Bernoulli **11**(5), 2005, 793–813

Distance-reducing Markov bases for sampling from a discrete sample space

AKIMICHI TAKEMURA* and SATOSHI AOKI**

Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan.
*E-mail: *takemura@stat.t.u-tokyo.ac.jp; **aoki@stat.t.u-tokyo.ac.jp*

We study Markov bases for sampling from a discrete sample space equipped with a convenient metric. Starting from any two states in the sample space, we ask whether we can always move closer by an element of a Markov basis. We call a Markov basis *distance-reducing* if this is the case. The particular metric we consider in this paper is the L_1-norm on the sample space. Some characterizations of L_1-norm-reducing Markov bases are derived.

Keywords: contingency tables; Graver bases; Gröbner bases; L_1-norm; Markov chain Monte Carlo; toric ideals

1. Introduction

A Markov basis for sampling from a discrete conditional distribution is usually studied in the framework of a toric ideal and its Gröbner basis (Sturmfels 1995; Diaconis and Sturmfels 1998; Dinwoodie 1998). For the case of a $3 \times 3 \times K$ contingency table with fixed two-dimensional marginals, in Aoki and Takemura (2003a) we used more elementary approach to derive a unique minimal Markov basis. The approach was based on exhaustive consideration of sign patterns when the L_1-norm (1-norm) between two contingency tables with the same marginals is minimized. In order to prove that a candidate set $\mathcal{B}$ of moves is a Markov basis, we have shown that the 1-norm between two contingency tables can always be decreased by an element of $\mathcal{B}$.

In order to study a minimal Markov basis and its uniqueness for other models, in Takemura and Aoki (2004) we considered whether two elements of the same fibre (reference set) are mutually accessible by a set of lower-degree moves and derived some results on the characterization of minimal Markov bases. Note that the notion of mutual accessibility is not directly related to any metric on the fibres. Therefore although the approaches in Aoki and Takemura (2003a) and Takemura and Aoki (2004) were similar, they were different in explicit consideration of the metric on the fibres.

In this paper we explicitly consider the 1-norm on the fibres in the general framework of Takemura and Aoki (2004) and derive some characterizations of 1-norm-reducing Markov bases. If a Markov basis is 1-norm-reducing, then the diameter of each fibre, which is regarded as a transition graph of the Markov chain with respect to the Markov basis, is easily bounded from above. The diameter of the graph is an important factor in various

1350–7265

results on the convergence rate of Markov chains; see, for example, Diaconis and Sturmfels (1998, Section 2.3), Diaconis and Saloff-Coste (1998, Section 4) and Bubley and Dyer (1997). Therefore the distance-reducing property of Markov bases is also relevant from the viewpoint of the convergence rate of the Markov chain. In addition, the idea of a distance reduction has already appeared implicitly in some *ad hoc* arguments on Markov chains. For example, it has been used to show the irreducibiliy of Markov chains in Guo and Thompson (1992, Appendix) and Aoki and Takemura (2003a, Appendix).

This paper is organized as follows. In Section 2 we summarize some preliminary material on moves and Markov bases. In Section 3.1 we study general properties of distance-reducing Markov bases. In Section 3.2 we derive some properties of 1-norm-reducing Markov bases. In Section 4 we give examples and conclude with a discussion.

2. Preliminaries on Markov bases

In this section we summarize preliminary material on Markov bases. We employ the framework of Takemura and Aoki (2004), although we also use well-known results on toric ideals and Gröbner basis; see, for example, Sturmfels (1995) and Hibi (2003).

2.1. Notation

Here we summarize the notation of Takemura and Aoki (2004). Let $\mathcal{I}$ be a finite set with $|\mathcal{I}|$ elements. With multi-way contingency tables in mind, an element of $\mathcal{I}$ is called a *cell* and denoted by $\mathbf{i} \in \mathcal{I}$. A non-negative integer $x_{\mathbf{i}} \in \mathbb{N} = \{0, 1, \ldots\}$ denotes the frequency of a cell $\mathbf{i}$. $n = \sum_{\mathbf{i}\in\mathcal{I}} x_{\mathbf{i}}$ denotes the sample size. Let $a(\mathbf{i}) \in \mathbb{N}^{\nu}$, $\mathbf{i} \in \mathcal{I}$, denote ν-dimensional fixed column vectors consisting of non-negative integers. A ν-dimensional sufficient statistic $\mathbf{t}$ is given by $\mathbf{t} = \sum_{\mathbf{i}\in\mathcal{I}} a(\mathbf{i})x_{\mathbf{i}}$.

Let the cells and the vectors $a(\mathbf{i})$ be appropriately ordered. Then

$$\mathbf{x} = \{x_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}} \in \mathbb{N}^{|\mathcal{I}|}$$

denotes an $|\mathcal{I}|$-dimensional column vector of cell frequencies (*frequency vector*) and

$$A = \{a(\mathbf{i})\}_{\mathbf{i}\in\mathcal{I}} = (a_{j\mathbf{i}})_{j=1,\ldots,\nu,\mathbf{i}\in\mathcal{I}}$$

denotes a $\nu \times |\mathcal{I}|$ matrix. Then the sufficient statistic $\mathbf{t}$ is written as $\mathbf{t} = A\mathbf{x}$.

We use the notation $|\mathbf{x}| = n = \sum_{\mathbf{i}} x_{\mathbf{i}}$ to denote the sample size of $\mathbf{x}$. It is the 1-norm of $\mathbf{x}$. $\mathbf{x} \geqslant \mathbf{0}$ means that the elements of $\mathbf{x}$ are non-negative and $\mathbf{x} \geqslant \mathbf{y}$ means $\mathbf{x} - \mathbf{y} \geqslant \mathbf{0}$. For a given frequency vector $\mathbf{x}$, its support is the set of cells with positive frequencies: $\text{supp}(\mathbf{x}) = \{\mathbf{i} | x_{\mathbf{i}} > 0\}$. Following Sturmfels (1995), we call the set of $\mathbf{x}$s for a given $\mathbf{t}$,

$$\mathcal{F}_{\mathbf{t}} = \{\mathbf{x} \geqslant \mathbf{0} | A\mathbf{x} = \mathbf{t}\},$$

a $\mathbf{t}$-*fibre* in this paper.

As in Takemura and Aoki (2004), we assume that the $|\mathcal{I}|$-dimensional row vector $(1, 1, \ldots, 1)$ is a linear combination of the rows of A. This assumption is standard in the theory of toric ideals (Hibi 2003, Section 4.1). Under this assumption the sample size n is determined from the sufficient statistic $\mathbf{t}$ and all elements of $\mathcal{F}_{\mathbf{t}}$ have the same sample size. Somewhat abusing the notation, we write $n = |\mathbf{t}|$ to denote the sample size of elements of $\mathcal{F}_{\mathbf{t}}$.

Let $\mathbb{Z} = \{0, \pm 1, \ldots\}$. An $|\mathcal{I}|$-dimensional vector of integers $\mathbf{z} \in \mathbb{Z}^{|\mathcal{I}|}$ is called a *move* if it is in the kernel of A:

$$A\mathbf{z} = \mathbf{0}.$$

For a move $\mathbf{z}$, the positive part $\mathbf{z}^+$ and the negative part $\mathbf{z}^-$ are defined by

$$z_{\mathbf{i}}^+ = \max(z_{\mathbf{i}}, 0), \qquad z_{\mathbf{i}}^- = -\min(z_{\mathbf{i}}, 0),$$

respectively. Then $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$. The positive part $\mathbf{z}^+$ and the negative part $\mathbf{z}^-$ have the same value of sufficient statistic $\mathbf{t} = A\mathbf{z}^+ = A\mathbf{z}^-$. The sample size of $\mathbf{z}^+$ (or $\mathbf{z}^-$) is called the *degree* of $\mathbf{z}$ and denoted by

$$\deg \mathbf{z} = |\mathbf{z}^+| = |\mathbf{z}^-|.$$

We also write $|\mathbf{z}| = \sum_{\mathbf{i}} |z_{\mathbf{i}}| = 2 \deg \mathbf{z}$, which is the 1-norm of $\mathbf{z}$.

We say that a move $\mathbf{z}$ is *applicable* to $\mathbf{x} \in \mathcal{F}_{\mathbf{t}}$ if $\mathbf{x} + \mathbf{z} \in \mathcal{F}_{\mathbf{t}}$, that is, adding $\mathbf{z}$ to $\mathbf{x}$ does not produce a negative cell. Clearly $\mathbf{z}$ is applicable to $\mathbf{x}$ if and only if $\mathbf{x} \geqslant \mathbf{z}^-$. Therefore if $\mathbf{z}$ is applicable to $\mathbf{x}$ then $|\mathbf{z}^+| \leqslant |\mathbf{x}|$.

Let $\mathcal{B} = \{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$ be a finite set of moves. Let $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$. We say that $\mathbf{y}$ is *accessible* from $\mathbf{x}$ by $\mathcal{B}$ and denote it by

$$\mathbf{x} \sim \mathbf{y} (\operatorname{mod} \mathcal{B}),$$

if there exists a sequence of moves $\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_k}$ of $\mathcal{B}$ and $\epsilon_j = \pm 1$, $j = 1, \ldots, k$, such that $\mathbf{y} = \mathbf{x} + \sum_{j=1}^k \epsilon_j \mathbf{z}_{i_j}$ and

$$\mathbf{x} + \sum_{j=1}^h \epsilon_j \mathbf{z}_{i_j} \in \mathcal{F}_{\mathbf{t}}, \qquad h = 1, \ldots, k-1, \tag{1}$$

that is, we can move from $\mathbf{x}$ to $\mathbf{y}$ by moves of $\mathcal{B}$ without causing negative cells on the way. Obviously the notion of accessibility is symmetric and transitive. Therefore accessibility by $\mathcal{B}$ is an equivalence relation and each $\mathcal{F}_{\mathbf{t}}$ is partitioned into disjoint equivalence classes by moves of $\mathcal{B}$. We call these equivalence classes $\mathcal{B}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$. Since the notion of accessibility is symmetric, we also say that $\mathbf{x}$ and $\mathbf{y}$ are mutually accessible by $\mathcal{B}$ if $\mathbf{x} \sim \mathbf{y} (\operatorname{mod} \mathcal{B})$. Let $\mathbf{x}$ and $\mathbf{y}$ be elements from two different $\mathcal{B}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$. We say that a move

$$\mathbf{z} = \mathbf{x} - \mathbf{y}$$

connects these two equivalence classes.

