数理解析研究所講究録
第 1889 巻 2014 年 42-51

Markov chain Monte Carlo methods for the regular two-level fractional factorial designs and cut ideals

Satoshi Aoki*† ‡, Hidefumi Ohsugi§†and Takayuki Hibi¶†

Abstract

It is known that a Markov basis of the binary graph model of a graph G corresponds to a set of binomial generators of cut ideals $I_{\widehat{G}}$ of the suspension $\widehat{G}$ of G. In this note, we give another application of cut ideals to statistics. We show that a set of binomial generators of cut ideals is a Markov basis of some regular two-level fractional factorial design. As application, we give a Markov basis of degree 2 for designs defined by at most two relations. This note is a summary of the paper [2].

1 Introduction

In the paper [2], following the Markov chain Monte Carlo approach in the designed experiments by [3], we give a new results on the correspondence between the regular two-level design and the algebraic concept, namely *cut ideals* defined in [10]. This note is a summary of the paper [2]. Because the Markov bases are characterized as the generators of well-specified toric ideals and are studied not only by statisticians but also by algebraists, it is valuable to connect statistical models to known class of toric ideals. In this note, we give a fundamental fact that the generator of cut ideals can be characterized as the Markov bases for the testing problems of log-linear models for the two-level regular fractional factorial designs.

2 Markov chain Monte Carlo method for regular two-level fractional factorial designs

In this section we introduce Markov chain Monte Carlo methods for testing the fitting of the log-linear models for regular two-level fractional factorial designs with count observations. Suppose we have nonnegative integer observations for each run of a regular

*Graduate School of Science and Engineering (Science Course), Kagoshima University.
†JST, CREST.
‡Email: aoki@sci.kagosgima-u.ac.jp
§Department of Mathematics, College of Science, Rikkyo University.
¶Department of Pure and Applied Mathematics, Graduate School of Information Science and Technology, Osaka University.

Table 1: Design and number of defects y for the wave-solder experiment

	Factor							y		
Run	**A**	**B**	**C**	**D**	**E**	**F**	**G**	1	2	3
1	1	1	1	1	1	1	1	13	30	26
2	1	1	1	2	2	2	2	4	16	11
3	1	1	2	1	1	2	2	20	15	20
4	1	1	2	2	2	1	1	42	43	64
5	1	2	1	1	2	1	2	14	15	17
6	1	2	1	2	1	2	1	10	17	16
7	1	2	2	1	2	2	1	36	29	53
8	1	2	2	2	1	1	2	5	9	16
9	2	1	1	1	2	2	1	29	0	14
10	2	1	1	2	1	1	2	10	26	9
11	2	1	2	1	2	1	2	28	173	19
12	2	1	2	2	1	2	1	100	129	151
13	2	2	1	1	1	2	2	11	15	11
14	2	2	1	2	2	1	1	17	2	17
15	2	2	2	1	1	1	1	53	70	89
16	2	2	2	2	2	2	2	23	22	7

fractional design. For simplicity, we also suppose that the observations are counts of some events and only one observation is obtained for each run. This is natural for the settings of Poisson sampling scheme, since the set of the totals for each run is the sufficient statistics for the parameters. We begin with an example.

Example 1 (Wave-soldering experiment). Table 1 is a 1/8 fraction of a full factorial design (i.e., a 2^{7-3} fractional factorial design) defined from the defining relation

$$\mathbf{ABDE} = \mathbf{ACDF} = \mathbf{BCDG} = \mathbf{I}, \tag{1}$$

and response data analyzed in [4] and reanalyzed in [7]. In Table 1, the observation y is the number of defects arising in a wave-soldering process in attaching components to an electronic circuit card. In Chapter 7 of [4], he considered seven factors of a wave-soldering process: (A) prebake condition, (B) flux density, (C) conveyer speed, (D) preheat condition, (E) cooling time, (F) ultrasonic solder agitator and (G) solder temperature, each at two levels with three boards from each run being assessed for defects. The aim of this experiment is to decide which levels for each factors are desirable to reduce solder defects.

