Efficient Markov bases for $\mathbb{Z}$-polytope sampling

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

at Massey University
Manawatū
New Zealand

Michael McVeagh
August 2021

Efficient Markov bases for $\mathbb{Z}$-polytope sampling

Michael McVeagh

Abstract

In this thesis we study the use of lattice bases for fibre sampling, with particular attention paid to applications in volume network tomography. We use a geometric interpretation of the fibre as a $\mathbb{Z}$-polytope to provide insight into the connectivity properties of lattice bases.

Fibre sampling is used when we are interested in fitting a statistical model to a random process that may only be observed indirectly via the underdetermined linear system $\mathbf{y} = A\mathbf{x}$. We consider the observed data $\mathbf{y}$ and random variable of interest $\mathbf{x}$ to contain count data. The likelihood function for such models requires a summation over the fibre $\mathcal{F}_{\mathbf{y}}$, the set of all non-negative integer vectors $\mathbf{x}$ satisfying this equation for some particular $\mathbf{y}$. This can be computationally infeasible when $\mathcal{F}_{\mathbf{y}}$ is large.

One approach to addressing this problem involves sampling from $\mathcal{F}_{\mathbf{y}}$ using a Markov Chain Monte Carlo algorithm, which amounts to taking a random walk through $\mathcal{F}_{\mathbf{y}}$. This is facilitated by a Markov basis: a set of moves that can be used construct such a walk, which is therefore a subset of the kernel of the configuration matrix A.

Algebraic algorithms for finding Markov bases based on the theory of Gröbner bases are available, but these can fail when the configuration matrix is large and the calculations become computationally infeasible. Instead, we propose constructing a sampler based on a type of lattice basis we call a column partition lattice basis, defined by a matrix U. Constructing such a basis is computationally much cheaper than constructing a Gröbner basis.

It is known that lattice bases are not necessarily Markov bases. We give a condition on the matrix U that guarantees that it is a Markov basis, and show for a certain class of configuration matrices how a U matrix that is a Markov basis can be constructed.

Construction of lattice bases that are Markov bases is facilitated when the configuration matrix is unimodular, or has unimodular partitions. We consider configuration matrices from volume network tomography, and give classes of traffic network that have configuration matrices with these desirable properties.

If a Markov basis cannot be found, one alternative is to sample from some larger set that includes $\mathcal{F}_{\mathbf{y}}$. We give some larger sets that can be used, subject to certain conditions.

Efficient Markov bases for $\mathbb{Z}$-polytope sampling

Michael McVeagh

Acknowledgements

First I would like to thank my main supervisor Martin Hazelton, whose guidance, knowledge, and Blackadder quotes proved invaluable. I would like to thank my other supervisors, Bruce van Brunt and Chris Tuffley. Their different supervision styles and areas of expertise perfectly complemented Martin's and always led to productive discussion, unless someone brought up grammar or optimal figure placement.

Thank you to my office mate Hammed and all the other Maths and Stats PhD students, to my Farsi teachers, and to the PhD futsal group for their friendship and support over the last four years.

I am grateful for financial support from Massey University in the form of a Vice-Chancellor's Doctoral Scholarship and a Covid 19 Scholarship Extension, from the School of Fundamental Sciences, and from the Royal Society of New Zealand Marsden Fund which partially supported my work.

I want to thank my parents Tim and Heather for always encouraging me.

Last, I wish to thank my sons Emil and Linus, and especially my partner Mareike, for their love, support, and encouragement throughout my studies and beyond.

Contents

List of Figures

Chapter 1

Introduction

In this thesis we are interested in methods for sampling from mathematical objects called $\mathbb{Z}$-polytopes. These consist of a set of points in some multi-dimensional space, each with non-negative integer co-ordinates. The points are bounded by a collection of hyperplanes. An example of a three-dimensional $\mathbb{Z}$-polytope is shown in Figure 1.1. $\mathbb{Z}$-polytopes generally lack a convenient representation for sampling directly, so we must instead turn to a technique called *Markov Chain Monte Carlo* (MCMC).

Markov Chain Monte Carlo can be implemented by taking a random walk through the $\mathbb{Z}$-polytope, collecting the points visited along the way. These visited points then become the sample. Constructing a walk with desirable properties is facilitated if we have a *Markov basis*, a set of moves that can be used to take a step from one point in the $\mathbb{Z}$-polytope to another.

In order for a given set of moves to qualify as a Markov basis, it must be capable of constructing a walk that can potentially visit every point in the $\mathbb{Z}$-polytope. Whether a set constitutes a Markov basis therefore depends on the structure of the $\mathbb{Z}$-polytope in question. In this thesis we are interested in how taking a geometric view might enable Markov basis construction and identification. The population from which we wish to sample is the non-negative integer vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$. We interpret this set as a $\mathbb{Z}$-polytope; the geometry of this $\mathbb{Z}$-polytope may provide insight into Markov basis construction and identification. In particular, we study a type of collection of moves we call a *column partition lattice basis*, and evaluate their potential use as Markov bases. The moves in a column partition lattice bases have a simple geometrical interpretation: they are moves in co-ordinate directions when the $\mathbb{Z}$-polytope is projected onto a subset of the axes.

Our motivation for studying $\mathbb{Z}$-polytope sampling comes from the study of statistical linear inverse problems. These are described in Section 1.1, where an example from network tomography is presented. The structure of the $\mathbb{Z}$-polytopes that arise in statistical linear inverse problems is determined by an underdetermined linear system $A\mathbf{x} = \mathbf{y}$, where A is a binary matrix and $\mathbf{x}$ and $\mathbf{y}$ are count vectors, so they contain only non-negative integers. The $\mathbb{Z}$-polytopes of interest are related to a translate of the kernel of A, where the translation is determined by the vector $\mathbf{y}$. Accordingly, whether or not a particular

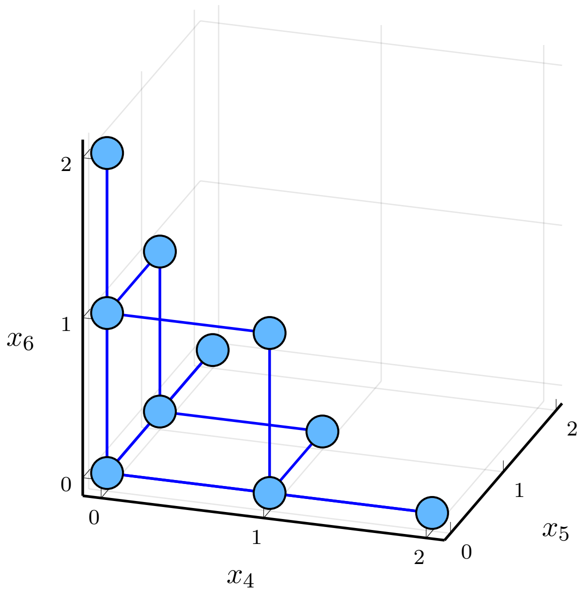


Figure 1.1: A three-dimensional $\mathbb{Z}$-polytope.

collection of moves allows construction of a random walk that can visit every point in the $\mathbb{Z}$-polytope is determined by properties of this linear system, which is to say of the matrix A and count vector $\mathbf{y}$.

After the description of statistical linear inverse problems in Section 1.1, Section 1.2 gives an overview of Markov Chain Monte Carlo and takes a closer look at what a Markov basis for $\mathbb{Z}$-polytope sampling might contain.

In Section 1.3 we lay out the problems this thesis addresses. The most important of these can be summarised as: given a configuration matrix A, how can we efficiently find a Markov basis for A? We then cover some existing work on the problem of finding Markov bases. This work tends to be algebraic rather than geometric. In particular, the Fundamental Theorem of Markov Bases gives a correspondence between a Markov basis and a generating set of an ideal in a polynomial ring. This approach is summarised in more detail later in Section 2.4.

Section 1.4 introduces lattice bases. Their potential use as Markov bases is the main topic of this thesis. This section covers some previous work on the use of lattice bases in polytope sampling and discusses what advantages and disadvantages they might might provide.

Section 1.5 describes three areas where statistical linear inverse problems arise: network tomography (Section 1.5.1), contingency table resampling (Section 1.5.2), and mark-recapture modelling in ecology (Section 1.5.3). Our main focus in this thesis is on network tomography.

Finally, Section 1.6 gives an overview of the rest of thesis.

1.1 A traffic flow model

Suppose that we are interested in building a statistical model for traffic flow on some road network. The network consists of a set of cities, the roads connecting them, and a collection of journeys that cars might make on those roads. Our model must provide a means for estimating the number of cars making each potential journey.

We represent this network as a directed graph where the nodes are the cities and the links are the roads. The potential journeys that cars on the network might make are represented as paths on the graph. It is usually infeasible to count traffic on paths directly: we can only observe the number of cars traversing each link during some time interval. If we wish to evaluate a statistical model for traffic counts on the network's paths, we can use a likelihood function. This requires a set of data against which to evaluate the model: in this case, a collection of traffic counts on each of the links in the network.

We collect the observed link traffic counts into the vector $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, where n is the number of links in the network and each entry y_i is a count of cars observed on the ith link. We similarly collect a combination of potential path traffic counts into a vector $\mathbf{x} \in \mathbb{Z}^r_{\geq 0}$, where there are r paths and each x_i is the count of cars making the ith journey. We know that multiple possible $\mathbf{x}$ vectors of path traffic counts can generate the same $\mathbf{y}$ vector of link traffic counts, so we will define the $\mathbf{y}$-*fibre* to be the set of potential count vectors $\mathbf{x}$ that can result in $\mathbf{y}$ being observed, and denote it $\mathcal{F}_\mathbf{y}$. Then the likelihood function for a path traffic count model is given by

$$\begin{aligned} \mathcal{L}(\theta) &= f(\mathbf{y}|\theta) \\ &= \sum_{\mathbf{x} \in \mathbb{Z}^r_{\geq 0}} f(\mathbf{y}|\mathbf{x}, \theta) f(\mathbf{x}|\theta) \\ &= \sum_{\mathbf{x} \in \mathcal{F}_\mathbf{y}} f(\mathbf{x}|\theta), \end{aligned} \tag{1.1.1}$$

where θ is a parameter vector for the model. In order to evaluate this, we need to be able to determine the elements of $\mathcal{F}_\mathbf{y}$ from a given $\mathbf{y}$.

The relationship between a network's path traffic counts and traffic counts on the network's links is given by the equation

$$A\mathbf{x} = \mathbf{y}, \tag{1.1.2}$$

where A is the *link-path incidence matrix*. There are n links in the network and r potential paths so A has n rows and r columns. The entries of A are given by $a_{ij} = 1$ if the ith link is traversed by car making the jth journey, and 0 otherwise.

The set of potential path traffic counts is defined by

$$\mathcal{F}_\mathbf{y} = \{\mathbf{x} \in \mathbb{Z}^r_{\geq 0} : A\mathbf{x} = \mathbf{y}\}.$$

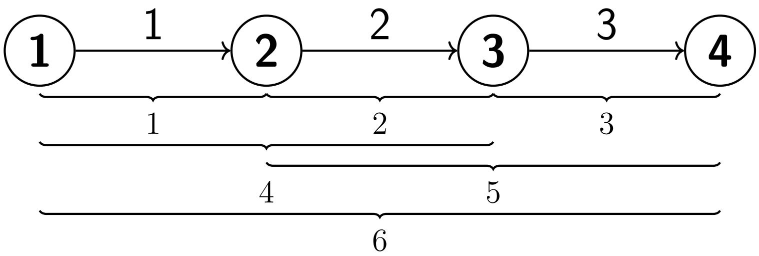


Figure 1.2: A representation of the three-link linear network, a traffic network with four cities connected linearly by three links. The underbraces show the six allowed paths.

Example 1.1.1 (Three-link linear network)**.** An example network, the three-link linear network, is shown in Figure 1.2. We are interested in east-bound traffic only, so we only show links connecting the cities in one direction. There are six potential journeys that cars may make on this network.

On the three-link linear network we have three traffic link counts, but there are six paths. This means that given some collection of traffic counts on links, multiple combinations of traffic counts on paths are possible. For example, suppose that one car is observed on each link in the network. In terms of path traffic counts, this could mean for example that there were three different cars that each drove a path consisting of just one of the links; or that there was one car that drove the entire length of the network.

We record these path traffic counts in the vector $\mathbf{y}$: in this case, we have $\mathbf{y} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$, and $\mathcal{F}_{\mathbf{y}}$ is given by

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix}1\\1\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix} \right\}.$$

In this case $\mathcal{F}_{\mathbf{y}}$ is small so we can enumerate it and we can evaluate the likelihood function (equation (1.1.1)) directly for these data. However, as the entries in $\mathbf{y}$ grow, and especially as the complexity of the network under study increases, the increased size of $\mathcal{F}_{\mathbf{y}}$ will make enumerating it computationally infeasible, making direct calculation of the likelihood function impossible. Instead of summing over every element of $\mathcal{F}_{\mathbf{y}}$ as in equation (1.1.1), we can perform statistical inference by taking a representative sample of the population $\mathcal{F}_{\mathbf{y}}$.

1.2 Markov Chain Monte Carlo

The lack of a convenient representation of $\mathcal{F}_{\mathbf{y}}$ means that direct sampling is impossible. Rejection sampling is impractical: the problems we are interested in are typically sufficiently large that it results in tiny acceptance rates. As a consequence, we resort to Markov Chain Monte Carlo, in which we aim to sample from a probability distribution $f_{X|Y}(\mathbf{x}|\mathbf{y},\theta)$ with support $\mathcal{F}_{\mathbf{y}}$ by constructing a Markov chain with invariant distribution equal to that target $f_{X|Y}$.

The particular form of $f_{X|Y}$ is not usually a major problem, since we can sample from some alternative proposal distribution q and then correct the sampling probabilities by applying the standard Metropolis-Hastings acceptance probability. We will not be particularly concerned with the functional form of q (it will generally suffice to think of it as a uniform distribution on some subset of $\mathcal{F}$). Our focus will be on the support of this distribution, so as to ensure that the sampler converges to its target distribution.

The most popular algorithm for sampling from fibres is random walk Metropolis-Hastings. This involves generating a random walk through the fibre using a Markov chain where the state space is $\mathcal{F}_{\mathbf{y}}$. A key step in this algorithm is generating a proposed next step $\mathbf{x}^\dagger \in \mathcal{F}_{\mathbf{y}}$ from the current state $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$. This uses a finite set of moves $\mathcal{B}$: for some $\mathbf{z} \in \mathcal{B}$, we can generate the proposed next step with $\mathbf{x}^\dagger = \mathbf{x} \pm \mathbf{z}$.

The sampling algorithm is set up so that the stationary distribution of the Markov chain matches some desired distribution. This is achieved by setting the acceptance probability α of a proposed next state $\mathbf{x}^\dagger$, given that $\mathbf{x}$ is the current state, to

$$\begin{aligned}\alpha &= \min\left\{1, \frac{f_{X|Y}(\mathbf{x}^\dagger|\mathbf{y},\theta)q(\mathbf{x}|\mathbf{x}^\dagger)}{f_{X|Y}(\mathbf{x}|\mathbf{y},\theta)q(\mathbf{x}^\dagger|\mathbf{x})}\right\} \\ &= \min\left\{1, \frac{f_X(\mathbf{x}^\dagger|\theta)q(\mathbf{x}|\mathbf{x}^\dagger)}{f_X(\mathbf{x}|\theta)q(\mathbf{x}^\dagger|\mathbf{x})}\right\}.\end{aligned}$$

Here, q is a proposal distribution supported within $\mathcal{F}_{\mathbf{y}}$ [45, 26].

A move, then, is required to be something that when applied to $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$ produces $\mathbf{x}^\dagger \in \mathcal{F}_{\mathbf{y}}$. The moves are therefore elements of $\ker_{\mathbb{Z}}(A)$, the integer elements of the kernel of the configuration matrix A.

In order for the sampler to converge to the desired distribution, the generated Markov chain must be irreducible. Assuming moves are selected at random, this will be the case if it is possible to eventually access every element of $\mathcal{F}_{\mathbf{y}}$ from every other element of $\mathcal{F}_{\mathbf{y}}$ using only the moves in $\mathcal{B}$ and following a walk that never leaves $\mathcal{F}_{\mathbf{y}}$.

Given a matrix A, a vector $\mathbf{y}$ and a pair of points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, if some set of moves $\mathcal{B}$ enables generation of a walk that starts at $\mathbf{x}_1$ and reaches $\mathbf{x}_2$ without leaving $\mathcal{F}_{\mathbf{y}}$, we say that $\mathbf{x}_1$ and $\mathbf{x}_2$ are *connected* by $\mathcal{B}$. If $\mathcal{B}$ connects all pairs of points in $\mathcal{F}_{\mathbf{y}}$, we say $\mathcal{B}$ connects $\mathcal{F}_{\mathbf{y}}$ and is a *Markov sub-basis* for A and $\mathbf{y}$, or for $\mathcal{F}_{\mathbf{y}}$.

If $\mathcal{B}$ connects $\mathcal{F}_{\mathbf{y}}$ for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, we call $\mathcal{B}$ a *Markov basis* for A.

Example 1.2.1. Let A be the link-path incidence matrix of the two-link linear network shown in Figure 1.3. Suppose that in some data set, three cars were observed on the first link in the network and four cars on the second. Then A and $\mathbf{y}$ are given by

$$A = \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix}3\\4\end{bmatrix}.$$

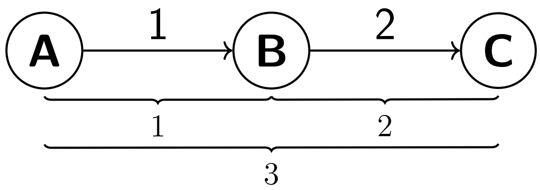


Figure 1.3: The two-link linear network from Example 1.2.1. The underbraces show the allowed paths.

The set of count vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$ is given by

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} \right\}.$$

We require a set of moves that enables us to move between the elements of $\mathcal{F}_{\mathbf{y}}$. Let

$$\mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \in \ker_{\mathbb{Z}}(A) \quad \text{and} \quad \mathcal{B} = \{\mathbf{z}\} \subseteq \ker_{\mathbb{Z}}(A).$$

Indexing the elements of $\mathcal{F}_{\mathbf{y}}$ by their first entry, we have

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_0 + \mathbf{z}, \\ \mathbf{x}_2 &= \mathbf{x}_0 + 2\mathbf{z}, \\ \mathbf{x}_3 &= \mathbf{x}_0 + 3\mathbf{z}. \end{aligned}$$

We can move from $\mathbf{x}_0$ to any other element of $\mathcal{F}_{\mathbf{y}}$ by adding an integer multiple of $\mathbf{z}$. Then $\mathcal{B}$ connects $\mathbf{x}_0$ to each other element of $\mathcal{F}_{\mathbf{y}}$, and therefore connects every pair of elements of $\mathcal{F}_{\mathbf{y}}$ to each other. It follows that $\mathcal{B}$ is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$.

If some set $\mathcal{B}$ contains moves that can be used to walk between two elements of some $\mathcal{F}_{\mathbf{y}}$, this set is not guaranteed to connect these elements. It could be that every walk between the two elements must step outside of $\mathcal{F}_{\mathbf{y}}$ at some point, as illustrated by the following example from Hazelton [25].

Example 1.2.2. Consider the triangular traffic network shown in Figure 1.4. Any path on the graph is included in the network except for the path $3 \to 1$, consisting of only the third link. The link-path incidence matrix is given by

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

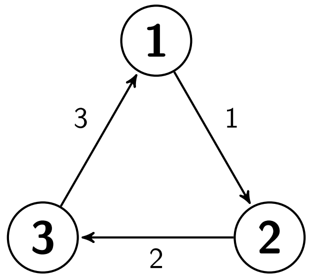


Figure 1.4: The triangular network from Example 1.2.2.

Setting $\mathbf{y} = \begin{bmatrix}4 & 4 & 4\end{bmatrix}^\intercal$ produces the fibre

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix}0\\4\\0\\0\\4\end{bmatrix}, \begin{bmatrix}1\\3\\0\\1\\3\end{bmatrix}, \begin{bmatrix}2\\2\\0\\2\\2\end{bmatrix}, \begin{bmatrix}3\\1\\0\\3\\1\end{bmatrix}, \begin{bmatrix}4\\0\\0\\4\\0\end{bmatrix}, \begin{bmatrix}0\\2\\1\\1\\3\end{bmatrix}, \begin{bmatrix}1\\1\\1\\2\\2\end{bmatrix}, \begin{bmatrix}2\\0\\1\\3\\1\end{bmatrix}, \begin{bmatrix}0\\0\\2\\2\\2\end{bmatrix} \right\}.$$

One potential set of moves is given by

$$\mathcal{B} = \left\{ \begin{bmatrix}1\\1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\2\\-1\\-1\\1\end{bmatrix} \right\}.$$

The set $\mathcal{B}$ makes up an integer basis for the integer kernel. The difference between every pair of elements of $\mathcal{F}_{\mathbf{y}}$ can be written as a sum of moves in $\mathcal{B}$, so a walk between any pair of points is possible. However, if one of the points is $\begin{bmatrix}0 & 4 & 0 & 0 & 4\end{bmatrix}^\intercal$, the walk necessarily leaves the polytope. Figure 1.5 shows $\mathcal{F}_{\mathbf{y}}$ plotted on the x_1 and x_5 axes and shows how this can happen. Superimposed on the polytope are some moves in $\mathcal{B}$ being used to connect elements of $\mathcal{F}_{\mathbf{y}}$. Neither of the moves in $\mathcal{B}$ can be applied to the point at $\begin{bmatrix}4 & 0 & 0 & 4 & 0\end{bmatrix}^\intercal$ without leaving the fibre.

The most important concern when selecting a collection of moves for use in fibre sampling is connectivity, but we also want the sampler to be efficient to run. Ideally we would like the sampled states to be independent, but because the proposed next steps in the walk are generated by a Markov chain they are dependent on the current state. Instead we must try to construct the sampler so that it minimises serial dependence as much as possible: we say that a sampler that does this exhibits *good mixing*. The choice of moves strongly affects the sampler's mixing, as illustrated by the following example.

Example 1.2.3. Consider the transport network in Figure 1.6. The nodes labelled 1 and 2 are origins for traffic, and the nodes labelled 3, 4, and 5 are destinations. Travel is

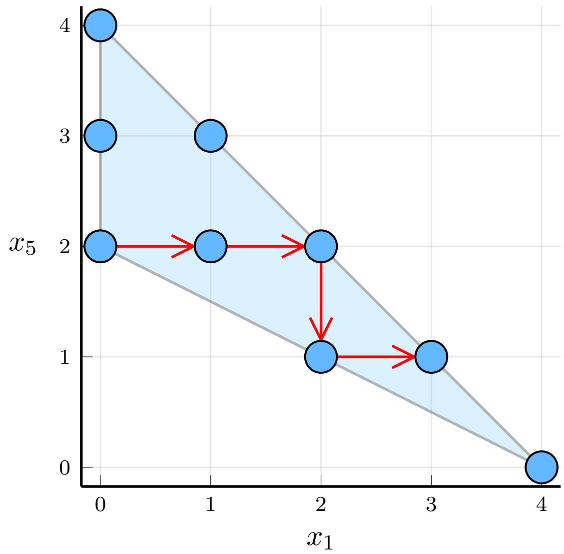


Figure 1.5: The projection of $\mathcal{F}_{\mathbf{y}}$ from Example 1.2.2 onto the x_1 and x_5 axes showing the inaccessible vertex.

