

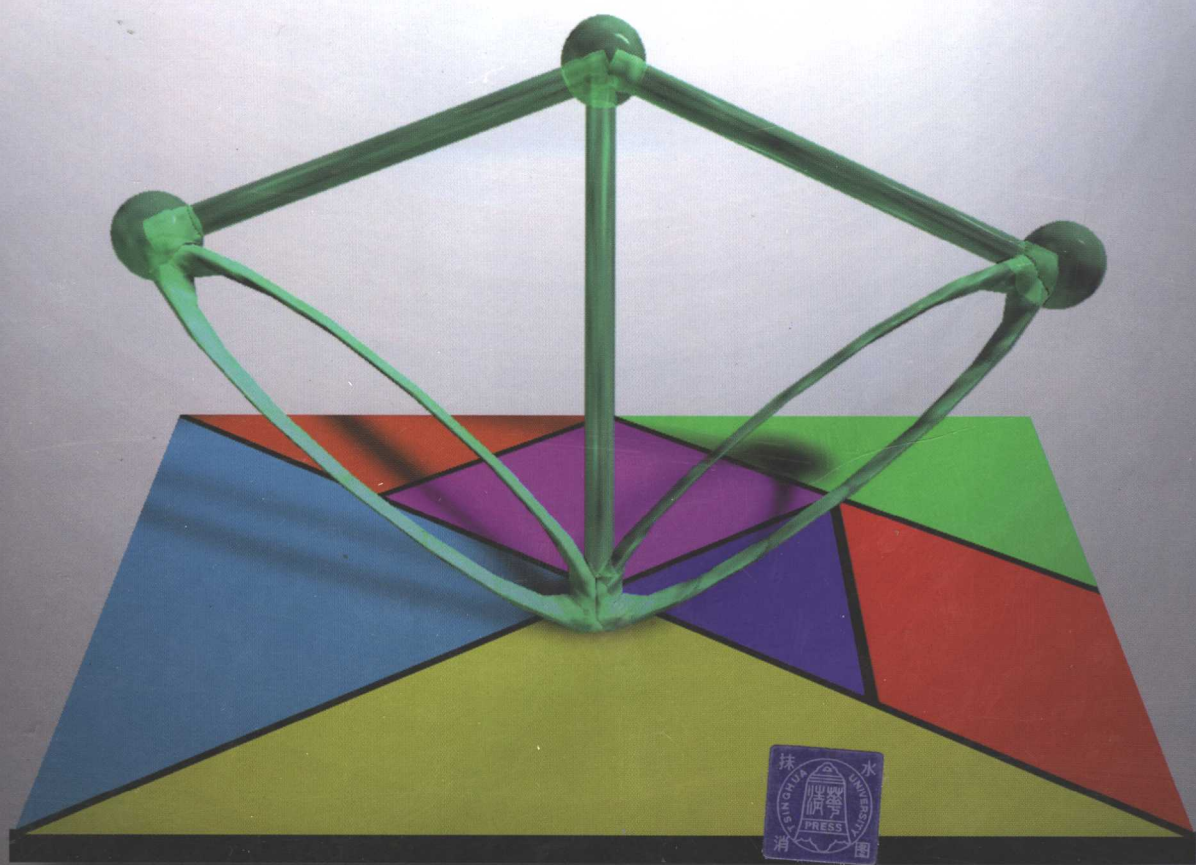
计算机科学组合学丛书

计算机算法导引

——设计与分析

(第2版)

卢开澄 编著



清华大学出版社

计算机科学组合学丛书

计算机算法导引

——设计与分析(第2版)

卢开澄 编著

清华大学出版社
北 京

内 容 简 介

本书为《计算机算法导引——设计与分析》的第2版。书中内容分3部分:第1部分是基本算法,按方法论区分,包含优先策略与分治策略、动态规划、概率算法、并行算法、搜索法、数据结构等;第2部分是若干专题,包括排序算法、计算几何及计算数论、线性规划;第3部分是复杂性理论与智能型算法,其中,智能型算法主要介绍了遗传算法和模拟退火算法。

本书可作为计算机系本科学生及研究生教材,数学系师生和科研工作者也可将其作为参考书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

