

摘要

针对同时包含定性和定量因子的试验, Deng, Hung 和 Lin(2015)提出了边际耦合设计, 这类设计比分片拉丁超立方体设计更加经济、可行. 很多学者都在改进边际耦合设计的优良性上做出了成果, 但是它们大多只考虑定性因子是等水平的情况. 混水平定性因子边际耦合设计的构造研究较少, 其构造远比定性因子是等水平的情况更加复杂. 因而, 如何构造混水平定性因子边际耦合设计是一个值得探讨的课题. 同时, 如何保证混水平边际耦合设计的定量部分实现二维或更高维投影分层性也是研究难点. 从保证定量因子部分具有低维投影分层性的角度构造混水平定性因子边际耦合是值得进一步关注的课题. 本学位论文将针对上述课题展开研究.

对于混水平定性因子边际耦合设计的构造, 本文提供了两种思路. 首先, 基于完全因析设计构造出一系列混水平定性因子边际耦合设计. 这些新设计中, 有的设计可以容纳任意不同水平的两个定性因子; 有的设计可以容纳多种不同水平的定性因子; 有的设计实现了定量部分具有二维和三维投影分层性. 其次, 基于可分解正交表构造出混水平定性因子边际耦合设计, 对定性因子部分的设计分三种情况进行了讨论. 两种思路下所得到的设计有各自优点, 定性部分的设计是一个完全因析设计时, 该设计可以容纳多种不同水平的定性因子, 除此之外, 部分算法可以实现定量部分的二维和高维投影分层性. 而定性部分的设计是一个可分解的混水平正交表时, 相同试验次数下, 可以容纳定性因子数量更多.

最后, 分别利用组合设计与本文构造的设计建立回归模型, 通过预测的均方误差对比发现: 边际耦合设计的拟合效果优于组合设计的拟合效果.

关键词: 混水平正交表; 边际耦合设计; 可分解正交表; 空间填充性

Abstract

For experiment that include both qualitative and quantitative factors, Deng, Hung, and Lin (2015) proposed marginally coupled designs(MCDs), which are more economical and feasible than sliced Latin hypercube designs. Many researchers have made efforts to improve the performance of MCDs, but most of them only consider cases where qualitative factors are equal-level. There is relatively little research on the construction of MCDs for mixed-level qualitative factors, and their construction is much more complex than the construction of MCDs for equal-level qualitative factors. Therefore, how to construct MCDs with mixed-level qualitative factors is a subject worth exploring. At the same time, it is also a challenge to construct MCDs with mixed-level qualitative factors when the designs for quantitative factors achieve two- or higher-dimensional projections. It is worth considering the construction of MCDs with mixed-level qualitative factors so that the designs for quantitative factors have low-dimensional projections. This thesis will focus on the above subjects.

This paper provides two ideas for the construction of MCDs with mixed-level qualitative factors. Firstly, a series of MCDs with mixed-level qualitative factors are constructed based on full factorial designs. Some of the constructed MCDs can accommodate two qualitative factors at any different level; some designs can accommodate qualitative factors with various levels; in the constructed MCDs, some the designs for quantitative factors achieve two- or three-dimensional projections. Secondly, based on resolvable orthogonal arrays and the qualitative factors in three cases, MCDs with mixed-level qualitative factors are constructed. The designs obtained under the two ideas above have their own advantages. The advantage of the first idea is that the constructed MCDs can accommodate a variety of qualitative factors with different level types, when the designs for qualitative factors are full factorial designs. In addition, some algorithms can achieve two- or higher-dimensional projections. Compared to the first idea, the MCDs under the second idea can accommodate more qualitative factors with the same run size, when the designs for qualitative factors are resolvable orthogonal arrays.

Finally, regression models are established using combination designs and the MCDs constructed in this paper. By comparing the mean square error of predictions, it is found that the fitting effect of MCDs is better than that of combination design.