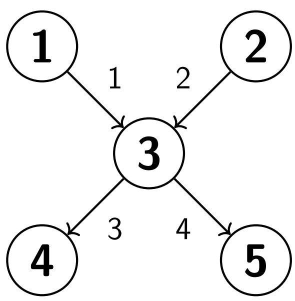
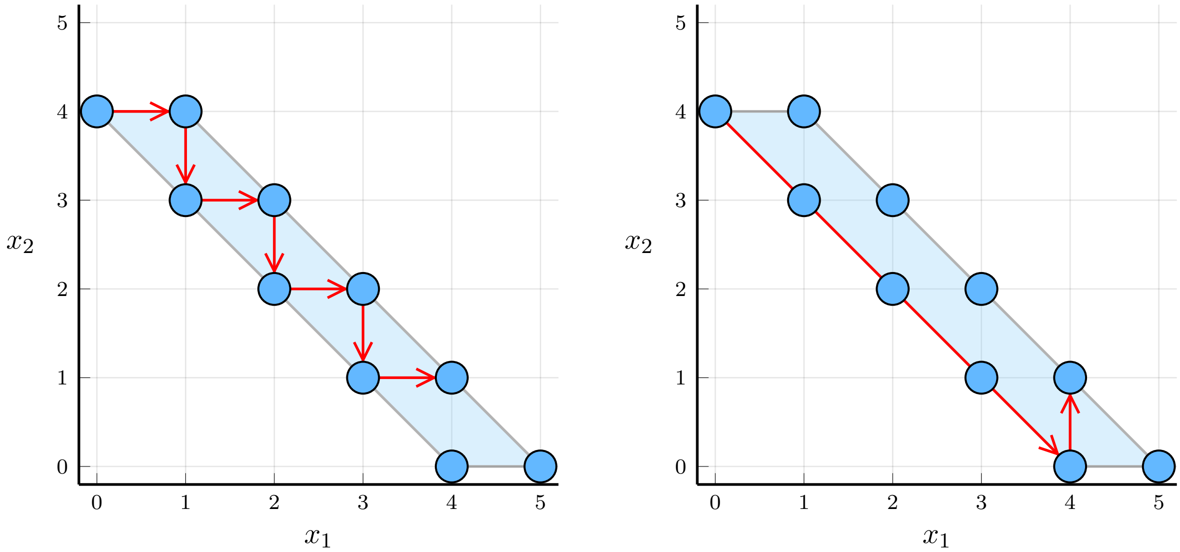


Figure 1.6: The transport network in Example 1.2.3.



(a) Using moves in $\mathcal{B}_1$.

(b) Using moves in $\mathcal{B}_2$.

Figure 1.7: The fibre $\mathcal{F}_{\mathbf{y}}$ from Example 1.2.3 plotted on the x_1 and x_2 axes comparing the mixing properties of two sets of moves.

permitted between any origin/destination pair. The link-path incidence matrix for this network is

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

Suppose the link traffic counts are given by $\mathbf{y} = \begin{bmatrix} 5 & 5 & 4 & 1 \end{bmatrix}^\intercal$. Then the fibre is given by

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 4 \\ 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 0 \\ 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \\ 4 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 0 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 1 \\ 1 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 4 \\ 1 \end{bmatrix} \right\}.$$

In this case $\mathcal{F}_{\mathbf{y}}$ is two-dimensional. Figure 1.7 shows $\mathcal{F}_{\mathbf{y}}$ plotted on the x_1 and x_2 axes.

If we construct our walk using as steps integer multiples of the elements of an integer basis for the integer kernel of A, given by

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\},$$

then when travelling around the fibre we are forced to zig zag. In Figure 1.7a, it takes seven steps to travel between the two selected elements of $\mathcal{F}_{\mathbf{y}}$.

If instead we use integer multiples of

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\},$$

then travel around $\mathcal{F}_{\mathbf{y}}$ is much swifter. In Figure 1.7b, it takes only two steps with $\mathcal{B}_2$ to make the same journey that took seven steps with $\mathcal{B}_1$. In fact, using $\mathcal{B}_2$ it takes at most three steps to move between any two states. This suggests that a sampler that uses as steps integer multiples of moves in $\mathcal{B}_2$ would have better mixing than one that uses integer multiples of moves in $\mathcal{B}_1$.

1.3 Finding Markov bases

Having a process for finding Markov bases and sub-bases for any particular problem is a critical research area. There are several related problems in which we are interested. First, recognising and constructing Markov bases: given a configuration matrix, how can we find a Markov basis? Can we recognise that some proposed set is a Markov basis? Second, given an algorithm for finding a Markov basis, can this algorithm be run on configuration matrices encountered in real world problems efficiently? And third, are there other desirable properties that a Markov basis might exhibit? What are these properties, and how can we recognise when a proposed Markov basis has them?

Algorithms already exist for finding Markov bases for a given configuration matrix, based on the algebraic work of Diaconis and Sturmfels in [19] and summarised in Section 2.4. But as we will see they are not without problems. Briefly, the Fundamental Theorem of Markov Bases, stated in this thesis as Theorem 2.4.5, identifies the kernel of the configuration matrix with an ideal in a polynomial ring. The problem of finding a Markov basis for a configuration matrix then becomes the problem of finding a generating set for this ideal. Most commonly, a particular type of generating set called a Gröbner basis is found.

The software package `4ti2` [44] is standard for finding Markov bases. It implements an algorithm [15] that is based on the work of Diaconis and Sturmfels [19]. While it is quick to find Markov bases for small problems, the time required for finding bases increases dramatically as the problem grows larger. Yoshida [50] observed that there are problems where it can never compute a Markov basis, and that computing Markov bases is NP-hard in general. Particular examples of problems where `4ti2` could not find a Markov basis in a reasonable amount of time were encountered by Schofield and Bonner [39], del Campo et al. [37], and Dinwoodie and Chen [21].

One of the examples in [21] where `4ti2` failed to find a Markov basis was for a three-way contingency table of size $7 \times 7 \times 2$. A database of Markov bases found by `4ti2` is maintained at [30] and is limited to bases for similar sized contingency tables. The largest four-way contingency table for which it provides a Markov basis is $3 \times 3 \times 3 \times 3$. The basis provided contains 303921 elements, which is a very large collection of moves. The database also contains Markov bases for some five-way and six-way tables with various marginals, but none has any dimension in which it is of length greater than two.

Because of the difficulty of finding complete Markov bases, various other sampling approaches have been tried. Dobra [22] observes full Markov bases found with the algebraic approach are particularly bad for sparse contingency tables, and proposes the use of dynamic Markov bases. Instead of determining a Markov basis before beginning the random walk, a walk using a dynamic Markov basis works by proposing a set of moves at each step in the walk that connect the current state $\mathbf{x}$ to nearby candidate states in $\mathcal{F}_{\mathbf{y}}$.

1.4 Lattice bases

Other approaches to MCMC fibre sampling have been proposed using lattice bases.

Definition 1.4.1 (Lattice basis [3])**.** A set $\mathcal{B}$ forms a *lattice basis* of $\ker_{\mathbb{Z}}(A)$ if every $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$ can be uniquely expressed as an integer combination of elements of $\mathcal{B}$.

One approach due to Aoki and Takemura [3] is to run a Markov chain without using a Markov basis. Proposing a move in their random walk involves randomly generating a set of multipliers from a distribution supported on $\mathbb{Z}$ with randomly chosen signs for each of the lattice basis elements, and summing them. The moves in their walk are therefore linear combinations of the elements of the lattice basis.

While this method is capable of generating all differences between elements of $\mathcal{F}_{\mathbf{y}}$, and therefore connecting $\mathcal{F}_{\mathbf{y}}$, it is not without problems. Suppose, for example, that the random walk is at a point in the fibre whose co-ordinates in some subset of dimensions are an extremum. In order to move from this point, the correct sign must be assigned to each of the moves that have been assigned non-zero multipliers: if there are k such moves, then the probability that the right sign is assigned to the relevant moves is $\frac{1}{2^k}$, and it is easy to see that in high dimensional problems, the walk may become stuck.

By contrast, our approach involves using a lattice basis directly. In this thesis, we are interested in the use of lattice bases for selecting moves in a $\mathbb{Z}$-polytope random walk sampler, where the steps in the walk are integer multiples of the lattice basis elements. Lattice bases have a particularly clear geometric interpretation, and are therefore a very attractive choice when developing samplers that respect the overall geometry of the $\mathbb{Z}$-polytope being sampled. We attempt to solve connectivity problems while still achieving good mixing by setting up the lattice basis so that it takes into account the geometry of the polytope. This generally involves including moves that sample in advantageous directions [25].

Little research has been done on connectivity of lattice bases. Tebaldi and West [45] developed the original version of the lattice polytope sampler we use. They claimed that the sampler always generates an irreducible Markov chain, but as we will see this is not the case. Aoki and Takemura [3] observe that while a Markov basis must always contain a lattice basis, configuration matrices for two-way contingency tables have no lattice bases that are Markov bases. They include little discussion of lattice bases beyond this.

Hazelton and co-authors [25, 26] developed a sampler which, during the random walk, continuously updates the lattice basis used in response to the local geometry of the projected polytope at the current state of the walk. Adapting the configuration matrix means finding a different lattice basis: this is sufficient in some cases to avoid some problems with connectivity of lattice bases. The method given for deciding how to adapt the basis also produces good mixing of the sampler [25].

Schofield and Bonner [39] found a method for constructing a single lattice basis that is a Markov basis, however it requires that the configuration matrix contains the identity matrix as a maximal submatrix. Their application was in capture-recapture studies in ecology, where it generally is the case that the identity matrix is a maximal submatrix of the configuration matrix [39]; this is not true for every application.

Samplers found using the algebraic approach may suffer due to inefficiency. For some configuration matrices, $\mathbf{y}$ vectors may be found such that MCMC sampling using a basis that does not take into account the polytope's geometry is arbitrarily inefficient. For example, if the vector $\mathbf{y}$ is such that for some index i, the entry $y_i = 0$, then all of the moves that involve altering an entry x_j in $\mathbf{x}$ such that $a_{ij} = 1$ can never be applied to any element of $\mathcal{F}_{\mathbf{y}}$. A sampler that must continually propose and reject such moves will not be very efficient.

We may find Markov chains with better mixing by tailoring the basis used to a particular $\mathbf{y}$ vector of interest. For example, if the polytope is long and thin, a basis with many moves in the dimensions in which the polytope is thin may decrease efficiency and slow mixing, as many opportunities to move in the longer dimensions are wasted.

1.5 Applications

Statistical linear inverse problems occur in a range of fields. Often the nature of the problem under study can imply some kind of structure on the associated configuration matrix, which may be useful when studying their Markov bases. If it is found that a configuration matrix having a particular structure implies that a collection of moves constructed in a certain way is a Markov basis, it is useful to know to which kind of problems this can be applied. We will now give some examples of applications.

1.5.1 Network tomography

Network tomography is our main motivation for studying polytope sampling and is the subject of Chapter 5. Generally it involves determining internal properties of a network based on observed properties. The networks may be computer and electronic networks [32, 17] or traffic networks [25, 26, 35, 45, 49].

We are interested particularly in estimating traffic counts on paths in traffic networks, as in the running example in Section 1.1, above. This particular type of network tomography is known as *volume network tomography.* A traffic network is represented by a graph, and a collection of paths on the graph. Given a vector of observed link traffic counts $\mathbf{y}$, we want to sample from the set $\mathcal{F}_{\mathbf{y}}$ of potential path traffic counts that could have led to $\mathbf{y}$. The traffic count vectors on links $\mathbf{y}$ and on paths $\mathbf{x}$ are related by $A\mathbf{x} = \mathbf{y}$, where A is the link-path incidence matrix of the network.

The configuration matrices in network tomography are frequently *unimodular.* A square matrix is unimodular if its determinant is ± 1. It is usual for configuration matrices to extend this definition to say that a rectangular matrix is unimodular if its invertible maximal submatrices all have determinant ± 1. This provides a large part of the motivation for our focus on unimodular matrices. Unimodularity of link-path incidence matrices is discussed in Section 5.2.

1.5.2 Contingency tables resampling

A contingency table is a multi-dimensional array containing count data, together with marginal totals. Markov Chain sampling methods are used to sample from the population of tables that share the same marginal totals. Contingency table resampling is the most common application and testing ground for theory around Markov bases, appearing as the motivating example (shown in Table 1.1) in Diaconis and Sturmfels' paper on the Fundamental Theorem of Markov Bases [19]. Their stated aim here was to test the independence of birthdays and deathdays.

Diaconis and Sturmfels proved that a set of moves defined by choosing two rows and two columns of a two-way contingency table, and modifying the four entries where they intersect with the moves

$$\begin{matrix} + & - \\ - & + \end{matrix} \quad \text{and} \quad \begin{matrix} - & + \\ + & - \end{matrix} \tag{1.5.1}$$

produces an irreducible Markov chain whose state space is the space of all two-way tables with the given marginal totals. These moves therefore constitute a Markov basis.

For example, for a small 2×3 contingency table, the contingency table representations of the moves making up a full Markov basis are given by

Month	Month of death												
of birth	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Total
Jan	1	0	0	0	1	2	0	0	1	0	1	0	6
Feb	1	0	0	1	0	0	0	0	0	1	0	2	5
March	1	0	0	0	2	1	0	0	0	0	0	1	5
April	3	0	2	0	0	0	1	0	1	3	1	1	12
May	2	1	1	1	1	1	1	1	1	1	1	0	12
June	2	0	0	0	1	0	0	0	0	0	0	0	3
July	2	0	2	1	0	0	0	0	1	1	1	2	10
Aug	0	0	0	3	0	0	1	0	0	1	0	2	7
Sept	0	0	0	1	1	0	0	0	0	0	1	0	3
Oct	1	1	0	2	0	0	1	0	0	1	1	0	7
Nov	0	1	1	1	2	0	0	2	0	1	1	0	9
Dec	0	1	1	0	0	0	1	0	0	0	0	0	3
Total	13	4	7	10	8	4	5	3	4	9	7	8	82

Table 1.1: The contingency table from Diaconis and Sturmfels [19] showing the months of birth and death of 82 descendants of Queen Victoria.

+		−	
−		+	

	+	−	
	−	+	

+	−		
−	+		

Diaconis and Sturmfels also looked at Markov bases for three-way contingency tables. They found that for $3 \times 3 \times 3$ tables, the reduced Gröbner basis contains 110 basic moves. In fact, Markov bases for three-way contingency tables with two-way margins were found to be arbitrarily complicated by De Loera and Onn [16].

Rapallo and Yoshida [38] looked at Markov bases for contingency tables which have upper bounds on their cell entries as well as lower bounds, and found that the set of moves in equation (1.5.1) is sufficient to connect $\mathcal{F}_{\mathbf{y}}$ as long as none of the upper bounds is zero. However, if some table entries are forced zeroes, a larger basis is required. Further work on contingency tables appears for example in [20, 34, 33, 16, 38, 3].

The configuration matrix A for a contingency table is a cell-margin incidence matrix for the contingency table. The vector $\mathbf{x}$ records the entries in the table and the vector $\mathbf{y}$ records the marginal totals. For the 2×3 table, this matrix is given by

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

and $\mathbf{x}$ and $\mathbf{y}$ are represented on the table by

x_1	x_2	x_3	y_1
x_4	x_5	x_6	y_2
y_3	y_4	y_5	

.

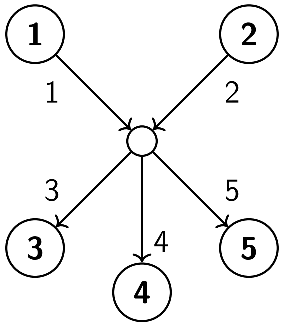


Figure 1.8: The traffic network corresponding to the 2×3 two-way contingency table.

The configuration matrix of a two-way contingency table is not full rank, and we usually remove a dependent row.

Two-way contingency tables are equivalent to a type of traffic network called a *star network*, an example of which is shown in Figure 1.8. Any entry in the table corresponds to a path in the network: the entry's row corresponds to the origin node and the column corresponds to the destination node. Thus the cell entry x_4 which appears in row 2 (marginal y_2), column 1 (marginal y_3) corresponds to the path on the star network that makes up column 4 of the configuration matrix, which travels from origin node 2 to destination node 3. For this reason, their polytope representations are sometimes called transportation polytopes.

1.5.3 Capture-recapture models

Another application comes from ecology. Identifying individual animals is an important part of capture-recapture studies, used for example in estimating population sizes [36]. However, observed data may be inaccurate due to misidentification of animals: a captured animal may be misidentified as one previously captured, when in reality two different animals were observed.

In Link et al. [36], a vector $\mathbf{y}$ of counts of animals with each possible recorded observation history is related to the vector $\mathbf{x}$ of counts of animals with each possible true observation history. Then $\mathbf{x}$ and $\mathbf{y}$ are related by $A\mathbf{x} = \mathbf{y}$, where A is a binary matrix containing a 1 in the entry at the ith row and jth column if true history i gives rise to recorded history j, and 0 otherwise. The exact form of the matrix is given by particulars such as the model being employed and the number of observation periods made. For example, a catalogue of animals may or may not exist: if it does not, then an animal can never be falsely identified the first time it is observed.

Statistical inference should be done for the count vector $\mathbf{x}$, the entries of which are counts partitioned by misidentification status. The statistical model being built can then take into account the possibility of misidentification in the data, leading to more accurate results [8, 39, 36].

Example 1.5.1. Suppose we are performing a capture-recapture study. We make observations over two periods during each of which each animal in the population is observed with some probability. An animal's observation histories are written as a string of digits which each represent the observation status of that animal during a survey: a 1 means the animal was observed, a 0 means the animal was not observed, and a 2 means the animal was observed but misidentified. The true capture histories are given by:

10: The animal is observed during the first period, but not the second. The recorded history is 10.

01: The animal is observed during the second period, but not the first. The recorded history is 01.

11: The animal is observed during both periods. The recorded history is 11.

12: The animal is observed during both periods, but is not recognised as the same animal. One animal with history 10 and one animal with history 01 are recorded.

There are 3 possible recorded histories and 4 possible true histories, so the configuration matrix A is 3×4 and is given by

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

This assumes that no catalogue of animals is available. If such a catalogue does exist, then animals may also be misidentified in the first observation period. Then the true histories are given by 01, 10, 11, 20, 02, 12, 21 and 22, and the configuration matrix is given by

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

The identity matrix occurs as a maximal submatrix in configuration matrices in capture-recapture models when every possible recorded capture history is also a potential true history [39].

1.6 Thesis overview

The aim of this thesis is to increase knowledge around finding bases suitable for use in random walk fibre samplers, with a focus on lattice bases. The way this problem is usually approached is with the algebraic methods introduced in Section 1.3 and discussed further in Section 2.4. This involves a correspondence between a Gröbner basis of an ideal in a polynomial ring, and a Markov basis for a configuration matrix.

By contrast, our approach involves lattice bases. It is generally much quicker to find a lattice basis than to find a Markov basis using the algebraic methods, however lattice bases are not always themselves Markov bases. We are therefore interested in recognising when a lattice basis is a Markov basis.

The type of lattice basis we use involves partitioning the columns of the configuration matrix, where different choices of column partition produce different lattice bases. We are also interested in which choices of column partition produce lattice bases that are Markov bases, and which do not.

One of our tools is to represent fibres with a $\mathbb{Z}$-polytope, defined in Section 2.2. The $\mathbb{Z}$-polytopes associated with any particular configuration matrix have much in common geometrically. By looking at the geometry of a configuration matrix's $\mathbb{Z}$-polytopes, we hope to gain insight into which collections of moves might constitute a Markov basis.

We focus mainly on configuration matrices that have at least one maximal square submatrix that is unimodular, and so invertible over the integers. Configuration matrices with this property seem to be common in network tomography applications (see Chapter 5 for more information). The existence of a unimodular maximal submatrix means that a lattice basis can be constructed that does not suffer from a particular kind of disconnectedness, which is described in Section 2.5.1.

Some configuration matrices may have only unimodular invertible maximal submatrices, so we say the matrix is itself said to be unimodular. Unimodular configuration matrices are the subject of Section 3.5. Unimodular configuration matrices have other nice properties, and we will take a special interest in them.

Chapter 2 gives a review of previous work on the problem of finding Markov bases. We begin in Section 2.2 by looking at geometric representations of fibres. We show the correspondence between different lattice bases and different projections of the $\mathbb{Z}$-polytope representation of a fibre.

In Section 2.3, different kinds of collections of moves are defined. These include Markov bases and sub-bases, the Graver basis, Gröbner bases, and lattice bases.

Gröbner bases and the Fundamental Theorem of Markov Bases are the subject of Section 2.4. A Gröbner basis is a type of generating set for an ideal in a polynomial ring. The Fundamental Theorem of Markov Bases lays out a correspondence between a generating set of an ideal and Markov bases. In this section we review some of the algebra around ideals and Gröbner bases.

Section 2.5 introduces column partition lattice bases. This is a type of lattice basis formed from a partition of the columns of the configuration matrix, and it is the focus of this thesis. A column partition lattice basis is not necessarily a Markov basis. We give some examples of what can go wrong, and provide a geometric view of each.

Chapter 3 is all about column partition lattice bases. In Section 3.2 we show how they are constructed, and in Section 3.3 we show the relationship between a column partition lattice basis and a projection of a $\mathbb{Z}$-polytope onto a lower dimensional space — the connectedness of a column partition lattice basis is closely related to the geometry of the corresponding projection of associated $\mathbb{Z}$-polytopes.

A column partition lattice basis is defined by the columns of a matrix, which we call U. Section 3.4 gives more information about U matrices and their columns. Section 3.5 looks at the impact unimodularity of a configuration matrix has upon its column partition lattice bases. A key result in this section concerns the relationship between the vectors that make up a column partition lattice basis and a type of integer kernel element called a circuit. For unimodular configuration matrices, the union of the column partition bases equals the set of circuits, which is also equal to the Graver basis, which is known to be a Markov basis.

Chapter 4 looks at connectivity of column partition lattice bases. We focus on column partition lattice bases formed from unimodular column partitions. In Section 4.2 we build on the method of Schofield and Bonner [39], who found that a column partition lattice basis could be constructed when the configuration matrix has the identity matrix as a maximal submatrix. We show that their condition can weakened, and the same line of reasoning guarantees that the column partition lattice basis constructed is a Markov basis.

In Section 4.4 we give a still weaker condition on the matrix U that guarantees that the column partition lattice basis it defines is a Markov basis, and we conjecture that this is also a necessary condition. For both Section 4.2 and Section 4.4 we provide a proof of connectivity using the Fundamental Theorem of Markov Bases. We also show how the geometry of the associated $\mathbb{Z}$-polytopes is affected by the conditions we give, and use this to build a proof via distance reduction. We conclude this chapter with some brief remarks on how column partition lattice bases that are not themselves Markov bases might be combined to form a Markov basis.

Chapter 5 concerns our main motivating application, network tomography. The configuration matrices of interest here are the link-path incidence matrices of traffic networks. We focus on investigating how the properties of the network affect the determinants of the maximal submatrices of the link-path incidence matrices. In particular, we look at properties of the underlying graph of the network, and the influence of rules regarding which paths are allowed routes for traffic in the network (which we term a *routing policy*).

We begin this chapter with a review of some relevant graph theory, including the definition and an example of a link-path incidence matrix.

In Section 5.2 we look at what is currently known about the determinants of submatrices of link-path incidence matrices, in particular maximal submatrices. We find that unimodularity is a common property of link-path incidence matrices arising in real world problems, but it is not always guaranteed. Examples of traffic networks where there are no unimodular maximal submatrices are given. We show by example that determinants of maximal submatrices are unbounded in general.

