SCIENTIA SINICA Mathematica

### 论文



# 具有特定纯净效应折中设计的若干理论及构造

王东莹1\*, 张润楚2,3

- 1. 吉林财经大学统计学院, 长春 130117;
- 2. 南开大学数学科学学院, 天津 100002;
- 3. 东北师范大学数学与统计学院, 长春 130024

E-mail: wangdy798@126.com, rczhang\_nk@126.com

收稿日期: 2018-04-26; 接受日期: 2019-05-02; 网络出版日期: 2019-12-26; \* 通信作者

**摘要** 许多因析试验中, 试验者只关心指定的一部分因子效应的估计效果. 针对此类问题, Addelman (1962) 首次提出了折中设计的方法, 并定义纯净折中设计以保证指定的因子效应被有效地估计出来, 但此类纯净折中设计的分辨度限定为 IV. 本文研究了四类全新的折中设计, 指定因子效应的集合分别记为  $\{G_1,G_1\times G_1\}$ 、 $\{G_1,G_1\times G_1,G_2\times G_2\}$ 、 $\{G_1,G_1\times G_1,G_1\times G_2\}$  和  $\{G_1,G_1\times G_2\}$ . 相应地, 可以定义四种类型的纯净折中设计, 即指定因子效应全部纯净的设计. 本文研究四类纯净折中设计的存在性和结构特性, 给出一些理论结果和具体的构造方法. 构造结果表明新型的纯净折中设计的分辨度为 III 或 IV, 并且与 Addelman (1962) 提出的纯净折中设计相比具有更多的纯净两阶交互效应.

关键词 折中设计 纯净效应 部分因析设计

MSC (2020) 主题分类 62K15, 62K05

### 1 引言

试验是人类探索和发现未知世界的有力工具,被广泛地应用于各个科学领域.近几十年来,试验设计作为指导试验的专门科学和技术手段得到了迅猛发展.因析设计是试验设计的一个重要研究方向,致力于解决多因子试验的问题,在各类设计相关问题的研究中起着重要作用.本文的研究内容将围绕因析设计展开.

因析试验中,试验者的研究目的不尽相同,当然也存在如下情形: 试验者仅关心一部分重要的因子效应的估计效果,而非全部因子效应的估计情况. 为实施此类试验并有效地估计指定的因子效应,试验者需要特殊的设计以确保对指定因子效应的最佳估计. 我们称此类设计为折中设计,其概念由Addelman<sup>[1]</sup> 首次提出. 文献 [1] 研究了三类折中设计,确保了所有的主效应和一些指定的两因子交互效应的不相关估计. 文献 [2] 给出了第四类折中设计. 这里,第一类至第四类折中设计中待估两因子交互效应的集合分别被指定为

英文引用格式: Wang D Y, Zhang R C. Some theoretical results and constructions of compromise designs with specified sets of clear effects (in Chinese). Sci Sin Math, 2021, 51: 763-772, doi: 10.1360/SCM-2018-0298

(1) 
$$\{G_1 \times G_1\}$$
,  
(2)  $\{G_1 \times G_1, G_2 \times G_2\}$ ,  
(3)  $\{G_1 \times G_1, G_1 \times G_2\}$ ,  
(4)  $\{G_1 \times G_2\}$ ,

其中  $\{G_1; G_2\}$  是正规  $2^{m-p}$  设计中 m 个因子的一种划分,  $G_1$  中含  $m_1$  个因子,  $G_2$  中含  $m-m_1$  个因子;  $G_i \times G_j$  (i,j=1,2) 是  $G_i$  与  $G_j$  交互生成的两因子交互效应的集合. 文献 [3] 讨论了第一类至第四类纯净折中设计 (clear compromise plan, CCP), 即全部主效应及对应于 (1.1) 的两阶交互效应均为纯净的, 同时研究了 CCP 的存在性和结构特性. 文献 [4] 将关于折中设计的研究推广至混水平  $4^m 2^n$  设计的情形.

值得注意的是, CCP 要求设计的分辨度为 *IV*, 以确保试验中所有因子的主效应都可以被纯净地估计出来. 但在实际试验中, 如果试验者只关心如何最优地估计一部分指定的因子效应, 对于其余因子效应的估计效果不做要求, 并且指定的因子效应可能仅包括一部分主效应和两阶交互效应, 此时不必要求设计的分辨度达到 *IV*. 所以, 选取最优的折中设计时可将分辨度的条件放宽. 另一方面, 当试验次数确定时, 分辨度为 *III* 的设计相较于分辨度为 *IV* 的设计包含的设计列更多. 反之, 当设计列数确定时, 分辨度为 *III* 的设计需要更少的试验次数. 为了确保试验经济且有效, 分辨度 *III* 的设计更有可能被选中. 于是, 我们在文献 [5] 中将折中设计的分辨度放宽为 *III* 和 *IV*, 并在此范围内选择最优的折中设计. 文献 [5] 给出了新的折中设计分类方式, 定义了以下四种待估因子效应的集合:

(1) 
$$\{G_1, G_1 \times G_1\},$$
  
(2)  $\{G_1, G_1 \times G_1, G_2 \times G_2\},$   
(3)  $\{G_1, G_1 \times G_1, G_1 \times G_2\},$   
(4)  $\{G_1, G_1 \times G_2\},$ 

其中  $G_i$  和  $G_i \times G_j$  (i,j=1,2) 分别为指定的主效应集合和两阶交互效应集合. 这里, 对于第一类折中设计, 虽然  $G_2$  中的因子不出现在  $\{G_1,G_1\times G_1\}$  中,但  $G_2$  的因子安排方法也会影响指定因子效应的估计效果, 不能忽略. 对于其余三类折中设计也是如此, 所以应当同时考虑用来安排  $G_1$  和  $G_2$  中因子的设计列. 为简明起见,  $\{G_1;G_2\}$  也用来表示设计列的一种划分, 分别对应用来安排  $G_1$  和  $G_2$  中因子的两部分设计列. 此时, 相应的折中设计记为  $T=\{G_1;G_2\}$ , 若设计 T 对应于 (1.2) 中某一种因子效应的集合全部纯净, 则称之为纯净折中设计  $(clear\ compromise\ design,\ CCD)$ , 分别记为第一类至第四类 CCD.