计算机算法导引——设计与分析/卢开澄编著. —2版. —北京:清华大学出版社,2006.1

(计算机科学组合学丛书)

ISBN 7-302-11501-X

I. 计… II. 卢… III. ①电子计算机—算法设计—高等学校—教材 ②电子计算机—算法分析—高等学校—教材 IV. TP301.6

中国版本图书馆 CIP 数据核字(2005)第 087654 号

出 版 者: 清华大学出版社
<http://www.tup.com.cn> 地 址: 北京清华大学学研大厦
社 总 机: 010-62770175 邮 编: 100084
客 户 服 务: 010-62776969

组稿编辑: 张 民

文稿编辑: 孙建春

封面设计: 卢开澄 傅瑞学

印 刷 者: 北京鑫丰华彩印有限公司

装 订 者: 三河市新茂装订有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 26.75 字数: 624 千字

版 次: 2006 年 1 月第 2 版 2006 年 1 月第 1 次印刷

书 号: ISBN 7-302-11501-X/TP·7549

印 数: 1~3000

定 价: 38.00 元

计算机科学组合学丛书序

电子计算机的出现是 20 世纪的大事,它改变了我们这个世界的面貌。可以毫不夸张地说,它的影响遍及世界的所有角落,几乎无处不感觉到它的存在。数学更不例外。严格地说,电子计算机本身就是近代数学的辉煌成就。将计算机与数学割裂开来,既不合理也不可能。组合学也就是在计算机科学蓬勃发展的刺激下面崛起的,从而成为近若干年来最活跃的数学分支。它研究的问题有的可追溯到欧拉和哈密尔顿等 18 世纪的数学家,但它成为一新的分支还是近若干年的事。它从与计算机科学相结合中获得了广阔的发展空间,从而也为计算机科学奠定了理论基础。

什么是计算机科学?有的学者将它定义为研究算法的一门学科。研究算法无疑是计算机科学的重要领域,也是本丛书的核心内容,贯穿始终。组合学家在 20 世纪 70 年代初建立的算法复杂性的 NP 理论,至今仍然令无数计算机科学工作者与数学工作者为之折腰。

计算机科学里的组合学内容十分广泛。本丛书涉及组合分析、图论、组合算法、近代密码学、组合优化、编码理论及算法复杂性等七部分。

组合分析是算法的理论基础。组合分析之与组合算法犹如数学分析之与计算数学,众所周知,前者是后者的理论根基。

图论原本是组合数学这个“家族”的主要成员,只因它已成长壮大,故自立门户独立出去。

算法复杂性的 NP 理论是近 30 年的一大成就。研究表明对于一类叫做 NPC 类的困难问题,至今都不存在有效算法,但它们难度相当,只要其中任何一个找到多项式解法,则全体都获得解决;或证明它们根本不存在有效办法。不论是前者还是后者都还看不见露到海平面上的桅杆塔,它吸引了众多的有志之士。密码学是其中十分引人入胜的分支。如若设计好的密码,对它的破译等价于某一 NPC 类困难问题,无疑这样的密码将是牢不可破的。

在计算机网络深入普及的信息时代,信息本身就是时间,就是财富。信息的传输通过是脆弱的公共信道,信息储存于“不设防”的计算机系统中,如何保护信息的安全使之不被窃取及不至于被篡改或破坏,已成为当今被普遍关注的重大问题。密码是有效而且可行的办法。在计算机网络的刺激下,近代密码学便在算法复杂性理论的基础上建立起来了。密码作为一种技术,自从人类有了战争,不久便有了它。但作为一门学科则是近 20 多年的事。甚至于它已成为其他学科的基础。密码也从此走出“军营”,进入百姓家。

实际中的“优化”问题是大量的,半个多世纪以来它曾经几度辉煌。近来在计算机科学的影响下,又出现了若干闪光点,十分耀眼,引人注目。

实际上密码也是一种编码。如果说密码学研究的编码是保证通信的保密与安全,则编码理论研究的是通信中如何纠错与检错。计算机纠错码是既实用,理论上又饶有趣味的分支。

本丛书是作者在清华大学计算机科学与技术系长期工作的总结。它不是一部“长篇”记述,而是互相关联又彼此相对独立,因此难免有少量交叉。它们涉及的面如此之广,囿于作者的水平,缺点和错误在所难免,敬请读者不吝指正。谢谢。

