967,296$ 。该范围内的素数个数经计算为 $\pi(N) = 203,280,221$ 。本文所使用的素数库包含了所有这些素数。

3. 方法

本研究采用计算数论的研究方法，核心流程包括素数数据库准备、搜索算法执行、结果验证三个步骤。

3.1 素数数据库的准备

本研究所需的素数数据库通过以下方式获得：

- **生成：** 使用经过验证的筛法程序生成全部 2^{32} 以内的素数。
- **存储：** 为便于高效读取，素数被存储为 SQLite 数据库文件（primes.db），包含单表 primes(prime)，共 203,280,221 条记录。
- **加载：** 程序运行时将数据库索引加载至内存，以实现 $O(1)$ 复杂度的素数存在性查询。

3.2 搜索算法

本研究的核心是一个实用且直接的二阶段搜索策略。该策略的总体目标是：对于每一个项数 k ($k \geq 2$)，首先寻找最小公差的 AP，并在此基础上进一步寻找最小首项的 AP。

算法理论基础:

1. 对于一个 k 项的素数 AP, 令 P_n 为大于等于 k 的最小素数, 则其公差 d 必须是基础公差 $D_0 = \prod_{p < k} p$ 的整数倍, 即 $d = a \times D_0$, 其中 a 为正整数。
2. 首项 P 必须不小于 P_n 。

算法详细描述:

阶段一: 寻找最小公差 AP

1. 初始化: 对于给定的 k , 计算基础公差 D_0 。令倍数 $a = 1$ 。
2. 遍历首项: 从首项 $P = P_n$ 开始, 从小到大遍历素数数据库中的每一个素数。
3. 遍历公差倍数: 若未有结果, 则递增 a (从 1 开始), 计算候选公差 $d = a \times D_0$, 重复第 2 步。
4. 验证: 对于每一组 (P, d) , 验证序列 $P, P+d, \dots, P+(k-1)d$ 是否全部为素数。若序列末项 $P+(k-1)d$ 超出搜索范围 N , 则该候选 AP 自动排除。
5. 终止与记录: 一旦找到有效 AP, 该 AP 即为最小公差的 AP。记录此 AP 为 AP_min_d 。若此 AP 的首项 P 恰好等于 P_n , 则它同时也是最小首项的 AP, 搜索终止; 否则进入阶段二。

阶段二: 寻找最小首项 AP

6. 初始化: 记录阶段一找到的最小公差倍数 a_min 。
7. 搜索: 再次从首项 $P = P_n$ 开始遍历素数数据库, 对于每个候选首项 P , 遍历 $a = a_min+1, a_min+2, \dots$, 验证序列。若序列末项超出搜索范围 N , 则该候选 AP 自动排除。

8. 终止与记录：一旦找到一个有效 AP，则立即记录它为最小首项的 AP AP_{min_start} ，并终止搜索（因为首项 P 是从小到大遍历的，第一个找到的 AP 即为所求）。若遍历完所有可能仍未找到，则 AP_{min_d} 同时也是最小首项的 AP。

该算法结构简单、逻辑清晰，旨在确保搜索结果在给定范围内的完备性和唯一性，避免因算法复杂度过高而引入未发现的错误。

3.3 技术实现

- 编程：搜索策略由作者设计，具体的 Python 代码实现由 DeepSeek AI 辅助完成。作者将策略逻辑提供给 AI，由 AI 生成对应的代码，作者负责运行和验证。

- 硬件：计算任务在个人计算机上完成。

4. 结果与讨论

4.1 数据总览

本研究在 $N = 2^{32}$ 范围内，统一完成了对 $k=2$ 至 16 的素数等差数列的最小公差与最小首项搜索。所有结果均汇总于表 1。为便于分析，表 2 同时列出了各 k 下的计算耗时及相关参数。

表 1: $k=2$ 至 16 最小公差 AP 与最小首项 AP 汇总 (范围: 2^{32})

项数 k	最小公差 AP 首项 P	最小公差 AP 公差 d	最小首项 AP 首项 P	最小首项 AP 公差 d
2	2	1	2	1
3	3	2	3	2
4	5	6	5	6
5	5	6	5	6
6	7	30	7	30
7	7	150	7	150
8	199	210	11	1,210,230
9	199	210	11	32,671,170
10	199	210	11	224,494,620
11	60,858,179	2,310	157	203,032,830
12	3,415,447,36	6,930	157	203,032,83