文献 [5] 通过定义全新的分类模式,即部分混杂效应个数型 (partial aliased effect number pattern, P-AENP),来衡量折中设计的好与坏,且针对第一类折中设计给出一些理论结果.本文将同时考虑四类 CCD,讨论各类 CCD 的存在性和结构特性,同时给出具体的构造方法.本文的结构安排如下:第 2节给出预备知识和基本符号;关于四类 CCD 的理论结果和构造方法在第 3 至 5 节给出;第 6 节对本文进行总结,给出 32-run 和 64-run CCD 的设计表.

#### 2 预备知识

首先, 简要介绍一些有用的记号. 本文研究的设计均为两水平正规部分因析设计, 记为  $2^{m-p}$  设计, 由选自  $H_k$  (k=m-p) 的 m 个设计列组成, 其中  $H_k$  表示试验次数为  $2^k$  饱和设计 (即  $2^{(2^k-1)-(2^k-k-1)}$ 

设计). 文献 [6,7] 将 Yates 序下饱和设计记为

$$H_k = \{1, 2, 12, 3, 13, 23, 123, \dots, 12 \cdots k\}_{2^k},$$

它由 k 个独立列生成,每个独立列包含  $2^k$  个分量,各分量取值为 1 或 -1. 不失一般性,  $H_k$  的 k 个独立列可选定为

$$\mathbf{1}_{2^k} = (1, -1, 1, -1, \dots, 1, -1)^{\mathrm{T}}, \quad \mathbf{2}_{2^k} = (1, 1, -1, -1, \dots, 1, 1, -1, -1)^{\mathrm{T}},$$

以此类推, 其中符号 T 表示转置. 与此同时, 记  $\mathbf{I}_{2^k} = (1,1,1,1,\dots,1,1,1,1)^{\mathrm{T}}$ . 为简明起见, 称  $\{\mathbf{1},\mathbf{2},\dots,\mathbf{k}\}$  为定义列, 除此之外的任意列均由一定数量的定义列生成, 称之为生成列. 例如,  $\mathbf{12}$  为生成列, 由定义列  $\mathbf{1}$  和  $\mathbf{2}$  对应分量相乘得到.

对于任意一组独立列, 若  $\{\alpha_1, \ldots, \alpha_f\} \subset H_k$ , 则称

$$\{\mathbf{I}, \alpha_1, \alpha_2, \dots, \alpha_f, \alpha_1\alpha_2, \dots, \alpha_{f-1}\alpha_f, \dots, \alpha_1\alpha_2 \cdots \alpha_f\}$$

为  $\{\alpha_1, \alpha_2, \ldots, \alpha_f\}$  生成的闭集合. 例如, **123** 和 **145** 是  $H_k$  (k > 5) 中的两个独立列, 它们生成的闭集合为  $\{\mathbf{I}, \mathbf{123}, \mathbf{145}, \mathbf{2345}\}$ . 特别地, 对于任意  $i \leq k$ , 令  $\{\mathbf{I}, H_i\}$  表示由定义列  $\{\mathbf{1}, \ldots, \mathbf{i}\}$  生成的闭集合, 即

$$\{\mathbf{I}, H_i\} = \{\mathbf{I}, \mathbf{1}, \mathbf{2}, \mathbf{12}, \mathbf{3}, \mathbf{13}, \mathbf{23}, \mathbf{123}, \dots, \mathbf{12} \cdots \mathbf{i}\};$$

 $\{I, \overline{H}_i\}$  表示由  $\{i+1,...,k\}$  生成的闭集合, 其中

$$\overline{H}_i = \{i+1, i+2, (i+1)(i+2), \dots, (i+1)(i+2) \cdots k\}.$$

此外, 我们还用到记号 M(k). 对于给定的试验次数为  $2^k$  (k=m-p), M(k) 表示存在分辨度不低于 V 的  $2^{m-p}$  设计的条件下, m 的最大值. 本文研究分辨度 III 和 IV 的  $2^{m-p}$  纯净折中设计, 所以仅考虑  $M(k)+1 \le m \le 2^k-1$  范围内的  $2^{m-p}$  设计.

### 3 四类纯净折中设计的存在性

本节将围绕  $2^{m-p}$  设计讨论纯净折中设计存在性的问题. 记  $T = \{G_1; G_2\}$  为一个  $2^{m-p}$  设计, 其中  $m = m_1 + m_2$ ,  $m_i = \#\{G_i\}$ , i = 1, 2. 文献 [3] 证明了第二类 CCP 不存在, 但该结论针对的是分辨 度为 IV 的  $2^{m-p}$  设计. 这里指出, 若同时考虑分辨度为 III 和 IV 的  $2^{m-p}$  设计, 第二类 CCD 也不存在. 以下定理证明了这一结论.

定理 3.1  $2^{m-p}$  设计中不存在第二类 CCD.

证明 根据 CCD 的定义, 需要对分辨度为 III 和 IV 的  $2^{m-p}$  设计进行讨论. 而对于分辨度 IV 的情形, 文献 [3, 推论 1] 已证, 因此, 我们仅证明分辨度 III 的情形下结论成立.