前 言

算法设计与分析是一个与计算机科学紧密相关的数学分支,它还很年轻,远未定型。由于它所包含的内容浩如烟海,故不一而足,描述算法的方式也迥异,不像普天之下的微积分学那样都是相似形。

众多的算法彼此独立,读起来往往使人有只见树木不见森林之感。若依方法论来区分,有的又无法概括到。本书为《计算机算法导引——设计与分析》的第2版,所讨论的内容分3部分:第1部分是基本方法,按方法论区分,包含优先策略与分治策略、动态规划、概率算法、并行算法、搜索法等;第2部分是若干专题,有排序算法、计算几何及计算数论、线性规划等;第3部分是复杂性理论与智能型算法等,智能型算法主要介绍遗传算法和模拟退火算法。

囿于个人的水平,疏漏和缺点在所难免,望读者多多指正。

卢开澄
2004年9月

目 录

第 1 部分 基本算法

第 1 章 数学准备	3
1.1 母函数	3
1.2 递推关系	5
1.3 Fibonacci 数列	9
1.3.1 Fibonacci 数列是典型的递推关系	9
1.3.2 问题的解	10
1.4 线性常系数递推关系举例	11
1.5 其他类型的递推关系举例	13
习题	18
第 2 章 优先策略与分治策略	20
2.1 优先策略:求最短树的 Kruskal 算法	20
2.2 求最短树的 Prim 算法	22
2.3 求最短路径的 Dijkstra 算法	24
2.4 文件存储问题	25
2.5 有期限的任务安排问题	27
2.6 数据压缩和 Huffman 树	29
2.7 分治策略与二分查找	33
2.8 整数乘法	34
2.9 矩阵乘积的 Strassen 算法	35
2.10 矩阵乘积的 Winograd 算法	38
2.11 布尔矩阵乘积的分段预处理方法	39
2.12 归并排序法	41
2.13 快速排序法	43
2.14 求序列中的第 k 个元素	48
习题	50
第 3 章 动态规划	53
3.1 最短路径问题	53
3.2 最佳原理	55

3.3	流动推销员问题	65
3.3.1	算法及例题	65
3.3.2	复杂性估计	67
3.4	矩阵链乘问题	68
3.5	最长公共子序列	70
3.6	图的任意两点间的最短距离	72
3.7	同顺序流水作业的任务安排问题	74
3.8	可靠性问题	76
3.9	最佳二分树	78
3.9.1	二分树的一些性质	78
3.9.2	最佳二分树的构成	81
	习题	88
第4章	概率算法	91
4.1	生日问题	91
4.2	概率算法举例	92
4.3	随机数的产生器	94
4.3.1	线性同余式法	94
4.3.2	离散对数法	95
4.3.3	BBS 法	96
4.3.4	素数法	96
4.4	素数的概率判定算法	96
4.4.1	关于素数的若干定理	96
4.4.2	Fermat 数	98
4.4.3	Miller-Rabin 的素数概率测试法	98
4.5	定理证明的数学准备	99
4.5.1	数论的基本知识	99
4.5.2	群论的基本知识	101
4.5.3	中国剩余定理	104
4.5.4	$x^n \equiv 1 \pmod{p}$ 的解	105
4.6	定理 A 的证明	107
4.7	定理 B 的证明	109
	习题	111
第5章	并行算法	113
5.1	并行计算机和并行算法的基本概念	113

5.2	递推关系的并行计算	116
5.3	图的并行算法举例	118
5.4	矩阵乘积的并行计算	121
5.5	分布计算	124
5.6	快速傅里叶变换	125
5.6.1	FFT 问题的背景	125
5.6.2	预备定理	125
5.6.3	快速算法	127
5.6.4	傅里叶逆变换	133
5.6.5	计算结果的重排	133
5.6.6	复杂性估计	134
5.7	卷积及其应用	136
5.7.1	卷积	136
5.7.2	多项式的一种快速乘法	137
5.8	数论变换	138
5.9	排序网络	140
5.9.1	引论	141
5.9.2	0-1 原理	142
5.9.3	B_n 型网络	143
5.9.4	M_n 归并网络	145
5.10	Batcher 奇偶归并网络	146
5.11	脉动阵列的并行处理	148
5.11.1	矩阵和向量乘法的并行处理	148
5.11.2	矩阵乘法的并行处理	150
5.11.3	带状矩阵的并行乘法	151
	习题	153
第 6 章	搜索法	154
6.1	引论	154
6.2	DFS 搜索法	155
6.3	无向图的 DFS 算法	157
6.4	有向图的 DFS 算法	160
6.5	互通块问题	163
6.6	强连通块问题	164
6.7	BFS 算法	168
6.8	拓扑排序	169