Enforcing some kind of structure on the network's graph can result in more well-behaved link-path incidence matrices. In Section 5.3 we consider networks on a kind of a graph called a *polytree*, which is a tree with directed links. We find that polytrees have link-path incidence matrices that are not only unimodular, but *totally unimodular*, meaning that all invertible submatrices have determinant ± 1.

In Section 5.4 we turn our attention to traffic networks on symmetric directed graphs. These appear to closely model real-world traffic networks, in that the connections in the graph are bidirectional: if there exists a link connecting node A to node B, then there is also a node connecting node B to node A.

Symmetric directed graphs for which the underlying graph is a tree do not in general have unimodular configuration matrices, but we show that given certain assumptions on the routing on the network, unimodular partitions can be found. We speculate that this result can be extended to all symmetric directed graphs.

Chapter 6 concerns an extension of the concept of a Markov basis to what we term an $\mathbf{m}$-Markov basis. For a collection of moves to be a Markov basis for a configuration matrix A requires that for all $\mathbf{y}$, they are capable of constructing a walk between any pair of points in $\mathcal{F}_{\mathbf{y}}$ such that the walk never visits a point with a negative entry in any co-ordinate, which is to say it need never leave the $\mathbb{Z}$-polytope. For $\mathbf{m}$-Markov bases, this condition is relaxed: an $\mathbf{m}$-Markov basis for a configuration matrix A and a non-negative integer vector $\mathbf{m}$ is a collection of moves such that for all $\mathbf{y}$, it is capable of constructing a walk that never visits any point $\mathbf{x}$ such that for any i, $x_i \leq -m_i$. Any Markov basis is therefore a $\mathbf{0}$-Markov basis.

An $\mathbf{m}$-Markov basis can be used to sample from some fibre $\mathcal{F}_{\mathbf{y}}$ by construct a walk through this slightly larger $\mathbb{Z}$-polytope, and discarding from the sample any elements that have any negative co-ordinates.

Our main theme is the Minus One Conjecture (Conjecture 6.3.2). The claim of the Minus One Conjecture is this: let A be a unimodular configuration matrix, let U be a column partition lattice basis, and let $\mathbf{y}$ be given. Then for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, U can be used to construct a walk between $\mathbf{x}_1$ and $\mathbf{x}_2$ that never visits a point with an entry less than -1 in the first n co-ordinates — these are the co-ordinates in the A_1 part of the configuration matrix. In real world problems this may represent a small portion of the total number of co-ordinates. Typically these are chosen to be the co-ordinates with high expected values — the busiest routes the a traffic network, for example [25]. Our hope is that by restricting stepping outside the $\mathbb{Z}$-polytope to these n dimensions, the proportion of the sample that needs to be discarded is minimised.

We include proofs of a few cases of the Minus One Conjecture in Section 6.4. The most important of these is when the configuration matrix, or the U matrix, is a *network matrix* (defined in Definition 5.3.2). This is an important result because all link-path incidence matrices for polytrees, and because it may provide a pathway to a proof for the full minus one conjecture: network matrices are key building blocks for all totally unimodular matrices.

Chapter 2

Polytopes, bases, and algebra

2.1 Introduction

In this chapter we look in more detail at some of the previous work on the problem of finding Markov bases. We first look at $\mathbb{Z}$-polytopes, in Section 2.2. The $\mathbb{Z}$-polytopes of interest are geometric interpretations of $\mathcal{F}_{\mathbf{y}}$, the $\mathbf{y}$-fibre from which we wish to sample.

In Section 2.3 we define some types of bases that may be of use. The bases of interest consist of elements in the integer kernel of the configuration matrix that span that integer kernel over $\mathbb{Z}$. This is a minimum requirement for a set of moves to be a Markov basis. Types of bases we cover include the Graver basis, Gröbner bases, and lattice bases.

Section 2.4 covers Gröbner bases in more detail. Gröbner bases representing the kernel of the configuration matrix are the most commonly used type of Markov bases. This section includes a statement of the Fundamental Theorem of Markov Bases (Theorem 2.4.5).

In Section 2.5 we turn to a type of basis that is the focus of this thesis, which we call a *column partition lattice basis*. Column partition lattice bases are lattice bases that are formed using a particular technique based on partitioning the columns of the configuration matrix. A single configuration matrix may have many different suitable column partitions, and therefore many different column partition lattice bases.

Column partition lattice bases are not necessarily Markov bases. In Section 2.5.1 we look at some examples of column partition lattice bases that are not Markov bases, and give a geometric interpretation of why this is.

Section 2.5.2 gives an example due to Schofield and Bonner [39] of a column partition lattice basis that is a Markov basis.

2.2 $\mathbb{Z}$-polytopes

In this section we give a review of $\mathbb{Z}$-polytopes as geometric representations of $\mathbf{y}$-fibres. Recall from Section 1.1 that given a vector $\mathbf{y}$, the $\mathbf{y}$-fibre is the set of non-negative integer vectors $\mathbf{x} \in \mathbb{Z}^r_{\geq 0}$ satisfying the linear equation $A\mathbf{x} = \mathbf{y}$.

Definition 2.2.1 (Fibre)**.** Let A be an $n \times r$ matrix of rank n and let $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be a count vector. Then the $\mathbf{y}$-fibre, denoted $\mathcal{F}_\mathbf{y}$, is the set

$$\mathcal{F}_\mathbf{y} = \{\mathbf{x} \in \mathbb{Z}^r_{\geq 0} : A\mathbf{x} = \mathbf{y}\}.$$

These $\mathbb{Z}$-polytopes are the intersection of a translate of the kernel of A with the non-negative orthant and the integer lattice.

Any vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$ is equal to some particular solution $\mathbf{x}_0$ plus some element of the kernel of A, which has dimension $r - n$. The set of solutions $\{\mathbf{x} : A\mathbf{x} = \mathbf{y}\}$ is equal to a translate of the kernel of A by $\mathbf{x}_0$, and is therefore an $(r - n)$-dimensional affine subspace of $\mathbb{R}^r$.

We are interested in the set of non-negative solutions to this system, which geometrically is the intersection of this affine space with the non-negative orthant. In our applications, configuration matrices A have the property that there is a positive vector in the row space, so $\ker(A)$ intersects the non-negative orthant at the origin only. The intersection of a translate of $\ker(A)$ with the non-negative orthant is therefore finite, and is an $(r - n)$-dimensional polytope.

The set $\mathcal{F}_\mathbf{y}$ contains only the integral elements of this polytope, so it is the intersection of this polytope with the integer lattice [26].

Definition 2.2.2 ($\mathbb{Z}$-polytope)**.** A *$\mathbb{Z}$-polytope* is the intersection of a polytope with the integer lattice.

The $\mathbb{Z}$-polytopes we are interested in have a convex underlying polytope that lies entirely in the non-negative orthant.

We give an example of the construction of a $\mathbb{Z}$-polytope from a configuration matrix and vector $\mathbf{y}$. The example is low dimensional to aid visualisation.

Example 2.2.3. Consider the configuration matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix},$$

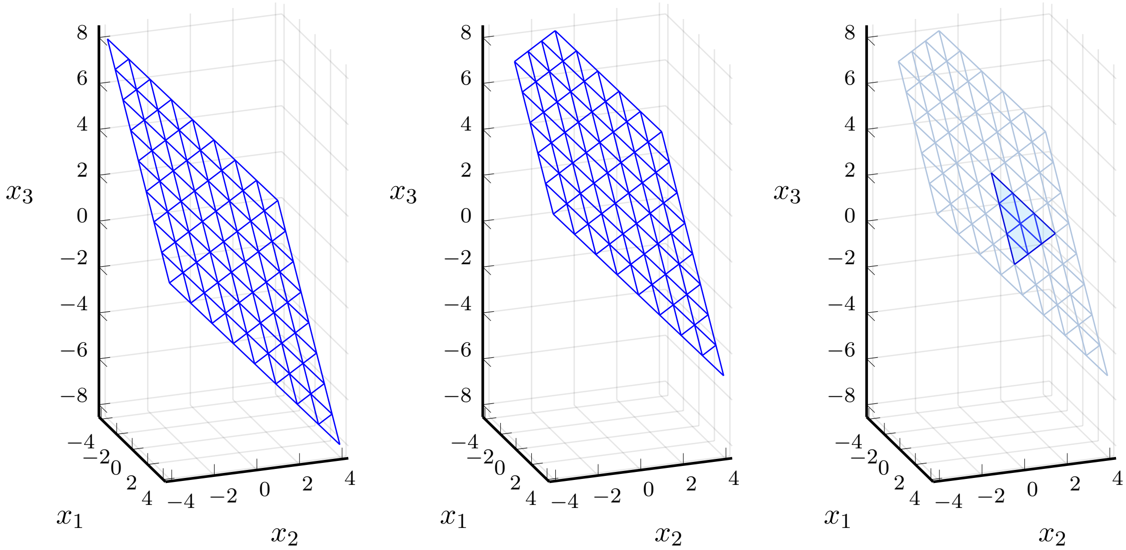
and the vector $\mathbf{y} = \begin{bmatrix} 3 \end{bmatrix}^\intercal$. The solitary row vector $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ is orthogonal to the kernel of A, which is shown in Figure 2.1a.

A particular solution to $A\mathbf{x} = \mathbf{y}$ is given by $\mathbf{x}_0 = \begin{bmatrix} 3 & 0 & 0 \end{bmatrix}^\intercal$. The corresponding translate of $\ker(A)$ is shown in Figure 2.1b, and the polytope comprising the intersection of this translate with the non-negative orthant is shown in Figure 2.1c. In this case the polytope is a triangle.

Figure 2.2a zooms in on the non-negative orthant to show the polytope. From there, we intersect the polytope with the integer lattice to get the $\mathbb{Z}$-polytope representation of $\mathcal{F}_\mathbf{y}$. The set $\mathcal{F}_\mathbf{y}$ is given by

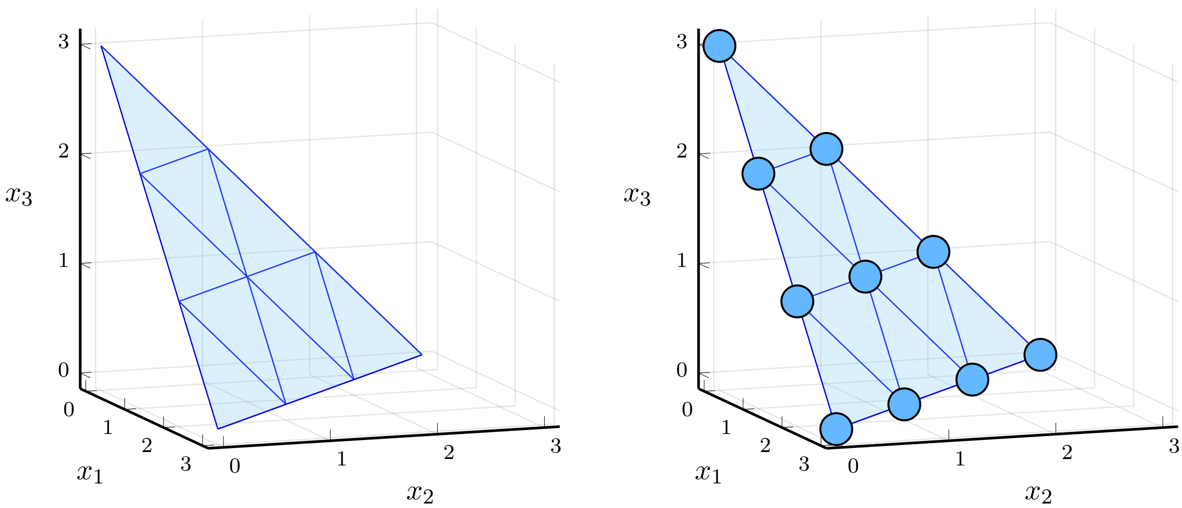
$$\mathcal{F}_\mathbf{y} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \right\},$$

and these are the points shown in Figure 2.2b.



(a) The kernel of A. (b) The kernel translated. (c) The non-negative part.

Figure 2.1: The construction of the polytope in Example 2.2.3.



(a) The polytope. (b) The ℤ-polytope.

Figure 2.2: The construction of the ℤ-polytope in Example 2.2.3.

2.3 Bases

In linear algebra, a basis is a set of vectors in some vector space such that each vector in the vector space can be written uniquely as a linear combination of the elements of the basis. The size of the basis is equal to the dimension of the vector space.

For the kinds of bases we discuss, this is not necessarily the case. A Markov basis for a configuration matrix A requires a spanning set of the kernel of A, but as we will see, more moves may be necessary. We will now describe some of the bases relevant to our work.

2.3.1 Markov bases and sub-bases

Markov bases and sub-bases are collections of moves that enable construction of a walk through a fibre that is capable of visiting every point in that fibre, while remaining at all times within the fibre. The moves in the basis or sub-basis are therefore vectors in the integer kernel of the configuration matrix.

We first define what it means for two points in a fibre to be *connected*.

Definition 2.3.1. Let $A \in \{0,1\}^{n\times r}$ be a configuration matrix and let $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be a non-negative integer vector. Let $\mathcal{B} \in \ker_{\mathbb{Z}}(A)$ be a collection of vectors. We say that $\mathcal{B}$ *connects* two points $\mathbf{x}_1, \mathbf{x} \in \mathcal{F}_{\mathbf{y}}$ if there is a sequence of moves $\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_k} \in \mathcal{B}$ and a sequence of signs $\epsilon_1, \ldots, \epsilon_k \in \{\pm 1\}$ such that

$$\mathbf{x}_1 + \sum_{j=1}^{k} \epsilon_j \mathbf{z}_{i_j} = \mathbf{x}_2, \tag{2.3.1}$$

and such that for all $m = 1, \ldots, k$,

$$\mathbf{x}_1 + \sum_{j=1}^{m} \epsilon_j \mathbf{z}_{i_j} \in \mathcal{F}_{\mathbf{y}}. \tag{2.3.2}$$

The condition in equation (2.3.1) means that the moves take the walk from $\mathbf{x}_1$ to $\mathbf{x}_2$. The condition in equation (2.3.2) means that the walk remains within $\mathcal{F}$ for its duration.

If a collection of moves $\mathcal{B}$ connects all of the points in some given fibre, then we say that $\mathcal{B}$ is a *Markov sub-basis* for that fibre.

Definition 2.3.2 (Markov sub-basis)**.** Let $A \in \{0,1\}^{n\times r}$ be a configuration matrix and let $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be a non-negative integer vector. Let $\mathcal{B} \in \ker_{\mathbb{Z}}(A)$ be a collection of vectors. We say that $\mathcal{B}$ is a *Markov sub-basis* for $A \in \{0,1\}^{n\times r}$ and $\mathbf{y}$, or equivalently for $\mathcal{F}_{\mathbf{y}}$, if for each pair of points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, there is a sequence of moves $\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_k} \in \mathcal{B}$ and a sequence of signs $\epsilon_1, \ldots, \epsilon_k \in \{\pm 1\}$ such that

$$\mathbf{x}_1 + \sum_{j=1}^{k} \epsilon_j \mathbf{z}_{i_j} = \mathbf{x}_2,$$

and such that for all $m = 1, \ldots, k$,

$$\mathbf{x}_1 + \sum_{j=1}^{m} \epsilon_j \mathbf{z}_{i_j} \in \mathcal{F}_{\mathbf{y}}.$$

Given a configuration matrix A and collection of moves $\mathcal{B}$, we say that $\mathcal{B}$ is a Markov basis for A if $\mathcal{B}$ is a Markov sub-basis for every fibre for A.

Definition 2.3.3 (Markov basis)**.** Let $A \in \{0,1\}^{n \times r}$ be a configuration matrix and let $\mathcal{B} \in \ker_{\mathbb{Z}}(A)$ be a collection of vectors. We say that $\mathcal{B}$ is a *Markov basis* for $A \in \{0,1\}^{n \times r}$, if for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, for each pair of points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, there is a sequence of moves $\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_k} \in \mathcal{B}$ and a sequence of signs $\epsilon_1, \ldots, \epsilon_k \in \{\pm 1\}$ such that

$$\mathbf{x}_1 + \sum_{j=1}^{k} \epsilon_j \mathbf{z}_{i_j} = \mathbf{x}_2,$$

and such that for all $m = 1, \ldots, k$,

$$\mathbf{x}_1 + \sum_{j=1}^{m} \epsilon_j \mathbf{z}_{i_j} \in \mathcal{F}_{\mathbf{y}}.$$

2.3.2 The Graver basis

The *Graver* basis is known to be a Markov basis [3]. Defining the Graver basis requires the definition of a *conformal decomposition*.

Definition 2.3.4. Let $\mathbf{u} \in \mathbb{Z}^r$. A *conformal decomposition* of $\mathbf{u}$ is an expression $\mathbf{u} = \mathbf{v} + \mathbf{w}$ where $\mathbf{v}, \mathbf{w} \in \mathbb{Z}^r \setminus \{\mathbf{0}\}$ and $|u_i| = |v_i| + |w_i|$ for all $i \in \{1, \ldots, r\}$.

Definition 2.3.5 (Graver basis [43])**.** Given a matrix $A \in \mathbb{Z}^{n \times r}$, the *Graver basis* of A, denoted $\mathcal{G}_A$, is the set of all nonzero $\mathbf{g} \in \ker_{\mathbb{Z}}(A)$ such that $\mathbf{g}$ does not have a conformal decomposition $\mathbf{g} = \mathbf{v} + \mathbf{w}$ with $\mathbf{v}$ and $\mathbf{w}$ in $\ker_{\mathbb{Z}}(A)$.

Clearly if for some A we have $\mathbf{g} \in \mathcal{G}_A$, then $-\mathbf{g} \in \mathcal{G}_A$. When enumerating Graver basis elements, we will in general only write down one of the elements $\mathbf{g}, -\mathbf{g}$.

Using the Graver basis to construct a walk in a $\mathbb{Z}$-polytope allows travel between any pair of points such that no step moves in the wrong direction in any co-ordinate [3]. Too see this, let A and $\mathbf{y}$ be given. Consider a pair of points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$. Then $\mathbf{z} = \mathbf{x}_2 - \mathbf{x}_1 \in \ker_{\mathbb{Z}}(A)$, and steps in any walk connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ must sum to $\mathbf{z}$. If $\mathbf{z} \in \mathcal{G}_A$, then we are done. If not, then $\mathbf{z}$ can be conformally decomposed into Graver basis elements. Suppose that this decomposition is given by

$$\mathbf{z} = \sum_{\mathbf{g} \in \mathcal{G}_A} c_{\mathbf{g}} \mathbf{g},$$

where each $\mathbf{c}_{\mathbf{g}} \in \mathbb{Z}_{\geq 0}$. The fact that the decomposition is conformal means that for $\mathbf{g}$ with $c_{\mathbf{g}} \neq 0$, there is no mismatch between the signs of entries in $\mathbf{g}$ and those of $\mathbf{z}$. Therefore each step in the walk travels in the correct direction in every co-ordinate.

Example 2.3.6. Let A be the link-path incidence matrix of the three-link linear network. Then A is given by

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

The Graver basis consists of the following vectors (and their negations):

$$\mathcal{G}_A = \left\{ \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

The first three elements of $\mathcal{G}_A$ make up a lattice basis U for $\ker_{\mathbb{Z}}(A)$.

Suppose the difference between two points $\mathbf{x}_1, \mathbf{x}_2$ in some fibre is given by the integer kernel element $\mathbf{x}_2 - \mathbf{x}_1 = \begin{bmatrix} -3 & -2 & 0 & 2 & -1 & 1 \end{bmatrix}^\intercal$. Writing $\mathbf{x}_2 - \mathbf{x}_1$ in terms of U we have

$$\mathbf{x}_2 - \mathbf{x}_1 = \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}.$$

For a lattice basis such as U, every element of the integer kernel can be written uniquely as an integer combination of elements of U, so these moves are necessary when using U to walk from $\mathbf{x}_1$ to $\mathbf{x}_2$. The second move $\begin{bmatrix} 0 & -1 & 0 & 1 & 1 & -1 \end{bmatrix}^\intercal$ travels in the wrong direction in the fifth co-ordinate.

The vector $\mathbf{x}_2 - \mathbf{x}_1$ can be conformally decomposed into Graver basis elements like so:

$$\mathbf{x}_2 - \mathbf{x}_1 = \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

These moves can be used to construct a walk from $\mathbf{x}_1$ to $\mathbf{x}_2$ such that the walk never travels in the wrong direction in any co-ordinate.

2.3.3 Gröbner bases

Gröbner bases are the current state of the art in Markov bases. The Fundamental Theorem of Markov Bases, discussed in Section 2.4, relates the integer kernel of a configuration matrix to an ideal in a polynomial ring, and Markov bases to generating sets of this ideal. Gröbner bases are generating sets that have particular properties.

Polynomial rings are a very well studied field. The correspondence established by the Fundamental Theorem allows the importation of much of the machinery of abstract algebra to the study of Markov bases. In particular, algorithms are available for finding Gröbner bases which can then be employed in the search for Markov bases.

Gröbner bases are covered more fully in Section 2.4.3, but briefly, Gröbner bases are defined relative to a term order $>$ and a choice of ordering of the co-ordinates, which allows one to order the points in the polytope. Different term orders and different orderings of the co-ordinates will produce different Gröbner bases. A geometric interpretation of Gröbner bases is that for any point $\mathbf{x}_1$ in the polytope that is not $>$-minimal, there exists a point $\mathbf{x}_2$ such that $\mathbf{x}_1 > \mathbf{x}_2$ and it is possible to move from $\mathbf{x}_1$ to $\mathbf{x}_2$.

Example 2.3.7. Consider the link-path incidence matrix of the three-link linear network, given by

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

A Markov basis for A based on a Gröbner basis for the ideal generated by $\ker_{\mathbb{Z}}(A)$ is given by

$$G = \left\{ \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

This basis uses lexicographic ordering with the term ordering $t_6 > t_5 > \cdots > t_1$.

Let $\mathbf{y} = \begin{bmatrix} 3 & 3 & 3 \end{bmatrix}^\mathsf{T}$, and consider $\mathbf{x}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \in \mathcal{F}_{\mathbf{y}}$. The second move $\mathbf{u}_2$ and third move $\mathbf{u}_3$ in G can each be subtracted from $\mathbf{x}_1$ to form

$$\mathbf{x}_2 = \mathbf{x}_1 - \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \mathbf{x}_1 - \mathbf{u}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Both of these vectors $\mathbf{x}_2, \mathbf{x}_3$ are such that $\mathbf{x}_1 > \mathbf{x}_2, \mathbf{x}_3$ under the given term ordering.

Using a different term ordering produces a different basis: for example, using the term ordering $t_1 > t_2 > \cdots > t_6$ produces the basis

$$G_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} \right\}.$$

2.3.4 Lattice bases

Lattice bases were introduced in Section 1.4 and defined in Definition 1.4.1. A lattice basis $\mathcal{B}$ for a configuration matrix A is a collection of vectors such that any element of $\ker_{\mathbb{Z}}(A)$ can be uniquely expressed as a sum of integer multiples of elements of $\mathcal{B}$. Although it is a basis for the integer kernel, a lattice basis does not necessarily contain only integer valued vectors.