对于分辨度为 III 的  $2^{m-p}$  设计, 长度为 3 的定义字可能有四种形式:  $G_1 \times G_1 \times G_1 \times G_1 \times G_2 \times G$ 

文献 [8] 已证得当  $m > 2^{k-1}$  (k = m - p) 时,  $2^{m-p}$  设计不具有纯净的两阶交互效应. 那么, 关于 CCD 存在性的讨论自然就落入  $M(k) + 1 \le m \le 2^{k-1}$  的参数范围之内. 进一步, 对于  $2^{k-2} + 2 \le m$ 

 $\leq 2^{k-1}$  的情形, 文献 [8] 证明了分辨度为 IV 的  $2^{m-p}$  设计中不存在纯净的两因子交互效应, 所以, CCD 一定是分辨度 III 的设计. 以下定理和推论详细阐述了  $2^{k-2} + 2 \leq m \leq 2^{k-1}$  的情形下 CCD 的存在性.

**定理 3.2** 若  $2^{k-2} + 2 \le m \le 2^{k-1}$ ,  $m_1 = 2$ , 则不存在第一类、第三类和第四类 CCD.

证明 反证法. 假定  $2^{m-p}$  设计  $T = \{G_1; G_2\} = \{\alpha_1, \alpha_2; \beta_1, \dots, \beta_{m-2}\}$  是第一类、第三类或第四类 CCD, 并且  $2^{k-2} + 2 \le m \le 2^{k-1}$ . 此时, 可以证明  $\alpha_1, \alpha_2$  和  $\alpha_1\alpha_2$  是纯净的. 当 T 为第一类或第三类 CCD 时, 由定义可知该结论显然成立. 因此, 只需证明 T 为第四类 CCD 时,  $\alpha_1\alpha_2$  是纯净的. 如果  $\alpha_1\alpha_2 = \beta_i$  或  $\alpha_1\alpha_2 = \beta_i\beta_j$   $(i \ne j)$ , 将导致  $G_1 \times G_2$  中的某些效应不是纯净的. 所以, 两种别名关系都不成立,  $\alpha_1\alpha_2$  是纯净的.

若  $\alpha_1$ 、 $\alpha_2$  和  $\alpha_1\alpha_2$  均为纯净效应,则有

$$\{\alpha_1\alpha_2, \alpha_1\beta_i, \alpha_2\beta_i, \alpha_1\alpha_2\beta_i, 1 \le i \le m-2\} \subseteq \overline{T},$$

其中  $\overline{T} = H_k \setminus T$  为 T 的补设计. 于是,  $\overline{T}$  至少包含 3m-5 个不同的列, 由此可得

$$3m - 5 \leqslant 2^k - m - 1,$$

等价于  $m \leq 2^{k-2} + 1$ , 产生矛盾.

根据定义, 对于第一类 (或第三类) CCD, 将  $G_1$  的部分列移动至  $G_2$ , 折中设计的结构发生改变, 但仍是第一类 (或第三类) CCD, 差别在于  $G_1$  中的设计列更少了. 结合定理 3.2 可得到以下推论.

推论 3.1 若  $2^{k-2} + 2 \le m \le 2^{k-1}$ ,  $m_1 > 2$ , 则不存在第一类和第三类 CCD.

我们可以通过删除第四类 CCD 中  $G_1$  的一列或多列来获取新的第四类 CCD. 由定理 3.2 可知, 当  $2^{k-2}+3 \le m \le 2^{k-1}$  且  $m_1=3$  时, 不存在第四类 CCD. 类似地, 当  $2^{k-2}+4 \le m \le 2^{k-1}$  且  $m_1=4$  时, 不存在第四类 CCD, 以此类推. 于是得到以下推论.

推论 3.2 若  $2^{k-2} + 2 \le m \le 2^{k-1}$   $(k \ge 4), 2 < m_1 \le m/2$ , 则不存在第四类 CCD.

证明 反证法. 假定  $2^{m-p}$  设计  $T = \{G_1; G_2\} = \{\alpha_1, \dots, \alpha_{m_1}; \beta_1, \dots, \beta_{m-m_1}\}$  为第四类 CCD, 其中  $2 < m_1 \le m/2$ .  $2^{k-2} + 2 \le m \le 2^{k-1}$ .

首先、考虑  $m_1 = 3$  的情形。 $\{G_1, G_1 \times G_2\}$  中的效应均为纯净的、意味着

$$\{\alpha_1\alpha_2, \alpha_1\alpha_3, \alpha_2\alpha_3, \alpha_1\alpha_2\beta_j, \alpha_1\alpha_3\beta_j, \alpha_2\alpha_3\beta_j, j = 1, \dots, m - 3\}$$

两两之间不存在别名关系且与 $\overline{T}$ 中的不同列别名. 又因为 $\overline{T}$ 包含 $2^k-1-m$ 个设计列, 所以有

$$3 + 3(m-3) \le 2^k - 1 - m$$
,

等价于  $m \leq 2^{k-2} + 1$ , 产生矛盾.

同样地, 对于  $4 \le m_1 < m/2$  的情形, 由第四类 CCD 的定义可知,  $G_1 \times G_2$  中的设计列两两不同 且与 T 的不同列别名. 于是有

$$m_1(m-m_1) \leq 2^k - 1 - m$$
,

即

$$m \leqslant \frac{2^k}{m_1 + 1} + m_1 - 1.$$

一方面, 若  $2^{k-2}+2 < m \le 2^{k-1}$ , 则有  $4 \le m_1 \le 2^{k-2}$ . 此范围内,  $2^k/(m_1+1)+m_1-1$  的最大值

$$\operatorname{Max}\left\{\frac{2^k}{m_1+1}+m_1-1\right\} \leqslant \operatorname{Max}\left\{\frac{2^k}{5}+3, \frac{2^k}{2^{k-2}+1}+2^{k-2}-1\right\} < 2^{k-2}+3 \quad (k \geqslant 4).$$

于是有  $m \leq 2^{k-2} + 2$ , 产生矛盾. 另一方面, 若  $m = 2^{k-2} + 2$ , 则  $4 \leq m_1 \leq 2^{k-3} + 1$   $(k \geq 5)$ . 此范围内, 我们有

$$\operatorname{Max}\left\{\frac{2^k}{m_1+1}+m_1-1\right\} \leqslant \operatorname{Max}\left\{\frac{2^k}{5}+3, \frac{2^k}{2^{k-3}+2}+2^{k-3}+1\right\} < 2^{k-2}+2.$$

于是,  $m \leq 2^{k-2} + 1$ , 产生矛盾.