6.9 min-max 搜索法	170
6.10 流动推销员问题的分支定界法	171
6.11 同顺序加工任务安排问题	175
习题	177
第7章 数据结构	179
7.1 “堆”和“堆集排序法”	179
7.1.1 堆	179
7.1.2 堆集排序法	182
7.1.3 优先级队和二进制堆	183
7.2 2-3 树	186
7.3 2-3-4 树	189
7.4 红黑树	191
7.4.1 RB 树性质	191
7.4.2 插入	192
7.4.3 删除	195
7.5 B-树	197
7.5.1 B-树性质	197
7.5.2 B-树的插入	199
7.5.3 B-树的删除	201
7.6 关于高度的均衡树	203
7.6.1 AVL 树——关于高度均衡的二分树	203
7.6.2 关于高度均衡的二分树的插入和删除	207
7.7 哈希表	210
7.7.1 什么是哈希表	210
7.7.2 哈希函数的构造方法	211
7.7.3 解决冲突的方法	212
7.7.4 哈希算法的分析(线性探测法分析)	214
7.7.5 二重哈希法	216
习题	217

第2部分 若干专题

第8章 排序算法	221
8.1 排序	221
8.2 下界估计	221
8.3 二分插入排序法	224

8.4	下溢排序法	226
8.5	Ford-Johnson 归并插入排序法	229
8.5.1	算法的非形式化描述	229
8.5.2	一般情形的讨论	230
8.5.3	算法分析	231
8.6	外存排序	233
8.6.1	外存归并排序法	233
8.6.2	三条带的外存归并排序法	235
8.6.3	阶式归并法	238
第9章	计算几何及计算数论	240
9.1	关于线段问题	240
9.2	凸包问题与 Voronoi 问题	244
9.2.1	凸包问题	244
9.2.2	Voronoi 图	247
9.2.3	Voronoi 图的构造法	248
9.2.4	Voronoi 图的应用简介	249
9.2.5	Voronoi 图的拓广	249
9.3	串匹配	250
9.3.1	搜索法	250
9.3.2	KMP 算法	251
9.3.3	BM 算法	253
9.3.4	RK 算法	254
9.4	数论的算法问题	255
9.4.1	求最大公因数	255
9.4.2	因数分解之一: Pollard ρ 法	257
9.4.3	Dixon 随机平方因数分解法	260
9.4.4	椭圆曲线因数分解法	261
9.5	大数模幂运算	270
9.6	$N \bmod M$	273
9.6.1	Barrett 归约	273
9.6.2	模乘算法	274
9.6.3	Montgomery 模幂运算	277
9.6.4	n 是偶数的情况	280
第10章	线性规划	282
10.1	问题的提出	282
10.2	线性规划的几何意义	284

10.3	单纯形法理论基础	287
10.4	单纯形法及单纯形表格	291
10.5	改善的单纯形法表格	297
10.6	对偶原理	300
10.6.1	对偶概念	300
10.6.2	对偶问题的经济意义	301
10.6.3	对偶问题的性质	302
10.6.4	对偶定理	303
10.6.5	影子价格	304
10.7	对偶单纯形法	307
10.8	退化情况及其他	311
10.8.1	退化情况	312
10.8.2	退化情况的循环不已与 Bland 法则	313
10.9	Dantzig-Wolfe 分解算法	314
10.10	整数规划	322
10.10.1	问题的提出	322
10.10.2	0-1 规划和 DFS 搜索法	324
10.10.3	分支定界法	333
10.11	Klee 与 Minty 举例	335