Example 2.3.8. Consider again the triangular network from Example 1.2.2. A link-path incidence matrix for this network is given by

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

We have $\dim(\ker(A)) = 3$, so any lattice basis has three elements. One lattice basis is given by

$$\mathcal{B} = \left\{ \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}.$$

If we were to multiply out the denominators, we would have the set

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -2 \end{bmatrix} \right\}.$$

The integer kernel contains the element

$$\mathbf{z} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 0 \end{bmatrix},$$

which cannot be expressed as an integer combination of elements of $\mathcal{B}_2$. This means that $\mathcal{B}_2$ is not a lattice basis. However, $\mathbf{z}$ can be expressed as an integer combination of elements of the lattice basis $\mathcal{B}$:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

.

A lattice basis are not necessarily a Markov basis, but every Markov basis must include a lattice basis. The following list of inclusions is from Aoki, Hara, and Takemura [3]:

a lattice basis $\subseteq$ a minimal Markov basis $\subseteq$ a reduced Gröbner basis $\subseteq$ the Graver basis.

The connectivity properties of lattice bases are the focus of this thesis.

2.4 Algebra and connectivity

In this section we introduce the Fundamental Theorem of Markov Bases [19]. The Fundamental Theorem of Markov Bases gives a correspondence between a Markov basis and a generating set of an ideal in a polynomial ring. It is useful for determining the connectedness of any set $\mathcal{F}_{\mathbf{y}}$ via any particular basis, and together with Buchberger's algorithm [10, 13] can be used to find Markov bases [19, 20] for a given configuration matrix.

We begin in Section 2.4.1 by reviewing some of the relevant algebra. This includes the definitions of polynomial rings, ideals, and generating sets.

In Section 2.4.2 we show how polynomials may represent vectors such as elements of the kernel of a matrix, or a fibre, before giving a statement of the Fundamental Theorem (Theorem 2.4.5). We show via examples how it may be used to demonstrate connectedness.

In Section 2.4.3 we take a more detailed look at Gröbner bases, and Theorem 2.4.15 gives Buchberger's algorithm for finding a Gröbner basis. We then demonstrate how they may be used to find a Markov basis for a given configuration matrix.

2.4.1 Ideals

Recall the definitions of a *polynomial ring*, an *ideal* in a polynomial ring, and its *generators*.

Definition 2.4.1 (Polynomial ring [13])**.** Let k be a field, and let $T = \{t_1, \ldots, t_r\}$ be a set of indeterminates. Then the *polynomial ring* $k[T]$ is the set of polynomials in the indeterminates T with co-efficients in the field k with respect to the binary operations $+$ (polynomial addition) and $\times$ (polynomial multiplication).

Example 2.4.2. The polynomial ring $\mathbb{Q}[t_1, t_2]$ contains polynomials in the indeterminates t_1 and t_2 with rational co-efficients. It contains polynomials such as

$$2t_1^2t_2 + t_1t_2 + 4t_2^3 \quad \text{and} \quad \frac{2}{3}t_1 + t_2^2.$$

They are added and multiplied using the usual operations of polynomial addition and multiplication.

Each of the terms in a given polynomial is a *monomial*. Give a non-negative integer vector $\mathbf{x}$, we will use the notation $T^{\mathbf{x}}$ to mean the monomial given by the elementwise exponentiation $t_1^{x_1}t_2^{x_2}\ldots t_r^{x_r}$: for example, if $T = \{t_1, t_2, t_3\}$ and $\mathbf{x} = \begin{bmatrix}1 & 0 & 2\end{bmatrix}^\intercal$, then $T^{\mathbf{x}} = t_1t_3^2$.

Definition 2.4.3 (Ideal [13])**.** Let k be a field, let $T = \{t_1, \ldots, t_r\}$ be a set of indeterminates, and let $k[T]$ be the polynomial ring over k in the indeterminates T. An *ideal* of $k[T]$ is a subset I of $k[T]$ such that:

1. $0 \in I$.
2. $\forall f, g \in I : f + g \in I$.
3. $\forall f \in I, h \in k[T] : hf \in I$.

Definition 2.4.4 (Generators [13])**.** Let $k[T]$ be a polynomial ring, and let I be an ideal of $k[T]$. We say a set $\{f_1, \ldots, f_s\} \subset I$ *generates* I if for all $f \in I$, we can write:

$$f = \sum_{i=1}^{s} h_i f_i \quad \text{where each} \quad h_i \in k[T].$$

Then we write $I = \langle f_1, \ldots, f_s\rangle$.

2.4.2 The Fundamental Theorem of Markov Bases

The Fundamental Theorem of Markov Bases uses a correspondence between Markov bases of configuration matrices and generating sets of an associated ideal in a certain polynomial ring. We use the set of indeterminates $T = \{t_1, \ldots, t_r\}$ and identify a vector $\mathbf{x} \in \mathbb{Z}_{\geq 0}^r$ with the monomial formed by taking the elementwise exponentiation: $T^{\mathbf{x}} = t_1^{x_1} \cdots t_r^{x_r}$. Then any element $\mathbf{x}$ of $\mathcal{F}_{\mathbf{y}}$ can be identified with the monomial $T^{\mathbf{x}}$. With each configuration matrix A we associate an ideal in $k[T]$, which we write as I_A, and define as

$$I_A = \langle T^{\mathbf{x}_1} - T^{\mathbf{x}_2} : A\mathbf{x}_1 = A\mathbf{x}_2 \rangle,$$

where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{Z}^r_{\geq 0}$ (and so $\mathbf{x}_1 - \mathbf{x}_2 \in \ker_{\mathbb{Z}}(A)$).

Any element $\mathbf{u}$ of a basis $\mathcal{B}$ can be identified with a monomial difference: we split $\mathbf{u}$ into its positive and negative parts, $\mathbf{u}^+$ and $\mathbf{u}^-$, so that $\mathbf{u} = \mathbf{u}^+ - \mathbf{u}^-$, and both $\mathbf{u}^+$ and $\mathbf{u}^-$ are in $\mathbb{Z}^r_{\geq 0}$. The monomial difference representing $\mathbf{u}$ is therefore $T^{\mathbf{u}+} - T^{\mathbf{u}^-}$. With each basis $\mathcal{B}$ we associate an ideal in $k[T]$, which we write as $I_{\mathcal{B}}$, and define as

$$I_{\mathcal{B}} = \langle T^{\mathbf{u}^+} - T^{\mathbf{u}^-} : \mathbf{u} \in \mathcal{B} \rangle.$$

Given $\mathbf{y}$ and $\mathcal{B}$, the Fundamental Theorem of Markov Bases says that two elements $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$ are connected by $\mathcal{B}$ if and only if their representation as a monomial difference is in the ideal $I_{\mathcal{B}}$ generated by the monomial difference representations of the elements of $\mathcal{B}$; that is

$$T^{\mathbf{x}_1} - T^{\mathbf{x}_2} \in I_{\mathcal{B}}.$$

So if for each $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, there exist $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathcal{B}$ and $f_1, \ldots, f_k \in k[T]$, such that

$$T^{\mathbf{x}_1} - T^{\mathbf{x}_2} = \sum_{i=1}^{k} f_i (T^{\mathbf{u}_i^+} - T^{\mathbf{u}_i^-}),$$

then $\mathcal{B}$ is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$. If $\mathcal{B}$ is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$ for all allowed $\mathbf{y}$ for some configuration matrix A, then $\mathcal{B}$ is a Markov basis for A.

Theorem 2.4.5 (Fundamental Theorem of Markov Bases [19, 20])**.** *A finite set of moves $\mathcal{B}$ is a Markov basis for A if and only if the set of monomial differences $\{T^{\mathbf{u}^+} - T^{\mathbf{u}^-} : \mathbf{u} \in \mathcal{B}\}$ generates the ideal I_A.*

A path connecting two points $\mathbf{x}$ and $\mathbf{x}_n$ in $\mathcal{F}_{\mathbf{y}}$ is given by a telescoping series that evaluates to $T^{\mathbf{x}_1} - T^{\mathbf{x}_2}$. Following Dinwoodie [20], suppose that there is a path from $\mathbf{x}$ to $\mathbf{x}_n$ using moves in $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. The path can be written

$$\mathbf{x}, \;\; \mathbf{x}_1 = \mathbf{x} - \epsilon_1 \mathbf{u}_{i_1}, \;\; \mathbf{x}_2 = \mathbf{x} - \epsilon_1 \mathbf{u}_{i_1} - \epsilon_2 \mathbf{u}_{i_2}, \;\; \ldots, \;\; \mathbf{x}_n$$

where ϵ_i represents the sign $+$ or $-$ and $\mathbf{x}_i \in \mathcal{F}_{\mathbf{y}}$ for each $1 \leq i \leq n$. With the polynomial notation, we can write

$$\begin{aligned} T^{\mathbf{x}} - T^{\mathbf{x}_n} &= (T^{\mathbf{x}} - T^{\mathbf{x}_1}) + (T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) + \cdots + (T^{\mathbf{x}_{n-1}} - T^{\mathbf{x}_n}) \\ &= \epsilon_1 T^{\mathbf{x} - \mathbf{u}_{i_1}^{\epsilon_1}} (T^{\mathbf{u}_{i_1}^+} - T^{\mathbf{u}_{i_1}^-}) + \cdots + \epsilon_n T^{\mathbf{x}_{n-1} - \mathbf{u}_{i_n}^{\epsilon_n}} (T^{\mathbf{u}_{i_n}^+} - T^{\mathbf{u}_{i_n}^-}) \end{aligned}$$

where $\mathbf{u}^{\epsilon_i}$ means $\mathbf{u}^+$ if $\epsilon_i = +$ and $\mathbf{u}^-$ if $\epsilon_i = -$. This shows that the monomial difference $T^{\mathbf{x}_1} - T^{\mathbf{x}_n}$ is in $I_{\mathcal{B}}$.

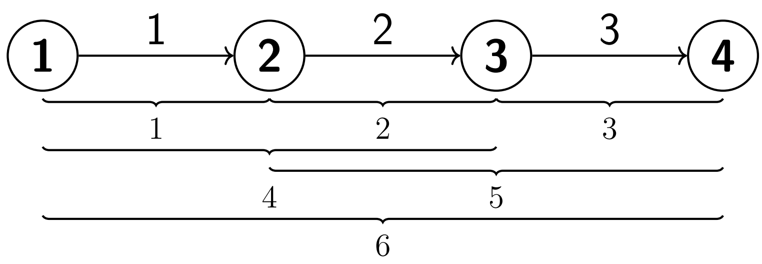


Figure 2.3: The three-link linear network from Example 2.4.6. The underbraces show allowed paths.

Example 2.4.6. Consider the three-link linear traffic network shown in Figure 2.3
The link-path incidence matrix A is given by

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

A lattice basis $\mathcal{B}$ is given by

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\} \subset \ker_{\mathbb{Z}}(A).$$

If one car is observed on each link of the network, then $\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\intercal}$, producing the fibre

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Then we can see that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$$

Indexing the elements $\mathbf{u}_i \in \mathcal{B}$ and $\mathbf{x}_i \in \mathcal{F}_{\mathbf{y}}$ from the left gives

$$\begin{aligned}\mathbf{x}_0 &= \mathbf{x}_1 + \mathbf{u}_2\\ &= \mathbf{x}_2 + \mathbf{u}_1\\ &= \mathbf{x}_3 + \mathbf{u}_3.\end{aligned}$$

Because each element of $\mathcal{F}_{\mathbf{y}}$ is connected to $\mathbf{x}_0$, they must all be connected to each other, and so $\mathcal{F}_{\mathbf{y}}$ is connected by $\mathcal{B}$. In algebraic terms, $\mathcal{B}$ is represented with the monomial differences $t_1t_2 - t_4$, $t_2t_3 - t_5$, and $t_1t_2t_3 - t_6$; and the elements of $\mathcal{F}_{\mathbf{y}}$ by the monomials $t_1t_2t_3$, t_1t_5, t_3t_4, and t_6. We can show that $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ are connected to $\mathbf{x}_0$ by writing the monomial difference representing the two elements as a sum of the monomial differences representing the basis elements multiplied by some element of the polynomial ring:

$$\begin{aligned}T^{\mathbf{x}_0} - T^{\mathbf{x}_1} &= t_1t_2t_3 - t_1t_5 = t_1(t_2t_3 - t_5) = t_1(T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-}),\\ T^{\mathbf{x}_0} - T^{\mathbf{x}_2} &= t_1t_2t_3 - t_3t_4 = t_3(t_1t_2 - t_4) = t_3(T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}),\\ T^{\mathbf{x}_0} - T^{\mathbf{x}_3} &= t_1t_2t_3 - t_6 = t_1t_2t_3 - t_6 = T^{\mathbf{u}_3^+} - T^{\mathbf{u}_3^-}.\end{aligned}$$

We can similarly show that these other elements $T^{\mathbf{x}_1}, T^{\mathbf{x}_2}$, and $T^{\mathbf{x}_3}$ are connected to each other:

$$\begin{aligned}T^{\mathbf{x}_2} - T^{\mathbf{x}_1} &= t_3t_4 - t_1t_5\\ &= -t_3(t_1t_2 - t_4) + t_1(t_2t_3 - t_5)\\ &= -t_3(T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}) + t_1(T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-}),\\ T^{\mathbf{x}_2} - T^{\mathbf{x}_3} &= t_3t_4 - t_6\\ &= -t_3(t_1t_2 - t_4) + (t_1t_2t_3 - t_6)\\ &= -t_3(T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}) + (T^{\mathbf{u}_3^+} - T^{\mathbf{u}_3^-}),\\ T^{\mathbf{x}_1} - T^{\mathbf{x}_3} &= t_1t_5 - t_6\\ &= -t_1(t_2t_3 - t_5) + (t_1t_2t_3 - t_6)\\ &= -t_1(T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-}) + (T^{\mathbf{u}_3^+} - T^{\mathbf{u}_3^-}).\end{aligned}$$

This shows that the monomial differences representing any two elements of $\mathcal{F}_{\mathbf{y}}$ are in the ideal generated by the monomial differences representing the basis elements, and so $\mathcal{F}_{\mathbf{y}}$ is connected by $\mathcal{B}$ by the Fundamental Theorem of Markov Bases. Therefore $\mathcal{B}$ is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$.

Note that this does not necessarily mean that $\mathcal{B}$ is a Markov basis, though. For $\mathcal{B}$ to be a Markov basis, it must connect $\mathcal{F}_{\mathbf{y}}$ for all allowed $\mathbf{y}$.

2.4.3 Gröbner bases and Buchberger's algorithm

The correspondence between Markov bases and generating sets of ideals given by the Fundamental Theorem of Markov Bases means that ideas from abstract algebra can be imported and used in order to understand and find Markov bases. Hilbert's Basis Theorem guarantees that a basis of finite size exists [13].

Theorem 2.4.7 (Hilbert's Basis Theorem)**.** *Let k be a field, let $T = \{t_1, \dots, t_r\}$ be a set of indeterminates, and let $k[T]$ be the polynomial ring over k in the indeterminates T. Every ideal $I \subseteq k[T]$ has a finite generating set. In other words, $I = \langle g_1, \dots, g_t \rangle$ for some $g_1, \dots, g_t \in I$.*

This means that a finite basis for the ideal of monomial differences I_A exists, and therefore that a finite Markov basis exists for any configuration matrix.

Such a finite basis can be found using a process called *Buchberger's algorithm*, which produces a particularly useful kind of basis called a *Gröbner basis.* The definition of a Gröbner basis is in terms of a monomial ordering, which is defined as follows:

Definition 2.4.8 (Monomial ordering [13])**.** Let k be a field, let $T = \{t_1, \dots, t_r\}$ be a set of indeterminates, and let $k[T]$ be the polynomial ring over k in indeterminates T. A *monomial ordering* $>$ on $k[T]$ is a relation $>$ on $\mathbb{Z}^r_{\geq 0}$, or equivalently, a relation on the set of monomials $T^{\mathbf{x}}$, $\mathbf{x} \in \mathbb{Z}^r_{\geq 0}$, satisfying:

1. $>$ is a total (or linear) ordering on $\mathbb{Z}^r_{\geq 0}$, meaning it can be used to compare every pair of elements in $\mathbb{Z}^r_{\geq 0}$.

2. If $\mathbf{x}_1 > \mathbf{x}_2$ and $\mathbf{x}_3 \in \mathbb{Z}^r_{\geq 0}$, then $\mathbf{x}_1 + \mathbf{x}_3 > \mathbf{x}_2 + \mathbf{x}_3$.

3. $>$ is a well-ordering on $\mathbb{Z}^r_{\geq 0}$. This means that every nonempty subset of $\mathbb{Z}^n_{\geq 0}$ has a smallest element under $>$. In other words, if $A \subseteq \mathbb{Z}^n_{\geq 0}$ is nonempty, then there is $\alpha \in A$ such that $\beta > \alpha$ for every $\beta \neq \alpha$ in A.

An example of a monomial ordering is *lexicographic order.* In lexicographic order, monomials are compared by comparing the values of each entry in the term in some order. Take as an example the elements of $\mathcal{F}_{\mathbf{y}}$ from Example 2.4.6. These are

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We can order the elements of $\mathcal{F}_{\mathbf{y}}$ with the term ordering $t_1 > t_2 > t_3 > t_4 > t_5 > t_6$, which gives the order

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} > \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Under $>$, the greatest term is t_1, and we can see that under this term ordering every term with $x_1 = 1$ precedes every term with $x_1 = 0$. In monomial terms, this ordering is

$$T^{\mathbf{x}_0} > T^{\mathbf{x}_1} > T^{\mathbf{x}_2} > T^{\mathbf{x}_3}.$$

In any polynomial, we can order the terms to find the *leading term*.

Definition 2.4.9 (Leading term [13])**.** Let k be a field, let $T = \{t_1, \ldots, t_r\}$ be a set of indeterminates, and let $k[T]$ be the polynomial ring over k in the indeterminates T. In any polynomial $f \in k[T]$, the leading term $\mathrm{LT}(f)$ with respect to some ordering $>$ is the term that is maximal with respect to $>$ out of all terms in f.

Applying this definition to elements of $\mathcal{B}$ from Example 2.4.6, using again the lexicographic order with term ordering $t_1 > t_2 > t_3 > t_4 > t_5 > t_6$ means that

$$\begin{aligned}\mathrm{LT}(T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}) &= \mathrm{LT}(t_1 t_2 - t_4)\\ &= t_1 t_2,\\ \mathrm{LT}(T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-}) &= \mathrm{LT}(t_2 t_3 - t_5)\\ &= t_2 t_3.\end{aligned}$$

We can now define a Gröbner basis.

Definition 2.4.10 (Gröbner basis [13])**.** A *Gröbner basis* with respect to a term ordering $>$ is a subset $G = \{g_1, \ldots, g_n\}$ of an ideal I such that

$$\langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_n)\rangle = \langle \mathrm{LT}(I)\rangle,$$

where $\mathrm{LT}(g)$ means the leading term of the polynomial under $>$.

Dinwoodie [20] and Diaconis and Sturmfels [19] give a technique for finding a Markov basis. They use another set of indeterminates $S = \{s_1, \ldots, s_n\}$ and identify the vector $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ with the monomial formed by taking the elementwise exponentiation $S^{\mathbf{y}} = s_1^{y_1} \cdots s_n^{y_n}$.

Let $>$ be any term ordering on $T \cup S$ such that for all $s \in S$ and all $t \in T$, $s > t$. Then a Gröbner basis for I_A can be found by finding a Gröbner basis for the ideal

$$I = \langle T^{\mathbf{e}_i} - S^{A\mathbf{e}_i} : i = 1, \ldots, r\rangle,$$

and taking from this Gröbner basis only those elements for which all terms are in $k[T]$.

This basis may be calculated using Buchberger's Algorithm, given for example in Cox et al. [13], which requires the following definition.

Definition 2.4.11 (S-polynomial [13])**.** The *S-polynomial* of polynomials f and g in $k[T]$ is given by

$$S(f, g) = \frac{T^\gamma}{\mathrm{LT}(f)} f - \frac{T^\gamma}{\mathrm{LT}(g)} g$$

where T^γ is the least common multiple of $\mathrm{LT}(f)$ and $\mathrm{LT}(g)$.

The practical effect of taking an S-polynomial is to find a multiple of each of f and g such that the leading terms cancel when the difference $f - g$ is taken.

For example, consider the lattice basis from Example 2.4.6.

Example 2.4.12. Two of the lattice basis elements given for the three-link linear traffic network in Example 2.4.6 have monomial difference representations

$$\begin{aligned} T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-} &= t_1 t_2 - t_4, \\ T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-} &= t_2 t_3 - t_5. \end{aligned}$$

Under the term ordering $t_1 > \cdots > t_6$, the S-polynomial $S(T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}, T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-})$ is given by

$$\begin{aligned} S(T^{\mathbf{u}_2^+} - T^{\mathbf{u}_2^-}, T^{\mathbf{u}_1^+} - T^{\mathbf{u}_1^-}) &= t_1(t_2 t_3 - t_5) - t_3(t_1 t_2 - t_4) \\ &= t_1 t_2 t_3 - t_1 t_5 - t_1 t_2 t_3 + t_3 t_4 \\ &= t_3 t_4 - t_1 t_5. \end{aligned}$$

Buchberger's algorithm requires the division algorithm, which requires the definition of the *multidegree* of a polynomial. The multidegree is defined as follows:

Definition 2.4.13 (Multidegree [13])**.** Let $f = \sum_{\mathbf{x}} a_{\mathbf{x}} T^{\mathbf{x}}$ be a non-zero polynomial in $k[T]$ and let $>$ be a monomial order. The *multidegree* of f is

$$\text{multidegree}(f) = \max\{\mathbf{x} \in \mathbb{Z}^n_{\geq 0} : a_{\mathbf{x}} \neq 0\},$$

where the maximum is taken with respect to the term ordering.

We now give the division algorithm:

Theorem 2.4.14 (The Division Algorithm in $k[T]$ [13])**.** *Let $>$ be a monomial order on $\mathbb{Z}^r_{\geq 0}$, and let $F = \{f_1, \dots, f_s\}$ be a collection of polynomials in $k[T]$. Then every $f \in k[T]$ can be written as*

$$f = q_1 f_1 + \cdots + q_s f_s + r,$$

where $q_i, r \in k[T]$ and either $r = 0$ or r is a linear combination, with coefficients in k, of monomials, none of which is divisible by any of $\mathrm{LT}(f_1), \dots, \mathrm{LT}(f_s)$*. We call r a* remainder *of f on division by F. Furthermore, if $q_i f_i \neq 0$, then*

$$\text{multidegree}(f) \geq \text{multidegree}(q_i f_i).$$

We can now give Buchberger's algorithm.

Theorem 2.4.15 (Buchberger's Algorithm [13])**.** *Let $I = \langle f_1, \dots, f_s \rangle \neq \{0\}$ be a polynomial ideal. Then a Gröbner basis for I can be constructed from the set of generators $F = \{f_1, \dots, f_s\}$ in a finite number of steps by the following algorithm:*

1. *Set $G = F$.*

2. *For each pair* $\{p, q\} \in G$*, find a remainder* r *of* $S(p, q)$ *when divided by the elements of* G*. If* $r \neq 0$*, add* r *to* G*.*
3. *Repeat the previous step until* $r = 0$ *for all pairs* $\{p, q\}$ *in* G*.*
4. *Return* G*, a Gröbner basis of* I*.*

Example 2.4.16. Consider the link-path incidence matrix A of a three-link linear traffic network, where the allowed paths are each single edge, and each pair of adjacent edges. Then A is given by

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

Using indeterminates T to represent the $\mathbf{x}$ vectors and S to represent the $\mathbf{y}$ vectors, the ideal I as defined by its generators is

$$I = \langle s_1 - t_1, s_2 - t_2, s_3 - t_3, s_1 s_2 - t_4, s_2 s_3 - t_5 \rangle.$$

We use lexicographical ordering with $s_1 > s_2 > s_3 > t_1 > t_2 > t_3 > t_4 > t_5$. There are ten pairs of distinct generators, which means at least ten S-polynomials to check.