综上, 定理得证.

## 4 $2^{k-2}+1 \leq m \leq 2^{k-1}$ 时 CCD 的构造方法

本节将分别给出  $2^{k-2}+2 \le m \le 2^{k-1}$  和  $m=2^{k-2}+1$  情形下的第一类、第三类和第四类 CCD 的具体构造方法. 首先, 考虑  $2^{k-2}+2 \le m \le 2^{k-1}$  的情形. 基于上一节的讨论, 我们只须针对  $m_1=1$  构造第一类、第三类和第四类. 这些设计通过以下定理中的设计 T 获得, 删除  $\overline{H}_1$  中任意  $2^{k-1}-m$  列即可. 定理很容易验证, 所以省略了证明过程.

定理 **4.1**  $2^{2^{k-1}-(2^{k-1}-k)}$  设计  $T = \{1; \overline{H}_1\}$  是一个第三类 CCD.

结合定理 3.2 和 4.1 及推论 3.1 和 3.2 可得到  $m_1$  可能取到的最大值, 具体结果由以下推论给出.

**推论 4.1** 对于第一类、第三类和第四类 CCD, 若  $2^{k-2}+2 \le m \le 2^{k-1}$  ( $k \ge 4$ ), 则  $\max(m_1) = 1$ . 接下来、考虑  $m = 2^{k-2}+1$  的特殊情形、以下为我们构造的第一类、第三类和第四类 CCD.

**定理 4.2** 令 T 为  $2^{2^{k-2}+1-(2^{k-2}+1-k)}$  设计,则

- (i)  $T = \{\mathbf{1}, \mathbf{2}; \mathbf{1} \cdot \overline{H}_2\}$  是第一类 CCD, 其分辨度为 IV;
- (ii)  $T = \{1, 2, 3; \{1, 23\} \cdot \overline{H}_3\}$  是第一类 CCD, 其分辨度为 IV;
- (iii)  $T = \{1, 2; \overline{H}_2\}$  和  $T' = \{1, 2; 12 \cdot \overline{H}_2\}$  是第三类 CCD, 前者分辨度为 III, 后者分辨度为 IV;
- (iv)  $T = \{1, 2, 3; \{I, 123\} \cdot \overline{H}_3\}$  是第三类 CCD, 其分辨度为 III;
- (v)  $T = \{\mathbf{1}, \mathbf{12} \cdot \overline{H}_2; \mathbf{2}\}$  是第四类 CCD, 其分辨度为 IV.

按照 CCD 的定义容易验证定理 4.2 中的结论成立, 在此我们省略证明过程.

对于  $2^{m-p}$  设计  $m=2^{k-2}+1$ , 参见定理 4.2(v), 我们推断第四类 CCD 中  $Max(m_1)=2^{k-2}$ . 同时, 利用定理 4.2(iii), 可证第一类或第三类 CCD 中  $Max(m_1)=3$ .

推论 4.2 对于第一类和第三类 CCD, 若  $m = 2^{k-2} + 1$   $(k \ge 5)$ , 则  $Max(m_1) = 3$ .

证明 根据第一类和第三类 CCD 的定义,  $G_1$  和  $G_1 \times G_1$  中的效应均为纯净的, 可知  $G_1 \times G_1$  和  $G_1 \times G_2$  两两不同且与 T 中的不同列别名. 于是有

$$\binom{m_1}{2} + m_1(2^{k-2} + 1 - m_1) \le 2^k - 1 - (2^{k-2} + 1),$$

等价于

$$m_1 - 3 \leqslant \frac{2}{2^{k-1} - m_1 - 2}.$$

然而  $m_1 \leq 2^{k-2} + 1$ ,  $k \geq 5$ , 可推得  $2/(2^{k-1} - m_1 - 2) < 1$ , 于是  $m_1 \leq 3$ . 另一方面, 定理 4.2(iii) 说明 存在 CCD 满足  $m_1 = 3$ . 综上可知  $\max(m_1) = 3$ .

对于  $M(k) + 1 \le m < 2^{k-2} + 1$  的情形, 满足  $m_1 \le 3$  的第一类和第三类 CCD 及满足  $m_1 \le 2^{k-2}$  的第四类 CCD 可通过删除定理 4.2 给出设计的部分列来构造. 我们利用以下例子来说明该方法.

**例 4.1** 对于试验次数为 64、因子个数为 15 (即  $2^{15-9}$ ) 的第三类 CCD, 可通过删除定理 4.2(iii) 设计 T' 中  $\mathbf{12} \cdot \overline{\mathbf{H}}_2$  的任意两列得到, 如设计

$$T_0 = \{1, 2; 12 \cdot \{3, 4, 34, 5, 345, 6, 36, 46, 346\}, 56, 356, 456, 3456\}.$$

该设计同构于  $2_{IV}^{15-9}$  MaxC2 设计 ({1,2,3,4,5,6,123,124,134,234,125,135,235,145,2456}, 参见 文献 [9, 表 4A.5], 将 1,2,3,4,5,6 重新标记为 6,2456,1245,45,2345,25,即得上述第三类 CCD). 所以,  $T_0$  也是一个 MaxC2 设计,包含 27 个纯净的两因子交互效应.