$\mathcal{B}$-equivalence classes can be considered in terms of a graph. Consider an undirected graph $G = G_{\mathbf{t}, \mathcal{B}}$ with the set of vertices $V = \mathcal{F}_{\mathbf{t}}$ and the edges between $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ if $\mathbf{y} = \mathbf{x} \pm \mathbf{z}$ for some $\mathbf{z} \in \mathcal{B}$. $\mathcal{B}$-equivalence classes are connected components of G.

In the following let

$$\mathcal{B}_n = \{\mathbf{z} | \deg \mathbf{z} \leqslant n\}$$

denote the set of moves with degree less than or equal to n.

2.2. Markov bases

A set of finite moves $\mathcal{B} = \{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$ is a *Markov basis* if, for all $\mathbf{t}$, $\mathcal{F}_\mathbf{t}$ itself constitutes one $\mathcal{B}$-equivalence class, that is, the graph $G_{\mathbf{t},\mathcal{B}}$ is connected for all $\mathbf{t}$. A Markov basis $\mathcal{B}$ is *minimal* if no proper subset of $\mathcal{B}$ is a Markov basis. A minimal Markov basis always exists, because from any Markov basis we can remove redundant elements one by one, until none of the remaining elements can be removed any further. If a set of minimal bases coincide except for sign changes of their elements, then we have a minimal basis which is unique.

A move $\mathbf{z}$ is *indispensable* if $\mathbf{z}$ or $-\mathbf{z}$ belongs to every Markov basis. $\mathbf{z}$ is indispensable if and only if it is the difference of elements of a two-element fibre

$$\mathbf{z} = \mathbf{x} - \mathbf{y}, \qquad \{\mathbf{x}, \mathbf{y}\} = \mathcal{F}_\mathbf{t} \text{ for some } \mathbf{t}.$$

A minimal basis is unique if and only if the set of indispensable moves forms a Markov basis. Ohsugi and Hibi (2003) have shown the important fact that the set of indispensable moves is the intersection of all reduced Gröbner bases.

Let $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ be a move of degree n. We call $\mathbf{z}$ *non-replaceable by lower-degree moves* if

$$\mathbf{z}^+ \not\sim \mathbf{z}^- (\bmod \mathcal{B}_{n-1}), \tag{2}$$

that is, if $\mathbf{z}$ connects different $\mathcal{B}_{n-1}$-equivalence classes of $\mathcal{F}_\mathbf{t} \ni \mathbf{z}^+$. Clearly an indispensable move is non-replaceable by lower-degree moves. From the argument in Takemura and Aoki (2004) we see that the set

$$\mathcal{B}_{\mathrm{MF}} = \{\mathbf{z} | \mathbf{z} \text{ is non-replaceable by lower-degree moves}\}$$

forms a Markov basis, and moreover $\mathcal{B}_{\mathrm{MF}}$ is the union of all minimal Markov bases. From the definition of $\mathcal{B}_{\mathrm{MF}}$, each $\mathcal{F}_\mathbf{t}$ is a single $\mathcal{B}_{|\mathbf{t}|-1}$-equivalence class for any $\mathbf{t} = A\mathbf{z}^+ = A\mathbf{z}^-$, $\mathbf{z} \notin \mathcal{B}_{\mathrm{MF}}$. Conversely, to construct minimal Markov bases, we only have to involve moves $\mathbf{z}$ connecting different $\mathcal{B}_{|\mathbf{t}|-1}$-equivalence classes of $\mathcal{F}_\mathbf{t}$ where $\mathbf{t} = A\mathbf{z}^+ = A\mathbf{z}^-$. Takemura and Aoki (2004) state that for all minimal Markov bases $\mathcal{B}$, the set of sufficient statistics

$$\{\mathbf{t} | \mathbf{t} = A\mathbf{z}^+ = A\mathbf{z}^-, \mathbf{z} \in \mathcal{B}\} = \{\mathbf{t} | \mathcal{F}_\mathbf{t} \text{ is not a single } \mathcal{B}_{|\mathbf{t}|-1}\text{-equivalence class}\}$$

is common, and is equal to $\{\mathbf{t} | \mathbf{t} = A\mathbf{z}^+ = A\mathbf{z}^-, \mathbf{z} \in \mathcal{B}_{\mathrm{MF}}\}$. Considering these facts, we call $\mathcal{B}_{\mathrm{MF}}$ a *minimum fibre Markov basis*. In the case of contingency tables with fixed marginals, the minimum-fibre Markov basis is invariant in the sense of Aoki and Takemura (2003b).

Consider a sum of two moves $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$. We say that there is no cancellation of signs in this sum if

$$\mathrm{supp}(\mathbf{z}^+) = \mathrm{supp}(\mathbf{z}_1^+) \cup \mathrm{supp}(\mathbf{z}_2^+), \qquad \mathrm{supp}(\mathbf{z}^-) = \mathrm{supp}(\mathbf{z}_1^-) \cup \mathrm{supp}(\mathbf{z}_2^-).$$

In this case we also say that $\mathbf{z}$ is a conformal sum of $\mathbf{z}_1$ and $\mathbf{z}_2$. Similarly, we say that there is no cancellation of signs in the sum of m moves $\mathbf{z} = \mathbf{z}_1 + \ldots + \mathbf{z}_m$ (or $\mathbf{z}$ is a conformal sum of m moves) if

$$\text{supp}(\mathbf{z}^+) = \text{supp}(\mathbf{z}_1^+) \cup \cdots \cup \text{supp}(\mathbf{z}_m^+), \qquad \text{supp}(\mathbf{z}^-) = \text{supp}(\mathbf{z}_1^-) \cup \cdots \cup \text{supp}(\mathbf{z}_m^-).$$

A move $\mathbf{z}$ is *primitive* if it cannot be written as a sum of two non-zero moves $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ with no cancellation of signs. Clearly a move $\mathbf{z}$, which is non-replaceable by lower-degree moves, is primitive. Note that $\mathbf{z}$ is primitive if it cannot be written as $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ such that $|\mathbf{z}| = |\mathbf{z}_1| + |\mathbf{z}_2|$ and $0 < |\mathbf{z}_1|, |\mathbf{z}_2| < |\mathbf{z}|$. The set of primitive moves is called the *Graver basis* and is a Markov basis. If a move $\mathbf{z}$ is not primitive, then we can recursively decompose $\mathbf{z}$ into a conformal sum of moves. This implies that any move $\mathbf{z}$ can be written as a conformal sum

$$\mathbf{z} = \mathbf{z}_1 + \ldots + \mathbf{z}_m, \tag{3}$$

where $\mathbf{z}_1, \ldots, \mathbf{z}_m$ are (not necessarily distinct) non-zero elements of the Graver basis. If there is no cancellation of signs, whenever $\mathbf{z}$ is applicable to some $\mathbf{x}$, $\mathbf{z}$ can be replaced by $\mathbf{z}_1, \ldots, \mathbf{z}_m$ in arbitrary order without causing negative cells on the way.

In Sturmfels (1995, Chapter 4), the notion of the conformal sum is discussed in expressing a move with non-negative rational linear combination of circuits. It seems that this notion of the conformal sum is not very useful in the framework of the present paper.

Let $\{u_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}$ be the set of indeterminates and let $K[\{u_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}]$ denote the polynomial ring in the indeterminates $\{u_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}$ over a field K. A frequency vector $\mathbf{x} = \{x_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}$ can be identified with the monomial

$$\prod_{\mathbf{i}\in\mathcal{I}} u_{\mathbf{i}}^{x_{\mathbf{i}}}.$$

Let $\{t_j\}_{j=1,\ldots,\nu}$ denote the set of indeterminates corresponding to the sufficient statistic $\mathbf{t}$ and let $K[\mathbf{t}]$ denote the polynomial ring in $t_1, \ldots, t_\nu$ over K. The toric ideal I_A is the kernel of the ring homomorphism

$$u_{\mathbf{i}} \mapsto \prod_{j=1}^{\nu} t_j^{a_{j\mathbf{i}}}.$$

Given a term order $\prec$ on $K[\{u_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}]$ let $\mathcal{B}_\prec$ denote the reduced Gröbner basis of I_A with respect to $\prec$. It is well known (Sturmfels 1995; Diaconis and Sturmfels 1998) that $\mathcal{B}_\prec$ is a Markov basis. The union of all reduced Gröbner bases for all possible term orders is called the universal Gröbner basis and is contained in the Graver basis.

3. Distance-reducing Markov bases

In this section we first define the notion of distance reduction by a set of moves when the sample space is equipped with an appropriate metric. Then we derive some characterizations of 1-norm-reducing Markov bases.

3.1. Distance reduction by a set of moves

Consider a metric $d(\mathbf{x}, \mathbf{y})$ on a fibre $\mathcal{F}_{\mathbf{t}}$. Although we are mainly concerned with the 1-norm in this paper, here we consider a general metric. A metric $d = d_{\mathbf{t}}$ on $\mathcal{F}_{\mathbf{t}}$ can be defined in various ways. If a metric d is defined on the whole sample space $\mathbb{N}^{|\mathcal{I}|}$, we consider the restriction of d for each $\mathcal{F}_{\mathbf{t}}$, that is, $d_{\mathbf{t}}(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$. If d is a norm on the set $\mathbb{Z}^{|\mathcal{I}|}$ of integer frequency vectors, such as the 1-norm $|\mathbf{z}|$, d is defined by $d_{\mathbf{t}}(\mathbf{x}, \mathbf{y}) = d(\mathbf{x} - \mathbf{y})$. For notational simplicity we henceforth suppress the subscript $\mathbf{t}$ on $d_{\mathbf{t}}$.

We now introduce the notion of a distance reduction by a set of moves. Let $\mathcal{B}$ denote a set of moves. We call $\mathcal{B}$ *d-reducing for* $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ if there exists an element $\mathbf{z} \in \mathcal{B}$ and $\epsilon = \pm 1$ such that $\epsilon\mathbf{z}$ is applicable to $\mathbf{x}$ or $\mathbf{y}$ and we can decrease the distance, that is,

$$\begin{aligned} & \mathbf{x} + \epsilon\mathbf{z} \in \mathcal{F}_{\mathbf{t}} \quad \text{and} \quad d(\mathbf{x} + \epsilon\mathbf{z}, \mathbf{y}) < d(\mathbf{x}, \mathbf{y}), \\ \text{or} \quad & \mathbf{y} + \epsilon\mathbf{z} \in \mathcal{F}_{\mathbf{t}} \quad \text{and} \quad d(\mathbf{x}, \mathbf{y} + \epsilon\mathbf{z}) < d(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{4}$$

We simply call $\mathcal{B}$ *d-reducing* if $\mathcal{B}$ is d-reducing for every $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ and for every $\mathbf{t}$.

