数 学 年 刊
1 (2) 1980

一类二元序列相关性的估计

戴宗铎 倪录群 杨君辉 陈文德
(中国科学院数学研究所)

§1. 引 言

精密机床使用光栅控制是一项有效的新技术. 其核心问题之一是所谓零光栅的刻线分布设计. 它提出了一个数学问题可叙述如下. 设 n, m 是正整数, $n>m$; 我们称序列

$$\underline{a}=(a_1,\ a_2,\ \cdots,\ a_n)$$

为(0, 1)序列, 其中 $a_i=0$ 或 1, $\sum\limits_{i=1}^{n}a_i=m$. 序列 $\underline{a}$ 的自相关函数定义为

$$\mathscr{C}(\tau)=\sum_{i=1}^{n-\tau}a_ia_{i+\tau},\quad 1\leqslant\tau\leqslant n-1,$$

$$P(\underline{a})=\max_{1\leqslant\tau\leqslant n-1}\mathscr{C}(\tau),$$

$P(\underline{a})$ 称为序列 $\underline{a}$ 的相关值. 零光栅刻线分布的最优化设计要求对给定的 n, m, 构造一个(0, 1)序列, 使 $P(\underline{a})$ 达到最小.

某研究所曾向我们提出一个更具体的问题: 要求寻找一个 $n=68$, $m=30$ 的序列 $\underline{a}$, 使 $P(\underline{a})$ 小于等于 10. 但通过对 $P(\underline{a})$ 下界的数学分析, 我们证明了这样的序列并不存在. 同时, 使用计算机找到了一组序列, 其 $P(\underline{a})=11$. 因此, 这组序列就 $n=68$, $m=30$ 而言, $P(\underline{a})$ 达到了最小, 所以它们已为最优.

本文将给出一般有限二元序列 $\underline{a}(n,\ m)$ 的相关值 $P(\underline{a})$ 下界的一个估计, 从而为高指标光栅的设计提供了一个判定标准. 同时简要地讨论用计算机寻找 $\underline{a}$, 使 $P(\underline{a})$ 尽量小的方法.

§2. $P(\underline{a})$ 的估计

设 $\underline{a}=(a_1,\ a_2,\ \cdots,\ a_n)$, 其中 $a_i=0$ 或 1, $\sum\limits_{i=1}^{n}a_i=m$. 如果序列 $\underline{a}$ 的 $b_1,\ b_2,\ \cdots,\ b_m$ 的位置是 1, $1\leqslant b_j\leqslant n$, $1\leqslant j\leqslant m$, 那么序列 $\underline{a}$ 和序列 $I=(b_1,\ b_2,\ \cdots,\ b_m)$, $1\leqslant b_j\leqslant n$, $b_i\neq b_j$ $(1\leqslant i\neq j\leqslant m)$, 一一对应. 又记

$$d_{t,s}=b_s-b_t,\quad 1\leqslant t\leqslant s\leqslant m.$$

令

$$\mathscr{D}=\{d_{t,s}\,|\,1\leqslant t<s\leqslant m\},$$

$$\mathscr{C}_\tau=\{d_{t,s}\,|\,d_{t,s}\in\mathscr{D},\ d_{t,s}=\tau\}.$$

本文1978年6月7日收到, 1979年12月17日修改.

于是
$$\mathscr{D}=\bigcup_{\tau=1}^{n-1}\mathscr{C}_\tau.$$

如用符号 $|A|$ 表示有限集合 A 的元素的个数，我们有
$$|\mathscr{D}|=m(m-1)/2, \tag{1}$$
$$\mathscr{C}(\tau)=\sum_{i=1}^{n-\tau}a_i a_{i+\tau}=\sum_{i,\,i+\tau\in I}a_i a_{i+\tau}=\sum_{b_s-b_t=\tau}a_{b_t}a_{b_s}=\sum_{d_{t,s}=\tau}1=|\mathscr{C}_\tau|,$$
即
$$|\mathscr{C}_\tau|=\mathscr{C}(\tau). \tag{2}$$

定理 设 $\underline{a}=(a_1, a_2, \cdots, a_n)$, $a_i=0$ 或 1, $\sum\limits_{i=1}^{n}a_i=m$, 记 $P(\underline{a})$ 为 P, 令
$$s_k=km-k(k+1)/2,\quad 1\leqslant k\leqslant m-1,$$
$$s_k=P\cdot q_k+r_k,\qquad 0\leqslant r_k<P.$$
$$\Sigma_1(m, k, P)=(s_k+r_k)(q_k+1),$$
$$\Sigma_2(n, k)=k(k+1)n-k(k+1)(2k+1)/3.$$
那么下面结论成立:

(I)
$$P\geqslant m(m-1)/2(n-1) \tag{3}$$
(II)
$$\Sigma_1(m, k, P)\leqslant\Sigma_2(n, k) \tag{4}$$
当 $k>\left[\dfrac{m}{2}\right]$ 时，恒有不等号成立.