### 5 $\mathbf{M}(k) + 1 \le m \le 2^{k-2} + 1$ 时 CCD 构造方法

由于第三类 CCD 同时为第一类和第二类 CCD, 所以直接给出第三类 CCD 的构造方法, 具体分为两种情形:

- (I)  $G_1$  具有  $m_1$  (≤ k) 个独立列;
- (II)  $G_1$  具有  $k_1$  ( $< m_1$ ) 个独立列和  $m_1 k_1$  个生成列.

对于情形 (II),  $m_1 - k_1$  个生成列是由  $k_1$  个独立列生成的, 每个生成列都对应一个定义关系. 于是,  $G_1$  可以看作一个  $2^{m_1-(m_1-k_1)}$  设计, 分辨度不低于 V, 故  $k_1 < m_1 \le M(k_1)$ .

首先, 考虑情形 (I). 不失一般性, 记  $G_1 = \{1, 2, ..., \mathbf{m}_1\}$ , 其中  $1, 2, ..., \mathbf{m}_1$  为  $H_k$  中的一部分定义列. 此外, 满足某一特定条件且容量达到最大的子集, 我们称之为符合该条件的最大子集. 为方便构造, 我们先介绍两个引理.

**引理 5.1** 令  $A_f$  为 {**I**,  $H_f$ } 中满足如下条件的最大子集:  $A_f$  中任意两列生成的交互作用列不与 {**1**, **2**, . . . , **f**, **12**, . . . , (**f** - **1**)**f**} 中的任一列发生混杂. 那么,  $A_f$  为一个闭子集, 其容量为  $2^{f-\lceil \log_2(f+1) \rceil}$ .

证明 首先, 使用归纳法证明  $A_f$  为一个闭子集. 当 f=1 时,  $A_1=\{\mathbf{I}\}$ , 显然为闭子集. 接下来, 假定 f=t 时结论成立, 试证  $A_{t+1}$  也是一个闭子集. 一方面, 如果不存在  $\beta\in H_{t+1}\setminus H_t$ , 使得  $\beta\cdot A_t$  中的列都由三个或三个以上的定义列生成, 那么有  $A_{t+1}=A_t$ . 另一方面, 如果存在这样的  $\beta\in H_{t+1}\setminus H_t$ , 则有

$$A_t \cup \beta \cdot A_t \subseteq A_{t+1}$$
.

显然有  $A_t \cap \beta \cdot A_t = \emptyset$ , 否则将存在  $\alpha_1, \alpha_2 \in A_t$ , 使得  $\alpha_1 = \beta \alpha_2$ , 等价于,  $\beta = \alpha_1 \alpha_2 \in A_t$ , 产生矛盾. 进一步,  $\beta \cdot A_t$  包含于  $H_{t+1} \setminus H_t$ , 且与  $A_t$  同构, 因此,  $\beta \cdot A_t$  中的任意列都不能添加到  $A_{t+1}$ , 亦即

$$A_t \cup \beta \cdot A_t = A_{t+1}$$
.

综合两种情形可知, At+1 为闭子集.

下面证明  $A_f$  的容量为  $2^{f-\lceil \log_2(f+1) \rceil}$ . 这里用长度为 f 的二元字符串来表示  $A_f$  中的列,具体如下:若生成该列的定义列中有  $\mathbf{i}$ ,则相应字符串的第 i 个分量取值为 1,否则取值为 0. 假定闭子集  $A_f$  具有  $2^{a_f}$  个列,那么  $2^{a_f}-1$  非零字符串可被看作为分辨度为 III 的  $2^{f-a_f}$  设计的定义字 (defining pencil  $[^{10]}$ ). 因此,根据分辨度为 III 的  $2^{f-a_f}$  设计存在的必要条件,对于  $a_f$  的约束条件是  $f \leq 2^{f-a_f}-1$ ,当且仅当  $2^{f-a_f}$  为分辨度为 III 的饱和设计时等号成立。于是,有  $a_f \leq f-\lceil \log_2(f+1) \rceil$ ,其中  $\lceil \log_2(f+1) \rceil$  表示不低于  $\log_2(f+1)$  的最小整数。所以,满足引理所述条件的最大子集  $A_f$  的容量为  $2^{f-\lceil \log_2(f+1) \rceil}$ .

下面给出  $A_f$  的构造方法. 为方便阐述, 将  $\{1,2,\ldots,f\}$  记为  $\{\alpha_1,\alpha_2,\ldots,\alpha_f\}$ .

注 5.1  $A_f$  表示由  $\{\alpha_{2^{i_0}}\alpha_{2^{i_1}}\cdots\alpha_{2^{i_s}}\alpha_{\sum_{t=0}^s 2^{i_t}}; \sum_{t=0}^s 2^{i_t}\leqslant f, 1\leqslant s\leqslant \lceil\log_2(f+1)\rceil-1, 0\leqslant i_0<\cdots< i_s\leqslant \lceil\log_2(f+1)\rceil-1\}$ 产生的闭子集. 那么,  $A_f$  的容量为  $2^{f-\lceil\log_2(f+1)\rceil}$ , 且满足引理 5.1 的结论.

我们以 f=6 为例进一步阐述注 5.1 的构造方法. 此时, $\lceil \log_2(f+1) \rceil - 1 = 2$ ,且有  $1 \le s \le 2$ , $0 \le i_0 < i_1 \le 2$ , $2^{i_0} + 2^{i_1} \le 6$ . 满足这些条件仅有三种情形:  $i_0 = 0$ ,  $i_1 = 1$ ;  $i_0 = 0$ ,  $i_1 = 2$ ;  $i_0 = 1$ ,  $i_1 = 2$ . 相应地,有  $\alpha_{2^0}\alpha_{2^1}\alpha_{2^0+2^1} = \alpha_1\alpha_2\alpha_3 = 123$ , $\alpha_{2^0}\alpha_{2^2}\alpha_{2^0+2^2} = \alpha_1\alpha_4\alpha_5 = 145$ ,和  $\alpha_{2^1}\alpha_{2^2}\alpha_{2^1+2^2} = \alpha_2\alpha_4\alpha_6 = 246$ . 因此,注 5.1 给出了一个由  $\{123,145,246\}$  生成的闭子集  $A_6 = \{I,123,145,2345,246,1346,1256,356\}$ .