Key words: mixed-level orthogonal arrays; marginally coupled designs; resolvable arrays; space-filling

目录

摘要	I
ABSTRACT	II
第一章 绪论	1
§1.1 研究背景及现状	1
§1.2 本文主要工作及内容安排	3
§1.3 预备知识	4
§1.3.1 符号说明	4
§1.3.2 定义及命题	4
第二章 混水平定性因子边际耦合设计的构造方法	10
§2.1 基于完全因析设计构造边际耦合设计	10
§2.2 基于可分解正交表构造边际耦合设计	22
第三章 数值模拟	30
§3.1 函数介绍	30
§3.1.1 真实模型	30
§3.2 不同设计的模拟比较	31
§3.2.1 边际耦合设计与组合设计的比较	32
§3.2.2 具有分层性与不具有分层性比较	35
第四章 总结与展望	43
参考文献	44
致谢	47

第一章 绪论

§ 1.1 研究背景及现状

试验是人类认识自然、发现规律的重要途径之一，它被广泛运用于工业、农业等各个领域，具有非常重要的理论和应用价值。按照试验实现方式，试验可以分为实体试验和计算机试验。实体试验是在试验室或者生产现场实现的，这类试验总会存在试验误差；而计算机试验是通过复杂的计算机代码来实现的，可以避免试验误差。现实生活中那些多因素、非线性和真实模型未知的情况，科学的试验设计可以让仿真模型的响应值达到最优或者近似最优。

按因子的个数，试验可以分为单因子试验和多因子试验。一般地，单因子试验的水平数可以适当多取。然而，随着因子个数的增加，试验的复杂程度也相应增加。按照试验的轮次，试验可以分为单一试验和序贯试验。单一试验是指根据一轮试验的数据，分析数据即可得到最终想要的目标，于是就终止试验。然而，在做完试验的一轮分析之后，往往需要追加一些试验，以求修改、验证之前的做法，从而通过一轮又一轮试验得到或者逼近最优解，这种试验方法称为序贯试验^[1]。按照因素的约束条件，试验可以分为无约束试验和混料试验。所谓无约束试验指的是，因子可以自由选取试验的值，不受其他因素约束。在一些试验中，因素之间的取值会互相影响，最典型的的就是混料试验。按照试验的实现方式，可以分为实体试验和计算机试验。传统的实体试验是在实验室或者工农业生产现场进行。由于在实体试验中总是存在随机误差，故实体试验需要遵循三个基本原则：重复、随机化、分区组。有些实体试验需要耗费大量的人力、物力和时间成本。随着现代科学技术的发展，试验所需要付出人力、物力和时间的代价越来越大，这就导致传统的实体试验在某些领域不太理想，因此，有些时候需要计算机试验来代替实体试验。除此之外，计算机试验的优点是没有随机误差，即输入相同的变量值，输出的响应值也相同。但是，如果盲目输入大量样本数据进行计算机试验，计算代价也是非常大的。因此，有效的试验设计在计算机试验研究中很重要。现实生活中绝大多数的研究对象往往都是多因素影响、非线性和未知其真实模型的，一个理想的计算机试验设计应该保证设计的样本点尽可能均匀地分布在整个空间上，这类设计被称为空间填充设计。空间填充设计非常适合计算机试验设计。三类最常用的空间填充设计为拉丁超立方体设计、均匀设计以及最大最小或者最小最大距离设计^[2]。

计算机试验往往包含大量的输入变量，而有限个设计点有时不能很好地覆盖整个高维输入空间。一种合理的方式是去寻求具有低维空间填充性的设计，通常利用正交表或者类似的结构来构造这类设计。McKay, Beckman 和 Conover^[3]首次提出了拉丁

超立方体设计, 这种设计被广泛运用到计算机试验的原因是该设计是一个强度为 1 的正交表, 水平数与试验次数相等, 实现了一维的最大分层: 当投影到任意一个维度时, 等间隔的设计点确保了在取值范围内均匀分布. 为了实现更高维的投影均匀性, Tang^[4]提出基于正交表的拉丁超立方体设计. 正交性作为定量因子试验设计的一个重要性质, 许多学者都对正交拉丁超立方体的构造进行研究, 如 Sun, Liu 和 Lin^[5], Lin, Mukerjee 和 Tang^[6], Wang, Sun, Lin 等^[7]. 为了实现设计在全空间上的均匀性, Johnson, Moore 和 Ylvisaker^[8]提出最大最小距离准则用作衡量一个设计在全空间上的投影均匀性, Joseph 和 Hung^[9]提出正交最大最小距离拉丁超立方体设计, Wang, Qian 和 Xu^[10]提出最优最大最小 L_1 距离拉丁超立方体设计.