We call $\mathcal{B}$ *strongly d-reducing for* $\mathbf{x}, \mathbf{y}$ if there exist elements $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{B}$ and $\epsilon_1, \epsilon_2 = \pm 1$ such that $\mathbf{x} + \epsilon_1\mathbf{z}_1 \in \mathcal{F}_{\mathbf{t}}$, $\mathbf{y} + \epsilon_2\mathbf{z}_2 \in \mathcal{F}_{\mathbf{t}}$ and

$$d(\mathbf{x} + \epsilon_1\mathbf{z}_1, \mathbf{y}) < d(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad d(\mathbf{x}, \mathbf{y} + \epsilon_2\mathbf{z}_2) < d(\mathbf{x}, \mathbf{y}), \tag{5}$$

and call $\mathcal{B}$ *strongly d-reducing* if $\mathcal{B}$ is strongly d-reducing for every $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ and for every $\mathbf{t}$. Clearly if $\mathcal{B}$ is strongly d-reducing, then $\mathcal{B}$ is d-reducing.

A basic argument that we have employed in Aoki and Takemura (2003a) is the following obvious fact.

Proposition 1. *Let a metric d be given on each fibre $\mathcal{F}_{\mathbf{t}}$. A set of finite moves $\mathcal{B}$ is a Markov basis if it is d-reducing.*

If $\mathcal{B}$ is d-reducing, then for every $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$, there exist $k > 0$, $\epsilon_l = \pm 1$, $\mathbf{z}_l \in \mathcal{B}$, $\mathbf{x}_l \in \mathcal{F}_{\mathbf{t}}$, $\mathbf{y}_l \in \mathcal{F}_{\mathbf{t}}$, $l = 1, \ldots, k$, with the following properties:

(i) $\mathbf{x}_k = \mathbf{y}_k$;
(ii) $d(\mathbf{x}_l, \mathbf{y}_l) < d(\mathbf{x}_{l-1}, \mathbf{y}_{l-1})$, $l = 1, \ldots, k$, where $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{y}_0 = \mathbf{y}$;
(iii) $(\mathbf{x}_l, \mathbf{y}_l) = (\mathbf{x}_{l-1} + \epsilon_l\mathbf{z}_l, \mathbf{y}_{l-1})$ or $(\mathbf{x}_l, \mathbf{y}_l) = (\mathbf{x}_{l-1}, \mathbf{y}_{l-1} + \epsilon_l\mathbf{z}_l)$, $l = 1, \ldots, k$.

Given the above sequence of frequency vectors, we can move from $\mathbf{x}$ to $\mathbf{x}_k = \mathbf{y}_k$ and then, reversing the moves, we can move from $\mathbf{y}_k$ to $\mathbf{y}$. Thus $\mathbf{y}$ is accessible from $\mathbf{x}$ by $\mathcal{B}$. Note that in this sequence of moves the distances $d(\mathbf{x}, \mathbf{x}_1), \ldots, d(\mathbf{x}, \mathbf{x}_k), d(\mathbf{x}, \mathbf{y}_{k-1}) \ldots, d(\mathbf{x}, \mathbf{y})$ might not be monotone increasing.

On the other hand, if $\mathcal{B}$ is strongly d-reducing, then starting from $\mathbf{y}$, we can always decrease the distance by moving from the side of $\mathbf{y}$, that is, we can find $k > 0$ and $\mathbf{y} = \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1}, \mathbf{y}_k = \mathbf{x}$ in $\mathcal{F}_{\mathbf{t}}$ such that $\mathbf{y}_l = \mathbf{y}_{l-1} + \epsilon_l\mathbf{z}_l$, $\epsilon_l = \pm 1$, $\mathbf{z}_l \in \mathcal{B}$, $l = 1, \ldots, k$, and $d(\mathbf{x}, \mathbf{y}_{k-1}), d(\mathbf{x}, \mathbf{y}_{k-2}), \ldots, d(\mathbf{x}, \mathbf{y})$ is strictly increasing.

Given a Markov basis $\mathcal{B}$, each fibre $\mathcal{F}_{\mathbf{t}}$ can be considered as a connected graph $G_{\mathbf{t},\mathcal{B}}$ as discussed in Section 2. Associated with a Markov basis $\mathcal{B}$, the *shortest path metric* $d_{\mathcal{B}}(\mathbf{x}, \mathbf{y})$

is defined to be the minimum number of steps needed to move from $\mathbf{x}$ to $\mathbf{y}$ by moves of $\mathcal{B}$. Trivially, $\mathcal{B}$ is itself $d_{\mathcal{B}}$-reducing. As discussed in Introduction, the diameter of $G_{\mathbf{t},\mathcal{B}}$ is relevant for the convergence rate of the Markov chain based on the Markov basis $\mathcal{B}$. However, direct computation of the diameter is generally difficult, whereas the largest 1-norm between two states in $\mathcal{F}_{\mathbf{t}}$ is clearly bounded from above by $2|\mathbf{t}|$. Suppose that $|\mathbf{x}-\mathbf{y}| = 2|\mathbf{t}|$ for $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ and $\mathcal{B}$ is 1-norm-reducing. Applying the moves from $\mathcal{B}$, we can decrease $|\mathbf{x}-\mathbf{y}|$ such that $|\mathbf{x}-\mathbf{y}| > |(\mathbf{x}+\epsilon\mathbf{z})-\mathbf{y}|$ or $|\mathbf{x}-\mathbf{y}| > |\mathbf{x}-(\mathbf{y}+\epsilon\mathbf{z})|$, where $\mathbf{z} \in \mathcal{B}$ and $\epsilon = \pm 1$. It should be noted that, in each case, the distance can be decreased by at least 2 and therefore $|(\mathbf{x}+\epsilon\mathbf{z})-\mathbf{y}| \leqslant 2|\mathbf{t}|-2$ or $|\mathbf{x}-(\mathbf{y}+\epsilon\mathbf{z})| \leqslant 2|\mathbf{t}|-2$ holds, since the 1-norm of every move is even from the relation $|\mathbf{z}| = 2\deg(\mathbf{z})$. This implies that if $\mathcal{B}$ is 1-norm-reducing, then the diameter of $G_{\mathbf{t},\mathcal{B}}$ is bounded from above by $|\mathbf{t}|$.

3.2. Some results on 1-norm-reducing Markov bases

The 1-norm $|\mathbf{z}| = \sum_{\mathbf{i}\in\mathcal{I}} |z_{\mathbf{i}}|$ on the set $\mathbb{Z}^{|\mathcal{I}|}$ of integer frequency vectors is a natural norm to consider for Markov bases. In this section we investigate Markov bases from the viewpoint of the 1-norm reduction.

Here we briefly summarize the results of this section. First we discuss the basic importance of the Graver basis in the investigation of the 1-norm reduction (Propositions 2–4). Then we introduce three closely related notions of 1-norm irreducibility of a move and discuss implications among them (Proposition 5). Based on these notions, conditions for unique minimality of (strongly) 1-norm-reducing Markov bases are given (Propositions 6 and 7). In Proposition 8 we introduce a Markov basis $\mathcal{B}_{\mathrm{LDI}}$ containing $\mathcal{B}_{\mathrm{MF}}$, and in Proposition 9 we discuss the case where $\mathcal{B}_{\mathrm{LDI}}$ consists of the indispensable moves.

We now start with the Graver basis.

Proposition 2. *The Graver basis is strongly 1-norm-reducing.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ be in the same fibre. As in (3), express $\mathbf{x}-\mathbf{y}$ as a conformal sum of non-zero elements of the Graver basis:

$$\mathbf{x}-\mathbf{y} = \mathbf{z}_1 + \ldots + \mathbf{z}_m.$$

Then $|\mathbf{x}-\mathbf{y}| = |\mathbf{z}_1| + \ldots + |\mathbf{z}_m|$. Now $\mathbf{z}_1$ can be subtracted from $\mathbf{x}$ and at the same time added to $\mathbf{y}$ to give

$$|(\mathbf{x}-\mathbf{z}_1)-\mathbf{y}| = |\mathbf{x}-(\mathbf{y}+\mathbf{z}_1)| = |\mathbf{z}_2| + \ldots + |\mathbf{z}_m| < |\mathbf{x}-\mathbf{y}|.$$

□

Note that the Graver basis is rich enough that we can take $\mathbf{z}_1 = \mathbf{z}_2$ in the definition of strong distance reduction in (5).