引理 5.2  $\diamond B_f$  为  $H_f$  中的子集, 且满足如下条件:

- (i)  $\{I, B_f\}$  是闭子集;
- (ii)  $B_f$  中的每一列都是由四个或四个以上定义列生成,则  $B_f$  的最大容量为  $2^{f-\lceil \log_2 f \rceil 1} 1$ .

引理 5.2 可以通过与引理 5.1 类似的方法加以证明. 同时,  $B_f$  中的列与某个分辨度为 IV 的  $2^{f-(f-\lceil \log_2 f \rceil-1)}$  设计的定义字存在一一对应的关系, 所以具有  $2^{f-\lceil \log_2 f \rceil-1}-1$  个设计列的  $B_f$  容易构造出来.

结合引理 5.1 和 5.2 给出的集合  $A_{m_1}$  和  $B_{m_1}$  ( $f=m_1$ ), 我们可以构造第三类 CCD. 具体构造方法可分为  $m_1 < k$  和  $m_1 = k$  两种情形, 分别由定理 5.1 和 5.2 给出.

定理 5.1 对于  $G_1 = \{\mathbf{1}, \mathbf{2}, \dots, \mathbf{m}_1\}$ ,  $m_1 < k$ , 令  $G_2 = B_{m_1} \cup (\bigcup_{\alpha \in A_{m_1}} \alpha \cdot \overline{H}_{m_1})$ . 若  $A_{m_1} \times B_{m_1}$  中任意两阶交互作用列都不与  $\{\mathbf{1}, \mathbf{2}, \dots, \mathbf{m}_1\}$  中的列发生混杂, 则  $T = \{G_1; G_2\}$  是第三类 CCD.

证明 首先证明  $G_1 \times G_1$  的两阶交互效应都是纯净的. 显然,  $G_1 \times G_1$  中的两阶交互效应不与  $G_1$  和  $G_2$  中的主效应发生混杂. 因此, 只需证明  $G_1 \times G_1$  中的两阶交互效应不与  $G_2 \times G_2$  和  $G_1 \times G_2$  中的两阶交互效应发生混杂.

首先,  $G_2 \times G_2$  中共有五种两阶交互效应, 形如  $\beta \times \beta'$ 、 $\beta \times \alpha\gamma$ 、 $\alpha\gamma \times \alpha\gamma'$ 、 $\alpha\gamma \times \alpha'\gamma$  和  $\alpha\gamma \times \alpha'\gamma'$  ( $\alpha, \alpha' \in A_{m_1}, \beta, \beta' \in B_{m_1}, \gamma, \gamma' \in \overline{H}_{m_1}$ ), 其中  $\beta \times \alpha\gamma$ 、 $\alpha\gamma \times \alpha\gamma'$  和  $\alpha\gamma \times \alpha'\gamma'$  均与  $\overline{H}_{m_1}$  中的某列发生混杂, 故不与  $G_1 \times G_1$  ( $\subset H_{m_1}$ ) 中的两阶交互效应发生混杂. 又根据  $A_{m_1}$  和  $B_{m_1}$  的定义可知,  $\beta \times \beta'$  和  $\alpha\gamma \times \alpha'\gamma = \alpha \times \alpha'$  均不与  $G_1 \times G_1$  中的两阶交互效应发生混杂. 所以,  $G_1 \times G_1$  中的两阶交互效应不与  $G_2 \times G_2$  中的两阶交互效应发生混杂.

其次, 要证  $G_1 \times G_1$  中的两阶交互效应不与  $G_1 \times G_2$  中的两阶交互效应发生混杂, 只须证  $G_2 = B_{m_1} \cup (\bigcup_{\alpha \in A_{m_1}} \alpha \cdot \overline{H}_{m_1})$  中的列不与  $G_1 \times G_1 \times G_1$  中的列发生混杂.  $B_{m_1}$  中的列涉及四个或四个以上的定义列, 显然不与  $G_1 \times G_1 \times G_1$  中的列发生混杂. 由于  $\bigcup_{\alpha \in A_{m_1}} \alpha \cdot \overline{H}_{m_1} \subset H_k \setminus H_{m_1}$ ,  $G_1 \times G_1 \times G_1 \subset H_{m_1}$ , 前者中的任意列均不与后者中的列发生混杂. 故  $G_2$  中的列不与  $G_1 \times G_1 \times G_1$  中的列发生混杂, 上述结论成立.

至此我们证明了  $G_1 \times G_1$  中的效应都是纯净的. 类似地, 可以证明  $G_1$  和  $G_1 \times G_2$  中的效应都是纯净的, 故设计 T 为第三类 CCD.

**例 5.1** 特别地, 取 k = 6,  $m_1 < 6$ , 根据定理 5.1, 我们构造出如下第三类 CCD, 见表 1.

定理 5.2 对于  $G_1 = \{1, 2, ..., k\}$ , 令  $G_2 = B_k$ , 则  $T = \{G_1; G_2\}$  是一个第三类 CCD.

定理 5.1 的证明方法对于上述定理仍然适用,因此省略其证明过程,仅给出一个构造实例. 当 k=6 时,在同构的意义下可取  $B_6=\{1234,3456,1256\}$ ,那么  $2^{9-3}$  设计  $T=\{1,2,3,4,5,6;1234,3456,1256\}$  是一个第三类 CCD.