上面这些设计只考虑仅包含定量因子的计算机试验, 对于既包含定量又包含定性因子的计算机试验, Qian 和 Wu^[11]提出分片空间填充设计. 由于分片空间填充设计不具有一维投影的最大分层性, Qian^[12]首次提出了分片拉丁超立方体设计. 后续许多学者关于分片拉丁超立方体构造方法进行优化, 如 Huang, Yang 和 Liu^[13], 王乐^[14], 张静^[15]. 分片拉丁超立方体设计的缺点是: 当定性因子水平组合数量比较大时, 会导致试验次数变得很大. 针对这种问题, Deng, Hung 和 Lin^[16]首次提出边际耦合设计, 这种设计不仅在试验次数上更加经济有效, 而且具有较好空间填充性. Deng, Hung 和 Lin^[16]中给出了这种设计存在的充分必要条件和构造方法. 该论文中给出了五种构造方法来构造不同类型的边际耦合设计: (i) 定性因子的水平都为 s_1 , 试验次数为 s_1^2 的边际耦合设计; (ii) 定性因子的水平都为 s_1 , 试验次数为 λs_1^2 的边际耦合设计; (iii) 定性因子部分的设计 D_1 是一个混水平正交表的边际耦合设计; (iv) 定性因子的水平组合重复和不重复时, 边际耦合设计的构造; (v) 定性部分的因子都为 2 水平的边际耦合设计. 通过这些构造方法所得到的边际耦合设计中关于定量因子的设计在二维和高维上存在聚集点. 为了避免聚集点, 增强定量因子部分的设计在二维投影上的均匀性, He, Lin 和 Sun^[17]提出了三种构造边际耦合设计的方法. He, Lin, Sun 等^[18]还给出了对于定性因子是两水平时, 定量因子的设计不仅保证了二维和高维投影均匀性, 而且使得该设计能容纳更多的定量因子的边际耦合设计的构造方法. He, Lin 和 Sun^[19]将 He, Lin, Sun 等^[18]中定性因子为两水平推广到任何素数次幂的情况. 正交性是衡量拉丁超立方体设计的重要性质, 而边际耦合设计中的定量部分作为一个拉丁超立方体设计, 很少有人考虑其正交性. Zhou, Yang 和 Liu^[20]提出了 5 种构造正交边际耦合设计的方法, 该设计的定量因子部分是正交拉丁超立方体设计, 其中部分构造方法得到的设计使得定量因子在二维投影上具有分层性. 在一个边际耦合设计中, 为了实现定性因子的水平组合对应到定量部分是一个小的拉丁超立方体设计, Yang, Lin, Zhou 等^[21]给出了双耦合设计的定义及构造方法.

上述改进边际耦合设计的论文都是针对定性因子设计是等水平的情况, 而定性

因子是混水平的边际耦合设计很少有人研究. Deng, Hung 和 Lin^[16]中的命题 2 给出了当定性因子部分的设计 D_1 是 $OA(n, s_1^{q_1} s_2^{q_2}, 2)$ 且 $s_1 = \alpha s_2$ 时, 边际耦合设计 $D = (D_1, D_2)$ 存在的充分必要条件. 根据此充分必要条件构造的边际耦合设计具有一定局限性: (i) 定性因子的不同水平之间必须存在一定的倍数关系; (ii) 定性因子的水平只能有两类. 因此, 定性因子是混水平时, 水平更加灵活的边际耦合设计的构造非常值得研究. 本文通过完全因析设计得到混水平定性因子边际耦合设计的定性部分, 当定性因子包含两种不同水平 s_1 和 s_2 时, 它们之间有无倍数关系均可. 还实现了定性因子可以是多种不同水平. 接下来, 由可分解正交表构造混水平定性因子的边际耦合设计, 新设计的试验次数更少或者新设计可以容纳更多定性因子.