Because we only consider designs with a single observation for each run in [2], we focus on the totals for each run in Table 1. We also ignore the second observation in run 11, which is an obvious outlier as pointed out in [7]. Therefore the weighted total of run 11 is $(28 + 19) \times 3/2 = 70.5 \simeq 71$. By replacing 2 by -1 in Table 1, we rewrite $k \times p$ design matrix as D, where each element is $+1$ or -1. Consequently, we have

$$D = \begin{pmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & +1 & -1 & -1 \\ & & & \vdots & & & \\ -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 69 \\ 31 \\ 55 \\ \vdots \\ 212 \\ 52 \end{pmatrix}.$$

In [2], we consider designs of p factors with two-level. We write the observations as $\mathbf{y} = (y_1, \ldots, y_k)'$, where k is the run size and $'$ denotes the transpose. Write the design matrix $D = (d_{ij})$, where $d_{ij} \in \{-1, 1\}$ is the level of the j-th factor in the i-th run for $i = 1, \ldots, k, j = 1, \ldots, p$.

In this case it is natural to consider the Poisson distribution as the sampling model, in the framework of generalized linear models ([9]). The observations $\mathbf{y}$ are realizations from k Poisson random variables $Y_1, \ldots, Y_k$, which are mutually independently distributed with the mean parameter $\mu_i = E(Y_i), i = 1, \ldots, k$. We call the log-linear model written by

$$\log \mu_i = \beta_0 + \beta_i d_{i1} + \cdots + \beta_p d_{ip}, \; i = 1, \ldots, k \tag{2}$$

as the main effect model in [2]. The equivalent model in the matrix form is $(\log \mu_1 \; \cdots \; \log \mu_k)' = M\beta$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{1} = (1, \ldots, 1)'$ and

$$M = \begin{pmatrix} \mathbf{1} & D \end{pmatrix}. \tag{3}$$

We call the $k \times (p+1)$ matrix M a *model matrix* of the main effect model. The interpretation of the parameter β_j in (2) is the parameter contrast for the main effect of the j-th factor. To consider the models including various interaction effects, see [3].

To judge the fitting of the main effect model (2), we can perform various goodness-of-fit tests. In the goodness-of-fit tests, the main effect model (2) is treated as the null model, whereas the saturated model is treated as the alternative model. Under the null model (2), β is the nuisance parameter and the sufficient statistic for β is given by $M'\mathbf{y} = (\sum_{i=1}^k y_i, \sum_{i=1}^k d_{i1} y_i, \ldots, \sum_{i=1}^k d_{ip} y_i)'$. Then the conditional distribution of $\mathbf{y}$ given the sufficient statistics is written as

$$f(\mathbf{y} \mid M'\mathbf{y} = M'\mathbf{y}^o) = \frac{1}{C(M'\mathbf{y}^o)} \prod_{i=1}^k \frac{1}{y_i!}, \tag{4}$$

where $\mathbf{y}^o$ is the observation count vector and $C(M'\mathbf{y}^o)$ is the normalizing constant determined from $M'\mathbf{y}^o$ written as

$$C(M'\mathbf{y}^o) = \sum_{\mathbf{y} \in \mathcal{F}(M'\mathbf{y}^o)} \left(\prod_{i=1}^k \frac{1}{y_i!} \right), \tag{5}$$

and

$$\mathcal{F}(M'\mathbf{y}^o) = \{\mathbf{y} \mid M'\mathbf{y} = M'\mathbf{y}^o, \; y_i \text{ is a nonnegative integer for } i = 1, \ldots, k\}. \tag{6}$$

Note that by sufficiency the conditional distribution does not depend on the values of the nuisance parameters.

In [2], we consider various goodness-of-fit tests based on the conditional distribution (4). There are several ways to choose the test statistics. For example, the likelihood ratio statistic

$$T(\mathbf{y}) = G^2(\mathbf{y}) = 2\sum_{i=1}^{k} y_i \log \frac{y_i}{\hat{\mu}_i} \tag{7}$$

is frequently used, where $\hat{\mu}_i$ is the maximum likelihood estimate for μ_i under the null model (i.e., fitted value). Note that the traditional asymptotic test evaluates the upper probability for the observed value $T(\mathbf{y}^o)$ based on the asymptotic distribution χ^2_{k-p-1}. However, since the fitting of the asymptotic approximation may be sometimes poor, we consider Markov chain Monte Carlo methods to evaluate the p values. Using the conditional distribution (4), the exact p value is written as