表 1 64-run 第三类 CCD							
$\overline{m}$	32	17	17	11	10		
$m_1$	1	2	3	4	5		
$G_1$	$\{1\}$	$\{1,2\}$	$\{{f 1},{f 2},{f 3}\}$	$\{{f 1},{f 2},{f 3},{f 4}\}$	$\{{f 1, 2, 3, 4, 5}\}$		
$G_2$	$\overline{H}_1$	$\overline{H}_2$	$\{\mathbf{I},123\}\!\cdot\!\overline{H}_3$	$\{\mathbf{I},123\}\!\cdot\!\overline{H}_{4}$	$\{{f 2345}, \{{f I}, {f 123}, {f 145}, {f 2345}\} {f \cdot} \overline{H}_5\}$		

接下来, 考虑情形 (II):  $G_1$  由  $k_1$  ( $< m_1$ ) 个独立列和  $m_1 - k_1$  个生成列组成. 该情形下, 我们基于 定理 5.1 中给出的设计 T 来构造第三类 CCD, 具体细节如下.

对于  $G_1$ , 不妨设  $\{1,2,\ldots,k_1\}$  为  $k_1$  个独立列,  $\{\gamma_1,\ldots,\gamma_{m_1-k_1}\}\subset H_{k_1}$  为  $m_1-k_1$  个生成列. 定 理 5.1 的证明过程显示, 当  $A_{k_1}$  中产生的两阶交互作用列不与  $G_1 \times G_1$  中的两阶交互作用列别名时, 定 理构造的  $2^{m-p}$  设计 T 为第三类 CCD. 因此, 令  $C_0 = \{\gamma_1 \mathbf{1}, \dots, \gamma_{m_1-k_1} \mathbf{k}_1\} \cup \{\gamma_i \gamma_i, 1 \leq i < j \leq m_1-k_1\}$ ,  $C_{k_1}$  为  $A_{k_1}$  的子集且  $C_{k_1} \times C_{k_1}$  中的交互作用列不与  $A_{k_1} \cap C_0$  中的列别名, 则  $2^{m-p}$  设计  $T = \{G_1; G_2\}$ (其中  $G_1 = \{1, 2, \dots, \mathbf{k}_1, \gamma_1, \dots, \gamma_{m_1 - k_1}\}, G_2 = \bigcup_{\alpha \in C_{k_*}} \alpha \cdot \overline{H}_{k_1}$ ) 是一个第三类 CCD.

值得说明的是, 当试验次数较少时, 上述方法构造的设计所包含的设计列相对少一些. 该方法对 于构造试验次数较多的纯净折中设计更为灵活. 下面用一个具体的例子进一步解释该方法.

例 5.2 对于  $k_1 = 6$ ,  $m_1 = 7$ , k > 6, 注 5.1 给出了  $A_6 = \{ \mathbf{I}, \mathbf{123}, \mathbf{145}, \mathbf{2345}, \mathbf{246}, \mathbf{1346}, \mathbf{1256}, \mathbf{$ 356}. 若  $\gamma_1 = 1234$ , 由此可得  $C_0 = \{123, 124, 134, 234, 12345, 12346\}$ ,  $C_0 \cap A_6 = \{123\}$ . 根据 上述构造方法, 不妨设  $C_6 = \{I, 145, 246, 1256\}$ , 那么设计  $T_1 = \{1, 2, 3, 4, 5, 6, 1234; \{I, 145, 2345, 1256\}\}$ 246,1346,1256,356}· $\overline{H}_6$ } 是一个第三类 CCD. 与此同时, 我们可以构造更多的第三类 CCD, 如设  $\dagger \uparrow T_2 = \{1, 2, 3, 4, 5, 6, 12345; \{I, 123, 246, 1346\} \cdot \overline{H}_6\} \ (\gamma_1 = 12345) \ \exists I \ T_3 = \{1, 2, 3, 4, 5, 6, 123456; \}$  $\{\mathbf{I}, \mathbf{123}, \mathbf{145}, \mathbf{2345}, \mathbf{246}, \mathbf{1346}, \mathbf{1256}, \mathbf{356}\} \cdot \overline{H}_6\} \ (\gamma_1 = \mathbf{123456}).$ 

### 6 总结与讨论

本文重新定义了四类折中设计, 讨论了各类纯净折中设计的存在性, 并且给出了各类纯净折中设 计的构造方法. 下面说明 CCD 与 MaxC2 设计[1] 和 CCP [3] 之间的联系.

第一, 个别情形下 CCD 是 MaxC2 设计. 例如, 定理 4.2(iii) 给出的  $2^{2^{k-2}+1-(2^{k-2}+1-k)}$  设计 T'= $\{1, 2; 12 \cdot \overline{H}_2\}$ , 即为一个 MaxC2 设计. 此外, 删除  $12 \cdot \overline{H}_2$  的一些列, 可得到  $2^{8-3} \cdot 2^{14-8} \cdot 2^{15-9}$  和 2<sup>16-10</sup> 设计, 这些设计是第一类、第三类和第四类 CCD, 同时也是 MaxC2 设计.

第二, 若仅考虑分辨度为 IV 的设计, 基于 (1.2) 定义的 CCD 与文献 [1,3] 定义的 CCP 相同. 通 过本文的第 4 和 5 节给出的构造方法可得到分辨度为 IV 的 CCD, 因此, 这些方法同样适用于 CCP 的构造. 例如, 定理 4.2(v) 给出的设计  $T = \{1, 12 \cdot \overline{H}_2; 2\}$  是一个分辨度为 IV 的第四类 CCD, 同时 也是一个第四类 CCP. 该设计与文献 [11] 构造的第四类 CCP 相比,  $G_1$  所包含的列更多. 另一方面, 对于分辨度为 III 的 CCD, 它们与 CCP 相比,  $G_1$  包含更多的设计列, 具有更多的纯净两阶交互效应, 详见表 2 和 3.