Proposition 3. *A set of moves $\mathcal{B}$ is 1-norm-reducing if and only if, for every element $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ of the Graver basis, $\mathcal{B}$ is 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$.*

Proof. We only have to prove sufficiency. Let $\mathbf{x}$, $\mathbf{y} \in \mathcal{F}_\mathbf{t}$ be arbitrarily given and let $\mathbf{x} - \mathbf{y} = \mathbf{z}_1 + \ldots + \mathbf{z}_m$ be a conformal sum of elements of the Graver basis. By assumption $\mathcal{B}$ is 1-norm-reducing for $\mathbf{z}_1^+, \mathbf{z}_1^-$. Among four possible cases, without loss of generality, consider the case where $\mathbf{z} \in \mathcal{B}$ is applicable to $\mathbf{z}_1^+$ and

$$|(\mathbf{z}_1^+ + \mathbf{z}) - \mathbf{z}_1^-| < |\mathbf{z}_1^+ - \mathbf{z}_1^-| = |\mathbf{z}_1|. \tag{6}$$

Since $\mathbf{z}$ is applicable to $\mathbf{z}_1^+$, $\mathbf{z}^- \leqslant \mathbf{z}_1^+ \leqslant (\mathbf{x} - \mathbf{y})^+$. Furthermore, (6) implies that

$$\varnothing \neq \mathrm{supp}(\mathbf{z}^+) \cap \mathrm{supp}(\mathbf{z}_1^-) \subset \mathrm{supp}(\mathbf{z}^+) \cap \mathrm{supp}((\mathbf{x} - \mathbf{y})^-).$$

It follows that $\mathbf{z}$ is applicable to $\mathbf{x}$ and $|(\mathbf{x} + \mathbf{z}) - \mathbf{y}| < |\mathbf{x} - \mathbf{y}|$. □

Note that the same statement holds for strong 1-norm reduction with exactly the same proof:

Proposition 4. *A set of moves $\mathcal{B}$ is strongly 1-norm-reducing if and only if, for every element $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ of the Graver basis, $\mathcal{B}$ is strongly 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$.*

Suppose that $\mathcal{B}$ is a 1-norm-reducing Markov basis. Then any $\mathcal{B}' \supset \mathcal{B}$ is a 1-norm-reducing Markov basis as well. In view of this, it is of interest to consider the minimality of 1-norm-reducing Markov bases. A 1-norm-reducing Markov basis $\mathcal{B}$ is minimal if every proper subset of $\mathcal{B}$ is not a 1-norm-reducing Markov basis. For a 1-norm-reducing Markov basis $\mathcal{B}$, we can examine, for each individual $\mathbf{z}$ of $\mathcal{B}$, whether $\mathcal{B} - \{\mathbf{z}\}$ is a 1-norm-reducing Markov basis. If $\mathcal{B} - \{\mathbf{z}\}$ is still 1-norm-reducing, we remove $\mathbf{z}$, recursively, until none of the remaining elements can be removed any further. Then we arrive at a minimal 1-norm-reducing Markov basis. Therefore every 1-norm-reducing Markov basis $\mathcal{B}$ contains a minimal 1-norm-reducing Markov basis.

Exactly the same argument holds concerning minimality of strongly 1-norm-reducing Markov bases. Every strongly 1-norm-reducing Markov basis contains a minimal strongly 1-norm-reducing Markov basis.

In Takemura and Aoki (2004) we considered the minimality of Markov bases. Similar arguments can be applied to the question of minimality of 1-norm-reducing Markov bases.

In order to study this minimality question we introduce three closely related notions of degree reduction of a move $\mathbf{z}$ by other moves. We say that a move $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ is *1-norm-reducible by another move* $\mathbf{z}' \neq \pm\mathbf{z}$ if $\mathbf{z}'$ is applicable to $\mathbf{z}^+$ and $|\mathbf{z} + \mathbf{z}'| < |\mathbf{z}|$ or if $\mathbf{z}'$ is applicable to $\mathbf{z}^-$ and $|-\mathbf{z} + \mathbf{z}'| = |\mathbf{z} - \mathbf{z}'| < |\mathbf{z}|$. We say that a move $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ is *strongly 1-norm-reducible by a pair of (other) moves* $\mathbf{z}_1, \mathbf{z}_2 \neq \pm\mathbf{z}$ if $\mathbf{z}_1$ is applicable to $\mathbf{z}^+$ and $|\mathbf{z} + \mathbf{z}_1| < |\mathbf{z}|$ and furthermore $\mathbf{z}_2$ is applicable to $\mathbf{z}^-$ and $|\mathbf{z} - \mathbf{z}_2| < |\mathbf{z}|$. Finally, we say that $\mathbf{z}$ is *1-norm-reducible by a lower-degree move* $\mathbf{z}'$ if $|\mathbf{z}'| < |\mathbf{z}|$ and $\mathbf{z}$ is 1-norm-reducible by $\mathbf{z}'$.

Consider the implications of these notions. If $\mathbf{z}$ is strongly 1-norm-reducible by $\mathbf{z}_1, \mathbf{z}_2$, then $\mathbf{z}$ is clearly 1-norm-reducible by $\mathbf{z}_1$ (or $\mathbf{z}_2$). We now show that if $\mathbf{z}$ is 1-norm-reducible by a lower-degree move $\mathbf{z}'$, then $\mathbf{z}$ is strongly 1-norm-reducible either by the pair $\mathbf{z}', \mathbf{z} + \mathbf{z}'$ or by the pair $\mathbf{z}' - \mathbf{z}, \mathbf{z}'$. To show this, first consider the case where $\mathbf{z}'$ is applicable to $\mathbf{z}^+$

and $|\mathbf{z}+\mathbf{z}'| < |\mathbf{z}|$. Let $\mathbf{z}'' = \mathbf{z}+\mathbf{z}'$. Then $|\mathbf{z}-\mathbf{z}''| = |\mathbf{z}'| < |\mathbf{z}|$ and we only need to check that $\mathbf{z}''$ is applicable to $\mathbf{z}^-$. In fact

$$\mathbf{z}'' = \mathbf{z}^+ - \mathbf{z}^- + (\mathbf{z}')^+ - (\mathbf{z}')^- = (\mathbf{z}^+ - (\mathbf{z}')^-) + (\mathbf{z}')^+ - \mathbf{z}^-$$
$$\geqslant (\mathbf{z}')^+ - \mathbf{z}^-.$$

This implies that $(\mathbf{z}'')^- \leqslant \mathbf{z}^-$ and $\mathbf{z}''$ is applicable to $\mathbf{z}^-$. Similarly, if $\mathbf{z}'$ is applicable to $\mathbf{z}^-$, we can check that $\mathbf{z}$ is strongly 1-norm-reducible by the pair $\mathbf{z}'-\mathbf{z}$, $\mathbf{z}'$.

Based on the above observation, we define three notions of irreducibility of a move. We call $\mathbf{z}$ *1-norm-irreducible* if it is not 1-norm-reducible by any other move $\mathbf{z}' \neq \mathbf{z}$. We call $\mathbf{z}$ *strongly 1-norm-irreducible* if it is not strongly 1-norm-reducible by any pair of other moves. Finally, we call $\mathbf{z}$ *1-norm lower-degree irreducible* if it is not 1-norm-reducible by any lower degree move. We state the above implications of the properties of moves, as well as further implications among indispensability and primitiveness, in the following proposition.

Proposition 5. *For a move* $\mathbf{z}$*, the following implications hold:*

$$\begin{aligned} \textit{indispensable} &\Rightarrow \textit{1-norm-irreducible} \\ &\Rightarrow \textit{strongly 1-norm-irreducible} \\ &\Rightarrow \textit{1-norm lower-degree irreducible} \\ &\Rightarrow \textit{primitive}. \end{aligned} \tag{7}$$

Proof. If $\mathbf{z}$ is not primitive, then $\mathbf{z}$ is clearly 1-norm-reducible by a lower-degree move. This proves the last implication.

If $\mathbf{z}$ is 1-norm-reducible by $\mathbf{z}' \neq \pm\mathbf{z}$, then $\mathbf{z}^- \neq \mathbf{z}^+ + \mathbf{z}' \in \mathcal{F}_{\mathbf{t}}$ or $\mathbf{z}^+ \neq \mathbf{z}^- + \mathbf{z}' \in \mathcal{F}_{\mathbf{t}}$, where $\mathbf{t} = A\mathbf{z}^+$. Therefore $\mathcal{F}_{\mathbf{t}}$ is not a two-element fibre. Therefore $\mathbf{z}$ is not indispensable. This proves the first implication. □

We will need a number of lemmas.

Lemma 1. *If* $\mathbf{z}$ *is 1-norm-reducible by another move* $\mathbf{z}' \neq \pm\mathbf{z}$*, then there exists a primitive move* $\mathbf{z}'' \neq \mathbf{z}$, $|\mathbf{z}''| \leqslant |\mathbf{z}'|$*, such that* $\mathbf{z}$ *is 1-norm-reducible by* $\mathbf{z}''$.

Proof. If $\mathbf{z}'$ is itself primitive just let $\mathbf{z}'' = \mathbf{z}'$. If $\mathbf{z}'$ is not primitive write $\mathbf{z}'$ as a conformal sum $\mathbf{z}' = \mathbf{z}_1 + \ldots + \mathbf{z}_m$ of non-zero elements of the Graver basis. Of the two possible cases, without loss of generality, consider the case where $\mathbf{z}'$ is applicable to $\mathbf{z}^+$ and $|\mathbf{z}+\mathbf{z}'| < |\mathbf{z}|$. In this case $(\mathbf{z}')^- \leqslant \mathbf{z}^+$ and $\text{supp}((\mathbf{z}')^+) \cap \text{supp}(\mathbf{z}^-) \neq \varnothing$. Since $\text{supp}((\mathbf{z}')^+) = \text{supp}(\mathbf{z}_1^+) \cup \cdots \cup \text{supp}(\mathbf{z}_m^+)$, there exists some l such that $\text{supp}((\mathbf{z}_l)^+) \cap \text{supp}(\mathbf{z}^-) \neq \varnothing$. Furthermore $\mathbf{z}_l^- \leqslant (\mathbf{z}')^- \leqslant \mathbf{z}^+$ and $|\mathbf{z}_l| < |\mathbf{z}'| \leqslant |\mathbf{z}|$. This implies that $\mathbf{z}$ is 1-norm-reducible by $\mathbf{z}'' = \mathbf{z}_l \neq \mathbf{z}$. □

The same argument proves the following lemma.

Lemma 2. *If* $\mathbf{z}$ *is strongly 1-norm-reducible by a pair of moves* $\mathbf{z}_1$, $\mathbf{z}_2 \neq \pm\mathbf{z}$ *then there exists a pair of primitive moves* $\mathbf{z}_1'$, $\mathbf{z}_2' \neq \pm\mathbf{z}$, $|\mathbf{z}_1'| \leqslant |\mathbf{z}_1|$, $|\mathbf{z}_2'| \leqslant |\mathbf{z}_2|$, *such that* $\mathbf{z}$ *is strongly 1-norm-reducible by the pair* $\mathbf{z}_1'$, $\mathbf{z}_2'$.

We now state some results on the minimality of 1-norm-reducing Markov bases. First we show that a 1-norm-reducing basis has to contain all 1-norm-irreducible moves.

Lemma 3. *Let* $\mathbf{z}$ *be a 1-norm-irreducible move. Then either* $\mathbf{z}$ *or* $-\mathbf{z}$ *belongs to every 1-norm-reducing Markov basis.*

Proof. We argue by contradiction. Let $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ be 1-norm-irreducible and let $\mathcal{B}$ be a 1-norm-reducing Markov basis not containing $\mathbf{z}$ nor $-\mathbf{z}$. Since $\mathcal{B}$ is 1-norm-reducing, $\mathcal{B}$ is 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$. But this contradicts the 1-norm irreducibility of $\mathbf{z}$ in view of (4). □

We say that there exists a unique minimal 1-norm-reducing Markov basis if all minimal 1-norm-reducing Markov bases coincide except for sign changes of their elements.