项数 k	最小公差 AP 首项 P	最小公差 AP 公差 d	最小首项 AP 首项 P	最小首项 AP 公差 d
2	1			0
1	14,933,623	30,030	4,943	60,060
3				
1	1,575,825,24	90,090	146,141	54,444,390
4	1			
1	335,600,533	510,510	7,113,751	269,579,31
5				0
1	1,202,798,68	3,063,06	13,816,84	236,366,13
6	7	0	3	0

注：所有结果均通过二阶段搜索算法的 Phase 2 独立获得，每个 k 值独立搜索，确保完备性。 $k=12$ 的最小首项 AP ($P=157, d=203,032,830$) 与 $k=11$ 共享同一数列的前 12 项，此为搜索过程中 Phase 2 独立运行确认的结果。

表 2: 计算耗时与搜索空间参数

项数 k	耗时	基础公差 D_0	最小公差倍数 $m=d/D_0$	k 的素 数性
2	<0.01 秒	1	1	素数
3	<0.01 秒	2	1	素数
4	<0.01 秒	6	1	合数
5	<0.01 秒	6	1	素数
6	<0.01 秒	30	1	合数
7	7 时 59 分 3 秒	30	5	素数
8	<0.01 秒	210	1	合数
9	<0.01 秒	210	1	合数
10	<0.01 秒	210	1	合数
11	21 时 31 分 36 秒	210	11	素数
12	6 时 12 分 53 秒	2,310	3	合数
13	1 天 2 时 36 分	2,310	13	素数

项数 k	耗时	基础公差 D_0	最小公差倍数 $m=d/D_0$	k 的素 数性
	8 秒			
14	5 时 22 分 42 秒	30,030	3	合数
15	1 天 12 时 12 分 1 秒	30,030	17	合数
16	9 天 12 时 25 分 13 秒	30,030	102	合数

注： $k=12$ 的最小公差解首项较大，是由于对于 $a=1$ ($d=2310$) 和 $a=2$ ($d=4620$)，在 2^{32} 范围内不存在有效的 12 项 AP；直到 $a=3$ ($d=6930$) 才找到解。其最小首项解与 $k=11$ 共享同一数列（详见表 1 注释）。

4.2 典型序列示例

以下列出部分长度下的典型序列，以供核验。

$k=8$ 的最小首项 AP（首项=11，公差=1,210,230）：

11, 1210241, 2420471, 3630701, 4840931, 6051161, 7261391, 8471621

$k=11$ 的最小首项 AP（首项=157，公差=203,032,830）：

157, 203032987, 406065817, 609098647, 812131477, 1015164307,
1218197137, 1421229967, 1624262797, 1827295627, 2030328457

k=13 的最小首项 AP (首项=4,943, 公差=60,060) :

4943, 65003, 125063, 185123, 245183, 305243, 365303, 425363, 485423,
545483, 605543, 665603, 725663

k=16 的最小公差 AP (首项=1,202,798,687, 公差=3,063,060) :

1202798687, 1205861747, 1208924807, 1211987867, 1215050927,
1218113987, 1221177047, 1224240107, 1227303167, 1230366227,
1233429287, 1236492347, 1239555407, 1242618467, 1245681527,
1248744587

本文搜索的是“给定长度下最小参数”的 AP，而非“最长 CPAP”记录。两种研究路线的目标不同，本文提供的是在统一 2^{32} 范围内系统搜索的基准数据。

4.3 讨论

4.3.1 计算复杂度的结构特征

从表 2 可清晰观察到，算法的搜索耗时与项数 k 的数学性质呈现一定的关联。当 k 为素数时（尤其是 7、11、13），计算耗时往往高于其前后的合数项。例如， $k=7$ 耗时约 8 小时，而其前后的 $k=6$ 和 $k=8$ 均在瞬间完成； $k=11$ 耗时近一天， $k=13$ 耗时超过一天，而其间的合数项 $k=12$ 仅需 6 小时。