The remainders of five of these S-polynomials when divided by the elements of G are 0. For example,

$$\begin{aligned} S(s_1 - t_1, s_2 - t_2) &= s_2(s_1 - t_1) - s_1(s_2 - t_2) \\ &= s_2 s_1 - s_2 t_1 - s_1 s_2 + s_1 t_2 \\ &= s_1 t_2 - s_2 t_1. \end{aligned}$$

Dividing $S(s_1 - t_1, s_2 - t_2)$ by $s_1 - t_1$ yields

$$s_1 t_2 - s_2 t_1 = t_2(s_1 - t_1) - (s_2 t_1 - t_1 t_2),$$

and dividing the remainder $-(s_2 t_1 - t_1 t_2)$ by $s_2 - t_2$ yields $-t_1$ with no remainder.

The remainders of the other five S-polynomials when divided by the elements of G are all in $\{\pm(t_1 t_2 - t_4), \pm(t_2 t_3 - t_5), \pm(t_1 t_5 - t_3 t_4)\}$. For example,

$$\begin{aligned} S(s_2 - t_2, s_1 s_2 - t_4) &= s_1(s_2 - t_2) - (s_1 s_2 - t_4) \\ &= s_1 s_2 - s_1 t_2 - s_1 s_2 + t_4 \\ &= -(s_1 t_2 - t_4). \end{aligned}$$

Dividing $S(s_2 - t_2, s_1 s_2 - t_4)$ by $s_1 - t_1$ yields

$$-(s_1 t_2 - t_4) = -t_2(s_1 - t_1) - (t_1 t_2 - t_4).$$

This remainder $t_1 t_2 - t_4$ is not divisible by any element of G and so is added to G; so too are $t_2 t_3 - t_5$ and $t_1 t_5 - t_3 t_4$, the remainders of the divisions of other S-polynomials by the elements of G. With three new elements in G there are eighteen new S-polynomials to consider. Fortunately, all have a remainder of 0 when divided by elements of the updated set G. Hence a Gröbner basis for I is given by

$$G = \{s_1 - t_1, s_2 - t_2, s_3 - t_3, s_1s_2 - t_4, s_2s_3 - t_5, t_1t_2 - t_4, t_2t_3 - t_5, t_1t_5 - t_3t_4\}.$$

Taking only the elements with all terms in T, we have

$$\{t_1t_2 - t_4, t_2t_3 - t_5, t_1t_5 - t_3t_4\},$$

which is a Gröbner basis for I_A and hence the vectors

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

are a Markov basis for A.

This example illustrates how the algorithm works to produce a Markov basis. Although there exist improvements and optimisations for Buchberger's algorithm that are applicable to the ideals of interest (for example, in Hemmecke and Malkin [28]), they are based on Buchberger's algorithm. This example also serves to demonstrate how difficult Markov bases are to compute algebraically. Finding a Markov basis for this extremely simple 3×5 configuration matrix required polynomial long division of 25 S-polynomials by up to 8 basis elements.

2.5 Column partition lattice bases

In this section we introduce a type of lattice basis we will call a *column partition lattice basis*, with the intention of using them in $\mathbb{Z}$-polytope sampling. They will be covered in more detail in Chapter 3. Column partition lattice bases enjoy the advantage of being relatively computationally cheap to find compared to Gröbner bases as they require only a matrix inversion and a matrix multiplication. They also have a simple geometric interpretation: they are moves in co-ordinate directions when the $\mathbb{Z}$-polytope is projected onto a certain group of co-ordinates.

A column partition lattice basis is suitable for MCMC $\mathbb{Z}$-polytope sampling if it is a Markov basis or Markov subbasis. A particular column partition lattice basis may or may not be a Markov basis: Section 2.5.2 gives a method for finding a column partition lattice basis that is a Markov basis for a very specific set of configuration matrices, while Section 2.5.1 gives some examples that are not Markov bases. The issue of connectedness of column partition lattice bases is the subject of Chapter 4.

Using a column partition lattice basis as a collection of moves for MCMC $\mathbb{Z}$-polytope sampling is the approach favoured by Tebaldi and West [45] and Hazelton [25, 26].

Column partition lattice bases are a type of lattice basis. Lattice bases are defined in Definition 1.4.1, which says that a set $\mathcal{B}$ forms a lattice basis of $\ker_{\mathbb{Z}}(A)$ if every $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$ is uniquely written as an integer combination of elements of $\mathcal{B}$. The elements of any lattice basis $\mathcal{B}$ span the kernel of the configuration matrix. If the configuration matrix A is $n \times r$ and of full rank, then the kernel of A is of dimension $r - n$ and any lattice basis contains $r - n$ elements.

Following the work of Tebaldi and West [45], we can construct a column partition lattice basis with the following method: we partition the columns of the configuration matrix A into A_1 and A_2, where A_1 is an invertible maximal $n \times n$ submatrix and A_2 is $n \times (r - n)$. We correspondingly partition $\mathbf{x}$ into $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathbf{x}_2$ gives a set of co-ordinates of the solution $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$ which determine the value of $\mathbf{x}_1$: we have

$$\begin{aligned} A\mathbf{x} &= \mathbf{y} \\ A_1\mathbf{x}_1 + A_2\mathbf{x}_2 &= \mathbf{y} \\ A_1\mathbf{x}_1 &= \mathbf{y} - A_2\mathbf{x}_2 \\ \mathbf{x}_1 &= A_1^{-1}\mathbf{y} - A_1^{-1}A_2\mathbf{x}_2. \end{aligned}$$

The columns of the matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I_{r-n} \end{bmatrix}$$

give a lattice basis for the kernel of A and for the set $\mathcal{F}_{\mathbf{y}}$. This process is described more fully with examples in Section 3.2.

2.5.1 Connectivity problems

Having defined column partition lattice bases, we now turn to the question of whether or not they are Markov bases. It turns out that this is not always the case. There are a few kinds of problem that might occur. Some examples are presented here: we refer to them as parity errors, isolated spaces, and reduced dimension.

Parity errors

From Definition 1.4.1, a lattice basis is a collection of kernel elements such that every integer kernel element can be uniquely expressed as an integer combination of its elements. This does not require that the elements of the lattice basis themselves are integral. Column partition lattice bases are not guaranteed to contain only integer elements. When they are not, a walk through a $\mathbb{Z}$-polytope constructed with such a basis must use as steps multiples of the basis elements. Suppose a walk on some fibre $\mathcal{F}$ is at some point $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$, so $\mathbf{x}$ must be integral. A proposed next step $\mathbf{x}^{\dagger} = \mathbf{x} + \mathbf{z}$ must also be integral: this can only occur when the proposed move $\mathbf{z}$ is also integral.

If a column partition lattice basis is not integral, it may be the case that some stepping stone required to connect two points is missing — although in this projection the stepping stone has integer value co-ordinates, it is not an element of $\mathcal{F}_{\mathbf{y}}$ because of a non-integer

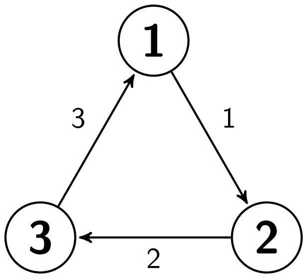


Figure 2.4: The triangular network from Examples 2.5.1 and 2.5.2.

in one of the hidden co-ordinates. We refer to problems of this type as *parity errors*. We illustrate this with an example from Hazelton and Bilton [26].

Example 2.5.1. Consider the traffic network consisting of three directed edges in a triangle, as shown in Figure 2.4. The allowed paths are: the path consisting of only the first link; the path consisting of only the second link; and the paths consisting of each pair of edges. The link-path incidence matrix for this network is given by

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

Suppose four cars are observed on each of the three links. Then the link count vector $\mathbf{y}$ is given by $\mathbf{y} = \begin{bmatrix} 4 & 4 & 4 \end{bmatrix}^\intercal$, and the set $\mathcal{F}_{\mathbf{y}}$ is given by

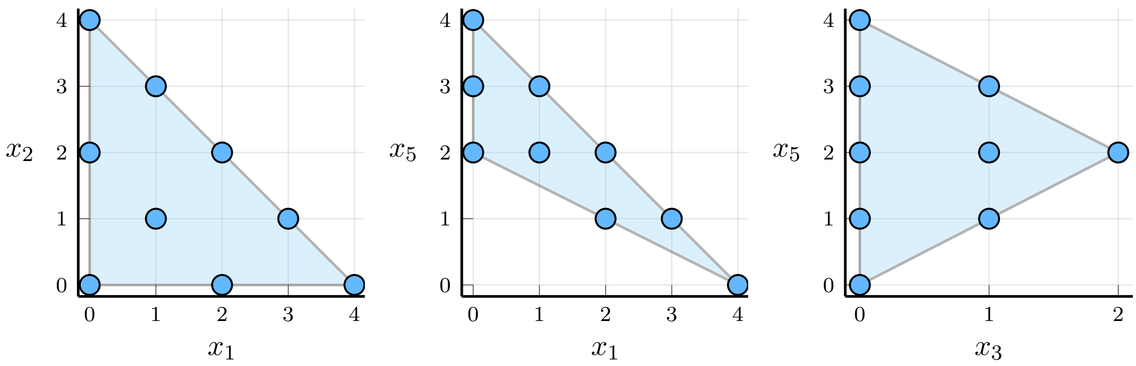
$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 4 \\ 0 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \\ 2 \\ 2 \end{bmatrix} \right\}.$$

The column partition lattice basis for the partition $\pi_1 = (\{3, 4, 5\}, \{1, 2\})$ is given by the matrix

$$U^{\pi_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

If this basis is used to construct a walk through $\mathcal{F}_{\mathbf{y}}$, the steps must be even integer multiples of the moves. In effect, this basis contains the moves

$$U_{\mathbb{Z}}^{\pi_1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}.$$



(a) Parity errors. (b) An isolated vertex. (c) A connected projection.

Figure 2.5: Three projections of the $\mathbb{Z}$-polytope in Examples 2.5.1 and 2.5.2.

This basis provides moves in co-ordinate directions when the $\mathbb{Z}$-polytope representation of $\mathcal{F}_\mathbf{y}$ is projected onto the x_1 and x_2 co-ordinates, as shown in Figure 2.5a. We can see that the elements of $\mathcal{F}_\mathbf{y}$ are divided into two connected cliques that are not connected to each other.

Isolated elements

A column partition lattice basis provides a limited collection of moves. It is possible that for some fibre $\mathcal{F}_\mathbf{y}$ and some column partition lattice basis U, there is some point $\mathbf{x} \in \mathcal{F}_\mathbf{y}$ such that no other elements of $\mathcal{F}_\mathbf{y}$ lie in any of the directions of the moves provided by U. A walk in $\mathcal{F}_\mathbf{y}$ using U either cannot access $\mathbf{x}$, or is permanently stuck at $\mathbf{x}$. We illustrate this with an example.

Example 2.5.2. Consider again the traffic network on the graph in Figure 2.4 from Example 2.5.1. A different column partition lattice basis using the partition $\pi_2 = (\{2,3,4\},\{1,5\})$ is given by

$$U^{\pi_2} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ -1 & -1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}.$$

This basis provides moves in co-ordinate directions when the $\mathbb{Z}$-polytope representation of $\mathcal{F}_\mathbf{y}$ is projected onto the x_1 and x_5 co-ordinates, as shown in Figure 2.5b. Consider the point $\mathbf{x} = \begin{bmatrix} 4 & 0 & 0 & 4 & 0 \end{bmatrix}^\intercal \in \mathcal{F}_\mathbf{y}$. In Figure 2.5b this is the point at $\begin{bmatrix} 4 & 0 \end{bmatrix}^\intercal$. This element $\mathbf{x}$ cannot have any integer multiple of a column of U^{π_2} added or subtracted from it without returning a vector with a negative entry, so using this lattice basis $\mathbf{x}$ is an isolated vertex.

The situations in Examples 2.5.1 and 2.5.2 can be rescued by selecting the column partition lattice basis given by

$$U^{\pi_3} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix},$$

which provides co-ordinate direction moves when the $\mathbb{Z}$-polytope is projected onto the x_3 and x_5 co-ordinates. The result is shown in Figure 2.5c; these moves are sufficient to connect all points in $\mathcal{F}_{\mathbf{y}}$. This column partition lattice basis is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$.

The isolated part of a $\mathbb{Z}$-polytope need not be only one vertex. The following example shows that larger isolated cliques are also possible.

Example 2.5.3. Consider the configuration matrix

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

This matrix has a repeated column, which may occur under some models used in capture-recapture studies in ecology. A column partition lattice basis for A is given by

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

This basis provides moves in co-ordinate directions when the $\mathbb{Z}$-polytope is projected onto the x_2, x_6, and x_7 dimensions. Setting $\mathbf{y} = \begin{bmatrix} 4 & 4 & 4 & 4 \end{bmatrix}^\intercal$ produces a fibre whose projected $\mathbb{Z}$-polytope representation can be thought of as an extension of the $\mathbb{Z}$-polytope in Example 2.5.2 along an axis orthogonal to x_1 and x_5. The result is shown in Figure 2.6. The points extending the isolated vertex at $\begin{bmatrix} 4 & 0 \end{bmatrix}^\intercal$ now make up a clique of connected points that are isolated from the rest of the $\mathbb{Z}$-polytope.

Reduced dimension

If the system involves a $\mathbf{y}$ vector that is sufficiently small, it is possible that the $\mathbb{Z}$-polytope representation of $\mathcal{F}_{\mathbf{y}}$ is squashed down to $d < r - n$ dimensions. Then in some projections there may not be enough room in the required dimensions to move from one point in $\mathcal{F}_{\mathbf{y}}$ to another.

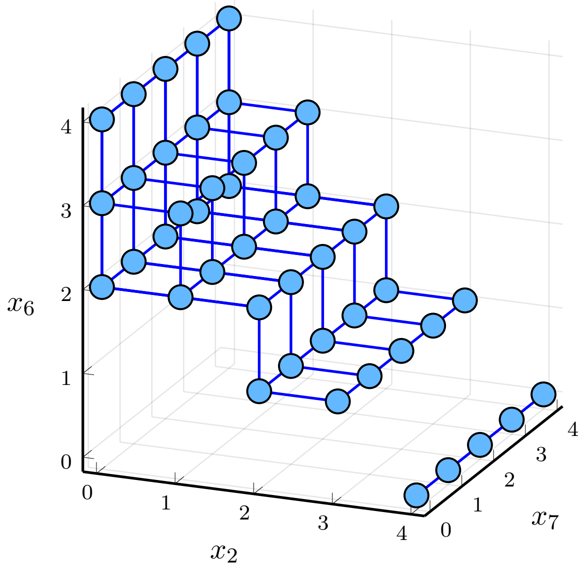


Figure 2.6: The $\mathbb{Z}$-polytope from Example 2.5.3 showing the isolated clique.

Example 2.5.4. Consider the 2×3 contingency table where each row and each of the first two columns contains entries that sum to 2, forcing the sum of the third column to be 0. The configuration matrix given by

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

where we have removed the dependent fifth row, and the vector of marginal totals is given by $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^{\intercal}$. The fibre $\mathcal{F}_{\mathbf{y}}$ is given by

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \\ 0 \\ 0 \end{bmatrix} \right\},$$

which are the contingency tables

2	0	0	2
0	2	0	2
2	2	0	

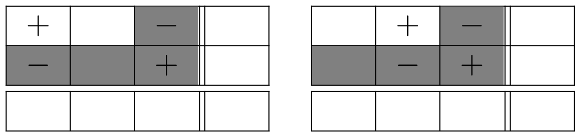
1	1	0	2
1	1	0	2
2	2	0	

0	2	0	2
2	0	0	2
2	2	0	

The column partition lattice basis for the partition $\pi = (\{3,4,5,6\},\{1,2\})$ is given by

$$U^{\pi_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}.$$

This basis gives co-ordinate direction moves when the $\mathbb{Z}$-polytope representation of $\mathcal{F}_{\mathbf{y}}$ is projected onto the x_1 and x_2 dimensions: The moves in U^{π_1} are represented on the contingency table by



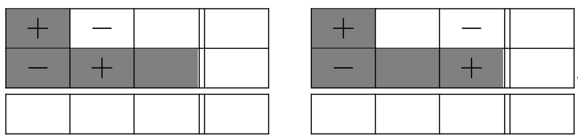
where the white cells represent the entries of $\mathbf{x}$ used as co-ordinates in the projected polytope (those in the A_1 part of A), and the grey cells represent the entries used to maintain the marginal totals (those in the A_2 part).

For all $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$, we have $x_3 = x_6 = 0$, and so none of the moves in U can be applied to any element of $\mathcal{F}_{\mathbf{y}}$. The problem is that the 0 in $\mathbf{y}$ forces the underlying polytope for the $\mathbb{Z}$-polytope representation of $\mathcal{F}_{\mathbf{y}}$ to be one dimensional, instead of the full two dimensions. The basis U^{π_1} provides co-ordinate direction moves when the $\mathbb{Z}$-polytope is projected onto the x_1 and x_2 dimensions, shown in Figure 2.7a.

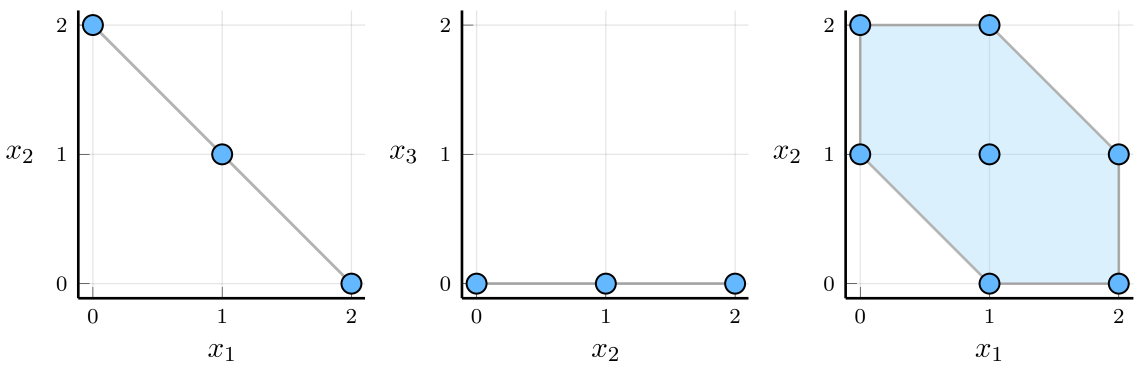
This problem can be avoided if the partition of A is chosen so that the projected polytope is long in the dimensions chosen as co-ordinate directions. Forming a column partition lattice basis with the partition $\pi_2 = (\{1,2,3,4\},\{5,6\})$ corresponds to projecting $\mathcal{F}_{\mathbf{y}}$ onto the x_2 and x_3 axes, and produces the basis U^{π_2}, given by

$$U^{\pi_2} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and represented on the contingency table by



The corresponding projection of $\mathbb{Z}$-polytope is shown in Figure 2.7b, which shows that the first move in U^{π_2} enables movement between the elements of $\mathcal{F}_{\mathbf{y}}$.



(a) Not full dimension. (b) A better orientation. (c) Full dimension.

Figure 2.7: The $\mathbb{Z}$-polytopes in Example 2.5.4.

This problem can also be avoided by restricting the use of column partition lattice bases of this type to classes of $\mathbf{y}$ such that the corresponding underlying projected polytope of $\mathcal{F}_{\mathbf{y}}$ is of full dimension, as in Figure 2.7c; or aligned with the axes, as in Figure 2.7b. That is, only use this lattice basis for sampling from fibres $\mathcal{F}_{\mathbf{y}}$ such that it is a Markov sub-basis for $\mathcal{F}_{\mathbf{y}}$.

2.5.2 A connected lattice basis

Schofield and Bonner [39] found a class of configuration matrices for which a column partition lattice basis that is guaranteed to be a Markov basis can be constructed.

Theorem 2.5.5 (Schofield and Bonner's Theorem 1 [39]). *Let A be a configuration matrix. Suppose that: (i) A contains only the values* 0 *and* 1*, and (ii) the columns of A contain all the columns of the identity matrix. Then there exists a lattice basis for A that is also a Markov basis for A.*

Their application was in capture-recapture studies in ecology, discussed in Section 1.5.3. They note that these conditions hold for many models in capture-recapture studies: in particular, condition (ii) holds when every observable history is also a true history in which there is no misidentification. This theorem may find use in other fields too: for example, in volume network tomography, the configuration matrix contains all the columns of the identity matrix when each link in the network is also an allowed path.

The lattice basis construct uses the identity matrix I_{r-n} as the A_1 part. In Section 4.2 we build on this work. We show that condition (i) can be replaced with the condition that A contains only non-negative integers. We then generalise the result to some configuration matrices that do not contain the identity matrix as a maximal submatrix. Instead, we require that the columns of A can be partitioned such that each column in the A_2 part is a non-negative integer sum of columns in the A_1 part. This is guaranteed if the matrix contains only non-negative integers, and A_1 is the identity matrix, but may hold under weaker conditions.

Chapter 3

The column partition lattice basis

3.1 Introduction

In Section 2.5 we introduced a type of basis we called a column partition lattice basis and proposed its use in $\mathbb{Z}$-polytope sampling. In this chapter we take a closer look at column partition lattice bases and demonstrate some of their properties. Of particular interest are properties that may affect the basis' utility in $\mathbb{Z}$-polytope sampling.

Section 3.2 describes a method for finding a column partition lattice basis. This method was previously covered in Section 2.5. Briefly, given a configuration matrix A and a partition π of columns of A, a column partition lattice basis is given by the columns of the matrix

$$U^\pi = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix},$$

where π partitions A into two matrices, A_1 and A_2, such that A_1 is square and invertible.

Section 3.3 gives a geometric interpretation of the choice of partition. Recall from Section 2.2 that given a suitable vector $\mathbf{y}$, the fibre

$$\mathcal{F}_\mathbf{y} = \{\mathbf{x} \in \mathbb{Z}^r_{\geq 0} : A\mathbf{x} = \mathbf{y}\}$$

is geometrically a $\mathbb{Z}$-polytope. Forming a column partition lattice basis U^π with the partition $\pi = (S_1, S_2)$ corresponds to projecting the $\mathbb{Z}$-polytope onto the subset of the co-ordinates in S_2. The moves in U^π are then steps in co-ordinate directions. The effect of both the choice of π and the value of $\mathbf{y}$ on the geometry of this projected polytope are explored.

In Section 3.4 we look closer at the vectors that make up a column partition lattice basis. We find that for any configuration matrix, the union of all column partition lattice bases corresponds to the set of integer kernel elements called *circuits*. The entries in any circuit of A are related to the determinants of the maximal submatrices of A — this fact can help avoid the problem of parity errors described in Section 2.5.1.

In this thesis we are particularly interested in *unimodular* configuration matrices. A unimodular matrix is one whose invertible maximal submatrices have a determinant of

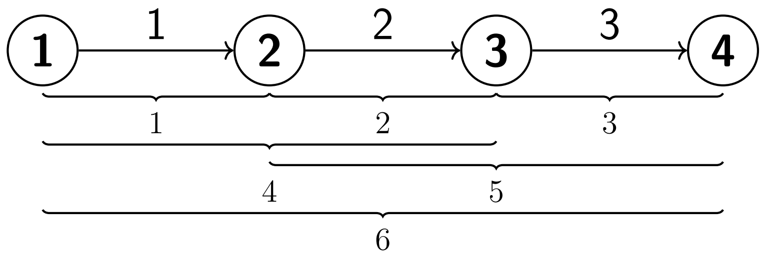


Figure 3.1: A three-link linear network. The underbraces show the allowed paths.