Proposition 6. *There exists a unique minimal 1-norm-reducing Markov basis if and only if 1-norm-irreducible moves form a 1-norm-reducing Markov basis.*

Proof. Since every 1-norm-irreducible move (or its sign change) belongs to every 1-norm-reducing Markov basis, if the set of 1-norm-irreducible moves is a 1-norm-reducing Markov basis, then it is clearly the unique minimal 1-norm-reducing Markov basis, ignoring the sign of each move.

Conversely, suppose that 1-norm-irreducible moves do not form a 1-norm-reducing Markov basis. Then every 1-norm-reducing Markov basis contains a 1-norm-reducible move. Let $\mathcal{B}$ be a minimal 1-norm-reducing Markov basis and let $\mathbf{z}_0 \in \mathcal{B}$ be 1-norm-reducible. Consider

$$\tilde{\mathcal{B}} = (\mathcal{B} \cup \mathcal{B}_{\text{Graver}}) - \{\mathbf{z}_0, -\mathbf{z}_0\},$$

where $\mathcal{B}_{\text{Graver}}$ is the Graver basis. We show that $\tilde{\mathcal{B}}$ is a 1-norm-reducing Markov basis. If this is the case $\tilde{\mathcal{B}}$ contains a minimal 1-norm-reducing Markov basis different from $\mathcal{B}$ even if we change the signs the elements.

Now by Propositions 1 and 3, it suffices to show that for every $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^- \in \mathcal{B}_{\text{Graver}}$, $\tilde{\mathcal{B}}$ is 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$. If $\mathbf{z}_0$ is not primitive, $\tilde{\mathcal{B}} \supset \mathcal{B}_{\text{Graver}}$ and $\tilde{\mathcal{B}}$ is 1-norm-reducing. Therefore let $\mathbf{z}_0$ be primitive. Each primitive $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^- \neq \mathbf{z}_0$ is already in $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{B}}$ is 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$. The only remaining case is $\mathbf{z} = \mathbf{z}_0$ itself, but by Lemma 1, $\mathbf{z}_0$ is 1-norm-reducible by a primitive $\mathbf{z}' \neq \pm\mathbf{z}_0$, $\mathbf{z}' \in \tilde{\mathcal{B}}$. □

Remark 1. In many examples we expect that the set of 1-norm-irreducible moves forms a Markov basis. However, it seems difficult to state a simple sufficient condition to guarantee this. The difficulty lies in eliminating the following possibility. Suppose that for some $\mathbf{t}$ with $|\mathbf{t}| = n$, the fibre $\mathcal{F}_\mathbf{t} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a three-element set with

$$\text{supp}(\mathbf{x}_3) \cap (\text{supp}(\mathbf{x}_1) \cup \text{supp}(\mathbf{x}_2)) = \emptyset, \qquad \text{supp}(\mathbf{x}_1) \cap \text{supp}(\mathbf{x}_2) \neq \emptyset.$$

Then $\mathbf{z}_1 = \mathbf{x}_1 - \mathbf{x}_3$ is 1-norm-reducible by $\mathbf{z}_2 = \mathbf{x}_2 - \mathbf{x}_3$ and vice versa. However, $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{B}_{\text{MF}}$ and if we throw away both $\mathbf{z}_1$ and $\mathbf{z}_2$, then $\mathcal{F}_\mathbf{t}$ is not connected. See Example 1 below. Also note that even if the set of 1-norm-irreducible moves forms a Markov basis, the Markov basis itself might not be 1-norm-reducing.

In Proposition 6 we considered the uniqueness of minimal 1-norm-reducing Markov bases. A parallel argument can be made concerning the uniqueness of minimal strongly 1-norm-reducing Markov bases in terms of strongly 1-norm-irreducible moves. The following proposition can be proved in exactly the same way as Proposition 6 using Lemma 2.

Proposition 7. *Let* $\mathbf{z}$ *be a strongly 1-norm-irreducible move. Then either* $\mathbf{z}$ *or* $-\mathbf{z}$ *belongs to every strongly 1-norm-reducing Markov basis. There exists a unique minimal strongly 1-norm-reducing Markov basis if and only if strongly 1-norm-irreducible moves form a strongly 1-norm-reducing Markov basis.*

Remark 2. In many examples we expect that the set of strongly 1-norm-irreducible moves forms a Markov basis. However, as in the case of 1-norm-reducing Markov bases, it seems difficult to state a simple sufficient condition to guarantee this. Suppose that for some $\mathbf{t}$ with $|\mathbf{t}| = n$, the fibre $\mathcal{F}_\mathbf{t} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ is a four-element set with

$$(\text{supp}(\mathbf{x}_1) \cup \text{supp}(\mathbf{x}_2)) \cap (\text{supp}(\mathbf{x}_3) \cup \text{supp}(\mathbf{x}_4)) = \emptyset,$$
$$\text{supp}(\mathbf{x}_1) \cap \text{supp}(\mathbf{x}_2) \neq \emptyset, \quad \text{supp}(\mathbf{x}_3) \cap \text{supp}(\mathbf{x}_4) \neq \emptyset.$$

Then $\mathbf{z}_{13} = \mathbf{x}_1 - \mathbf{x}_3$ is strongly 1-norm-reducible by the pair $-\mathbf{z}_{14} = \mathbf{x}_4 - \mathbf{x}_1$, $\mathbf{z}_{23} = \mathbf{x}_2 - \mathbf{x}_3$. Similarly, $\mathbf{x}_2 - \mathbf{x}_3$, $\mathbf{x}_1 - \mathbf{x}_4$, $\mathbf{x}_2 - \mathbf{x}_4$ are not strongly 1-norm-irreducible. However, these four moves belong to $\mathcal{B}_{\text{MF}}$ and if we throw away these four moves, then $\mathcal{F}_\mathbf{t}$ is not connected. See Example 1 below.

In Remarks 1 and 2 we noted the difficulty of establishing that the set of 1-norm-irreducible moves or the set of strongly 1-norm-irreducible moves forms a Markov basis. This corresponds to the fact that there are no general implications between the non-replaceability in (2) and (strong) 1-norm irreducibility. The following example illustrates this difficulty.

Example 1. As examples of Remarks 1 and 2, we consider a hierarchical model for $2 \times 2 \times 2 \times 2$ contingency tables, where the generating set is $12/13/23/34$. In this case, a frequency vector is

$$\mathbf{x} = (x_{1111}, x_{1112}, x_{1121}, x_{1122}, \ldots, x_{2211}, x_{2212}, x_{2221}, x_{2222})^\text{T}$$

and the sufficient statistic is

$$\mathbf{t} = (\{x_{ij\cdot\cdot}\}, \{x_{i\cdot k\cdot}\}, \{x_{\cdot jk\cdot}\}, \{x_{\cdot\cdot kl}\})^{\mathrm{T}}.$$

The corresponding matrix A is given by

$$A = \begin{bmatrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
1&1&0&0&1&1&0&0&0&0&0&0&0&0&0&0\\
0&0&1&1&0&0&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&0&0&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&1\\
1&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&1&1&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&1&1&0&0&0&0&0&0&1&1&0&0\\
0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{bmatrix}.$$

In Aoki and Takemura (2003b) we give a minimal Markov basis for this problem. We consider the case where the values of the sufficient statistic are given by

$$x_{ij\cdot\cdot} = x_{i\cdot k\cdot} = x_{\cdot jk\cdot} = 1, \quad 1 \leqslant i, j, k \leqslant 2,$$
$$x_{\cdot\cdot 11} = x_{\cdot\cdot 12} = 1, \qquad x_{\cdot\cdot 21} = 0, \qquad x_{\cdot\cdot 22} = 2.$$

In this case, the fibre corresponding to the above $\mathbf{t}$ contains four elements,

$$\mathbf{x}_1 = u_{1111}u_{1222}u_{2122}u_{2212}, \qquad \mathbf{x}_2 = u_{1112}u_{1222}u_{2122}u_{2211},$$
$$\mathbf{x}_3 = u_{1122}u_{1211}u_{2112}u_{2222}, \qquad \mathbf{x}_4 = u_{1122}u_{1212}u_{2111}u_{2222},$$

where $\{u_{ijkl}\}$ is the set of indeterminates. This corresponds to the situation in Remark 2.

Modifying the above example, we can make an example for Remark 1. In the above example, we have the restriction that $\mathbf{i} = 1212$ is a structural zero cell ($x_{1212} \equiv 0$) and omit the cell. In this case, a frequency vector $\mathbf{x}$ is treated as a 15-dimensional vector and A is 16×15. The corresponding fibre contains three elements,

$$\mathbf{x}_1 = u_{1111}u_{1222}u_{2122}u_{2212}, \qquad \mathbf{x}_2 = u_{1112}u_{1222}u_{2122}u_{2211}, \qquad \mathbf{x}_3 = u_{1122}u_{1211}u_{2112}u_{2222},$$

which corresponds to the situation of Remark 1.

We now show that the set of 1-norm lower-degree irreducible moves forms a Markov basis. In the following let $\mathcal{B}_{\mathrm{LDI}}$ denote the set of 1-norm lower-degree irreducible moves. In the case of contingency tables with fixed marginals, $\mathcal{B}_{\mathrm{LDI}}$ is invariant in the sense of Aoki and Takemura (2003b). We state the following proposition.