$$p = \sum_{\mathbf{y}\in\mathcal{F}(M'\mathbf{y}^o)} f(\mathbf{y} \mid M'\mathbf{y} = M'\mathbf{y}^o)\mathbf{1}(T(\mathbf{y}) \geq T(\mathbf{y}^o)), \tag{8}$$

where

$$\mathbf{1}(T(\mathbf{y}) \geq T(\mathbf{y}^o)) = \begin{cases} 1, & \text{if } T(\mathbf{y}) \geq T(\mathbf{y}^o), \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Of course, if we can calculate the exact p value of (8) and (9), it is best. Unfortunately, however, an enumeration of all the elements in $\mathcal{F}(M'\mathbf{y}^o)$ and hence the calculation of the normalizing constant $C(M'\mathbf{y}^o)$ is usually computationally infeasible for large sample space. Instead, we consider a Markov chain Monte Carlo method. Note that, as one of the important advantages of Markov chain Monte Carlo method, we need not calculate the normalizing constant (5) to evaluate p values.

To perform the Markov chain Monte Carlo procedure, we have to construct a connected, aperiodic and reversible Markov chain over the conditional sample space (6) with the stationary distribution (4). If such a chain is constructed, we can sample from the chain as $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(T)}$ after discarding some initial burn-in steps, and evaluate p values as

$$\hat{p} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(T(\mathbf{y}^{(t)}) \geq T(\mathbf{y}^o)).$$

Such a chain can be constructed easily by *Markov basis*. Once a Markov basis is calculated, we can construct a connected, aperiodic and reversible Markov chain over the space (6), which can be modified so that the stationary distribution is the conditional distribution (4) by the Metropolis-Hastings procedure. See [5] and [8] for details.

Markov basis is characterized algebraically as follows. Write indeterminates $x_1, \ldots, x_k$ and consider polynomial ring $K[x_1, \ldots, x_k]$ for some field K. Consider the integer kernel of the transpose of the model matrix M, $\mathrm{Ker}_{\mathbb{Z}}M'$. For each $\mathbf{b} = (b_1, \ldots, b_k)' \in \mathrm{Ker}_{\mathbb{Z}}M'$, define binomial in $K[x_1, \ldots, x_k]$ as

$$f_{\mathbf{b}} = \prod_{b_j>0} x_j^{b_j} - \prod_{b_j<0} x_j^{-b_j}.$$

Then the binomial ideal in $K[x_1, \ldots, x_k]$,

$$I(M') = \langle \{ f_{\mathbf{b}} \mid \mathbf{b} \in \mathrm{Ker}_{\mathbf{Z}} M' \} \rangle \,,$$

is called a toric ideal with the configuration M'. Let $\{f_{\mathbf{b}^{(1)}}, \ldots, f_{\mathbf{b}^{(s)}}\}$ be any generating set of $I(M')$. Then the set of integer vectors $\{\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(s)}\}$ constitutes a Markov basis. See [5] for detail. To compute a Markov basis for given configuration M', we can rely on various algebraic softwares such as 4ti2 ([1]). See the following example.

Example 2 (Wave-soldering experiment, continued). We analyze the data in Table 1. The fitted value under the main effect model is calculated as

$$\begin{aligned} \hat{\mu} = (&68.87,\ 19.70,\ 78.85,\ 147.59,\ 12.14,\ 54.77,\ 104.53,\ 54.54, \\ &75.31,\ 39.29,\ 75.00,\ 338.37,\ 27.83,\ 52.09,\ 208.47,\ 59.64)'. \end{aligned}$$

Then the likelihood ratio for the observed data is calculated as $T(\mathbf{y}^o) = G^2(\mathbf{y}^o) = 117.81$ and the corresponding asymptotic p value is less than 0.0001 from the asymptotic distribution χ^2_8. This result tells us that the null hypothesis is highly significant and is rejected, i.e., the existence of some interaction effects is suggested. To evaluate the p value by Markov chain Monte Carlo method, we have to calculate a Markov basis first. If we use 4ti2, we prepare the data file (configuration M') as

```
8 16
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
 1  1  1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1
 1  1  1  1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1
 1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1
 1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1
 1 -1  1 -1 -1  1 -1  1 -1  1 -1  1  1 -1  1 -1
 1 -1 -1  1  1 -1 -1  1 -1  1  1 -1 -1  1  1 -1
 1 -1 -1  1 -1  1  1 -1  1 -1 -1  1 -1  1  1 -1
```