± 1. Properties particular to column partition lattice bases for unimodular configuration matrices are examined in Section 3.5. We find that the main advantage unimodularity of a configuration matrix confers on its column partition lattice bases is that parity errors (described in Section 2.5.1) cannot occur. We also find that if a configuration matrix is unimodular, then a matrix that defines a column partition lattice basis for it is *totally unimodular*, meaning all invertible submatrices have determinant ± 1.

Additionally, we find that the union of the column partition lattice bases of a unimodular matrix is equal to the set of circuits, which for a unimodular configuration matrix is equal to its Graver basis. The Graver basis is known to be a Markov basis. This has important implications for the dynamic Markov basis of Hazelton et al. [27], discussed in Section 3.5.2. For comparison, Section 3.5 also shows by example that column partition lattice bases for non-unimodular matrices do not necessarily share these advantages.

3.2 Construction

A column partition lattice basis for a configuration matrix $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ is a lattice basis for the kernel of A. It is formed by partitioning the columns of A into two parts, A_1 and A_2. The A_1 part must be invertible; the A_2 part contains the balance of the columns. Under the partition π, the column partition lattice basis is the collection of columns of the matrix

$$U^\pi = \begin{bmatrix} -A_1^{-1} A_2 \\ I \end{bmatrix}.$$

Strictly speaking, U^π is a matrix. We may refer to U^π as a basis: in this case, we mean the basis comprising the collection of columns of U.

This formula for writing U^π assumes that A_1 consists of the first n columns of A. We will in general make this assumption, except when comparing two different column partition lattice bases for the same configuration matrix. If A_1 does not consist of the first n columns of A, then the rows of U^π as it appears above will need to be reordered to match the original column ordering of A. We may also omit the superscript π in U^π when only one column partition lattice basis is being considered.

We illustrate the process of finding a column partition lattice basis with an example.

Example 3.2.1. Consider a directed linear traffic network consisting of three links, where travel between any pair of nodes is allowed, as shown in Figure 3.1. The link-path incidence matrix A is given by

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

We partition A without permuting the columns: $\pi_1 = (\{1,2,3\},\{4,5,6\})$. The parts are given by

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

This produces

$$A_1^{-1}A_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

and the induced column partition lattice basis is given by the columns of the matrix

$$U^{\pi_1} = \begin{bmatrix} -1 & 0 & -1 \\ -1 & -1 & -1 \\ 0 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The moves in this basis are given by the set of vectors

$$\left\{ \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

In terms of the linear traffic network, we can interpret the last column of U^{π_1} as adding to some path count vector $\mathbf{x}$ one car that drives the path made of all three links, and compensating by removing three cars that each travel a path made of one distinct link of the network. Thus the value of $\mathbf{y}$, the counts of cars observed on each link in the network, is unchanged.

Different partitions of the configuration matrix will typically yield different column partition lattice bases. Taking another partition $\pi_2 = (\{4,5,6\},\{1,2,3\})$, we form

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so that

$$A_1^{-1}A_2 = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}.$$

Negating and appending the identity matrix produces the matrix

$$\begin{bmatrix} 0 & -1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We need to reverse the column permutation by reordering the rows of this matrix to match the original ordering, so the column partition lattice basis induced by this partition is given by the columns of the matrix

$$U^{\pi_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & -1 \end{bmatrix}.$$

The columns of this matrix make up a basis because the identity matrix in the first co-ordinates ensures independence.

3.3 Polytope geometry

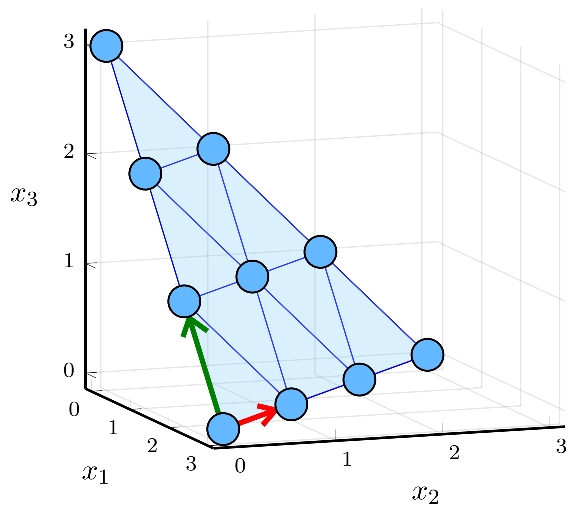
In Section 2.2, we saw that for a configuration matrix A, and a count vector $\mathbf{y}$, the $\mathbb{Z}$-polytope representation of $\mathcal{F}_{\mathbf{y}}$ is the intersection of a translate of the kernel of A with the non-negative orthant and the integer lattice. If A is of size $n \times r$, then the underlying polytope is an $r-n$ dimensional object in r dimensional space. Using the $r-n$ moves in a column partition lattice basis, we can move around this $\mathbb{Z}$-polytope in $r-n$ independent directions.

A column partition lattice basis is defined by the matrix

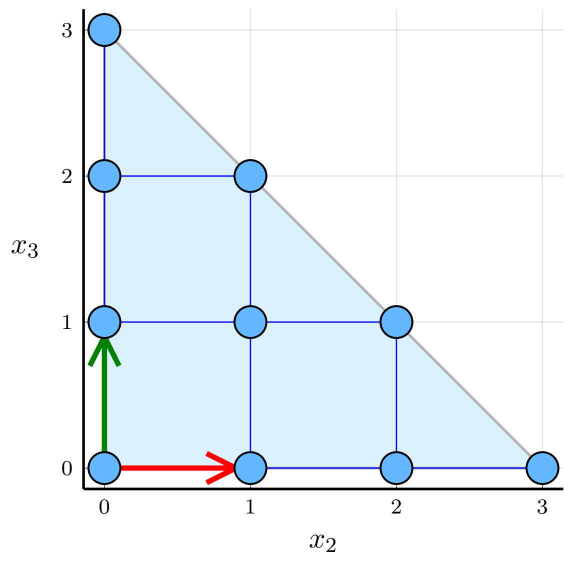
$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}.$$

We can select the ith column of U by multiplying by $\mathbf{e}_i$, so a move in U can be written in the form

$$\mathbf{u}_i = \begin{bmatrix} -A_1^{-1}A_2\mathbf{e}_i \\ \mathbf{e}_i \end{bmatrix}.$$



(a) The $\mathbb{Z}$-polytope.



(b) The projected $\mathbb{Z}$-polytope.

Figure 3.2: The projection of the $\mathbb{Z}$-polytope in Example 2.2.3, showing corresponding column partition lattice basis moves.

Taking into account only the A_2 co-ordinates, the move $\mathbf{u}_i$ is a unit step in the $(n+i)$th direction, while the other co-ordinates are held constant.

If we plot the $\mathbb{Z}$-polytope onto the A_2 co-ordinates, then the moves in U are equivalent to the set of steps in co-ordinate directions.

Observation 3.3.1. Given a configuration matrix A, a column partition π, and a vector $\mathbf{y}$, studying whether the column partition lattice basis U^π connects the elements of $\mathcal{F}_\mathbf{y}$ is equivalent to studying whether the points in the projection of the $\mathbb{Z}$-polytope onto the A_2 co-ordinates are connected by co-ordinate direction moves.

Figure 3.2 shows the two dimensional $\mathbb{Z}$-polytope from Example 2.2.3. In Figure 3.2a it is shown in its original three dimensions. In Figure 3.2b it is shown projected onto the x_2 and x_3 dimensions — there are three choices of pairs of co-ordinates upon which to project this $\mathbb{Z}$-polytope, but due to symmetry all appear identical. The column partition lattice basis elements corresponding to projecting onto x_2 and x_3 are shown in red and green.

3.3.1 Bounding hyperplanes

We can get a better understanding of whether or not co-ordinate direction moves in a projected $\mathbb{Z}$-polytope can connect its points by looking at the geometry of the underlying projected polytope. The projected $\mathbb{Z}$-polytope in Figure 3.2 is bounded by the x_1 axis where $x_3 = 0$, the x_3 axis where $x_1 = 0$, and by the diagonal line segment shown in black, where $x_2 = 0$. These lines correspond to the *bounding hyperplanes* of the $\mathbb{Z}$-polytope.

Definition 3.3.2 (Bounding hyperplane)**.** The ith *bounding hyperplane* of a polytope is the set $\{\mathbf{x} : A\mathbf{x} = \mathbf{y}, x_i = 0\}$.

The ith bounding hyperplane is an affine subspace that divides the space into sets that have $x_i < 0$ and $x_i > 0$. Therefore $\mathcal{F}_{\mathbf{y}}$ lies entirely on one side of any face. The projection of the underlying polytope is the intersection of the positive sides of the bounding hyperplanes, and part of a bounding hyperplane may make up a face of the underlying polytope.

In Figure 3.2, every integral point within these bounding hyperplanes is a point in the fibre, which is to say that this projection does not suffer from the problem of parity errors described in Section 2.5.1. In this case, whether or not the column partition lattice basis connects the fibre is entirely determined by the orientation and position of the bounding hyperplanes.

The orientation and position of the bounding hyperplanes of a projected polytope are given by the matrix U and the vector $A_1^{-1}\mathbf{y}$. The bounding hyperplanes for A_2 co-ordinates are orthogonal to the corresponding axis and intersect the origin: in Figure 3.2, the bounding hyperplane for $x_3 = 0$ is the x_1 axis. Together, these keep the projected polytope boxed in to the non-negative orthant. For the A_1 co-ordinates, the ith row vector of U is normal to the bounding hyperplane for $x_i = 0$, and together with the ith entry of $A_1^{-1}\mathbf{y}$ gives the position for the first bounding hyperplanes.

Theorem 3.3.3. *Let A be an $n \times r$ configuration matrix and let π partition the columns of A into A_1 and A_2. Let $C = A_1^{-1}A_2$, and let*

$$U = \begin{bmatrix} -C \\ I \end{bmatrix}$$

be a column partition lattice basis for A. A normal vector $\mathbf{n}_i$ to the ith bounding hyperplane of the projection of the polytope onto the A_2 co-ordinates is given by the ith row vector $\mathbf{c}_{i\cdot}$ of C, or

$$\mathbf{n}_i = \mathbf{c}_{i\cdot}.$$

The ith bounding hyperplane intersects the x_{n+j} axis at

$$x_{n+j} = \frac{(A_1^{-1}\mathbf{y})_i}{(\mathbf{c}_{i\cdot})_j}.$$

Proof. We have

$$\begin{aligned} A\mathbf{x} &= \mathbf{y} \\ A_1\mathbf{x}_1 + A_2\mathbf{x}_2 &= \mathbf{y} \\ A_1^{-1}(A_1\mathbf{x}_1 + A_2\mathbf{x}_2) &= A_1^{-1}\mathbf{y} \\ \mathbf{x}_1 + C\mathbf{x}_2 &= A_1^{-1}\mathbf{y} \end{aligned}$$

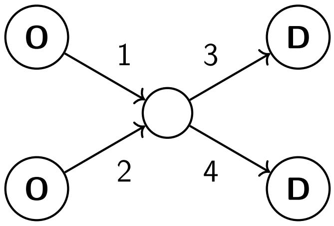


Figure 3.3: The graph of the traffic network in Example 3.3.4. Nodes marked O are origins for traffic, and nodes marked D are destinations.

Taking the ith row vector of this equation and setting $x_i = 0$ gives an equation for the ith bounding hyperplane:

$$\mathbf{c}_{i\cdot} \cdot \mathbf{x}_2 = (A_1^{-1}\mathbf{y})_i.$$

This affine space is equal to the translated kernel of the row vector $\mathbf{c}_{i\cdot}$ (considered as a matrix), which is orthogonal to the rowspace, which has the vector $\mathbf{c}_{i\cdot}$ as a basis. This vector $\mathbf{c}_{i\cdot}$ is therefore normal to the ith bounding hyperplane.

To find the position of this face, we must find a particular solution $\hat{\mathbf{x}}_2$ to

$$\mathbf{c}_{i\cdot} \cdot \mathbf{x}_2 = (A_1^{-1}\mathbf{y})_i.$$

Let j be the index of a non-zero entry of $\mathbf{c}_{i\cdot}$. Setting all except the $(n + j)$th element of $\hat{\mathbf{x}}_2$ to zero gives us the x_{n+j} intercept. We have

$$\hat{\mathbf{x}}_2 = \frac{(A_1^{-1}\mathbf{y})_i}{c_{i,j}}\mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth standard basis vector. Therefore, the ith bounding hyperplane intersects the x_{n+j} axis at

$$x_{n+j} = \frac{(A_1^{-1}\mathbf{y})_i}{c_{i,j}}.$$

□

If A is unimodular, then by Theorem 3.5.5 the non-zero elements of C are all ± 1 and we have $c_{i,j} = \pm 1$, and the x_{n+j} intercept is at $\pm A_1^{-1}\mathbf{y}$.

Example 3.3.4. Consider the traffic network in Figure 3.3, which has configuration matrix

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

If link traffic counts $\mathbf{y} = \begin{bmatrix}3 & 3 & 2 & 2\end{bmatrix}^\mathsf{T}$ are observed, then

$$\mathcal{F}_\mathbf{y} = \left\{ \begin{bmatrix}2\\0\\2\\1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\\0\\0\\2\end{bmatrix}, \begin{bmatrix}2\\0\\1\\2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\2\\1\\0\\1\\2\end{bmatrix}, \begin{bmatrix}1\\1\\0\\2\\2\\0\end{bmatrix}, \begin{bmatrix}0\\2\\0\\1\\2\\1\end{bmatrix} \right\}.$$

We choose the column partition such that

$$A_1 = \begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 1 & 1\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}.$$

Then the column partition lattice basis is given by

$$U = \begin{bmatrix}-1 & -1\\-1 & 0\\0 & -1\\1 & 1\\1 & 0\\0 & 1\end{bmatrix},$$

of which the C part is given by

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

By Theorem 3.3.3, C and $A_1^{-1}\mathbf{y} = \begin{bmatrix}3 & -1 & 2 & 2\end{bmatrix}^\mathsf{T}$ give us the geometry of the projection of the underlying polytope.

Figure 3.4 compares the bounding hyperplanes of the underlying projected polytope with the elements of $\mathcal{F}_\mathbf{y}$. The first bounding hyperplane, labelled x_1, has a normal vector $\mathbf{c}_{1\cdot} = \begin{bmatrix}1 & 1\end{bmatrix}$ and intercepts the x_5 axis at

$$\begin{aligned} x_5 &= \frac{(A_1^{-1}\mathbf{y})_1}{(\mathbf{c}_{1\cdot})_1} \\ &= \frac{3}{1} \\ &= 3. \end{aligned}$$

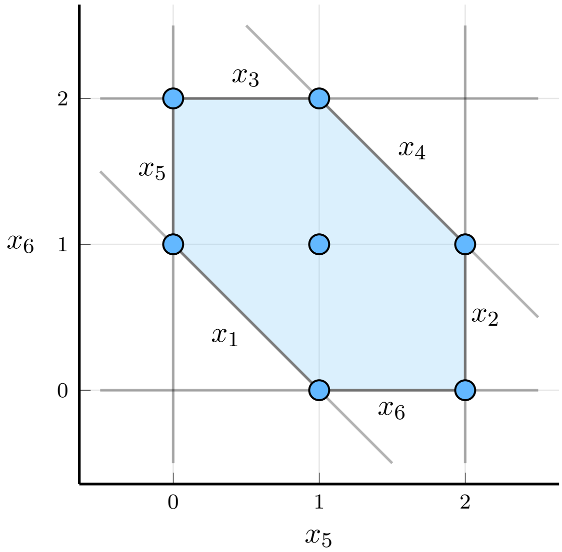


Figure 3.4: The set $\mathcal{F}_{\mathbf{y}}$ from Example 3.3.4 projected onto the x_5 and x_6 dimensions showing the bounding hyperplanes of the underlying polytope.

The second bounding hyperplane, labelled x_2, has a normal vector $\mathbf{c}_{2\cdot} = \begin{bmatrix}-1 & -1\end{bmatrix}$ and intercepts the x_5 axis at

$$\begin{aligned} x_5 &= \frac{(A_1^{-1}\mathbf{y})_2}{(\mathbf{c}_{2\cdot})_1} \\ &= \frac{-1}{-1} \\ &= 1. \end{aligned}$$

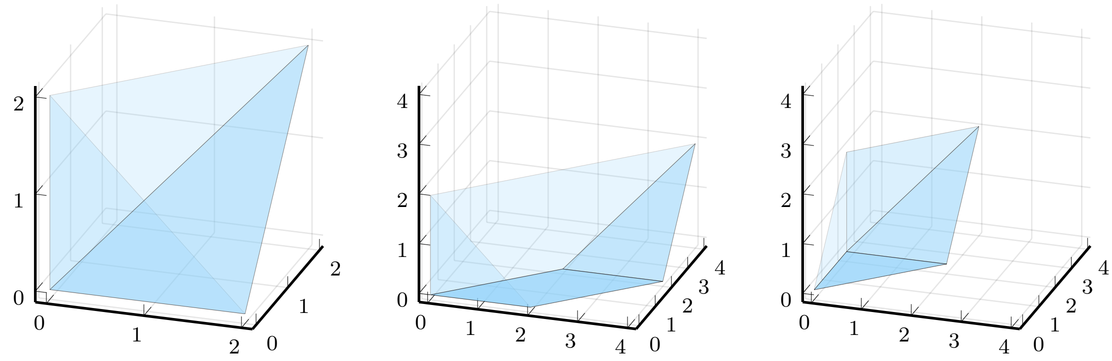
For a given configuration matrix and column partition, the C matrices and therefore the orientations of the bounding hyperplanes are the same for any $\mathbf{y}$. Changing the value of $\mathbf{y}$ changes the position of the faces relative to each other — this can result in underlying polytopes not only of different sizes, but of different shapes for the same configuration matrix.

Example 3.3.5. Consider again the three-link linear network in Figure 1.2, which has configuration matrix

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

We wish to project onto the x_1, x_2, and x_3 axes, so we choose

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$



(a) $\mathbf{y} = \begin{bmatrix}2 & 2 & 2\end{bmatrix}^\intercal$. (b) $\mathbf{y} = \begin{bmatrix}4 & 4 & 2\end{bmatrix}^\intercal$. (c) $\mathbf{y} = \begin{bmatrix}2 & 4 & 2\end{bmatrix}^\intercal$.

Figure 3.5: The underlying projected polytopes from Example 3.3.5 projected onto the x_1, x_2, and x_3 dimensions.

inducing the column partition lattice basis

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

The last three rows of U give normal vectors to the x_4, x_5, and x_6 bounding hyperplanes. The positions of these bounding hyperplanes are given by $A_1^{-1}\mathbf{y}$, so altering the values of $\mathbf{y}$ will alter the positions of the bounding hyperplanes relative to each other. Some underlying projected polytopes for different values of $\mathbf{y}$ for this system are shown in Figure 3.5. These serve to demonstrate that altering the value of $\mathbf{y}$ can change the geometry of a projected $\mathbb{Z}$-polytope quite dramatically.

3.3.2 Constructing matrices from particular projected polytopes

In this thesis we are interested in applying geometric insight to the search for Markov bases. In doing so, it can be useful to take a particular projected $\mathbb{Z}$-polytope with properties that we wish to study, and see from which configuration matrices and column partition lattice bases it might arise. Here we outline through a series of examples our technique for finding a configuration matrix and column partition from a $\mathbb{Z}$-polytope. The technique is based on writing down normal vectors to the bounding hyperplanes of the polytope we wish to study and collecting them as rows of a matrix. This matrix becomes the foundation of our $A_1^{-1}A_2$ matrix. This matrix is modified until it becomes an allowable configuration matrix under the model of interest, which despite the modifications still

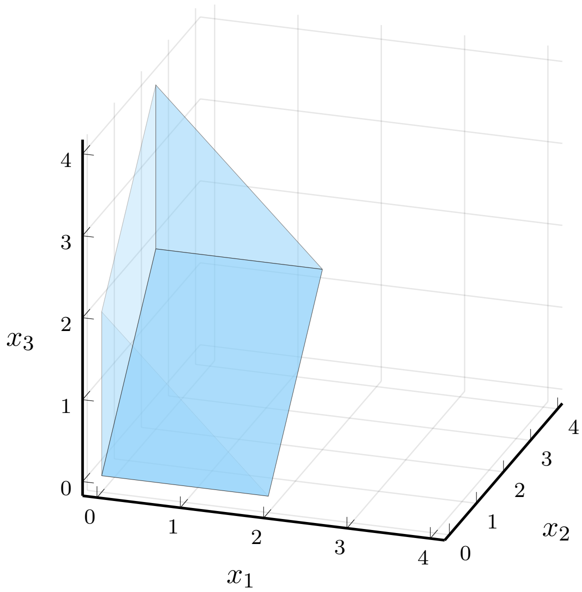


Figure 3.6: The underlying projected polytope from Example 3.3.6.

produces the required $\mathbb{Z}$-polytope. For most models, we require that the configuration matrix A is such that $A \in \{0,1\}^{n \times r}$. In each case, the column partition that produces the required projection is given by setting A_1 to be the first n columns of A.

Example 3.3.6. Consider an underlying projected polytope whose bounding hyperplanes are given by the rows of the matrix

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

Each of these rows must appear in $A_1^{-1}A_2$, and so this matrix M will become a submatrix of $A_1^{-1}A_2$. An example of a projected polytope that has bounding hyperplanes with these normal vectors is shown in Figure 3.6. The bounding hyperplanes have been positioned according to the corresponding entries of the vector $A_1^{-1}\mathbf{y}$, which are $\begin{bmatrix} 0 & 2 & 2 \end{bmatrix}^\intercal$. We initialise $A_1^{-1}A_2$ with the rows of M, and we have

$$A_1^{-1}A_2 = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

We prepend the identity matrix and perform row operations to produce a $\{0,1\}$ matrix.

Then

$$A_1^{-1}A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$
$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

This matrix has the problem that columns two and four are duplicates, which we will generally want to avoid. We avoid this by adding a row to $A_1^{-1}A_2$ to distinguish them, which means adding a dummy bounding hyperplane to the projected polytope. This dummy bounding hyperplane will not affect the projected $\mathbb{Z}$-polytope if the corresponding entry in $\mathbf{y}$ is such that the $\mathbb{Z}$-polytope lies on its positive side. We append to $A_1^{-1}A_2$ the row vector $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ to get

$$A_1^{-1}A_2 = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Note that appending a standard basis vector does not affect the total unimodularity of a matrix [40]. Prepending the identity matrix and performing row operations produces

$$A_1^{-1}A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix},$$
$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

This matrix has all entries in $\{0, 1\}$ and has no repeated or zero columns and is a potential configuration matrix for many applications. For example, this matrix is a link-path incidence matrix for a four-link linear traffic network.

Another problem we might encounter is that it is not always possible to perform row operations that produce a $\{0, 1\}$ matrix with a sufficient number of unique rows.