§ 1.2 本文主要工作及内容安排

本学位论文针对混水平定性因子边际耦合设计的构造和应用, 给出了一系列算法和数值模拟. 论文内容如下:

第一章是引言, 包括研究背景及意义、相关定义和符号说明.

第二章是混水平定性因子边际耦合设计的构造方法. 第一种思路是利用完全因析设计得到混水平定性因子边际耦合设计的定性部分. 相关算法得到定性因子部分的设计可以容纳任意两种不同水平的两个因子, 或者可以容纳多个不同水平的定性因子, 或者定量部分通过一个差阵和饱和正交表实现分层性. 第二种思路是利用正交表的可分解性构造来构造边际耦合设计的定性部分, 并分三种情况进行讨论. 两种思路下所得到的设计有各自优点, 定性部分的设计是一个完全因析设计时, 该设计可以容纳多种不同水平的定性因子, 除此之外, 部分算法可以实现定量部分的二维和高维投影分层性. 而定性部分的设计是一个具有可分解性的混水平正交表时, 试验次数会更少或者相同试验次数可以容纳更多定性因子.

第三章是数值模拟. 在建模过程中, 为了体现一个设计的空间填充性对拟合效果的影响, 使用两个不同的试验设计, 即边际耦合设计和组合设计, 采用相同的拟合模型对同一个真实模型进行拟合. 最后, 通过计算均方误差可知, 空间填充性更好的边际耦合设计可以使拟合精度更高.

第四章对本文的研究内容进行总结并提出未来可以继续研究方向.

§ 1.3 预备知识

§ 1.3.1 符号说明

符号	说明
$D_{n \times q}$	n 行 q 列的设计矩阵
d_{ij}	矩阵 D 第 i 行第 j 列的元素
$d_{i\cdot}$	矩阵 D 第 i 行所有元素
$d_{\cdot j}$	矩阵 D 第 j 列所有元素
$d_i^{(k)}$	列向量 d_i 的第 k 个元素
C_n	各分量取值为 c 的 n 维列向量
$GF(s)$	s 阶伽罗瓦域, 其元素来自集合 $\{\alpha_0 = 0, \alpha_1, \alpha_2, \dots, \alpha_{s-1}\}$
$\lfloor x \rfloor$	小于或者等于 x 的最大整数
$a \bmod b$	a 除以 b 取余
$\oplus$	矩阵运算符, Kronecker 和
$\otimes$	矩阵运算符, Kronecker 积

§ 1.3.2 定义及命题

定义 1.1^[22] 设计矩阵 $D_{n \times q}$, 其元素均来自集合 $\{0, 1, \dots, s-1\}$, 若 $D_{n \times q}$ 满足如下条件, 称 $D_{n \times q}$ 是一个试验次数为 n , 因子个数为 q , 因子水平为 s , 强度为 t 的等水平正交表, 记作 $OA(n, s^q, t)$:

- (i) 取任意一行, 集合 $\{0, 1, \dots, s-1\}$ 中的元素都会出现, 且出现次数相等;
- (ii) 取任意 t 列, 这 t 列所有水平组合都会出现, 且出现次数相等.

例 1.1 表 1-1 中的设计是一个等水平正交表 $OA(9, 3^4, 2)$, 试验次数 $n=9$, 因子个数 $q=4$, 因子水平数 $s=3$, 强度 $t=2$.

表 1-1 等水平正交表

$OA(9, 3^4, 2)$			
0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

定义 1.2^[22] 设计矩阵 $D_{n \times q} = (A_1, A_2, \dots, A_m)$, 若 $D_{n \times q}$ 满足如下条件, 称 $D_{n \times q}$ 为混水平正交表, 记作 $OA(n, s_1^{q_1} s_2^{q_2} \dots s_m^{q_m}, t)$:

- (i) 对于 $i=1,2,\dots,m$, 每个 A_i 均是一个正交表 $OA(n, s_i^{q_i}, 1)$ 且 $s_i \neq s_j, 1 \leq i \neq j \leq m$;
 (ii) 从 $D_{n \times q}$ 中任取 t 列, 这 t 列所有水平组合都会出现, 且出现次数相等.