Proposition 8. *If a move* $\mathbf{z}$ *is non-replaceable by lower-degree moves, then* $\mathbf{z}$ *is 1-norm lower-degree irreducible. Hence* $\mathcal{B}_{\mathrm{MF}} \subset \mathcal{B}_{\mathrm{LDI}}$ *and* $\mathcal{B}_{\mathrm{LDI}}$ *is a Markov basis.*

Proof. If $\mathbf{z}$ is 1-norm-reducible by a lower-degree move $\mathbf{z}'$, then we can move from $\mathbf{z}^+$ to $\mathbf{z}^-$ either by two lower-degree moves $\mathbf{z}'$ and $\mathbf{z}+\mathbf{z}'$ or by $\mathbf{z}'-\mathbf{z}$ and $\mathbf{z}'$. Therefore $\mathbf{z}^+ \sim \mathbf{z}^- (\mathrm{mod}\, \mathcal{B}_{\deg \mathbf{z}-1})$. This implies that if $\mathbf{z}$ is non-replaceable by lower-degree moves, then $\mathbf{z}$ is 1-norm lower-degree irreducible. □

Remark 3. Together with indispensability and primitiveness, the result of Proposition 8 can be summarized as follows (cf. Proposition 5). For a move $\mathbf{z}$, the following implications hold:

$$\begin{aligned}\text{indispensable} &\Rightarrow \text{non-replaceable by lower-degree moves}\\ &\Rightarrow \text{1-norm lower-degree irreducible}\\ &\Rightarrow \text{primitive.}\end{aligned}$$

Finally, we consider the case where the set of indispensable moves is 1-norm-reducing.

Proposition 9. *Suppose that the set* $\mathcal{B}_{\mathrm{IDP}}$ *of indispensable moves is 1-norm-reducing. Then* $\mathcal{B}_{\mathrm{IDP}} = \mathcal{B}_{\mathrm{LDI}}$.

Proof. Let $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ be a move with $\mathbf{t} = A\mathbf{z}^+$. It suffices to prove that if $\mathbf{z}$ is not indispensable, then $\mathbf{z}$ is reducible by a lower-degree move $\mathbf{z}'$. By assumption $\mathcal{B}_{\mathrm{IDP}}$ is 1-norm-reducing. Therefore, there exists an indispensable move $\mathbf{z}'$, such that $\mathbf{z}$ is 1-norm-reducible by $\mathbf{z}'$. Note that $\mathbf{z}' \notin \mathcal{F}_{\mathbf{t}}$. By Takemura and Aoki (2004, Lemma 2.1), it follows that $|\mathbf{z}'| < |\mathbf{z}|$. Therefore $\mathbf{z}$ is 1-norm-reducible by a lower-degree move $\mathbf{z}'$. □

The following example illustrating Remark 3 and Proposition 9 was communicated to us by Hidefumi Ohsugi in 2004.

Example 2. Consider the 6×9 matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1\\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Here $\mathcal{I} = \{1, \ldots, 9\}$. It can be easily checked that the set of indispensable moves of degree 2 given by

$$\mathcal{B}_{\mathrm{IDP}} = \{u_2u_6 - u_7u_8,\, u_1u_5 - u_7u_9,\, u_3u_9 - u_4u_8\}$$

forms the unique minimal Markov basis. Now consider the fibre $\mathcal{F}_{\mathbf{t}}$ with $\mathbf{t} = (1, 1, 1, 1, 1, 1)$:

$$\mathcal{F}_{(1,\ldots,1)} = \{u_1u_3u_5,\, u_3u_7u_9,\, u_4u_7u_8,\, u_2u_4u_6\}$$

is a four-element set and the graph $G_{(1,\ldots,1),\mathcal{B}_{\text{IDP}}}$ is the following linear tree:

$$u_1u_3u_5 - u_3u_7u_9 - u_4u_7u_8 - u_2u_4u_6.$$

Therefore by $\mathcal{B}_{\text{IDP}}$ we can move from $u_1u_3u_5$ to $u_2u_4u_6$ but the only way to move is via $u_3u_7u_9$ and $u_4u_7u_8$. Hence, $\mathbf{z} = u_1u_3u_5 - u_2u_4u_6$ is replaceable by lower-degree moves but it is 1-norm lower-degree irreducible. Now note the 1-norms:

$$6 = |u_1u_3u_5 - u_2u_4u_6| = |u_3u_7u_9 - u_2u_4u_6| = |u_1u_3u_5 - u_4u_7u_8|.$$

This shows that $\mathcal{B}_{\text{IDP}}$ is not 1-norm-reducing.

Remark 4. Many of the results of this section hold for any metric d on $\mathbb{N}^{|\mathcal{I}|}$ with the following property. If $\text{supp}(\mathbf{x}) \cap \text{supp}(\mathbf{y}) = \emptyset$, $\text{supp}(\mathbf{x}') \subset \text{supp}(\mathbf{x})$, $\text{supp}(\mathbf{y}') \subset \text{supp}(\mathbf{y})$, $\deg \mathbf{x}' < \deg \mathbf{x}$ and $\deg \mathbf{y}' < \deg \mathbf{y}$, then

$$d(\mathbf{x}', \mathbf{y}') < d(\mathbf{x}, \mathbf{y}).$$

4. Examples and discussion

4.1. Examples

In this section, we consider some standard models of contingency tables. In Takemura and Aoki (2004, Section 3), we have considered minimal Markov bases and their uniqueness for these models. In this paper, we investigate minimal 1-norm-reducing Markov bases and their uniqueness. To display each move $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$, we write a binomial $\prod_{\mathbf{i}} u_{\mathbf{i}}^{z^+_{\mathbf{i}}} - \prod_{\mathbf{i}} u_{\mathbf{i}}^{z^-_{\mathbf{i}}}$ in this section, where $\{u_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}$ is the set of indeterminates. For some cases, we use 4ti2 (Hemmecke and Hemmecke 2003) to compute the reduced Gröbner basis and the Graver basis.

4.1.1. One-way contingency tables

First we consider the simplest case of one-way contingency tables. Let $\mathbf{x} = \{x_i\}$ be an I-dimensional frequency vector and $A = (1, \ldots, 1)$ in an I-dimensional row vector. The sufficient statistic $\mathbf{t}$ is the sample size n. This corresponds to testing the homogeneity of mean parameters for I independent Poisson variables conditional on the total sample size n. In this case, a minimum-fibre Markov basis is the set of moves of degree 1, that is,

$$\mathcal{B}_{\text{MF}} = \{\mathbf{z} | \mathbf{z} = u_i - u_j,\ i \neq j\}.$$

As shown in Takemura and Aoki (2004, Section 3), each element of $\mathcal{B}_{\text{MF}}$ is dispensable. However, it is obvious that these degree 1 moves are 1-norm-irreducible. Furthermore, the Graver basis in this case obviously coincides with $\mathcal{B}_{\text{MF}}$. Therefore, $\mathcal{B}_{\text{MF}}$ is the unique minimal strongly 1-norm-reducing Markov basis.

We now consider the reduced Gröbner basis with respect to purely lexicographic or degree lexicographic order. Let $u_1 \succ \cdots \succ u_I$. Then the reduced Gröbner basis is easily shown to be

$$\mathcal{B}_{\prec} = \{\mathbf{z} | \mathbf{z} = u_i - u_I, \ 1 \leqslant i \leqslant I-1\}.$$

We see that in order to move from one frequency vector to another, the frequencies have to pass through the last cell I. For example, to move from $(n, 0, \ldots, 0)$ to $(0, n, 0, \ldots, 0)$ we need $2n$ steps of $\mathcal{B}_{\prec}$ and $\mathcal{B}_{\prec}$ is not 1-norm-reducing. In the case of $\mathcal{B}_{\mathrm{MF}}$, n steps of $u_1 - u_2 \in \mathcal{B}_{\mathrm{MF}}$ are sufficient for moving from $(n, 0, \ldots, 0)$ to $(0, n, 0, \ldots, 0)$. We see that the diameter of $G_{n,\mathcal{B}_{\prec}}$ is $2n$, whereas the diameter of $G_{n,\mathcal{B}_{\mathrm{MF}}}$ is n.

4.1.2. Two-way contingency tables

Next we consider $I \times J$ two-way contingency tables with fixed row and column sums. Let $\mathbf{x} = \{x_{ij}\}$ be an IJ-dimensional column vector of cell frequencies and

$$A = \begin{bmatrix} 1_I^{\mathrm{T}} \otimes E_J \\ E_I \otimes 1_J^{\mathrm{T}} \end{bmatrix},$$

where 1_I is the I-dimensional vector consisting of 1s, E_J is the $J \times J$ identity matrix and $\otimes$ denotes the Kronecker product. This is a standard example of testing the hypothesis that the rows and the columns are independent. In this case, it is known that the set of degree 2 moves,

$$\mathcal{B} = \{u_{ij}u_{i'j'} - u_{ij'}u_{i'j}, \ i \neq i', \ j \neq j'\},$$

is the unique minimal Markov basis (see Takemura and Aoki 2004, Section 3). Here we have the following proposition.

Proposition 10. *$\mathcal{B}$ is the unique minimal strongly 1-norm-reducing Markov basis.*

Proof. In this case, it is well known that the Graver basis consists of all the moves which are written as

$$\mathbf{z} = u_{i_1 j_1} u_{i_2 j_2} \cdots u_{i_s j_s} - u_{i_2 j_1} u_{i_3 j_2} \cdots u_{i_1 j_s}, \tag{8}$$

where $(i_1, j_1), (j_1, i_2), \ldots, (i_s, j_s), (j_s, i_1)$ is a circuit in the complete bipartite graph $K_{I,J}$. Since $z^+_{i_1 j_1}, z^+_{i_2 j_2} > 0$ and $z^-_{i_2 j_1}, z^-_{i_3 j_2} > 0$, $\mathbf{z}$ is strongly 1-norm-reducible by a pair of moves $\mathbf{z}', \mathbf{z}'' \in \mathcal{B}$ where

$$\mathbf{z}' = u_{i_1 j_1} u_{i_2 j_2} - u_{i_2 j_1} u_{i_1 j_2},$$
$$\mathbf{z}'' = u_{i_3 j_1} u_{i_2 j_2} - u_{i_2 j_1} u_{i_3 j_2}.$$

Therefore, for any element $\mathbf{z}$ of the Graver basis, $\mathcal{B}$ is strongly 1-norm-reducing for $\mathbf{z}^+, \mathbf{z}^-$. Moreover, since each element of $\mathcal{B}$ is indispensable, it is also 1-norm-irreducible from Proposition 5. Uniqueness also follows from Proposition 8. □

We briefly mention two-way tables with structural zeros. In Aoki and Takemura (2005) we obtained the unique minimal Markov basis for two-way contingency tables with arbitrary patterns of structural zeros. It should be noted that even in the presence of structural zeros the primitive moves have the form (8), because other moves can be decomposed as the conformal sum of primitive moves from the argument in Aoki and Takemura (2005). Therefore the Graver basis for this case is obtained from the from the Graver basis for the case of no structural zeros, by removing elements which contain structural zero cells in the support. From these considerations, it can easily be proved that the unique minimal Markov basis is at the same time the unique minimal strongly 1-norm-reducing Markov basis for two-way tables with arbitrary patterns of structural zeros.