Example 3.3.7. Consider a projected polytope with bounding hyperplanes that have normal vectors given by the rows of the matrix

$$M = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The only $\{0,1\}$ vector in the rowspace of M is $\begin{bmatrix}0 & 0\end{bmatrix}$, while we require at least two unique rows in the configuration matrix we are constructing. We can solve this by appending standard basis vectors:

$$A_1^{-1}A_2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} A_1^{-1}A &= \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \\ A &= \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{aligned}$$

which is a configuration matrix that meets our requirements. This particular matrix is the configuration matrix of a 2×3 contingency table.

The last problem we will consider may occur when the underlying projected polytope has a face with a normal vector with entries other than 0 or ± 1. By appending multiple copies of the same standard basis row vectors, we produce a matrix that can be transformed to a $\{0,1\}$ matrix through row operations. Again, provided we choose $\mathbf{y}$ carefully, these extra rows will not affect the $\mathbb{Z}$-polytope produced.

Example 3.3.8. Consider a projected polytope that has a bounding hyperplane with the normal vector given by the row of the matrix

$$M = \begin{bmatrix} 5 & 1 \end{bmatrix}.$$

We need to subtract 1 from the 5 four times to get it to a value in $\{0, \pm 1\}$. If we simply append the row $\begin{bmatrix}-1 & 0\end{bmatrix}$, we end up with a 4 in the A_1 part, and we are stuck with -1 in the A_2 part:

$$\begin{aligned} A_1^{-1}A &= \begin{bmatrix} 1 & 0 & 5 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \\ A &= \begin{bmatrix} 1 & 4 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}. \end{aligned}$$

Instead we append four copies of the row $\begin{bmatrix}-1 & 0\end{bmatrix}$, and one copy of the row $\begin{bmatrix}1 & 0\end{bmatrix}$. Then

$$A_1^{-1}A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 5 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

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

which is a configuration matrix with all $\{0,1\}$ entries, as required.

In each of these examples, adding a row vector to $A_1^{-1}A_2$ does not mean it no longer produces the polytope we are interested in. Choosing an appropriate entry for the $\mathbf{y}$ vector means that the corresponding bounding hyperplane has no effect on the polytope.

3.4 Properties of U

In this section we prove some properties of column partition lattice bases. In Section 3.4.1 we look at properties of the matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix},$$

whose columns define the column partition lattice basis. Theorem 3.4.1 relates the determinants of submatrices of $A_1^{-1}A_2$ to the determinants of maximal submatrices of A. Of particular interest are the determinants of 1×1 submatrices: these are just the entries of $A_1^{-1}A_2$. This tells us what kind of entries a U matrix might contain, and has important implications for the problem of parity errors described in Section 2.5.1.

We then show in Section 3.4.3 that each element of a column partition lattice basis is a scaled *circuit* of the configuration matrix. The circuits of a configuration matrix are a particular kind of minimal element of the integer kernel. The implications of these results on unimodular configuration matrices will be explored in Section 3.5.

3.4.1 Submatrix determinants

We show that the determinant of any square submatrix of $A_1^{-1}A_2$ is equal in absolute value to the determinants of a maximal submatrix of A divided by $\det(A_1)$. This includes the 1×1 submatrices, which are simply the entries of $A_1^{-1}A_2$.

Theorem 3.4.1. *Let* $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ *be an* $n \times r$ *configuration matrix. Let* M *be the* $k \times k$ *square submatrix of* $A_1^{-1}A_2$ *with row indices* $\mathcal{I} = \{i_1, \ldots, i_k\}$ *and column indices* $\mathcal{J} = \{j_1, \ldots, j_k\}$. *Let* $A_1^{\mathcal{I}\setminus\mathcal{J}}$ *be the maximal square submatrix of* A *constructed by taking* A_1 *and removing columns with indices* $\mathcal{I}$, *and appending columns from* A_2 *with indices* $\mathcal{J}$. *Then*

$$\det(M) = (-1)^p \frac{\det(A_1^{\mathcal{I}\setminus\mathcal{J}})}{\det(A_1)},$$

where

$$p = \frac{k(k-1)}{2} + \sum_{i\in\mathcal{I}}(n-i).$$

Proof. We form the matrix $PA_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}$ by taking the identity matrix I_n and replacing the columns with indices in $\mathcal{I}$ with the columns in $A_1^{-1}A_2$ with indices in $\mathcal{J}$. Here, P is the permutation matrix that moves the columns with indices $\mathcal{J}$ to positions with indices $\mathcal{I}$. By Cramer's rule $PA_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}$ has determinant $\det(M)$. We multiply by P^{-1}, which reorders the columns so that the match the order in which they appear in $A_1^{-1}A$. This requires switching

$$p = \frac{k(k-1)}{2} + \sum_{i\in\mathcal{I}}(n-i)$$

pairs of columns. Each switch multiplies the determinant by -1, so $A_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}$ has determinant $(-1)^p \det(M)$. This matrix $A_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}$ includes the columns of $I = A_1^{-1}A_1$ with indices in $\mathcal{I}$, and the columns of $A_1^{-1}A_2$ with indices in $\mathcal{J}$. Multiplying by A_1, we have

$$\begin{aligned}
A_1A_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}} &= A_1^{\mathcal{I}\setminus\mathcal{J}} \\
\det(A_1A_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}) &= \det(A_1^{\mathcal{I}\setminus\mathcal{J}}) \\
\det(A_1)\det(A_1^{-1}A_1^{\mathcal{I}\setminus\mathcal{J}}) &= \det(A_1^{\mathcal{I}\setminus\mathcal{J}}) \\
\det(A_1)(-1)^p\det(M) &= \det(A_1^{\mathcal{I}\setminus\mathcal{J}}) \\
\det(M) &= (-1)^p\frac{\det(A_1^{\mathcal{I}\setminus\mathcal{J}})}{\det(A_1)}.
\end{aligned}$$

□

Of particular interest are the determinants of the 1×1 submatrices, which are the entries of $A_1^{-1}A_2$. These are given by the following corollary.

Corollary 3.4.2. *Let* $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ *be an* $n \times r$ *configuration matrix, where* A_1 *is square and invertible. If* a_{ij} *is the element in the* i*th row and* j*th column of* $A_1^{-1}A_2$, *then*

$$a_{ij} = (-1)^{n-i}\frac{\det(A_1^{i\setminus j})}{\det(A_1)},$$

where $A_1^{i\setminus j}$ *is the maximal square submatrix of* A *obtained by taking* A_1 *and removing the* i*th column, and appending the* j*th column of* A_2.

Proof. We can apply Theorem 3.4.1, with $M = \left[a_{ij}\right]$. We have

$$\begin{aligned} a_{ij} &= \det(M) \\ &= (-1)^p \frac{\det(A_1^{i \setminus j})}{\det(A_1)}. \end{aligned}$$

In this case $k = 1$ and $\mathcal{I} = \{i\}$, so

$$\begin{aligned} p &= \frac{k(k-1)}{2} + \sum_{i \in \mathcal{I}} (n-i) \\ &= n - i, \end{aligned}$$

which produces

$$a_{ij} = (-1)^{n-i} \frac{\det(A_1^{i \setminus j})}{\det(A_1)}.$$

as required. □

The following example may help to provide clarity.

Example 3.4.3. Consider the matrix

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

If we partition A such that A_1 consists of the first three columns, then $\det(A_1) = 1$ and

$$A_1^{-1} A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{bmatrix}.$$

We choose M to be the 2×2 matrix in the bottom right hand corner of $A_1^{-1}A$, so

$$M = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and $\det(M) = -2$, whose row indices are $\mathcal{I} = \{2, 3\}$ and column indices are $\mathcal{J} = \{2, 3\}$.

We construct the matrix $A_1^{-1} A_1^{\mathcal{I} \setminus \mathcal{J}}$ by taking the first column of $I = A_1^{-1} A_1$ and the second and third columns of $A_1^{-1} A_2$. Then

$$A_1^{-1} A_1^{\mathcal{I} \setminus \mathcal{J}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix},$$

and $\det(A_1^{-1} A_1^{\mathcal{I} \setminus \mathcal{J}}) = -2$.

The matrix $A_1^{\mathcal{I}\setminus\mathcal{J}}$ is obtained by taking the first, fifth and sixth columns of A, so

$$A_1^{\mathcal{I}\setminus\mathcal{J}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

which has determinant -2.

We have

$$(-1)^p \frac{\det(A_1^{\mathcal{I}\setminus\mathcal{J}})}{\det(A_1)} = (-1)^p \frac{-2}{1}$$
$$= -2,$$

where no column switches are required so $p = 0$. This equals the previously calculated $\det(M) = -2$, matching Theorem 3.4.1.

We include a more detailed calculation for a 1×1 submatrix of some $A_1^{-1} A_2$.

Example 3.4.4. When we construct a column partition lattice basis for a configuration matrix A, we begin by row reducing A by finding

$$A_1^{-1} A = \begin{bmatrix} A_1^{-1} A_1 & A_1^{-1} A_2 \end{bmatrix}$$
$$= \begin{bmatrix} I & A_1^{-1} A_2 \end{bmatrix}.$$

The row operations transform A_1 to I, which has determinant $\det(I) = 1$, so the effect of the row operations on the determinant was to divide it by $\det(A_1)$. These row operations have the same effect on the determinants of all of the maximal submatrices of A: if N is a square maximal submatrix of A, then $A_1^{-1} N$ is a square maximal submatrix of $A_1^{-1} A$ and

$$\det(A_1^{-1} N) = \frac{\det(N)}{\det(A_1)}.$$

One class of square maximal submatrices of A can be constructed by taking A_1 and removing the ith column and appending the jth column of A_2. We will write this matrix as $A_1^{i\setminus j}$. After row reducing A, this matrix is of the form

$$A_1^{-1} A_1^{i\setminus j} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & a_{1j} \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & a_{2j} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & a_{i-1,j} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & a_{ij} \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & a_{i+1,j} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & a_{nj} \end{bmatrix},$$

where $\mathbf{a}_j = \begin{bmatrix} a_{1j} & \dots & a_{nj} \end{bmatrix}^\intercal$ is the jth column of $A_1^{-1}A_2$. We have

$$\begin{vmatrix} 1 & 0 & \cdots & 0 & a_{1j} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & a_{2j} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & a_{i-1,j} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & a_{ij} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & a_{i+1,j} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nj} & 0 & \cdots & 1 \end{vmatrix} = a_{ij}$$

by Cramer's rule. Then $n-i$ column switches are required to place the columns in the order in which they appear in $A_1^{-1}A_2$, so

$$\begin{aligned} a_{ij} &= (-1)^{n-i} \det\left(A_1^{-1}A_1^{i\backslash j}\right) \\ &= (-1)^{n-i} \frac{\det\left(A_1^{i\backslash j}\right)}{\det\left(A_1\right)}. \end{aligned}$$

Theorem 3.4.1 and Corollary 3.4.2 show a one-to-one correspondence between the maximal square submatrices of A and the square submatrices of $A_1^{-1}A_2$; and between the elements of $A_1 \times A_2$ (the ordered pairs of columns where one is from A_1 and one is from A_2), and the entries in $A_1^{-1}A_2$. If A is $n \times r$, then A_1 is $n \times n$, and A_2 and $A_1^{-1}A_2$ are both $n \times (r-n)$. We set $k = r - n$, and then the number of entries in $A_1^{-1}A_2$ is given by $n \times k$. This matches the numbers of pairs of columns in $A_1 \times A_2$.

The number of columns in A is $n+k$, so there are $\binom{n+k}{n}$ maximal square submatrices in A. For each $0 \le i \le \min(k,n)$, $A_1^{-1}A_2$ contains $\binom{k}{i} \times \binom{n}{i}$ square submatrices of size $i \times i$. The number of square submatrices in $A_1^{-1}A_2$ is therefore given by

$$\sum_{i=0}^{\min(n,k)} \binom{n}{i}\binom{k}{i},$$

producing the identity

$$\binom{n+k}{n} = \sum_{i=0}^{\min(n,k)} \binom{k}{i}\binom{n}{i}.$$

This is a special case of Vandermonde's identity,

$$\binom{k+q}{n} = \sum_{i=0}^{k} \binom{k}{i}\binom{q}{n-i},$$

where setting $q = n$ produces

$$\begin{aligned}\binom{k+n}{n} &= \sum_{i=0}^{k} \binom{k}{i}\binom{n}{n-i} \\ &= \sum_{i=0}^{k} \binom{k}{i}\binom{n}{i}.\end{aligned}$$

3.4.2 Maximal submatrix determinants

One of the problems for connectivity of column partition lattice bases was the problem of parity errors, explained in Section 2.5.1. Parity errors can be avoided if the column partition can be chosen such that the U matrix has all integer entries. Section 3.4.1 showed that the determinants of the submatrices of a column partition lattice basis matrix are related to the determinants of the maximal submatrices of the configuration matrix, divided by the determinant of the A_1 part under the partition used. This includes the 1×1 submatrices, the entries of U. The following theorem uses this fact to show how non-integer entries in U might be avoided.

Theorem 3.4.5. *Let $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ be a configuration matrix and a column partition thereof, and let U be the corresponding column partition lattice basis. Let $A_1^{i\backslash j}$ be the maximal square submatrix of A obtained by taking columns of A_1, removing the ith column, and appending the jth column of A_2. Let $\mathcal{A}_1$ be the set of all such matrices, so that*

$$\mathcal{A}_1 = \{A_1^{i\backslash j} : i = 1, \ldots, n; j = 1, \ldots, r-n\}.$$

Then U is integral if and only if $\det(A_1)$ *divides* $\det(A_1^{i\backslash j})$ *for all $A_1^{i\backslash j} \in \mathcal{A}_1$.*

Proof. The column partition lattice basis matrix U is made by appending I_n to $-A_1^{-1}A_2$. If $A_1^{-1}A_2$ is integral, then so is U. By Corollary 3.4.2, each entry a_{ij} of $A_1^{-1}A_2$ is given by

$$a_{ij} = (-1)^{n-i}\frac{\det(A_1^{i\backslash j})}{\det(A_1)}.$$

These matrices $A_1^{i\backslash j}$ are the elements of $\mathcal{A}_1$, so if $\det(A_1)$ divides $\det(A_1^{i\backslash j})$ for each $A_1^{i\backslash j} \in \mathcal{A}_1$, then U is integral. If there is $A_1^{i\backslash j} \in \mathcal{A}_1$ such that A_1 does not divide $A_1^{i\backslash j}$, then a_{ij} is not integral, and so U contains a non-integral entry. □

Example 3.4.6. Consider the configuration matrix

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

This is the configuration matrix discussed in Section 5.2 with some of the dependent rows removed. Checking with a computer, there are sixteen maximal square submatrices with determinant ± 2 and one with determinant -4. There are also 67 singular maximal square submatrices.

We choose as A_1 the maximal square submatrix with determinant -4,

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

Then the determinant of A_1 does not divide the determinant of any other invertible maximal submatrix, and the induced column partition lattice basis should contain non-integer values. This basis is

$$U_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

If instead we choose the maximal square submatrix

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

as the A_1 part, then $\det(A_1) = 2$, which divides the determinants of all other invertible maximal square submatrices. This induces the column partition lattice basis

$$U_2 = \begin{bmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which has all integral entries.

3.4.3 Circuits

Scaled basis elements

In Section 3.4.2 we saw that the matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}$$

which defines a column partition lattice basis may have non-integer entries. When constructing a random walk between the points in a fibre, we must step from integer point to integer point, so the moves used to step must themselves be integral. If U is used to construct a random walk through a fibre, then the steps used must be integer multiples of the columns of U where the denominators have been multiplied out. We will say that these integer multiples are *scaled* column partition lattice basis elements.

Definition 3.4.7 (Scaled column partition lattice basis vectors)**.** Let A be a configuration matrix, and let the columns of U be a column partition lattice basis for A. The corresponding *scaled* column partition lattice basis vectors are given by the matrix $U_{\mathbb{Z}}$, which is constructed by multiplying each column of U by the lowest common multiple of the denominators of the entries of that column.

The scaled column partition lattice basis vectors are the shortest allowed moves in co-ordinate directions. If some column partition lattice basis U has all integer entries, then $U = U_{\mathbb{Z}}$.

Example 3.4.8. Consider the configuration matrix

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

for which

$$U = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a column partition lattice basis. The corresponding scaled column partition lattice elements are given by the matrix

$$U_{\mathbb{Z}} = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

where each column of U has been multiplied by two to give the corresponding column of $U_{\mathbb{Z}}$.

The scaled elements of a column partition lattice basis do not necessarily constitute a lattice basis themselves. Considering again A and U from Example 3.4.8, take the integer kernel element $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$ given by

$$\mathbf{z} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

where $\mathbf{z}$ is expressed in terms of the basis as the sum of the first two columns of U.

In terms of columns of $U_{\mathbb{Z}}$, we have

$$\mathbf{z} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 2 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 0 \\ 2 \\ 0 \end{bmatrix}.$$

This means that $\mathbf{z}$ is not expressible as an integer combination of scaled basis elements, so by Definition 1.4.1, $U_{\mathbb{Z}}$ is not a lattice basis.

Circuits

The scaled elements of the column partition lattice bases of a configuration matrix correspond to elements of the integer kernel called *circuits*. The definition of a circuit requires the definition of the *support* of a vector.

Definition 3.4.9 (Support)**.** The *support* of a vector $\mathbf{u}$, written $\text{supp}(\mathbf{u})$, is the set of indices i such that $u_i \neq 0$.

When dealing with partitions of a matrix, we will also write $\text{supp}(A_i) = \{j : \mathbf{a}_j \in A_i\}$, where $\mathbf{a}_j$ is a column of A_i. This set $\text{supp}(A_i)$ is therefore the set of indices of the columns of A in A_i.

Definition 3.4.10 (Circuit [18])**.** A *circuit* of a matrix A is a vector $\mathbf{u} \in \ker_{\mathbb{Z}}(A)$ such that its entries u_i are relatively prime and the support $\text{supp}(\mathbf{u})$ is minimal with respect to inclusion.

We will write $\mathcal{C}_A$ for the set of circuits of the matrix A.

Theorem 3.4.11. *Let A be an $n \times r$ configuration matrix, and let Π be the collection of column partitions of A. Then the union over Π of the scaled column partition lattice bases of A is equal to the set $\mathcal{C}_A$, or*

$$\mathcal{C}_A = \bigcup_{\pi \in \Pi} U_{\mathbb{Z}}^{\pi},$$

where each $U_{\mathbb{Z}}^{\pi}$ is considered as the set of its columns, and each element is ordered with respect to the original ordering of the columns of A.

Proof. First, we show that $\bigcup_{\pi \in \Pi} U_{\mathbb{Z}}^{\pi} \subset \mathcal{C}_A$. Let $\mathbf{u} \in \mathbb{Q}^r$ be an element of a column partition lattice basis U of A under partition of the columns of A into A_1 and A_2. Let $\mathbf{u}_{\mathbb{Z}}$ be formed by multiplying $\mathbf{u}$ by the least common multiple of the denominators of entries in $\mathbf{u}$, so $\mathbf{u} \in \bigcup_{\pi \in \Pi} U_{\mathbb{Z}}^{\pi}$. Then $\text{supp}(\mathbf{u}_{\mathbb{Z}}) \subset \text{supp}(A_1) \cup \{i\}$ for some $i \in \text{supp}(A_2)$. The the entries of $\mathbf{u}_{\mathbb{Z}}$ are integers and relatively prime, so to show that $\mathbf{u}_{\mathbb{Z}} \in \mathcal{C}_A$ we need only show that $\text{supp}(\mathbf{u}_{\mathbb{Z}})$ is minimal with respect to inclusion.

Suppose that this is not the case: then there exists a non-zero $\mathbf{z} \in \ker(A)$ such that $\text{supp}(\mathbf{z}) \subsetneq \text{supp}(\mathbf{u}_{\mathbb{Z}}) \subset \text{supp}(A_1) \cup \{i\}$. We claim that $i \notin \text{supp}(\mathbf{z})$. If it were not, then $\mathbf{z} \in \ker(A)$, so

$$\begin{aligned} \mathbf{0} &= A\mathbf{z} \\ &= A_1\mathbf{z}_1 + A_2\mathbf{z}_2. \end{aligned}$$

If $i \notin \text{supp}(\mathbf{z})$, then $\mathbf{z}_2 = \mathbf{0}$, so

$$\mathbf{0} = A_1\mathbf{z}_1.$$

The matrix A_1 is invertible, so

$$\begin{aligned} \mathbf{z}_1 &= A_1^{-1}\mathbf{0} \\ &= \mathbf{0}. \end{aligned}$$

Then $\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{0}$ and so $\mathbf{z} = \mathbf{0}$, contradicting our assumption. Therefore, $i \in \text{supp}(\mathbf{z})$. Then $\mathbf{z}^{(i)} \neq 0$. Let $\mathbf{v} = \mathbf{u}_{\mathbb{Z}}^{(i)}\mathbf{z} - \mathbf{z}^{(i)}\mathbf{u}_{\mathbb{Z}}$. This vector $\mathbf{v}$ is a linear combination of kernel elements, so it must also be in the kernel. The subtraction cancels the ith elements and $\text{supp}(\mathbf{v}) \subseteq \text{supp}(A_1)$. Then $A_1\mathbf{v}_1 = \mathbf{0}$, and so $\mathbf{v} = \mathbf{0}$. Then $\text{supp}(\mathbf{z}) = \text{supp}(\mathbf{u}_{\mathbb{Z}})$, and so $\mathbf{u}_{\mathbb{Z}} \in \mathcal{C}_A$.

We now need to show that $\mathcal{C}_A \subset \mathcal{U}_{\mathbb{Z}}$. Diaconis and Sturmfels give a formula [18, page 15] that produces integer multiples of every circuit of an $n \times r$ integer matrix A. The formula is

$$k\mathbf{c}_{\tau} = \sum_{i=1}^{n+1} (-1)^i \det(A_{\tau \setminus \{\tau_i\}})\mathbf{e}_{\tau_i}. \tag{3.4.1}$$

The variable τ ranges over each $(n+1)$-element subset $\{\tau_1, \ldots, \tau_{n+1}\}$ of $\{1, \ldots, r\}$, and A_{σ} denotes the submatrix of A found by taking the columns with indices belonging to σ. The multiplier k is the greatest common divisor of $|\det(A_{\tau \setminus \{\tau_i\}})|$ for each $\tau_i \in \tau$.

Let $\mathbf{c} \in \mathcal{C}_A$, so it can be written in the form in equation (3.4.1). Then $|\text{supp}(\mathbf{c})| \leq n+1$. Choose some $\tau_k \in \tau$ such that $c_{\tau_k} \neq 0$. Since c_{τ_k} is non-zero, $\det(A_{\tau \setminus \{\tau_k\}})$ must be non-zero and so the columns of $A_{\tau \setminus \{\tau_k\}}$ are linearly independent.

Partition A such that $A_1 = A_{\tau \setminus \{\tau_k\}}$ and find the matrix U whose columns are the induced lattice basis. The column $\mathbf{u}_{\tau_k}$ represents a move in the τ_k dimension in this projection and has $\text{supp}(\mathbf{u}_{\tau_k}) \subseteq \tau$. The corresponding scaled column partition lattice basis element is found by multiplying this column by the least common multiple m of its denominators. This produces an integer vector $m\mathbf{u}_{\tau_k}$ whose entries are relatively prime and whose other entries are determined by the τ_kth entry, and so $m\mathbf{u}_{\tau_k}$ can only be $\mathbf{c}$. □

The elements of a column partition lattice basis correspond to the circuits in the following way:

Remark 3.4.12. The elements of a column partition lattice basis are circuits that are scaled so that the entry corresponding to a column of the A_2 partition is 1.