and run the command `markov`. Then we have a minimal Markov basis with 77 elements as follows.

```
77 16
 0  0  0  0  0  0  0  0  1  1 -1 -1 -1 -1  1  1
 0  0  0  0  0  1 -1  0  1  0  0 -1 -1 -1  1  1
 0  0  0  0  0  1  0 -1  0  1  0 -1 -1 -1  1  1
.....
```

Using this Markov basis, we can evaluate p value by Markov chain Monte Carlo method. After $50,000$ burn-in-steps from $\mathbf{y}^o$ itself as the initial state, we sample $100,000$ Monte Carlo sample by Metropolis-Hasting algorithm, which yields $\hat{p} = 0.0000$ again. Figure 1 is a histogram of the Monte Carlo sampling of the likelihood ratio statistic under the main effect model, along with the corresponding asymptotic distribution χ^2_8.

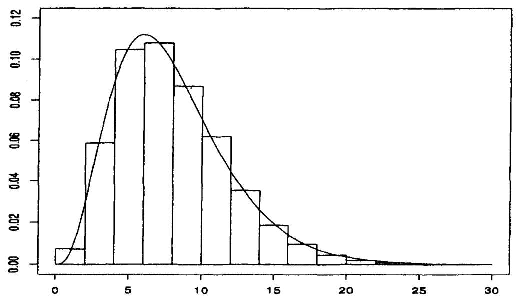


Figure 1: Asymptotic and Monte Carlo estimated distribution of likelihood ratio

3 Two-level regular fractional factorial designs and cut ideals

In this section, we show that a cut ideal for a finite connected graph can be characterized as the toric ideal $I(M')$ for a model matrix of the main effect model for some regular two-level fractional factorial designs.

3.1 Cut ideals

We start with the definition of the cut ideal. Consider a connected finite graph $G = (V, E)$. We also consider unordered partitions $A|B$ of the vertex set V. Let $\mathcal{P}(V)$ be the set of the unordered partitions of V, i.e., $\mathcal{P}(V) = \{A|B \mid A \cup B = V,\ A \cap B = \emptyset\}$. We introduce the sets of indeterminates $\{s_{ij} \mid \{i,j\} \in E\}$, $\{t_{ij} \mid \{i,j\} \in E\}$ and $\{q_{A|B} \mid A|B \in \mathcal{P}(V)\}$. Let $K[\mathbf{q}] = K[q_{A|B} \mid A|B \in \mathcal{P}(V)]$, $K[\mathbf{s},\mathbf{t}] = K[s_{ij}, t_{ij} \mid \{i,j\} \in E]$ be polynomial rings over a field K. For each partition $A|B \in \mathcal{P}(V)$, we define a subset $\mathrm{Cut}(A|B)$ of the edge set E as $\mathrm{Cut}(A|B) = \{\{i,j\} \in E \mid i \in A, j \in B \text{ or } i \in B, j \in A\}$. Define homomorphism of polynomial rings as

$$\phi_G:\ K[\mathbf{q}] \to K[\mathbf{s},\mathbf{t}],\quad q_{A|B} \mapsto \prod_{\{i,j\}\in \mathrm{Cut}(A|B)} s_{ij} \cdot \prod_{\{i,j\}\in E\setminus \mathrm{Cut}(A|B)} t_{ij}. \tag{10}$$

We may think of $\mathbf{s}$ and $\mathbf{t}$ as abbreviations for "separated" and "together", respectively. Then the cut ideal of the graph G is defined as $I_G = \mathrm{Ker}(\phi_G)$. We also use the following two examples given in [10].