(III)
$$P\geqslant\max_{1\leqslant k\leqslant m-1}\left\{\frac{s_k^2}{\dfrac{1}{3}k(k+1)(3n-2k-1)-s_k}\right\} \tag{5}$$
若上式右边在 $k>\left[\dfrac{m}{2}\right]$ 处取极大，则恒有不等号成立.

证

(I) 由 P 的定义，以及(2)，可知
$$|\mathscr{C}_\tau|\leqslant\max_{1\leqslant\tau\leqslant n-1}\mathscr{C}(\tau)=P,$$
所以
$$|\mathscr{D}|=\left|\bigcup_{\tau=1}^{n-1}\mathscr{C}_\tau\right|\leqslant(n-1)P,$$
应用(1)，即得
$$P\geqslant m(m-1)/2(n-1).$$

(II) 令
$$\mathscr{D}_k=\{d_{t,s}\,|\,s-t=k,\ d_{t,s}\in\mathscr{D}\},$$
如果记
$$d_{t,t+1}=d_t,$$
那么
$$\mathscr{D}_1=\{d_t,\ t=1, 2, \cdots, m-1\},$$
$$\mathscr{D}_2=\{d_{i,i+2}=d_i+d_{i+1},\ i=1, 2, \cdots, m-2\},$$
一般
$$\mathscr{D}_k=\{d_{i,i+k}=d_i+d_{i+1}+\cdots+d_{i+k-1},\ i=1, 2, \cdots, m-k\}.$$
则有
$$|\mathscr{D}_k|=m-k,$$

$$\left|\bigcup_{i=1}^{k}\mathscr{D}_i\right|=km-k(k+1)/2,$$

即

$$s_k=\left|\bigcup_{i=1}^{k}\mathscr{D}_i\right|.$$

为证明(4)式成立，只须证明

$$\frac{1}{2}\Sigma_1(m,\ k,\ P)\leqslant\sum_{x\in\bigcup_{i=1}^{k}\mathscr{D}_i}x\leqslant\frac{1}{2}\Sigma_2(n,\ k).\tag{6}$$

先证

$$\frac{1}{2}\Sigma_1(m,\ k,\ P)\leqslant\sum_{x\in\bigcup_{i=1}^{k}\mathscr{D}_i}x^*.\tag{7}$$

我们将 $\bigcup_{j=1}^{k}\mathscr{D}_j$ 中 s_k 个元素按大小次序排列为

$$e_1\leqslant e_2\leqslant\cdots\leqslant e_{s_k},$$

其中取值为 τ 的元素皆属于 $\mathscr{C}_\tau$，因此 e_i 中取值为 τ 的元素的个数 l_τ 满足

$$l_\tau\leqslant|\mathscr{C}_\tau|\leqslant P;$$

$$(e_1,\ e_2,\ \cdots,\ e_{s_k})=(\underbrace{1,\ \cdots,\ 1}_{l_1\text{个}},\ \underbrace{2,\ \cdots,\ 2}_{l_2\text{个}},\ \cdots,\ \underbrace{\tau,\ \cdots,\ \tau}_{l_\tau\text{个}}\cdots)$$

$$\sum_{\tau=1}^{n-1}l_\tau=s_k.$$

由于 $s_k=P\cdot q_k+r_k$，我们可以构造一个序列

$$(f_1,\ f_2,\ \cdots,\ f_{s_k})=(\underbrace{1,\ \cdots,\ 1}_{P\text{个}},\ \underbrace{2,\ \cdots,\ 2}_{P\text{个}},\ \cdots,\ \underbrace{q_k,\ \cdots,\ q_k}_{P\text{个}},\ \underbrace{q_k+1,\ \cdots,\ q_k+1}_{r_k\text{个}}),$$

因为 $l_\tau\leqslant P$，所以 $e_i\geqslant f_i$，于是

$$\sum_{x\in\bigcup_{i=1}^{k}\mathscr{D}_i}x=\sum_{i=1}^{s_k}e_i\geqslant\sum_{i=1}^{s_k}f_i=P\sum_{i=1}^{q_k}i+r_k(q_k+1)$$

$$=P\,\frac{q_k(q_k+1)}{2}+r_k\,(q_k+1)=\frac{1}{2}(s_k+r_k)(q_k+1).$$

于是(7)式得证*.