第 3 部分 复杂性理论与智能型算法

第 11 章	算法复杂性理论	341
11.1	图灵机	341
11.2	图灵机和算法	345
11.3	k 条带的图灵机	347
11.4	非确定型图灵机	348
11.5	停机问题	349
11.6	布尔表达式	351
11.7	布尔变量和网络	353
11.8	问题的转换	354
11.9	Cook 定理	356
11.10	几个 NP 完备的例子	360
11.11	复杂度类	368
11.12	近似解法	370
11.12.1	任务安排的近似算法	370
11.12.2	装箱问题的近似算法	374
11.12.3	流动推销员问题的近似算法	376

11.12.4	顶点覆盖问题的近似算法.....	384
11.13	近代密码学简介	385
11.13.1	密码概念.....	385
11.13.2	背包公钥密码.....	388
11.13.3	RSA 公钥密码	389
第 12 章	智能型算法	391
12.1	遗传算法.....	391
12.2	什么是遗传算法.....	398
12.3	TSP 问题	398
12.3.1	编码.....	398
12.3.2	初始“种群”的生成.....	398
12.3.3	杂交.....	400
12.3.4	变异算术.....	403
12.3.5	模式定理.....	404
12.4	模拟退火算法简介.....	405
	习题.....	412

第 1 部分

基本算法

第1章 数学准备

“计算机算法”顾名思义讨论的是一类与计算机有关的数学问题,它需要许多数学基础的支持,现在仅就递推关系与母函数作一些介绍。

首先介绍一些常用的符号和术语,例如 R 表示实数集合; R^+ 表示正的实数集合。

又如 $N = \{0, 1, 2, \dots\}$ 即包含 0 在内的正整数; $N^+ = \{1, 2, \dots\}$ 即正整数。

还经常用到的符号如 $\triangleq$ 表示“定义”, $\exists$ 表示存在,如 $\exists n_0 \in N$; $\forall$ 为所有的,如 $\forall n \geq n_0$,即对所有 $\geq n_0$ 的数 n 。

令 f 和 g 是由映射 N 到 R 的两个函数。请读者注意下面三个定义的区别。

定义 令 $f: N \rightarrow R^+$,

$O(f) \triangleq \{g: N \rightarrow R^+ \mid \exists c \in R^+, \exists n_0 \in N, \forall n \geq n_0, \text{使 } g(n) \leq cf(n)\}$

也就是 $O(f)$ 是函数 $g: N \rightarrow R^+$ 的集合,它以 $cf(n)$ 为上界。亦即若存在 $c \in R$, 且 $\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = c$, 则 $g \in O(f)$ 。基于上述概念,若 $f(n) = n^3, g(n) = n^2$, 则 $g \in O(f)$, 但 $f \notin O(g)$ 。

定义 令 $f: N \rightarrow R^+$,

$\Omega(f) \triangleq \{g: N \rightarrow R \mid \exists n_0 \in N, \exists c \in R^+, \forall n \geq n_0, \text{使 } cf(n) \leq g(n)\}$,

若 $\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = \infty$ 或 $\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = c > 0$, 则 $g \in \Omega(f)$ 。

定义 令 $f: N \rightarrow R$,

$\Theta(f) = O(f) \cap \Omega(f)$,

若 $\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = c, c \in R^+$, 则 $g \in \Theta(f)$ 。

其中 $c \in R^+$ 表示 $c \neq 0, c \neq \infty$, $g \in \Theta(f)$ 意味着 g 和 f 同阶。

O, Ω, Θ 有以下的性质(假定 $f, g, h: N \rightarrow R$):

(1) 若 $f \in O(g), g \in O(h)$, 则 $f \in O(h)$;

(2) $f \in O(g) \Leftrightarrow g \in \Omega(f)$;

(3) 若 $f \in \Theta(g)$, 则 $g \in \Theta(f)$;

(4) $O(f+g) = O(\max\{f, g\})$ 。

1.1 母函数

递推关系是计数的一个强有力工具,特别在作算法分析时是必需的。递推关系的求解主要是利用母函数。当然母函数尚有其他用处,但这里主要介绍在解递推关系上的应用。例如

$$\begin{aligned} & (1+a_1x)(1+a_2x)\cdots(1+a_nx) \\ &= 1 + (a_1+a_2+\cdots+a_n)x + (a_1a_2+a_1a_3+\cdots+a_{n-1}a_n)x^2 + \cdots + a_1a_2\cdots a_na^n \quad (1.1) \end{aligned}$$

x^2 项的系数 $a_1a_2+a_1a_3+\cdots+a_{n-1}a_n$ 中所有的项包含了从 n 个元素 $a_1, a_2, \dots, a_n$ 中