Example 3 (Complete graph on four vertices). Let $G = K_4$ be the complete graph on four vertices $V = \{1,2,3,4\}$. Then the edge set is $E = \{12, 13, 14, 23, 24, 34\}$. The map ϕ_{K_4} is specified by

$$\begin{array}{lclclcl}
q_{\emptyset|1234} & \mapsto & t_{12}t_{13}t_{14}t_{23}t_{24}t_{34} & \quad & q_{4|123} & \mapsto & t_{12}t_{13}s_{14}t_{23}s_{24}s_{34}\\
q_{1|234} & \mapsto & s_{12}s_{13}s_{14}t_{23}t_{24}t_{34} & & q_{12|34} & \mapsto & t_{12}s_{13}s_{14}s_{23}s_{24}t_{34}\\
q_{2|134} & \mapsto & s_{12}t_{13}t_{14}s_{23}s_{24}t_{34} & & q_{13|24} & \mapsto & s_{12}t_{13}s_{14}s_{23}t_{24}s_{34}\\
q_{3|124} & \mapsto & t_{12}s_{13}t_{14}s_{23}t_{24}s_{34} & & q_{14|23} & \mapsto & s_{12}s_{13}t_{14}t_{23}s_{24}s_{34}.
\end{array}$$

In this case, the cut ideal is a principal ideal given by

$$I_{K_4} = \left\langle q_{\emptyset|1234}q_{12|34}q_{13|24}q_{14|23} - q_{1|234}q_{2|134}q_{3|124}q_{4|123} \right\rangle .$$

Example 4 (4-cycle). Let $G = C_4$ be the 4-cycle with $V = \{1,2,3,4\},\ E = \{12, 23, 34, 14\}$. The map ϕ_{C_4} is derived from ϕ_{K_4} in Example 3 by setting $s_{13} = t_{13} = s_{24} = t_{24} = 1$ as

$$\begin{array}{rclrcl}
q_{\emptyset|1234} & \mapsto & t_{12}t_{14}t_{23}t_{34} & q_{4|123} & \mapsto & t_{12}s_{14}t_{23}s_{34} \\
q_{1|234} & \mapsto & s_{12}s_{14}t_{23}t_{34} & q_{12|34} & \mapsto & t_{12}s_{14}s_{23}t_{34} \\
q_{2|134} & \mapsto & s_{12}t_{14}s_{23}t_{34} & q_{13|24} & \mapsto & s_{12}s_{14}s_{23}s_{34} \\
q_{3|124} & \mapsto & t_{12}t_{14}s_{23}s_{34} & q_{14|23} & \mapsto & s_{12}t_{14}t_{23}s_{34}.
\end{array}$$

In this case, the cut ideal is given by

$$I_{C_4} = \left\langle q_{\emptyset|1234}q_{13|24} - q_{1|234}q_{3|124},\ q_{\emptyset|1234}q_{13|24} - q_{2|134}q_{4|123},\ q_{\emptyset|1234}q_{13|24} - q_{12|34}q_{14|23} \right\rangle .$$

Now we relates the cut ideals to the regular two-level fractional factorial designs. We express the map ϕ_G by $2^{|V|-1} \times 2|E|$ matrix $H = \{h_{A|B,e}\}$ where each row of H represents $A|B \in \mathcal{P}(V)$ and each two columns of H represents E as

$$h_{A|B,e} = \begin{cases} (1,0) & \text{if } e \in E \setminus \mathrm{Cut}(A|B) \\ (0,1) & \text{if } e \in \mathrm{Cut}(A|B). \end{cases}$$

Note that there are $|\mathcal{P}(V)| = 2^{|V|-1}$ unordered partitions of V. We also see that each two columns of H correspond to $\mathbf{t}$ and $\mathbf{s}$. Then the cut ideal, the kernel of ϕ_G of (10), is written as the toric ideal of the configuration matrix H'.

Example 5 (4-cycle, continued). For the case of $G = C_4$ of Example 4, the matrix H can be written as follows.

$$\begin{array}{c|cccccccc|}
 & t_{12} & s_{12} & t_{14} & s_{14} & t_{23} & s_{23} & t_{34} & s_{34} \\ \hline
q_{\emptyset|1234} & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
q_{3|124} & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
q_{4|123} & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
q_{12|34} & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
q_{14|23} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
q_{2|134} & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
q_{1|234} & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
q_{13|24} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline
\end{array} \qquad (11)$$

The kernel of H' coincides to the kernel of M' of (3) for the two-level design D of $|E|$ factors with $2^{|V|-1}$ runs, where the level of the factor X_e for the run $A|B \in \mathcal{P}(V)$ is given by the following map:

$$\begin{array}{rccl}
X_e : & \mathcal{P}(V) & \to & \{+1, -1\} \\
 & \cup\!\! \mid & & \cup\!\! \mid \\
 & A|B & \mapsto & \begin{cases} +1 & \text{if } e \in E \setminus \mathrm{Cut}(A|B) \\ -1 & \text{if } e \in \mathrm{Cut}(A|B) \end{cases}
\end{array} \qquad (12)$$

Example 6 (4-cycle, continued). For the case of $G = C_4$, the map X_e of (12) gives the design matrix D as follows.