这一现象的经验根源在于：当 k 为素数时，最小公差倍数 m 往往需要增大才能找到有效解（如 $k=7$ 时 $m=5$ ， $k=11$ 时 $m=11$ ， $k=13$ 时 $m=13$ ），而 m 的增大直接导致搜索空间的扩大。需要指出的是，本实验数据中， $k=15$ （合

数, $m=17$) 和 $k=16$ (合数, $m=102$) 的 m 值也显著增大, 这表明 m 的增大不仅与 k 的素数性有关, 还受到搜索范围边界效应的影响——随着 k 增大, 基础公差 D_0 的阶梯式增长使搜索空间的维度本身就在扩大。因此, “素数项数导致 m 跃升”应被理解为在当前搜索范围 (2^{32}) 内的一种观察性倾向, 而非绝对规律。更大范围的数据将有助于进一步厘清这一现象。

4.3.2 最小公差解的收敛性观察

观察最小公差序列, 可归纳出以下现象:

- 对于 $k=8, 9, 10$, 均存在公差为基础公差 (倍数 $m=1$) 的解, 且该解均以 $P=199$ 起始。这表明在此范围内, 连续三个长度可共享同一个极小公差解。
- 对于 $k=11$, 最小公差倍数跃升至 $m=11$ (即公差 $d = 11 \times 210 = 2310 = 2 \times 3 \times 5 \times 7 \times 11$), 首项也大幅跃升, 直接例证了素数项数往往需要“升级”公差的倾向性。
- 对于 $k=16$, 最小公差倍数达到 $m=102$, 凸显了随着长度增加, 寻找极小公差解的难度急剧上升。

这些数据点构成了一套离散的基准样本, 为研究“在给定有限范围内, 长度为 k 的 AP 其最小公差倍数的上界”这一问题, 提供了精确的初始参考值。

4.3.3 范围与最小首项关系的初步探讨

与本文数据集 (DOI: 10.57760/sciencedb.44689) 中所记录的更小范围 (5×10^8) 内的搜索结果相比, 一个显著趋势是: 对于 $k \geq 11$, 最小首项的解均发生了显著下降。这定量地展示了扩大搜索范围对优化最小首项的巨大潜力, 也提示了最小首项可能对搜索上界具有更强的依赖性。

上述观察共同指向一个更深刻的问题: 最小首项、最小公差与搜索范围之间是否存在精确的数量关系? 这将是作者下一步理论研究的核心目标。本文所提供的系统性数据, 恰为该理论探索奠定了坚实的实证基础。

5. 结论与未来工作展望

本文在 2^{32} 的素数范围内，运用一种实用的二阶段搜索策略，系统性地完成了对长度 2 至 16 的素数等差数列最小公差解与最小首项解的搜索，获得了一套精确、可复现的基准数据集。基于统一的搜索框架和完整的运行日志，本文揭示了计算耗时与项数素数性质之间的阶跃式增长关系，并初步探讨了最小参数随范围扩大而下降的趋势。

本工作的主要贡献在于：提出了一种基于基础公差的系统搜索框架，并获得了精确的基准数据，为后续理论研究奠定了实证基础。所获数据及搜索代码均已完整存档，可供独立核验。

后续工作正围绕以下方向推进：

- **理论建模：** 基于本文及更大范围的数据，分析最小首项、最小公差与搜索上界之间的统计关系，尝试建立数学模型或提出新的猜想，以期从理论层面深化对格林-陶定理的理解。

- **算法扩展：** 探索脱离预计算素数库的素性判定方案，将搜索范围进一步推向 2^{64} 乃至更大整数域。

- **结构推广：** 将基础公差框架与搜索结果，推广至其他素数分布模式（如孪生素数等差数列）的研究中。

这些由算法显式构造出的数列，不仅为格林-陶定理提供了具体的实例，也为探索素数分布的深层规律提供了新的计算工具与实证方向。

参考文献

[1] Green, B., & Tao, T. (2008). The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167(2), 481–547. DOI: 10.4007/annals.2008.167.481.