一般地, 当试验次数固定且 m 或  $m_1$  较大时, CCD 不一定存在. 此时, 折中设计对于指定的因子 效应估计效果的好与坏, 应通过更为合理的分类模式来衡量. 关于新的分类模式及其对应的最优折中 设计将在我们的后续工作中进行深入地研究.

最后, 为方便应用, 我们给出以下设计表, 表 2-4 分别列举 32-run 和 64-run CCD, 其中设计列由 其在饱和设计中的序号来表示、C表示纯净两阶交互效应的个数. 值得说明的是, 现有研究中文献 [11]

表 2 32-run 第三类 CCD

设计	$m_1$	$G_1$	$G_2$	C
$2^{7-2.4}$	4	$\{4, 8, 16, 29\}$	$\{1, 2, 3\}$	18
$2^{9-4}$	3	$\{1, 2, 4\}$	$\{8, 15, 16, 23, 24, 31\}$	21
$2^{16-11}$	1	{1}	$\{2,4,6,8,10,12,14,16,18,20,22,24,26,28,30\}$	15

表 3 64-run 第三类 CCD

设计	$m_1$	$G_1$	$G_2$	C
$2^{9-3}$	6	$\{1, 2, 4, 8, 16, 32\}$	$\{15, 51, 60\}$	33
$2^{10-4}$	5	$\{1, 2, 4, 8, 16\}$	$\{30, 32, 39, 57, 62\}$	35
$2^{17-11}$	3	$\{1, 2, 4\}$	$\{8, 15, 16, 23, 24, 31, 32, 39, 40, 47, 48, 55, 56, 63\}$	45
$2^{32-26}$	1	{1}	$\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32,$	31
			34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62	

表 4 64-run 第四类 CCD

设计	$m_1$	$G_1$	$G_2$	C
$2^{11-5}$	8	$\{5, 9, 17, 30, 33, 48, 53, 57\}$	$\{1, 2, 3\}$	31
$2^{17-11.6}$	16	$\{1, 2, 4, 7, 8, 11, 13, 14, 16, 19, 21, 22, 25, 26, 28, 32\}$	<b>{63</b> }	31

给出了相对广泛的  $2^{m-p}$  设计表. 我们构造的这些设计中, 除  $2^{7-2.4}$  和  $2^{17-11.6}$  之外其余设计均未被列出, 但它们比现有设计具有更多的纯净两阶交互效应. 表 2 和 3 给出的第三类 CCD, 所有的纯净两阶交互效应均属于  $\{G_1 \times G_1, G_1 \times G_2\}$ . 当设计参数相同时, 与 CCP [3] 相比这些 CCD 包含更多纯净两阶交互效应, 其  $G_1$  包含的设计列也更多. 此外, 根据表 2 和 3 中的 CCD, 我们可以通过将  $G_1$  中的列移动到  $G_2$ , 或将  $G_1 \cup G_2$  中的部分列删除的方法得到更多的 CCD.

#### 参考文献

- 1 Addelman S. Symmetrical and asymmetrical fractional factorial plans. Technometrics, 1962, 4: 47–58
- 2 Sun D X. Estimation capacity and related topics in experimental designs. PhD Thesis. Waterloo: University of Waterloo, 1993
- 3 Ke W M, Tang B, Wu H Q. Compromise plans with clear two-factor interactions. Statist Sinica, 2015, 15: 709-715
- 4 Zhao S L, Zhang R C. Compromise  $4^m 2^n$  plans with clear two-factor interactions. Acta Math Appl Sin Engl Ser, 2010, 26: 99–106
- 5 Ye S L, Wang D Y, Zhang R C. Partial aliased effect number pattern and selection of optimal compromise designs. Metrika, 2019, 82: 269–293
- 6 Li P F, Zhao S L, Zhang R C. A theory on constructing  $2^{n-m}$  designs with general minimum lower order confounding. Statist Sinica, 2011, 21: 1571–1589
- 7 Zhang R C, Cheng Y. General minimum lower order confounding designs: An overview and a construction theory. J Statist Plann Inference, 2010, 140: 1719–1730
- 8 Chen H G, Hedayat A S.  $2^{n-m}$  Designs with resolution III or IV containing clear two-factor interactions. J Statist Plann Inference, 1998, 75: 147–158
- 9 Wu C F J, Hamada M. Experiments: Planning, Analysis and Parameter Design Optimization. New York: John Wiley & Sons. 2010
- 10 Mukerjee R, Wu C F J. A Modern Theory of Factorial Designs. New York: Springer, 2006
- 11 Chen J, Sun D X, Wu C F J. A catalogue of two-level and three-level fractional factorial designs with small runs.

Internat Statist Rev, 1993, 61: 131-145

## Some theoretical results and constructions of compromise designs with specified sets of clear effects

Dongying Wang & Runchu Zhang

Abstract In many practical experiments, the experimenter is only interested in the estimation of a few of specified factor effects. The designs that are suitable to achieve such purposes are called compromise plans, which were first studied by Addelman (1962). Clear compromise plans allow the estimation of specified factor effects, but the resolution is limited to IV. In this paper, we study four classes of compromise designs, with specified effects denoting as  $\{G_1, G_1 \times G_1\}$ ,  $\{G_1, G_1 \times G_1, G_2 \times G_2\}$ ,  $\{G_1, G_1 \times G_1, G_1 \times G_2\}$  and  $\{G_1, G_1 \times G_2\}$ , respectively. Clear compromise design, that is compromise design with these specified effects being clear, can be of resolution III or IV. We give the existence and characteristics of clear compromise designs of classes one to four, and provide some theoretic results and constructions of clear compromise designs. Compared with the clear compromise plan, the proposed clear compromise design has more clear two factor interactions.

Keywords compromise design, clear effect, fractional factorial design

MSC(2020) 62K15, 62K05 doi: 10.1360/SCM-2018-0298