	X_{12}	X_{14}	X_{23}	X_{34}
$q_{\emptyset\|1234}$	1	1	1	1
$q_{3\|124}$	1	1	−1	−1
$q_{4\|123}$	1	−1	1	−1
$q_{12\|34}$	1	−1	−1	1
$q_{14\|23}$	−1	1	1	−1
$q_{2\|134}$	−1	1	−1	1
$q_{1\|234}$	−1	−1	1	1
$q_{13\|24}$	−1	−1	−1	−1

For this D, it is easily seen that $\mathrm{Ker}(M')$ coincides to $\mathrm{Ker}(H')$ if H is given by (11).

3.2 Regular designs and cut ideals

In Example 6, we obtain the toric ideal for the main effect model of the regular two-level fractional factorial designs defined by $X_{12}X_{14}X_{23}X_{34} = 1$ from the the cut ideal of $G = C_4$. In fact, there is a clear relation between finite connected graphs G and regular two-level designs D. As we have seen in Example 6, the cut ideal for G can be related to the design of $p = |E|$ factors with $k = 2^{|V|-1}$ runs. Since each factor of this design corresponds to the edge E of G, we write each factor X_{ij} for $\{i,j\} \in E$. Since there are 2^p runs in the full factorial design of p factors, the design obtained from G by the relation (12) is a $2^{|V|-1-p}$ fraction of the full factorial design of p factors. We show this fraction is specified as the regular fractional factorial designs.

Let $G = (V, E)$ be a finite connected graph with the edge set $E = \{e_1, \dots, e_p\}$. Then, the *cycle space* $\mathcal{C}(G)$ of G is a subspace of $\mathbb{F}_2^{|E|}$ spanned by

$$\left\{ \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_r} \in \mathbb{F}_2^{|E|} \;\middle|\; (e_{i_1}, \dots, e_{i_r}) \text{ is a cycle of } G \right\},$$

where $\mathbf{e}_j$ is the jth coordinate vector of $\mathbb{F}_2^{|E|}$. On the other hand, the *cut space* $\mathcal{C}^*(G)$ of G is a subspace of $\mathbb{F}_2^{|E|}$ defined by

$$\mathcal{C}^*(G) = \left\{ \sum_{e_j \in \mathrm{Cut}(A|B)} \mathbf{e}_j \in \mathbb{F}_2^{|E|} \;\middle|\; A|B \in \mathcal{P}(V) \right\}.$$

Fix a spanning tree T of G. For each $e \in E \setminus T$, the set $T \cup \{e\}$ has exactly one cycle C_e of G. Such a cycle C_e is called a *fundamental cycle* of G. Since T has $|V| - 1$ edges, there are $|E| - |V| + 1$ edges in $E \setminus T$. It then follows that there exists $|E| - |V| + 1$ fundamental cycles in G.

Theorem 7. *Let $G = (V, E)$ be a finite connected graph and let D be the design matrix of $|E|$ factors with $2^{|V|-1}$ runs defined by (12). Then D is a regular fractional factorial design with all relations*

$$X_{e_{i_1}}(A|B)X_{e_{i_2}}(A|B)\cdots X_{e_{i_m}}(A|B) = 1, \tag{13}$$

where $(e_{i_1}, \dots, e_{i_m})$ is a fundamental cycle of G.

Theorem 7 shows the relation of the cut ideals and regular two-level fractional factorial designs. For a given connected finite graph, we can consider corresponding regular two-level fractional factorial designs from Theorem 7. Unfortunately, however, the converse does not always hold. For given regular two-level fractional factorial designs (strictly, we should say that "for given designs and *models*"), it does not always exist corresponding connected finite graphs.