取两个组合的全体;同理, x^3 项系数 $a_1a_2a_3+a_1a_2a_4+\cdots+a_{n-2}a_{n-1}a_n$ 包含了从 n 个元素 $a_1, a_2, \cdots, a_n$ 中取 3 个组合的全体,以此类推。

若令 $a_1=a_2=\cdots=a_n=1$ 在(1.1)式中 x^2 项系数 $a_1a_2+a_1a_3+\cdots+a_{n-1}a_n$ 中每一个组合有 1 个贡献,其他各项以此类推,故有

$$(1+x)^n=1+C(n,1)x+C(n,2)x^2+\cdots+C(n,n)x^n \quad (1.2)$$

另一方面,

$$(1+x)^m(1+x)^n=(1+x)^{m+n}$$

$$\begin{aligned} \text{所以} \quad & [C(m,0)+C(m,1)x+\cdots+C(m,m)x^m][C(n,0)+C(n,1)x+\cdots+C(n,n)x^n] \\ & = [C(m+n,0)+C(m+n,1)x+\cdots+C(m+n,m+n)x^{m+n}] \end{aligned}$$

比较等号两端 x 项对应系数,可得一等式

$$C(m+n,r)=C(m,0)C(n,r)+C(m,1)C(n,r-1)+\cdots+C(m,r)C(n,0)$$

这里不过给出另一种比较简单的证明。

同样,对于 $(1+x)^n\left(1+\frac{1}{x}\right)^m$, (设 $n \geq m$), 用类似方法可得等式

$$C(m+n,m)=C(n,0)C(m,0)+C(n,1)C(m,1)+\cdots+C(n,m)C(m,m)$$

方法如下。

$$(1+x)^n\left(1+\frac{1}{x}\right)^m=x^{-m}(1+x)^{m+n} \quad (1.3)$$

$$\begin{aligned} \text{所以} \quad & [C(n,0)+C(n,1)x+\cdots+C(n,n)x^n][C(m,0)+C(m,1)x^{-1}+\cdots+C(m,m)x^{-m}] \\ & = x^{-m}[C(m+n,0)+C(m+n,1)x+C(m+n,2)x^2+\cdots+C(m+n,m+n)x^{m+n}] \end{aligned}$$

比较等号两端的常数项,即得公式(1.3)。

又如等式

$$(1+x)^m=C(n,0)+C(n,1)x+C(n,2)x^2+\cdots+C(n,n)x^n$$

令 $x=1$, 可得

$$C(n,0)+C(n,1)+\cdots+C(n,n)=2^n \quad (1.4)$$

(1.2)式等号的两端对 x 求导可得

$$n(1+x)^{n-1}=C(n,1)+2C(n,2)x+3C(n,3)x^2+\cdots+nC(n,n)x^{n-1} \quad (1.5)$$

等式(1.5)两端令 $x=1$, 得

$$C(n,1)+2C(n,2)+3C(n,3)+\cdots+nC(n,n)=n2^{n-1} \quad (1.6)$$

类似的办法可以得到

$$\begin{aligned} & C(n,1)x+2C(n,2)x^2+3C(n,3)x^3+\cdots+nC(n,n)x^n \\ & = nx(1+x)^{n-1} \end{aligned}$$

所以

$$\begin{aligned} & C(n,1)+2^2C(n,2)+3^2C(n,3)+\cdots+n^2C(n,n) \\ & = n(n+1)2^{n-2} \end{aligned} \quad (1.7)$$

还可以类似地推出一些等式,但通过上面一些例子已可见函数 $(1+x)^n$ 在研究序列 $C(n,0), C(n,1), \cdots, C(n,n)$ 的关系时所起的作用,对于其他序列也有同样结果,现引进母函数概念如下。

定义: 对于序列 $a_0, a_1, a_2, \cdots$, 构造一函数

$$G(x)=a_0+a_1x+a_2x^2+\cdots$$