其次证明

$$\sum_{x\in\bigcup_{i=1}^{k}\mathscr{D}_i}x\leqslant\frac{1}{2}\Sigma_2(n,\ k),\tag{8}$$

则有

$$\sum_{x\in\mathscr{D}_1}x=\sum_{t=1}^{m-1}d_t\leqslant n-1,$$

$$\sum_{x\in\mathscr{D}_2}x=(d_1+d_2)+(d_2+d_3)+\cdots+(d_{m-2}+d_{m-1})$$

$$=2\sum_{i=1}^{m-1}d_i-(d_1+d_{m-1})\leqslant 2(n-1)-2.$$

---

* (7) 式等号成立，当且仅当对每个 i: $e_i=f_i$，即

$$\bigcup_{i=1}^{k}\mathscr{D}_i=(\underbrace{1,\ \cdots,\ 1}_{P\text{个}},\ \underbrace{2,\ \cdots,\ 2}_{P\text{个}},\ \cdots,\ \underbrace{q_k,\ \cdots,\ q_k}_{P\text{个}},\ \underbrace{q_k+1,\ \cdots,\ q_k+1}_{r_k\text{个}}),$$

此条件记为 $A(k)$。

一般，当 $k\leqslant\left[\frac{m}{2}\right]$ 时，

$$\sum_{x\in\mathscr{D}_k}x=(d_1+d_2+\cdots+d_k)+(d_2+\cdots+d_{k+1})+\cdots+(d_{m-k}+\cdots+d_{m-1})$$

$$=k\sum_{i=1}^{m-1}d_i-(k-1)(d_1+d_{m-1})-(k-2)(d_2+d_{m-2})-\cdots-(d_{k-1}+d_{m+1-k})$$

$$\leqslant k(n-1)-2(1+2+\cdots+(k-1))=kn-k^2{}^{**}.$$

而当 $k>\left[\frac{m}{2}\right]$ 时，这时一定有 $m-k<k$，

$$\sum_{x\in\mathscr{D}_k}x=(d_1+\cdots+d_k)+(d_2+\cdots+d_{k+1})+\cdots+(d_{m-k}+\cdots+d_{m-1})$$

$$=(m-k)\sum_{i=1}^{m-1}d_i-(m-k-1)(d_1+d_{m-1})$$

$$-(m-k-2)(d_2+d_{m-2})-\cdots-(d_{m-k-1}+d_{k+1})$$

$$\leqslant(m-k)(n-1)-2(1+2+\cdots+(m-k-1))=(m-k)n-(m-k)^2.$$

但因为 $m-k<k$，易知有 $(m-k)n-(m-k)^2<kn-k^2$，于是当 $k>\left[\frac{m}{2}\right]$ 时，

$$\sum_{x\in\mathscr{D}_k}x<kn-k^2.$$

因此
$$\sum_{x\in\bigcup_{i=1}^{k}\mathscr{D}_i}x\leqslant\sum_{i=1}^{k}\sum_{x\in\mathscr{D}_i}x\leqslant\sum_{i=1}^{k}(in-i^2)=\frac{k(k+1)}{2}n-\frac{k(k+1)(2k+1)}{6}$$

这正是(8)式***. 由(7)(8)立即可得(6)，即定理结论(II)成立.

(III) 设 $s=Pq+r$，有

$$(s+r)(q+1)=s(q+1)+r(q+1)=s\left(\frac{s-r}{P}+1\right)+r(q+1)$$

$$=s\left(\frac{s}{P}+1\right)-\frac{sr}{P}+r(q+1),$$

因为 $\frac{s}{P}\leqslant q+1$，所以 $(s+r)(q+1)\geqslant s\left(\frac{s}{P}+1\right)$. 特别 $(s_k+r_k)(q_k+1)\geqslant s_k\left(\frac{s_k}{P}+1\right)$，再由(4)式可知

$$s_k\left(\frac{s_k}{P}+1\right)\leqslant k(k+1)n-\frac{k(k+1)(2k+1)}{3},\quad 1\leqslant k\leqslant m-1,$$

因此
$$P\geqslant\max_{1\leqslant k\leqslant m-1}\left\{\frac{s_k^2}{k(k+1)n-k(k+1)(2k+1)/3-s_k}\right\},$$

此即(5)式，至此定理证毕.