Proposition 8. *If a 2^{p-q} design corresponds to a finite graph by the relation (12), then we have $p \leq \binom{p-q+1}{2}$.*

Thus, obvious counterexamples for the converse are given since some regular 2^{p-q} designs satisfy $\binom{p-q+1}{2} < p$ (for example, $(p,q) = (5,3),(5,4),(6,4),(6,5)$ and so on). On the other hand, a necessary condition related with the resolution is as follows.

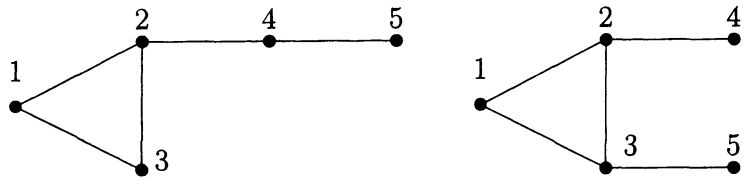
Proposition 9. *If a 2^{p-q} design of resolution IV or more corresponds to a finite graph by the relation (12), then we have $p \leq \lfloor (p-q+1)^2/4 \rfloor$.*

If the resolution of a design is V or more, then similar results are obtained by the results in [6]. From these considerations, an important question arises.

Question 10. *Characterize regular two-level fractional factorial designs that can correspond to a finite graph by the relation (12).*

A complete answer to this question is not yet obtained at present. We present several fundamental characterizations in the rest of this section. Note that the above correspondence is not one-to-one even if it exists. In fact, for any finite connected graph G, we can specify a design D uniquely by (13). However, for a given design G, we can consider several graphs satisfying the relation (13) if it exists.

Example 11 (2^{5-1} design with $X_{12}X_{13}X_{23} = 1$ of 5 factors). Consider 2^{5-1} fractional factorial design $X_{12}X_{13}X_{23} = 1$ of 5 factors, or, $\mathbf{ABC} = \mathbf{I}$ in the convention of designed experiment literature. There are several corresponding graphs that give this design such as follows.



Now we show two important special cases, designs corresponding to complete graphs and trees.

Corollary 12. *Let $G = K_n$ be the complete graph on $|V| = n$ vertices. Then, G is specified as the regular $2^{c_1-c_2}$ fractional factorial design of c_1 two-level factors by (13), where*

$$c_1 = \binom{n}{2}, \quad c_2 = \binom{n-1}{2}.$$

The defining relation of this design is written as $X_{1i}X_{1j}X_{ij} = 1$ for any pair (i,j) with $2 \leq i < j \leq n$.

Another important case is as follows.

Corollary 13. *Any spanning tree $G = (V, E)$ is specified as the full factorial design of $|V| - 1$ two-level factors by (13).*

4 Discussion

We apply known results on cut ideals to the regular two-level fractional factorial designs. See [2] for details.

References

[1] 4ti2 team. 4ti2 – A software package for algebraic, geometric and combinatorial problems on linear spaces. Available at `www.4ti2.de`.

[2] S. Aoki, T. Hibi and H. Ohsugi (2013). Markov chain Monte Carlo methods for the regular two-level fractional factorial designs and cut ideals. *Journal of Statistical Planning and Inference*, to appear.

[3] S. Aoki and A. Takemura (2010). Markov chain Monte Carlo tests for designed experiments. *Journal of Statistical Planning and Inference*, **140**, 817–830.

[4] L. W. Condra (1993). *Reliability Improvement with Design of Experiments.* Marcel Dekker, New York, NY., 1993.

[5] P. Diaconis and B. Sturmfels (1998). Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, **26**, 363–397.

[6] D. K. Garnick, Y. H. H. Kwong and F, Lazebnik (1993). Extremal graphs without three-cycles or four-cycles, *Journal of Graph Theory*, **17**, 633–645.

[7] M. Hamada and J. A. Nelder (1997). Generalized linear models for quality-improvement experiments. *Journal of Quality Technology*, 29:292–304.

[8] W. K. Hastings (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

[9] P. McCullagh and J. A. Nelder (1989). *Generalized linear models.* 2nd ed. Chapman & Hall, London.

[10] B. Sturmfels and S. Sullivant (2008). Toric geometry of cuts and splits. *Michigan Math. J.*, **57**, 689–709.