5B-04

直交計画を用いた実験における複素空間での計算の高速化に関する一考察

浮田 善文† 齋藤 友彦†† 松嶋 敏泰‡‡
† 横浜商科大学 商学部 経営情報学科
†† 湘南工科大学 工学部 情報工学科
‡‡ 早稲田大学 基幹理工学研究科

1 はじめに

実験計画法は様々な要因が目標となる特性値にどのような影響を与えているかなどを明らかにする統計的実験手法である．実験計画法のモデルは，全てのパラメータが独立となる複素空間上の直交基底関数モデルで表現することにより，プログラミングを容易化することができる．しかし，因子数が多く，主効果や交互作用効果を多く含む場合には，パラメータ推定に時間がかかることが問題点としてあげられる．そこで本稿では，直交計画を利用した大規模な実験において，パラメータ推定に高速フーリエ変換を利用することで計算の高速化が可能となることを示す．

2 準備

2.1 ガロア体

q 個の元からなる体をガロア体 [1] といい，$GF(q)$ で表す．なお，ガロア体の位数 q は，$q = p^m$ でなければならない．ただし，p は素数である．

例 1. $GF(2^2) = GF(4)$ は体を成す．その演算表を表 1 に示す．

表 1: $GF(4)$ 上の加法と乗法表 [1]

+	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

2.2 ガロア体上の直交基底関数

はじめに，長さ n のベクトルを考え，$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$，$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$，$\boldsymbol{b} = (b_1, b_2, \cdots, b_n)$ とする．ここで，$\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b} \in GF(q)^n$ としておく．

例 2. $q = 4, n = 2$ の場合，$GF(4)^2 = \{00, 01, 02, \cdots, 33\}$ であり，$|GF(4)^2| = 16$ となる．

A Study on Accelerating Computation in Complex Space for Experiments Using Orthogonal Designs
† Yoshifumi Ukita
Yokohama College of Commerce
†† Tomohiko Saito
Shonan Institute of Technology
‡‡ Toshiyasu Matsushima
Waseda University

このとき，$GF(q)^n$ の指標群 [2] は $\{\mathcal{X}_{\boldsymbol{a}} | \boldsymbol{a} \in GF(q)^n\}$ で表すことができ，次式が成立する．

$$\frac{1}{q^n} \sum_{\boldsymbol{x} \in GF(q)^n} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) \mathcal{X}^*_{\boldsymbol{b}}(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{a} = \boldsymbol{b} \\ 0, & \boldsymbol{a} \neq \boldsymbol{b} \end{cases} \tag{1}$$

ここで $\mathcal{X}^*_{\boldsymbol{b}}(\boldsymbol{x})$ は $\mathcal{X}_{\boldsymbol{b}}(\boldsymbol{x})$ の複素共役である．

例 3. $q = 4$ のとき，$\mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) = e^{i\pi \boldsymbol{a} \cdot \boldsymbol{x}/2}$ と書くことができる．ここで，内積 $\boldsymbol{a} \cdot \boldsymbol{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$ は $GF(q)$ 上で計算される．

このとき，任意の関数 $f : GF(q)^n \to C$ (C は複素数体) は，指標の線形結合により，

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in GF(q)^n} f_{\boldsymbol{a}} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) \tag{2}$$

で一意に表現可能で，係数 $f_{\boldsymbol{a}}$ は次式で与えられる．

$$f_{\boldsymbol{a}} = \frac{1}{q^n} \sum_{\boldsymbol{x} \in GF(q)^n} f(\boldsymbol{x}) \mathcal{X}^*_{\boldsymbol{a}}(\boldsymbol{x}) \tag{3}$$

3 実験計画法

本稿では，実験計画法のモデルを，全てのパラメータが独立となる直交基底関数モデル [3] で表現する．

3.1 必要な用語の説明 [3]

はじめに，説明に必要となる記号を定義しておく．長さ n のベクトル $\boldsymbol{a}$ の 0 でない成分の数を $\boldsymbol{a}$ のハミング重みといい，$H_w(\boldsymbol{a})$ で表すものとする．