例 1.2 表 1-2 中的设计是一个混水平正交表 $OA(8, 2^4 4^1, 2)$, 试验次数 $n=8$, 因子个数 $q=5$, 因子水平数 $s_1=2, s_2=4$, 强度 $t=2$.

表 1-2 混水平正交表

$OA(8, 2^4 4^1, 2)$				
0	0	0	0	0
0	1	0	1	1
1	0	1	1	2
1	1	1	0	3
0	1	1	1	0
0	0	1	0	1
1	1	0	0	2
1	0	0	1	3

定义 1.3^[22] 令 A 是一个等水平正交表 $OA(n, s^q, t)$ 且 $t \geq 2$, 若将 A 进行行排列, 可化为 $A = (A_1^T, A_2^T, \dots, A_{n/\alpha s}^T)^T$, 使得 A_i 是一个 $OA(\alpha s, s^q, 1)$, 其中 $1 \leq i \leq n/\alpha s$, 则称 A 是 α -可分解正交表. 当 $\alpha=1$ 时, 称 A 是完全可分解正交表.

例 1.3 表 1-3 中的设计 A 是一个等水平正交表 $OA(16, 4^4, 2)$, 试验次数 $n=16$, 因子个数 $q=4$, 因子水平数 $s=4$, 强度 $t=2$. 令 A_i 是由 A 的第 $4(i-1)+1, 4(i-1)+2, 4(i-1)+3, 4(i-1)+4$ 行组成, $i=1,2,3,4$, 则 A 可表示为 $A = (A_1^T, A_2^T, A_3^T, A_4^T)^T$, 显然, A_i^T 是一个强度为 1 的正交表 $OA(4, 4^4, 1)$, $1 \leq i \leq 4$. 故 A 是一个完全可分解正交表.

定义 1.4^[3] 设计矩阵 $D_{n \times p}$ 的每一列都是集合 $\{0, 1, \dots, n-1\}$ 中所有元素的一个排列, 称 $D_{n \times p}$ 为拉丁超立方体设计, 记作 $LHD(n, p)$.

例 1.4 表 1-4 中的设计是一个拉丁超立方体设计 $LHD(6, 3)$, 其中试验次数 $n=6$, 因子个数 $p=3$.

定义 1.5^[16] 设计矩阵 $D = (D_1, D_2)$, 其中 D_1 为定性因子部分的设计, D_2 为定量因子部分的设计, 如果设计 D 满足如下条件, 称 $D = (D_1, D_2)$ 是边际耦合设计:

(i) D_1 是一个 $OA(n, s_1^{q_1} s_2^{q_2} \dots s_m^{q_m}, t)$, D_2 是一个 $LHD(n, p)$;

(ii) 对于定性因子部分的设计 D_1 , 其任意一个因子的每一个水平, 对应到 D_2 中的行构成一个小的拉丁超立方体设计.

表 1-3 完全可分解正交表

$OA(16, 4^4, 2)$			
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
0	1	2	3
1	0	3	2
2	3	0	1
3	2	1	0
0	2	3	1
1	3	2	0
2	0	1	3
3	1	0	2
0	3	1	2
1	2	0	3
2	1	3	0
3	0	2	1

表 1-4 拉丁超立方体设计

$LHD(6, 3)$		
0	1	5
1	2	3
2	5	2
3	3	4
4	4	0
5	0	1

He, Lin 和 Sun(2017)的命题 1 给出了边际耦合设计 $D = (D_1, D_2)$ 存在的充分必要条件, 如下所示.

命题 1.1^[17] 设 D_1 是一个等水平正交表 $OA(n, s^m, 2)$, D_2 是一个拉丁超立方体设计 $LHD(n, p)$, 则设计 $D = (D_1, D_2)$ 是边际耦合设计的充分必要条件为 $(D_1, \lfloor d_j/s \rfloor)$ 是一个 $OA(n, s^m(n/s), 2)$, 其中 d_j 是 D_2 的第 j 列, $j=1, 2, \dots, p$, m 为任意正整数.