** 当 $k\leqslant\left[\frac{m}{2}\right]$ 时，等号成立（即 $\sum_{x\in\mathscr{D}_k}x=kn-n^2$），当且仅当 $b_1=1$, $b_m=n$，且 $d_1, d_2, \cdots, d_{k-1}, d_{m+1-k}, \cdots, d_{m-1}$ 皆等于 1. 此条件记为 $B(k)$.

当 $k>\left[\frac{m}{2}\right]$ 时，恒有不等号成立.

*** 当 $k\leqslant\left[\frac{m}{2}\right]$ 时，(8)式等号成立，当且仅当对每个 $i\leqslant k$，条件 $B(i)$ 都成立. 因而(4)式等号成立，当且仅当条件 $A(k)$，以及条件 $B(i)$, $i\leqslant k$ 皆成立. 当 $k>\left[\frac{m}{2}\right]$ 时，恒有不等号成立.

如 n、m 满足 $n\geqslant 2m$, $m>7$, 可以证明定理结论 (III) (5) 式右端的极大一定在 $\left[\frac{\sqrt{m}}{2}\right]\leqslant k\leqslant[\sqrt{m}+1]$ 的范围内达到. 于是我们只须在这范围内选择 k 值来估计相关值 $P(\underline{a})$.

下面用例子说明如何应用定理.

例 1　设序列 $\underline{a}$ 的 $n=34$, $m=15$, 那么 $P(\underline{a})\geqslant 5$.

证　方法 I: 设 $P=4$, 那么应用定理结论 (II) 中的 (4) 式; 取 $k=2$, $s_2=27=4\cdot 6+3$, $q_2=6$, $r_2=3$, (4) 式左端为

$$(s_2+r_2)(q_2+1)=210,$$

右端为

$$2(2+1)\cdot 34-\frac{2(2+1)(2\cdot 2+1)}{3}=194.$$

得出矛盾, 所以 $P(\underline{a})\geqslant 5$.

方法 II: 用定理 (III) 中 (5) 式, 记

$$f_k=\frac{s_k^2}{\frac{1}{3}k(k+1)(3n-2k-1)-s_k},$$

(5) 式即

$$P\geqslant\max_{1<k<m-1}\{f_k\},$$

取 $k=2$, 则 $s_2=27$, $f_2=729/167>4.3$, 所以 $P(\underline{a})\geqslant 5$.

我国几年前引进西德 Haiderhain 公司的零光栅, 其二元序列就是 $n=34$, $m=15$, $P(\underline{a})=5$. 可见, 它选的序列已达最优, 但由于 n 较小, 精度不够高.

例 2　设序列 $\underline{a}$ 的 $n=68$, $m=30$, 那么 $P(\underline{a})\geqslant 11$.

证　取 $k=3$, $s_3=84$, 易算得

$$f_3=84^2/704>10,$$

所以

$$P(\underline{a})\geqslant 11.$$

例 3　对于序列 $\underline{a}$,

(1) 当 $n=69$, $m=31$ 时, $P(\underline{a})\geqslant 11$;

(2) 当 $n=69$, $m=29$ 时, $P(\underline{a})\geqslant 10$.

证　(1) 当 $n=69$, $m=31$ 时, 取 $k=3$, $s_3=87$ 则 $f_3=87^2/713>10$, 所以 $P(\underline{a})\geqslant 11$.

(2) 当 $n=69$, $m=29$ 时, 取 $k=4$, $s_4=106$, 则 $f_4=106^2/1214=9.2$, 所以 $P(\underline{a})\geqslant 10$.

例 4　设序列 $\underline{a}$ 的 $n=69$, $m=30$, 则 $P(\underline{a})\geqslant 11$.

证　应用 (5) 式, 类似以上各例, 可知: $P\geqslant 10$. 如取 $P=10$, $k=4$, 算得 $s_4=110$, $q_4=11$, $r_4=0$, 以及 $\Sigma_1(30, 4, 10)=\Sigma_2(69, 4)=1320$; 如取 $k=5$, 则 $s_5=135$, $q_5=13$, $r_5=5$, 也有 $\Sigma_1(30, 5, 10)=\Sigma_2(69, 5)=1960$. 至此一时还不能由定理得出结论; 我们用 *, **, *** 作进一步分析. 因现在对 $k=4$, 5 均有

$$\Sigma_1(m, k, P)=\Sigma_2(n, k),$$

所以根据 *, **, ***, 必有条件 $A(k)$、条件 $B(k)$ 对 $k=4$、5 皆成立. 由条件 $A(4)$, 我们有

$$\bigcup_{i=1}^{4}\mathscr{D}_i=\{\underbrace{1,\ \cdots,\ 1}_{10个},\ \underbrace{2,\ \cdots,\ 2}_{10个},\ \cdots,\ \underbrace{11,\ \cdots,\ 11}_{10个}\}. \tag{9}$$