4.1.3. Three-way contingency tables with fixed one-dimensional marginals

Next we consider three-way contingency tables with fixed one-dimensional marginals. In this case, $\mathbf{x} = \{x_{ijk}\}$ is an IJK-dimensional frequency vector and

$$A = \begin{bmatrix} 1_I^{\mathrm{T}} \otimes 1_J^{\mathrm{T}} \otimes E_K \\ 1_I^{\mathrm{T}} \otimes E_J \otimes 1_K^{\mathrm{T}} \\ E_I \otimes 1_J^{\mathrm{T}} \otimes 1_K^{\mathrm{T}} \end{bmatrix}.$$

This corresponds to testing the hypothesis that the three factors are completely independent, that is, $p_{ijk} = \alpha_i \beta_j \gamma_k$. As Takemura and Aoki (2004) have shown, a minimal Markov basis for this case is not unique. The minimum-fibre Markov basis can be written as follows:

$$\begin{aligned}
\mathcal{B}_{\mathrm{MF}} &= \mathcal{B}_{\mathrm{IDP}} \cup \mathcal{B}^*, \\
\mathcal{B}_{\mathrm{IDP}} &= \{u_{ij_1k_1}u_{ij_2k_2} - u_{ij_1k_2}u_{ij_2k_1},\ j_1 \neq j_2,\ k_1 \neq k_2\} \\
&\quad \cup \{u_{i_1jk_1}u_{i_2jk_2} - u_{i_1jk_2}u_{i_2jk_1},\ i_1 \neq i_2,\ k_1 \neq k_2\} \\
&\quad \cup \{u_{i_1j_1k}u_{i_2j_2k} - u_{i_1j_2k}u_{i_2j_1k},\ i_1 \neq i_2,\ j_1 \neq j_2\}, \\
\mathcal{B}^* &= \{u_{i_1j_1k_1}u_{i_2j_2k_2} - u_{i_1j_1k_2}u_{i_2j_2k_1},\ u_{i_1j_1k_1}u_{i_2j_2k_2} - u_{i_1j_2k_1}u_{i_2j_1k_2}, \\
&\qquad u_{i_1j_1k_1}u_{i_2j_2k_2} - u_{i_1j_2k_2}u_{i_2j_1k_1},\ u_{i_1j_1k_2}u_{i_2j_2k_1} - u_{i_1j_2k_1}u_{i_2j_1k_2}, \\
&\qquad u_{i_1j_1k_2}u_{i_2j_2k_1} - u_{i_1j_2k_2}u_{i_2j_1k_1},\ u_{i_1j_2k_1}u_{i_2j_1k_2} - u_{i_1j_2k_2}u_{i_2j_1k_1}, \\
&\qquad i_1 \neq i_2,\ j_1 \neq j_2,\ k_1 \neq k_2\}.
\end{aligned}$$

Here, $\mathcal{B}_{\mathrm{IDP}}$ is the set of indispensable moves. $\mathcal{B}^*$ is the set of all degree 2 moves which connect all the elements of the four-element fibre

$$\begin{aligned}
\mathcal{F}_{i_1i_2j_1j_2k_1k_2} &= \{\mathbf{x} = \{x_{ijk}\} | x_{i_1\cdot\cdot} = x_{i_2\cdot\cdot} = x_{\cdot j_1\cdot} = x_{\cdot j_2\cdot} = x_{\cdot\cdot k_1} = x_{\cdot\cdot k_2} = 1\} \\
&= \{u_{i_1j_1k_1}u_{i_2j_2k_2},\ u_{i_1j_1k_2}u_{i_2j_2k_1},\ u_{i_1j_2k_1}u_{i_2j_1k_2},\ u_{i_1j_2k_2}u_{i_2j_1k_1}\}.
\end{aligned}$$

The minimal Markov basis in this case consists of $\mathcal{B}_{\mathrm{IDP}}$ and three moves for each $(i_1, i_2, j_1, j_2, k_1, k_2)$, which connects four elements of $\mathcal{F}_{i_1i_2j_1j_2k_1k_2}$ into a tree.

We now show that, for the $2 \times 2 \times 2$ case, each minimal Markov basis is not 1-norm-reducing and the minimum-fibre Markov basis is at the same time the unique minimal strongly 1-norm-reducing Markov basis. By 4ti2, the Graver basis for the $2 \times 2 \times 2$ case is calculated as

$$\mathcal{B}_{\text{Graver}} = \mathcal{B}_{\text{MF}} \cup \{u_{ijk}^2 u_{i'j'k'} - u_{ijk'} u_{ij'k} u_{i'jk},\ i \neq i',\ j \neq j' k \neq k'\}.$$

It is evident that each dispensable move in $\mathcal{B}_{\text{MF}}$ is 1-norm-irreducible, and therefore each minimal Markov basis is not 1-norm-reducing. On the other hand, a degree 3 move $u_{ijk}^2 u_{i'j'k'} - u_{ijk'} u_{ij'k} u_{i'jk}$ is strongly 1-norm-reducible by a pair of moves, $u_{ijk} u_{i'j'k'} - u_{ij'k'} u_{i'jk}$ and $u_{ijk} u_{ij'k'} - u_{ijk'} u_{ij'k}$ for example, which are in $\mathcal{B}_{\text{MF}}$.

Again by 4ti2, the reduced Gröbner basis for the $2 \times 2 \times 2$ case under the purely lexicographic or degree lexicographic term order

$$u_{111} \succ u_{112} \succ \cdots \succ u_{222}$$

is found to be

$$\mathcal{B}_{\prec} = \mathcal{B}_{\text{IDP}} \cup \{u_{121} u_{212} - u_{111} u_{222},\ u_{112} u_{221} - u_{111} u_{222},\ u_{122} u_{211} - u_{111} u_{222}\}.$$

This is a minimal Markov basis. Note that the four elements of $\mathcal{F}_{121212}$ are connected via $u_{111} u_{222}$. This is similar to the case of one-way contingency tables and the diameter of $G_{\mathbf{t},\mathcal{B}_{\prec}}$ is generally larger than the diameter $G_{\mathbf{t},\mathcal{B}_{\text{MF}}}$.

Note that two-way contingency tables and three-way contingency tables with fixed one-dimensional marginals are special cases of decomposable models considered in Dobra (2003). It is of interest to investigate the Markov basis given by Dobra (2003) in the framework of this paper.

4.1.4. Three-way contingency tables with fixed two-dimensional marginals

Next we consider three-way contingency tables with fixed two-dimensional marginals. In this case, $\mathbf{x} = \{x_{ijk}\}$ is an IJK-dimensional frequency vector and

$$A = \begin{bmatrix} 1_I^{\mathrm{T}} \otimes E_J \otimes E_K \\ E_I \otimes 1_J^{\mathrm{T}} \otimes E_K \\ E_I \otimes E_J \otimes 1_K^{\mathrm{T}} \end{bmatrix}.$$

This corresponds to testing no three-factor interactions of the log-linear model. As has been frequently stated, it is surprisingly difficult to obtain a Markov basis for this problem, except for small I, J, K.

For the case of $2 \times J \times K$ tables, Diaconis and Sturmfels (1998) have shown that the set of degree 4, 6, ..., $\min(J, K)$ moves, where the degree $2s$ move is written as

$$u_{1j_1k_1} u_{1j_2k_2} \cdots u_{1j_sk_s} u_{2j_2k_1} u_{2j_3k_2} \cdots u_{2j_1k_s} - u_{1j_2k_1} u_{1j_3k_2} \cdots u_{1j_1k_s} u_{2j_1k_1} u_{2j_2k_2} \cdots u_{2j_sk_s},$$

constitutes a Markov basis. Takemura and Aoki (2004) have shown that this is the unique minimal Markov basis for this problem. Our argument here is that this is also the unique minimal strongly 1-norm-reducing Markov basis. This is obvious from the fact that the above unique minimal Markov basis is also the Graver basis. See Sturmfels (1995, Corollary 14.12).

For the next simpler case, we consider $3 \times 3 \times 3$ tables. Aoki and Takemura (2003a) have shown that the unique minimal Markov basis for this problem consists of two types of moves,

$$u_{111}u_{122}u_{212}u_{221} - u_{112}u_{121}u_{211}u_{222}$$

and

$$u_{111}u_{123}u_{132}u_{212}u_{221}u_{233} - u_{112}u_{121}u_{133}u_{211}u_{223}u_{232}.$$

We show that the above basis $\mathcal{B}$ is also the unique minimal strongly 1-norm-reducing Markov basis. To show this, we have to check that, for each element $\mathbf{z}$ of the Graver basis, $\mathcal{B}$ is strongly 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$. From Sturmfels (1995, Theorem 14.13), augmenting the above two types by the following five types of moves gives the Graver basis:

$$u_{111}u_{123}u_{132}u_{222}u_{231}u_{313}u_{321} - u_{113}u_{122}u_{131}u_{221}u_{232}u_{311}u_{323},$$

$$u_{112}u_{121}u_{133}u_{222}u_{231}u_{311}u_{323}u_{332}^2 - u_{111}u_{123}u_{132}u_{221}u_{232}u_{312}u_{322}u_{331}u_{333},$$

$$u_{112}u_{123}u_{131}u_{213}u_{221}u_{232}u_{311}u_{322}u_{333} - u_{113}u_{121}u_{132}u_{211}u_{222}u_{233}u_{312}u_{323}u_{331},$$

$$u_{111}u_{123}u_{131}u_{213}^2u_{221}u_{231}u_{311}u_{322}u_{333} - u_{113}u_{122}u_{132}u_{211}^2u_{223}u_{233}u_{313}u_{321}u_{332},$$

$$u_{111}u_{123}^2u_{132}^2u_{213}u_{222}^2u_{231}u_{312}u_{321}u_{333} - u_{113}u_{122}^2u_{131}u_{133}u_{212}u_{221}u_{223}u_{232}u_{311}u_{323}u_{332}.$$

For each move $\mathbf{z}$ of the Graver basis, it is easy to see that $\mathcal{B}$ is strongly 1-norm-reducing for $\mathbf{z}^+$, $\mathbf{z}^-$, that is, we can apply elements of $\mathcal{B}$ to $\mathbf{z}^+$ and $\mathbf{z}^-$ and decrease $|\mathbf{z}^+ - \mathbf{z}^-|$.