[2] PrimeGrid. (n.d.). Prime arithmetic progressions. Retrieved August 2025, from <https://www.primegrid.com/>

[3] Caldwell, C. K. (n.d.). The Prime Pages. Retrieved August 2025, from <https://primes.utm.edu/>

[4] Andersen, J. K. (2008). The smallest known CPAP-k. Retrieved August 2025, from <https://primerecords.dk/cpap.htm>

[5] Ribenboim, P. (2004). The Little Book of Bigger Primes (2nd ed.). Springer-Verlag.

[6] Hardy, G. H., & Wright, E. M. (2008). An Introduction to the Theory of Numbers (6th ed.). Oxford University Press.

[7] 张益唐. (2014). Bounded gaps between primes. *Annals of Mathematics*, 179(3), 1121–1174. DOI: 10.4007/annals.2014.179.3.7.

[8] 王元. (1999). 王元论哥德巴赫猜想. 山东教育出版社.

[9] 潘承洞, 潘承彪. (1988). 素数定理的初等证明. 上海科学技术出版社.

[10] 曾飞戈. (2026). 素数等差数列的构造性猜想：基于基础公差的系统研究（预印本）. 国家预印本平台. DOI: 10.12383/202607200001V1.

数据可用性声明

本文的完整搜索结果及数据集已存放于中国科学数据银行，读者可通过以下 DOI 引用和获取：

- DOI: 10.57760/sciencedb.29083 (素数等差数列的最小参数计算：基于 2^{32} 范围的系统性搜索)

- DOI: 10.57760/sciencedb.44689 (素数等差数列最小参数数据集： 10^1 – 10^9 与 2^1 – 2^{32} 范围内的实例)

致谢

作者谨此向为本研究提供支持的各方致以诚挚的谢意：

感谢湖南人文科技学院数学系在学术交流方面给予的鼓励。

感谢朋友孙沛夫先生在计算硬件维护方面提供的可靠支持。

感谢曾翔医生、邓守荣医生、宋旭东店长及湖南长沙齿邦口腔连锁医院博林分院全体医护人员。他们精湛的医术和关怀，保障了本研究者写作期间的身体健康，为研究的顺利完成提供了重要条件。

本研究在核心的实现环节深度依赖公开 DeepSeek AI 提供的代码生成与文献调研辅助，这对于研究的完成至关重要。

本研究未接受任何外部资金资助，所有计算均在个人设备上完成。作者声明不存在任何利益冲突。

附录

附录 A: 数学符号与术语表

符号	含义	说明
	等差数列	
AP	(Arithmetic Progression)	本文特指素数等差数列
k	等差数列项数	$k \geq 2$
P	素数等差数列的首项	P 为素数
d	素数等差数列的公差	$d = m \times D_0, m \in \mathbb{Z}^+$
D_0	基础公差 (Base Difference)	$D_0 = \prod_{p < k} p$, p 遍历所有小于 k 的素数; 与 <i>primorial</i> 函数 $(k-1)\#$ 等价
$\pi(N)$	素数计数函数	不超过 N 的素数个数

附录 B: 完整搜索日志摘要

以下为关键节点的日志摘要，完整日志请参阅 DOI 存档。日志时间戳为程序运行时的本地时间。

[22:35:28] 开始加载数据库: primes.db

[02:54:30] 数据库加载成功! 耗时: 15541.84 秒

[02:54:30] 素数数量: 203,280,221

[02:54:30] 最大素数: 4294967291

[12:34:47] 开始搜索, 最大项数=100

[12:34:47] 数据范围: 4294967296

[12:34:48] 项数=6 结果 (耗时: 0.00 秒): 最小公差(30) [7,37,...], 最小首项
(30) [7,37,...]

[20:33:51] 项数=7 结果 (耗时: 7 时 59 分 3.11 秒): 最小公差(150) [7,157,...],
最小首项(150) [7,157,...]

[20:33:51] 项数=8 结果 (耗时: 0.00 秒): 最小公差(210) [199,409,...], 最小
首项(1210230) [11,1210241,...]

...

[01:32:39] 项数=16 结果 (耗时: 9 天 12 时 25 分 12.97 秒): 最小公差
(3063060) [1202798687,...], 最小首项(236366130) [13816843,...]

全部原始日志已存放于科学数据银行 (ScienceDB), 读者可通过

DOI: <https://doi.org/10.57760/sciencedb.29083> 引用和获取。