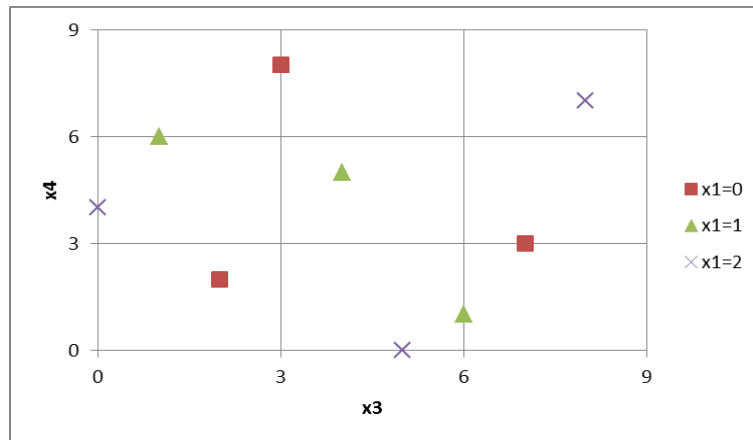
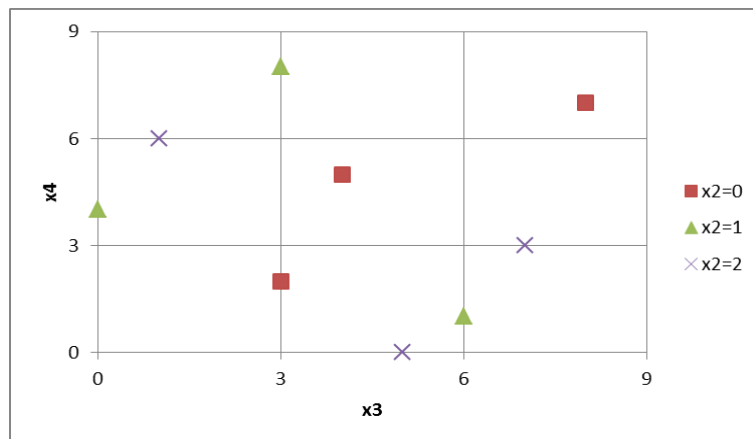
本文将命题 2.1 作简单推广, 可以得到如下命题 2.1.

命题 1.2 设 $D_1 = (D_{11}, D_{12}, \dots, D_{1m})$ 是一个混水平正交表 $OA(n, s_1^{q_1} s_2^{q_2} \dots s_m^{q_m}, 2)$, 其中 D_{1i} 是一个等水平正交表 $OA(n, s_i^{q_i}, 2)$, $i=1, 2, \dots, m$. 设 D_2 是一个拉丁超立方体设计 $LHD(n, p)$, 则设计 $D = (D_1, D_2)$ 是边际耦合设计的充分必要条件是对 $i=1, 2, \dots, m$, $(D_{1i}, \lfloor d_j/s_i \rfloor)$ 是一个 $OA(n, s_i^{q_i}(n/s_i), 2)$, 其中 d_j 是 D_2 的第 j 列, $j=1, 2, \dots, p$, m 为任意正整数.

例 1.5 表 1-5 中的设计是一个等水平定性因子边际耦合设计 $D=(D_1, D_2)$, 试验次数 $n=9$, D_1 包含了两个定性因子 x_1 和 x_2 , D_2 包含了两个定量因子 x_3 和 x_4 . 图 1-1 和图 1-2 是关于 x_3 和 x_4 的散点图, 两张图分别是按照 D_1 中第一列因子 x_1 和第二列因子 x_2 取相同水平时, D_2 中对应的坐标点进行不同标记. 将横纵坐标三等分后, 三个区间 $[0,3)$ 、 $[3,6)$ 、 $[6,9)$ 分别视作低、中、高三个水平, 可以看出, D_1 中任何一个定性因子的每一个水平, 对应任何一个定量因子的取值在低中高三个范围内都有点存在.

表 1-5 等水平定性因子边际耦合设计

D_1		D_2	
0	0	2	2
0	1	3	8
0	2	7	3
1	0	4	5
1	1	6	1
1	2	1	6
2	0	8	7
2	1	0	4
2	2	5	0

图 1-1 定量因子按 x_1 分为 3 类的散点图图 1-2 定量因子按 x_2 分为 3 类的散点图

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/966102211240010205>