4.1.5. Poisson regression

Here we consider a simple example of Poisson regression discussed in Diaconis *et al.* (1998). Let $\mathbf{x} = (x_0, x_1, \ldots, x_4)^{\mathrm{T}}$ and

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}.$$

For this problem, Takemura and Aoki (2004) have shown that a minimal Markov basis is not unique. The minimum-fibre Markov basis is given as follows:

$$\mathcal{B}_{\mathrm{MF}} = \mathcal{B}_{\mathrm{IDP}} \cup \mathcal{B}^*,$$

$$\mathcal{B}_{\mathrm{IDP}} = \{u_0u_3 - u_1u_2,\ u_0u_2 - u_1^2,\ u_1u_4 - u_2u_3,\ u_2u_4 - u_3^2\},$$

$$\mathcal{B}^* = \{u_0u_4 - u_1u_3,\ u_0u_4 - u_2^2,\ u_1u_3 - u_2^2\}.$$

Here $\mathcal{B}^*$ is the set of all degree 2 moves which connect all the elements of the three-element fibre

$$\mathcal{F} = \{u_0u_4,\ u_1u_3,\ u_2^2\}.$$

Therefore there are three minimal Markov bases for this case since any one move in $\mathcal{B}^*$ is not needed to construct a Markov basis.

Similarly to the $2 \times 2 \times 2$ tables with fixed one-dimensional marginals discussed above, $\mathcal{B}_{\mathrm{MF}}$ is the unique minimal strongly 1-norm-reducing Markov basis, and each minimal

Markov basis is not 1-norm-reducing. The latter statement is obvious since each element of $\mathcal{B}^*$ is 1-norm-irreducible. To show that $\mathcal{B}_{\mathrm{MF}}$ is strongly 1-norm-reducing, we compute the Graver basis for this problem by 4ti2. It is given as follows.

$$\mathcal{B}_{\mathrm{Graver}} = \mathcal{B}_{\mathrm{MF}} \cup \{u_0^3 u_4 - u_1^4,\ u_0^2 u_3 - u_1^3,\ u_0^2 u_4 - u_1^2 u_2,\ u_0 u_3^2 - u_1^2 u_4,\ u_2^3 - u_1^2 u_4,$$
$$u_0 u_3^2 - u_2^3,\ u_0 u_4^2 - u_2 u_3^2,\ u_3^3 - u_1 u_4^2,\ u_0 u_4^3 - u_3^4\}.$$

It is seen that each move above is strongly 1-norm-reducible by a pair of moves in $\mathcal{B}_{\mathrm{MF}}$.

4.1.6. Hardy–Weinberg model

Finally, we consider the Hardy–Weinberg model for four alleles, that is,

$$\mathbf{x} = (x_{11}, x_{12}, x_{13}, x_{14}, x_{22}, x_{23}, x_{24}, x_{33}, x_{34}, x_{44})^{\mathrm{T}}$$

and

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 \end{bmatrix}.$$

As is stated in Takemura and Aoki (2004), a minimal Markov basis for this case is not unique, and the minimum-fibre Markov basis is as follows:

$$\mathcal{B}_{\mathrm{MF}} = \mathcal{B}_{\mathrm{IDP}} \cup \mathcal{B}^*,$$
$$\mathcal{B}_{\mathrm{IDP}} = \{u_{i_1 i_1} u_{i_2 i_3} - u_{i_1 i_2} u_{i_1 i_3},\ u_{i_1 i_1} u_{i_2 i_2} - u_{i_1 i_2}^2\},$$
$$\mathcal{B}^* = \{u_{12} u_{34} - u_{13} u_{24},\ u_{12} u_{34} - u_{14} u_{23},\ u_{13} u_{24} - u_{14} u_{23}\},$$

where i_1, i_2, i_3 are all distinct, and $u_{ij} = u_{ji}$ for $i > j$. Here $\mathcal{B}^*$ is the set of all degree 2 moves which connect all the elements of the three-element fibre

$$\mathcal{F} = \{u_{12} u_{34},\ u_{13} u_{24},\ u_{14} u_{23}\}.$$

There are three minimal Markov bases for this case since any one move in $\mathcal{B}^*$ is not needed to construct a Markov basis. In this case, again, the minimum-fibre Markov basis is the unique minimal strongly 1-norm-reducing Markov basis, and each minimal Markov basis is not 1-norm-reducing. To show this, we give the Graver basis for this problem, computed by 4ti2. In this case, augmenting $\mathcal{B}_{\mathrm{MF}}$ by the following eight types of moves gives the Graver basis: 12 relations of $u_{12}^2 u_{33} - u_{13}^2 u_{22}$, 12 relations of $u_{12}^2 u_{34} - u_{13} u_{14} u_{22}$, 12 relations of $u_{12} u_{13} u_{24} - u_{11} u_{22} u_{34}$, 4 relations of $u_{11} u_{22} u_{33} - u_{12} u_{13} u_{24}$, 6 relations of $u_{11} u_{23} u_{24} - u_{22} u_{13} u_{14}$, 3 relations of $u_{12} u_{13} u_{24} u_{34} - u_{11} u_{22} u_{33} u_{44}$, 12 relations of $u_{12}^2 u_{34}^2 - u_{14}^2 u_{22} u_{33}$, 12 relations of $u_{12}^2 u_{33} u_{44} - u_{11} u_{23} u_{24} u_{34}$. It is easily seen that each move is strongly 1-norm-reducible by a pair of moves in $\mathcal{B}_{\mathrm{MF}}$.

4.2. Discussion

In this paper some results on the 1-norm-reducing Markov basis are obtained. Actually there remain more unsolved questions than answers in the framework of this paper. For example, one can ask under what conditions the minimum-fibre Markov basis $\mathcal{B}_{\mathrm{MF}}$ is 1-norm-reducing and, similarly, under what conditions $\mathcal{B}_{\mathrm{LDI}}$ is 1-norm-reducing. One can also ask when $\mathcal{B}_{\mathrm{LDI}}$ coincides with $\mathcal{B}_{\mathrm{MF}}$ or the universal Gröbner basis.

Another set of questions can be asked on reducing the distance in more than one step. We may call a Markov basis $\mathcal{B}$ *1-norm-reducing in* k *steps* if for every pair of states in the same fibre $\mathcal{F}_{\mathbf{t}}$, we can reduce the 1-norm by at most k moves from $\mathcal{B}$. Since the Graver basis is finite and for the 1-norm-reduction it suffices to move from the positive part to the negative part of each primitive move, every Markov basis $\mathcal{B}$ is 1-norm-reducing in k steps for some finite k. Then the natural question to ask is what is the minimum k such that $\mathcal{B}$ is 1-norm-reducing in k steps.

According to Proposition 3, we have to consider all the primitive moves to check whether a given Markov basis is 1-norm-reducing or not. However, it is generally difficult to compute the Graver basis even for a problem of moderate size. For example, in the case of $I \times J \times K$ three-way contingency tables with fixed one-dimensional marginals, we have shown in Takemura and Aoki (2004) that minimal Markov bases and the minimum-fibre Markov basis consist of two types of $2 \times 2 \times 2$ degree 2 moves only. In Section 4.1, for the simplest $2 \times 2 \times 2$ case, we have checked that the minimum-fibre Markov basis is the unique minimal strongly 1-norm-reducing Markov basis, but we do not know whether the same result holds for the general $I \times J \times K$ case, since the general form of the Graver bases is not known at present. In fact, we have found by 4ti2 that the Graver basis for the $2 \times 2 \times 3$ problem contains primitive moves of degree 4 such as

$$u_{111}u_{122}^2u_{213} - u_{221}u_{112}^2u_{123}.$$

Similarly, the Graver basis for the Hardy–Weinberg model of the four-allele problem suggests a complicated structure of the Graver basis for the general r-allele problem, while the minimum-fibre Markov basis is the same as in the four-allele case, as shown in Takemura and Aoki (2004).

Acknowledgement

The authors are grateful to a referee for very helpful comments and suggestions.

References

Aoki, S. and Takemura, A. (2003a) Minimal basis for connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals. *Aust. N. Z. J. Statist.*, **45**, 229–249.

Aoki, S. and Takemura, A. (2003b) Invariant minimal Markov basis for sampling contingency tables with fixed marginals. Technical Report METR 03-25, Department of Mathematical Engineering and Information Physics, University of Tokyo. Submitted for publication.

Aoki, S. and Takemura, A. (2005) Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *J. Statist. Comput. Simulation*. To appear.

Bubley, R. and Dyer, M. (1997) Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pp. 223–231. Los Alamitos, CA: IEEE Computer Society Press.

Diaconis, P. and Saloff-Coste, L. (1998) What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, **57**, 20–36.

Diaconis, P. and Sturmfels, B. (1998) Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, **26**, 363–397.

Diaconis, P., Eisenbud, D. and Sturmfels, B. (1998) Lattice walks and primary decomposition. In B.E. Sagan and R.P. Stanley (eds), *Mathematical Essays in Honor of Gian-Carlo Rota*, pp. 173–193. Boston: Birkhäuser.

Dinwoodie, I. H. (1998) The Diaconis–Sturmfels algorithm and rules of succession. *Bernoulli*, **4**, 401–410.

Dobra, A. (2003) Markov bases for decomposable graphical models. *Bernoulli*, **9**, 1093–1108.

Guo, S.W. and Thompson, E.A. (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrika*, **48**, 361–372.

Hemmecke, R. and Hemmecke, R. (2003) 4ti2 Version 1.1 — Computation of Hilbert bases, Graver bases, toric Gröbner bases, and more. Available at www.4ti2.de.

Hibi, T. (2003) *Gröbner Bases*. Tokyo: Asakura Shoten (in Japanese).

Ohsugi, H. and Hibi, T. (2003) Indispensable binomials of toric ideals. Submitted for publication.

Sturmfels, B. (1995) *Gröbner Bases and Convex Polytopes*. Providence, RI: American Mathematical Society.

Takemura, A. and Aoki, S. (2004) Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. *Ann. Inst. Statist. Math.*, **56**, 1–17.

Received April 2004 and revised August 2004