$F_1, F_2, \ldots, F_n$ をモデルに含まれる n 個の因子とする．本稿では，n 個の因子は全て同じ水準数を持つものとし，その水準数を q としておく．

ここで，集合 $A \subseteq \{0,1\}^n$ はモデルに含まれる一般平均，因子，因子間の交互作用を表す集合とする．因子 F_i は $H_w(\boldsymbol{a}) = 1, a_i = 1$ となる $\boldsymbol{a}$ により表され，因子 F_i と因子 F_j の交互作用は $H_w(\boldsymbol{a}) = 2, a_i = 1, a_j = 1$ となる $\boldsymbol{a}$ により表されるものとする．また，一般平均は $H_w(\boldsymbol{a}) = 0$ となる $\boldsymbol{a}$ により表されるものとする．

例えば，$A = \{0000, 1000, 0100, 0010, 0001, 1100\}$ の要素は順番に，一般平均，因子 F_1，因子 F_2，因子 F_3，因子 F_4，因子 F_1 と因子 F_2 の交互作用を示している．

3.2 直交基底関数モデル [3]

集合 A に対し，$I_A = \{(b_1a_1, b_2a_2, \ldots, b_na_n)|\boldsymbol{a} \in A, b_i \in \{0,1,\cdots,q-1\}\}$ と定義すると，集合 A に含まれる一般平均は $f_{\boldsymbol{0}}$，因子 F_i の主効果は $\{f_{\boldsymbol{a}}|H_w(\boldsymbol{a}) = 1, a_i \neq 0, \boldsymbol{a} \in I_A\}$，因子 F_i と因子 F_j の交互作用効果は $\{f_{\boldsymbol{a}}|H_w(\boldsymbol{a}) = 2, a_i \neq 0, a_j \neq 0, \boldsymbol{a} \in I_A\}$ により表現される．

水準組合せ $\boldsymbol{x}$ で実験を行うときの特性値を $t(\boldsymbol{x})$ とする．このとき，以下のモデルを仮定する．

$$t(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in I_A} f_{\boldsymbol{a}} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) + \epsilon \tag{4}$$

ここで ϵ は平均 0，分散 σ^2 のガウス確率変数である．

3.3 実験の形式

本稿においては，集合 A は事前に与えられるものとする．このもとで，まず実験する水準組合せの集合である実験計画 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ を決定する．その後，実験を行い，対応する特性値 $t(\boldsymbol{x}_1), \ldots, t(\boldsymbol{x}_N)$ を得る．実験後には，得られた N 個の入出力の対であるデータ集合 $\{(\boldsymbol{x}, t(\boldsymbol{x}))|\boldsymbol{x} \in \boldsymbol{X}\}$ から，各効果の推定を行う．本稿では，同じ大きさの実験計画の中で各効果の不偏推定量の分散を最小にすることが知られている直交計画を用いる．

3.4 直交計画 [3]

$v(\boldsymbol{a}) = \{i|a_i \neq 0, 1 \leq i \leq n\}$ とする．集合 A に対し，$k_A \times n$ 行列 G_A を

$$G_A = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{1,2} & \cdots & g_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k_A,1} & g_{k_A,2} & \cdots & g_{k_A,n} \end{bmatrix} \tag{5}$$

$g_{i,j} \in GF(q)(1 \leq i \leq k_A, 1 \leq j \leq n)$ とし，以下の二つの条件を満たす行列とする．

1. $\boldsymbol{g}_{\cdot j}$ を G_A の j 番目の列とすると，任意の $\boldsymbol{a}', \boldsymbol{a}'' \in A$ に対し，集合 $\{\boldsymbol{g}_{\cdot j}|j \in v(\boldsymbol{a}' + \boldsymbol{a}'')\}^*$ が $GF(q)$ 上で線形独立
2. $\boldsymbol{g}_{i\cdot}$ を G_A の i 番目の行とすると，集合 $\{\boldsymbol{g}_{i\cdot}|i \leq i \leq k_A\}$ が $GF(q)$ 上で線形独立