由条件 $B(4)$,

$$\sum_{x\in\mathscr{D}_4} x=4n-4^2=260 \tag{10}$$

而根据条件 $B(5)$, d_1, d_2, d_3, d_4, d_{26}, d_{27}, d_{28}, d_{29} 取值均应为1, 从而 $d_{1,5}=d_{26,30}=4$. 注意到 $|\mathscr{D}_4|=26$, 以及(9)式, 于是

$$\sum_{x\in\mathscr{D}_4} x\leqslant 4\times 2+11\times 10+10\times 10+9\times 4=256.$$

这与(10)式矛盾, 说明 $P>10$. 证毕.

§3. 序列的构造

对于给定的 n, m, 如何构造一个序列, 使 P 尽量小, 目前还没有一般性的办法. 但使用计算机, 运用一些技巧, 寻找这种序列是可能的. 如我们用**长城 203**, 找到了一批 $n=68$, $m=30$, $P=11$ 的序列. 我们使用的方法实质上是穷举法. 但如果从一个随意给出的初始序列出发穷举, 那么, 即使 n 不太大, 高速计算机要完成这穷举也是非常困难的. 我们提出如下几点:

(1) 初始序列应是较好的. 即先找一个序列, 其相关值尽量接近所要求的值. 这种初始序列可以用 m, n 值较小, 但较优的序列适当加长或者用几段这样的序列拼置而成.

(2) 然后上计算机, 对初始序列进行修改. 例如可以从序列的某一端开始, 按字典排列方式, 逐次移动取值为1的元.

(3) 程序还应该允许可以灵活地修改序列的某一部分. 这样容易结合人工的判断, 采取一定措施, 迅速得出所求序列.

最后, 我们列出5个 $n=68$, $m=30$, $P=11$ 的最优序列, 一并供参考. 下面给出的是序列 $\underline{a}$ 所对应的序列 $I=(b_1,\ b_2,\ \cdots,\ b_m)$.

(1, 2, 5, 6, 7, 9, 13, 15, 18, 20, 21, 22, 27, 30, 33, 38, 40, 41, 47, 48, 51, 52, 54, 56, 57, 62, 64, 66, 67, 68)

(1, 2, 5, 6, 7, 9, 13, 15, 18, 21, 26, 27, 30, 31, 33, 38, 40, 41, 47, 48, 51, 52, 54, 56, 57, 62, 64, 66, 67, 68)

(1, 2, 5, 6, 7, 9, 13, 15, 18, 21, 24, 25, 26, 28, 33, 38, 40, 41, 47, 48, 51, 52, 54, 56, 57, 62, 64, 66, 67, 68)

(1, 2, 5, 6, 7, 9, 13, 15, 18, 20, 26, 28, 29, 32, 33, 38, 40, 41, 47, 48, 51, 52, 54, 56, 57, 62, 64, 66, 67, 68)

(1, 2, 5, 6, 7, 9, 13, 15, 18, 20, 21, 22, 27, 30, 33, 38, 40, 41, 47, 48, 51, 52, 54, 56, 58, 63, 64, 66, 67, 68)

本文承万哲先老师帮助指导, 在此致谢.

THE ESTIMATION OF THE CORRELATIONS OF A CLASS OF BINARY SEQUENCES

DAI ZONGDUO　NI LUQUN　YANG JUNHUI　CHEN WENDE
(*Institute of Mathematica, Academia Sinica*)

ABSTRACT

Suppose $a(n, m)$ is a binary sequence

$$\underline{a}=(a_1, a_2, \cdots, a_n), \ a_i=0 \text{ or } 1;$$

$$m=\sum_{i=1}^{n} a_i;$$

$$\mathscr{C}(\tau)=\sum_{i=1}^{n-\tau} a_i \cdot a_{i+\tau};$$

$$P(\underline{a})=\max_{1\leqslant\tau\leqslant n-1} \mathscr{C}(\tau).$$

We call $P(\underline{a})$ the correlation of the sequence $\underline{a}(n, m)$. Given positive integers n and m, the binary sequence $\underline{a}(n, m)$ such that its $P(\underline{a})$ is the smallest is called the optimal sequence. It is useful in practice.

In this paper a lower bound of the correlation of the binary sequences $\underline{a}(n, m)$ for given positive integers n and m is estimated. At the end of this paper, we listed, as examples, several optimal sequences for $n=68$, $m=30$.