このとき，集合 A に対する直交計画 $\boldsymbol{X}_A$ は $\boldsymbol{X}_A = \{\boldsymbol{x}|\boldsymbol{g} = \boldsymbol{r}G_A, \boldsymbol{r} \in GF(q)^{k_A}\}$ で与えられ，実験回数 N は $N = |\boldsymbol{X}_A| = q^{k_A}$ である．

3.5 各効果の推定

実験回数 q^{k_A} の直交計画 X_A を用いるとき，$f_{\boldsymbol{a}}$ の不偏推定量 $\hat{f}_{\boldsymbol{a}}$ は次式で与えられる．

$$\hat{f}_{\boldsymbol{a}} = \frac{1}{q^{k_A}} \sum_{\boldsymbol{x} \in \boldsymbol{X}_A} \mathcal{X}^*_{\boldsymbol{a}}(\boldsymbol{x}) t(\boldsymbol{x}) \tag{6}$$

* $\boldsymbol{a}' = (a'_1, a'_2, \cdots, a'_n), \boldsymbol{a}'' = (a''_1, a''_2, \cdots, a''_n)$ とすると，ベクトル間の加法は，排他的論理和 $\oplus$ を用い，$\boldsymbol{a}' + \boldsymbol{a}'' = (a'_1 \oplus a''_1, a'_2 \oplus a''_2, \cdots, a'_n \oplus a''_n)$ で定義される．

4 複素空間での各効果の推定の高速化

本稿では，式 (6) で与えられる各効果の推定を，高速フーリエ変換 (Fast Fourier Transform: FFT) により行うことで計算の高速化を行う．なお，以下では，水準数 q は $q = 2^r$ (r: 正整数) とする．

実験回数 $N = q^{k_A}$ の直交計画について実験を行い，得られた $\{t(\boldsymbol{x})|\boldsymbol{x} \in \boldsymbol{X}\}$ から，各効果の推定を行う場合を考える．このとき，k_A 次元離散フーリエ変換 (Discrete Fourier Transform: DFT) であれば，複素乗算回数は q^{2k_A} となる．一方，k_A 次元ベクトル・ラディックス FFT であれば，複素乗算回数 V_M は，

$$V_M = \frac{q^{k_A} - 1}{2^{k_A}} q^{k_A} \log_2 q \tag{7}$$

で与えられ [4]，複素乗算回数の低減により，各効果を推定する計算の高速化が可能となることがわかる．

また，実際のプログラミングを行う場合についても考える．n 次元 FFT の計算には，効率的な数値計算のライブラリである NumPy [5] の numpy.fft.fftn 関数を使用することができる．このため，直交基底関数モデルは，ライブラリを利用したプログラミングに適したモデルであることがわかる．

5 おわりに

本稿では，実験計画法において，高速フーリエ変換を利用することで各効果の推定の計算の高速化が可能となることを示した．さらに，多次元 FFT の計算にライブラリを利用することにより，プログラミングが容易化できることを示した．

謝辞

本研究は JSPS 科研費 JP17K00316 の助成を受けたものです．

参考文献

[1] 平澤茂一, 西島利尚:符号理論入門, 培風館 (1999).

[2] Stein, E. M. and Shakarchi, R.: Fourier Analysis, Princeton University Press (2003).

[3] Y. Ukita, T. Saito, T. Matsushima and S. Hirasawa: A Note on a Sampling Theorem for Functions over $GF(q)^n$ Domain, *IEICE Trans. Fundam.*, Vol.E93-A, No.6, pp.1024-1031 (2010).

[4] 佐川雅彦, 貴家仁志：高速フーリエ変換とその応用, 昭晃堂 (1992).

[5] NumPy, available from ⟨https://numpy.org/⟩ (2021).