Circuits as polytope edges

One of the defining features of circuits is that their support is minimal by inclusion. For any underlying polytope or projected polytope, a vector directed along an edge of the polytope is a multiple of a circuit of the configuration matrix [9].

Example 3.4.13. Consider the three-link linear network. The configuration matrix is

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

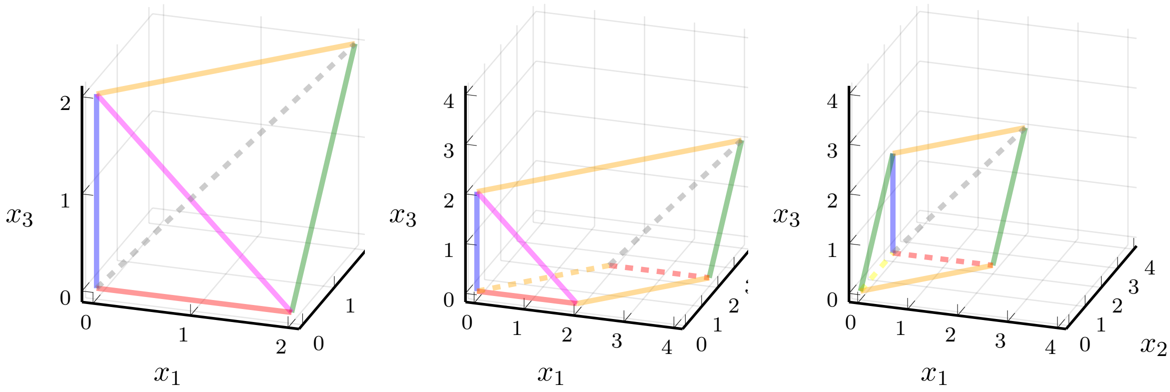
and the circuits are given by

$$\mathcal{C}_A = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}.$$

If any underlying polytope for this configuration matrix is projected onto the x_1, x_2, and x_3 co-ordinate subspace, the edges are given by the x_1, x_2, and x_3 parts of the circuits. The edges are therefore given by the vectors

$$\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Several potential underlying polytopes for A with different values of $\mathbf{y}$ are shown projected onto the x_1, x_2 and x_3 axes in Figure 3.7. These are the polytopes that appear in Example 3.3.5. Note that a circuit may appear as an edge more than once in a polytope, or it may not appear at all.



(a) $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$. (b) $\mathbf{y} = \begin{bmatrix} 4 & 4 & 2 \end{bmatrix}^\intercal$. (c) $\mathbf{y} = \begin{bmatrix} 2 & 4 & 2 \end{bmatrix}^\intercal$.

Figure 3.7: The edges of some of the underlying projected polytopes in Example 3.4.13.

3.5 Unimodularity

In this thesis we are particularly interested in *unimodular* matrices. Unimodular matrices occur frequently in statistical inverse problems: the configuration matrix for a two-way contingency table is unimodular, as are the link-path incidence matrices of many traffic networks (see Section 5.2, below). More examples are given in Section 1.5. In this section we look at some properties of column partition lattice bases of unimodular configuration matrices. In doing so, we demonstrate the effect of some of the theorems in Sections 3.3 and 3.4 on unimodular configuration matrices. We begin with the definition of a unimodular matrix.

Definition 3.5.1 (Unimodular matrix)**.** A square matrix is *unimodular* if its determinant is ± 1.

It is common to extend the definition of unimodularity to rectangular matrices: a rectangular matrix is unimodular if each of its invertible maximal square submatrices is unimodular.

3.5.1 Total unimodularity of U

Airoldi [2] and Hazelton [25] noted that for a particular class of unimodular matrices called *totally unimodular* matrices, any column partition lattice basis matrix will contain all entries in $\{0, \pm 1\}$, and will itself be totally unimodular.

Definition 3.5.2 (Total unimodularity)**.** A matrix is *totally unimodular* if the determinant of every invertible square submatrix is ± 1.

Totally unimodular matrices have some useful properties, listed for example in Schrijver [40]. We state these properties in Theorem 3.5.4. We will require the definition of an *Eulerian* matrix:

Definition 3.5.3 (Eulerian matrix)**.** A matrix $A \in \{0, \pm 1\}^{n \times r}$ is *Eulerian* if for each row and column, the sum of the entries is a multiple of two.

Theorem 3.5.4. *Let A be a matrix with all entries in $\{0, \pm 1\}$. Then the following statements are equivalent.*

1. *A is totally unimodular.*
2. *Every square Eulerian submatrix of A is singular (Camion [11]).*
3. *The sum of the entries of each square Eulerian submatrix of A is divisible by four (Camion [11]).*
4. *For every subset $\mathcal{A}$ of the columns of A, each column $\mathbf{a} \in \mathcal{A}$ can be assigned a multiplier $\epsilon_{\mathbf{a}} \in \{\pm 1\}$ such that*

   $$\sum_{\mathbf{a} \in \mathcal{A}} \epsilon_{\mathbf{a}} \mathbf{a} \in \{0, \pm 1\}^n$$

   (Ghouila-Houri [23]).
5. *The matrix $\begin{bmatrix} I & A \end{bmatrix}$ is unimodular.*

We claim that if a configuration matrix is unimodular, then any column partition lattice basis matrix will be totally unimodular and contain only $\{0, \pm 1\}$ entries. In fact, this property holds when all non-zero $n \times n$ minors of A have the same absolute value.

Theorem 3.5.5. *Let A be a configuration matrix, and suppose that the determinants of all maximal invertible submatrices of A are $\pm d$. Then the matrix*

$$U = \begin{bmatrix} -A_1^{-1} A_2 \\ I \end{bmatrix}$$

is totally unimodular.

Proof. We first show that $A_1^{-1} A_2$ is totally unimodular. Theorem 3.4.1 says that the determinant of a submatrix M of $A_1^{-1} A_2$ is given by

$$\det(M) = (-1)^p \frac{\det(A_1^{\mathcal{I} \setminus \mathcal{J}})}{\det(A_1)},$$

where

$$p = \frac{k(k-1)}{2} + \sum_{i \in \mathcal{I}} (n - i).$$

So the determinant of M is equal to plus or minus the ratio of the determinants of two maximal submatrices of the configuration matrix A. These determinants are all $\pm d$, so this ratio is ± 1. Therefore the determinant of any submatrix of $A_1^{-1} A_2$ is ± 1.

The matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}$$

is also totally unimodular since appending an identity matrix does not affect total unimodularity. □

We can also prove Theorem 3.5.5 using the fifth property of totally unimodular matrices listed in Theorem 3.5.4.

Alternative proof of Theorem 3.5.5. The maximal invertible square submatrices of A all have determinant $\pm d$. This includes A_1, and so multiplying A by A_1^{-1} has the effect of dividing the determinants of the maximal submatrices of A by $\pm d$. The invertible maximal submatrices of A all have determinant $\pm d$, and so the Hermite normal form $\begin{bmatrix} I & A_1^{-1}A_2 \end{bmatrix}$ is unimodular. Property 5 of totally unimodular matrices from Theorem 3.5.4 says that a matrix A is totally unimodular if the matrix $\begin{bmatrix} I & A \end{bmatrix}$ is unimodular. Substituting $A_1^{-1}A_2$ for A shows that $A_1^{-1}A_2$ is totally unimodular, and it follows that U is totally unimodular. □

We illustrate this with an example.

Example 3.5.6. Consider the configuration matrix

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

The determinants of the invertible maximal submatrices are all ± 2. This includes A_1 in the following partition, we will use to construct our column partition lattice basis:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix A_1^{-1} has all entries $\pm\frac{1}{2}$:

$$A_1^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Every entry in the matrix $A_1^{-1}A_2$ is equal to $\pm\frac{1}{2} \pm \frac{1}{2}$, so the matrix

$$U = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

has all entries equal to 0 or ± 1. This matrix is also totally unimodular.

3.5.2 Circuits and the Graver basis

We saw in Section 3.4.3 how the elements of a column partition lattice basis correspond to circuits of the configuration matrix: each column partition basis element is a circuit that has been scaled so that the entry corresponding to a column in A_2 under the current partition is 1. If the $n \times n$ minors of A are the same in absolute value, these entries are already 1, and so the set of circuits is equal to the column partition lattice bases.

Theorem 3.5.7. *Let A be a configuration matrix of rank n, and suppose the non-zero $n \times n$ minors of A are the same in absolute value. Let Π be the set of column partitions of A such that A_1 is square and invertible. Then*

$$\mathcal{C}_A = \bigcup_{\pi \in \Pi} U^\pi.$$

Proof. By Theorem 3.4.11,

$$\mathcal{C}_A = \bigcup_{\pi \in \Pi} U^\pi_{\mathbb{Z}}.$$

By Theorem 3.5.5, $U \in \{0, \pm 1\}^{r \times (r-n)}$, and so $U = U_{\mathbb{Z}}$. It follows that

$$\mathcal{C}_A = \bigcup_{\pi \in \Pi} U^\pi.$$

□

Example 3.5.8. Consider again the system from Example 3.5.6, where

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

The column partition $\pi = (\{1, 2, 3, 4\}, \{5, 6, 7\})$ induces the column partition lattice basis

$$U^\pi = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The set of circuits for this system is given by

$$\mathcal{C}_A = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\}.$$

Consider the column partitions

$$\pi_1 = (\{2,3,4,5\},\{1,6,7\})$$
$$\pi_2 = (\{3,4,5,6\},\{1,2,7\}).$$

The column partition lattice bases they induce are

$$U^{\pi_1} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \qquad U^{\pi_2} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

We can see that the union of $U^{\pi}, U^{\pi_1}, U^{\pi_2}$ contains every circuit.

Any other column partition lattice basis is a collection of these (possibly negated) vectors. For example, $\pi_3 = (\{4,5,6,7\},\{1,2,3\})$ induces

$$U^{\pi_3} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix},$$

which takes the third, fifth, and sixth elements of $\mathcal{C}_A$ as listed above.

Theorem 3.5.7 states that for configuration matrices whose $n \times n$ minors are the same in absolute value, the union of the column partition lattice bases is equal to the set of circuits of A. We can combine this fact usefully with the following property of such matrices given by Sturmfels [42].

Theorem 3.5.9. *Let A be a configuration matrix. If the non-zero $n \times n$ minors of A are the same in absolute value, then the set of circuits $\mathcal{C}_A$ equals the Graver basis $\mathcal{G}_A$.*

Together, these imply the following:

Theorem 3.5.10. *Let A be a configuration matrix such that all the non-zero $n \times n$ minors of A are the same in absolute value. Then the union of the column partition lattice bases is equal to the Graver basis:*

$$\mathcal{G}_A = \bigcup_{\pi \in \Pi} U^{\pi},$$

where Π is the set of column partitions of A such that A_1 is invertible.

It is not true in general that taking the union of the column partition lattice bases produces the Graver basis, as illustrated by the following example.

Example 3.5.11. Consider the matrix

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

This matrix has fifteen maximal submatrices of which four are singular and nine are unimodular; the other two have determinant ± 2. The circuits of this matrix are given by the columns of the matrix

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

The vector $\begin{bmatrix} -1 & 1 & -1 & -1 & 1 & 1 \end{bmatrix}^{\intercal}$ is also in the integer kernel, but can not be conformally composed of circuits. Therefore the set of circuits of A does not include all of the elements of the Graver basis of A.

The Graver basis and the adaptive sampler

Theorem 3.5.10 states that for a unimodular configuration matrix, the union of the column partition lattice bases is the Graver basis. The Graver basis is known to be a Markov basis [3].

One application for this fact can be found in the dynamic lattice basis sampler of Hazelton et al. [27]. This sampler uses a column partition lattice basis to sample from the fibre, but frequently changes which column partition is used to generate the lattice basis. This is done in response to the sampler's current position in the fibre. All column partitions have a non-zero probability of being selected, so this sampler can generate all column partition lattice bases. If the configuration matrix is unimodular, then all Graver basis elements are accessible by the sampler, and so irreducibility of the Markov chain is guaranteed. This methodology can be tailored to select geometrically advantageous column partition bases with high probability, hence facilitating quick mixing [27].

3.5.3 The Graver basis

Any element of the integer kernel of a configuration matrix can be expressed in terms of a column partition lattice basis. These elements include other circuits. One of the themes we explore later in this thesis is that of using a column partition lattice basis that is not necessarily a Markov basis to simulate moves in a known Markov basis. In Chapter 4 we use this idea to prove that a column partition lattice basis is a Markov basis by showing that it can simulate each move in a Graver basis. In Chapter 6, we use the idea of simulating a Graver basis move when considering random walks that may visit particular points outside of the fibre in order to visit every point in the fibre.

In this section we examine how Graver basis elements of a unimodular matrix can be expressed in terms of a column partition lattice basis. In doing so, we rely on the fact that for a unimodular matrix, each element of the Graver basis requires at most one copy of each element of any column partition lattice basis:

Theorem 3.5.12. *Let A be an $n \times r$ unimodular configuration matrix and let the columns of U define a column partition lattice basis U. Each element of $\mathcal{G}_A$, the Graver basis of A, is a sum of at most one signed copy of each element of U.*

Proof. The basis U is a column partition lattice basis, so the entries of any element $\mathbf{z}$ of $\ker_{\mathbb{Z}}(A)$ corresponding to the columns of A in the A_2 partition tell us which lattice basis elements need to be combined to get $\mathbf{z}$. Let $\mathbf{g}$ be an element of the Graver basis of A. Then since A is unimodular, all entries of $\mathbf{g}$ are in $\{0, \pm 1\}$. It is therefore as sum of at most one signed copy of each element of U, as required. □

This is not in general true for all configuration matrices as illustrated by the following example.

Example 3.5.13. Consider the configuration matrix

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

whose entries are all zero or one. A column partition lattice basis is given by

$$U = \begin{bmatrix} 4 & -1 \\ 0 & 1 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \\ 1 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The vector

$$\mathbf{c} = \begin{bmatrix} 0 & 4 & -1 & -1 & -1 & 1 & -1 & -4 \end{bmatrix}^{\mathsf{T}}$$

is circuit of A that is not in U, and the vector

$$\mathbf{g} = \begin{bmatrix} 2 & 2 & -1 & -1 & -1 & 1 & -1 & -2 \end{bmatrix}^{\mathsf{T}}$$

is in the Graver basis but is not a circuit. Expressing these two vectors in terms of U, we have $\mathbf{c} = \mathbf{u}_1 + 4\mathbf{u}_2$, and $\mathbf{g} = \mathbf{u}_1 + 2\mathbf{u}_2$. In each case we need more than one copy of some element of U to construct the required vector.

In fact, we cannot express $\mathbf{g}$ as a sum of plus or minus one copy of each vector in a subset of circuits of A either. The set of circuits of A is $\mathcal{C}_A = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{c}\}$. We have

$$\begin{aligned}\mathbf{g} &= \mathbf{u}_1 + 2\mathbf{u}_2 \\ &= \mathbf{c} - 2\mathbf{u}_2 \\ &= \frac{\mathbf{u}_1 + \mathbf{c}}{2}.\end{aligned}$$

The first property of Graver bases of unimodular configuration matrices that we will demonstrate is to show that if a signed sum of a certain combination of column partition lattice basis elements produces some Graver basis element, then a different signed sum of the same combination cannot produce a different Graver basis element.

Theorem 3.5.14. *Let A be an $n \times r$ unimodular configuration matrix and let U be a column partition lattice basis. Suppose $\mathbf{g} \in \mathcal{G}_A$ is a signed sum of some subset $U_\mathbf{g}$ of U, so that*

$$\mathbf{g} = \sum_{\mathbf{u} \in U_\mathbf{g}} \epsilon_\mathbf{u} \mathbf{u}$$

where $\epsilon_\mathbf{u} \in \{\pm 1\}$ for all $\mathbf{u} \in U_\mathbf{g}$. Then no other signed sum of the elements of $U_\mathbf{g}$ has all entries 0 or ± 1, and therefore none of them is in $\mathcal{G}_A$.

Proof. Without loss of generality, let $\epsilon_\mathbf{u} = 1$ for all $\mathbf{u} \in U_\mathbf{g}$. Let

$$\mathbf{z} = \sum_{\mathbf{u} \in U_\mathbf{g}} \mu_\mathbf{u} \mathbf{u}$$

for some other collection of multipliers $\mu_\mathbf{u}$ for each $\mathbf{u} \in U_\mathbf{g}$ with not all $\mu_\mathbf{u} = \epsilon_\mathbf{u}$. We need to show that $\mathbf{z} \notin \{0, \pm 1\}^r$.

We sum the elements of $U_\mathbf{g}$ with positive $\mu_\mathbf{u}$, and those with negative $\mu_\mathbf{u}$, to obtain

$$\mathbf{v} = \sum_{\mathbf{u}:\mu_\mathbf{u}=1} \mathbf{u} \quad \text{and} \quad \mathbf{w} = \sum_{\mathbf{u}:\mu_\mathbf{u}=-1} \mathbf{u},$$

so that $\mathbf{g} = \mathbf{v} + \mathbf{w}$ and $\mathbf{z} = \mathbf{v} - \mathbf{w}$.

From the definition of the Graver basis $\mathbf{v} + \mathbf{w}$ cannot be a conformal decomposition of $\mathbf{g}$, so there is some index i where the entries of $\mathbf{v}$ and $\mathbf{w}$ have opposite signs. When we take the difference $\mathbf{z} = \mathbf{v} - \mathbf{w}$, at index i there is a sum of two elements with the same sign, so $|z_i| > 1$ and $\mathbf{z} \notin \{0, \pm 1\}$. Therefore $\mathbf{z} \notin \mathcal{G}_A$, as required. □

For unimodular matrices, the Graver basis is equal to the set of circuits. The circuits are defined by having minimal support: for an $n \times r$ configuration matrix, each circuit has at most $n+1$ non-zero elements. We can use this fact to give a limit on the number of column partition basis elements we need to combine to produce any Graver basis element.

Theorem 3.5.15. *Let A be an $n \times r$ unimodular configuration matrix and let U be a column partition lattice basis. Let $\mathbf{g} \in \mathcal{G}_A$, and let $U_\mathbf{g}$ be the columns of U such that*

$$\mathbf{g} = \sum_{\mathbf{u} \in U_\mathbf{g}} \epsilon_\mathbf{u} \mathbf{u}$$

where each $\epsilon_\mathbf{u} \in \{\pm 1\}$. Then $U_\mathbf{g}$ has at most $\min(r-n, n+1)$ elements.

Proof. The set $U_\mathbf{g}$ is a subset of U, which has $r-n$ elements. Each element is used at most once, so $\mathbf{g}$ is a sum of at most $r-n$ elements of U.

The matrix A is unimodular, so by Theorem 3.5.9, $\mathbf{g}$ is a circuit and has at most $n+1$ non-zero elements. The basis U is a column partition lattice basis, so the elements of $\mathbf{g}$ corresponding to the A_2 partition of A give co-ordinates for expressing $\mathbf{g}$ in terms of U. This means that $\mathbf{g}$ is a sum of at most $n+1$ elements of U.

Combining these two upper limits means that $\mathbf{g}$ is a sum of at most $\min(r-n, n+1)$ elements of U, as required. □

Taken together, these three theorems mean that for a unimodular configuration matrix A, the elements of $\mathcal{G}_A$ can be found with the following method:

1. Take a subset of at most $\min(r-n, n+1)$ elements of U.
2. Check if the collection of multipliers $\epsilon = \pm 1$ of these elements that give a vector in $\{0, \pm 1\}^r$ is unique.
3. If so, this vector is an element of $\mathcal{G}_A$.

This also gives an upper bound for the size of $\mathcal{G}_A$.

Corollary 3.5.16. *Let $A \in \{0,1\}^{n \times r}$ be a unimodular configuration matrix, and let $m = \min(r-n, n+1)$. Then*

$$|\mathcal{G}_A| \leq \sum_{i=1}^{m} \binom{r-n}{i}.$$

3.5.4 Algebra

In Section 2.4 we saw how elements of $\ker_\mathbb{Z}(A)$ for a configuration matrix A can be represented as monomial differences. This includes elements of a column partition lattice basis whose elements are integral, for example a basis for any unimodular configuration matrix.

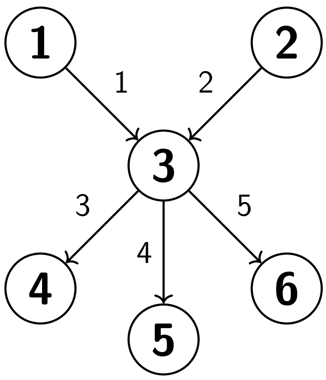


Figure 3.8: The graph from the transport network in Example 3.5.17.

Let $U \in \mathbb{Z}^{r\times(r-n)}$ be a column partition lattice basis for A, and let

$$F = \{T^{\mathbf{u}^+} - T^{\mathbf{u}^-} : \mathbf{u} \in U\}$$

be the monomial difference representations of U. This set F generates an ideal, $I_U \subseteq I_A$, with equality when U is a Markov basis.

The ideal I_U has a Gröbner basis. Recall from Definition 2.4.10 that a Gröbner basis is a subset $G = \{g_1, \ldots, g_n\}$ of an ideal I such that

$$\langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_n)\rangle = \langle \mathrm{LT}(I)\rangle.$$

An interesting question is: is F necessarily a Gröbner basis for I_U? It turns out that this is not the case, as shown by the following counterexample.

Example 3.5.17. Consider the traffic network shown on the graph in Figure 3.8 which has the unimodular configuration matrix

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

In this network, nodes 1 and 2 may be used as origins for traffic, and nodes $3, 4, 5$ and 6 may function as destinations.

A column partition lattice basis is given by

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

The monomial difference representations of U are given by

$$F = \{t_1t_3 - t_2t_6, t_1t_4 - t_2t_7, t_1t_5 - t_2t_8\}.$$

We claim that F is not a Gröbner basis for I_U under any term ordering.

There are eight indeterminates, so even considering only lex order there are 8! possible term orderings; fortunately we need not check them all. For any term ordering, the lead term of $T^{\mathbf{u}^+} - T^{\mathbf{u}^-}$ is either $T^{\mathbf{u}^+}$ or $T^{\mathbf{u}^-}$. There are three monomial differences in F, and therefore eight combinations of potential leading terms. We will show that for each of the eight, there is a polynomial $p \in I_U$ such that $\mathrm{LT}(p)$ is not divisible by $\mathrm{LT}(f)$ for any $f \in F$. This implies that F is not a Gröbner basis under any term ordering.

The polynomials we require are S-polynomials of pairs of elements

$$T^{\mathbf{u}_i^+} - T^{\mathbf{u}_i^-}, T^{\mathbf{u}_j^+} - T^{\mathbf{u}_j^-} \in F$$

such that

$$\gcd\left(\mathrm{LT}(T^{\mathbf{u}_i^+} - T^{\mathbf{u}_i^-}), \mathrm{LT}(T^{\mathbf{u}_j^+} - T^{\mathbf{u}_j^-})\right) \neq 1.$$

Recall from Definition 2.4.11 that the S-polynomial of two polynomials f and g in $k[T]$ under a given term ordering is given by

$$S(f,g) = \frac{T^\gamma}{\mathrm{LT}(f)} f - \frac{T^\gamma}{\mathrm{LT}(g)} g,$$

where T^γ is the least common multiple of $\mathrm{LT}(f)$ and $\mathrm{LT}(g)$.

The combinations of leading terms of elements of F and the required S-polynomials are given in the following table:

$t_1t_3 - t_2t_6$	$t_1t_4 - t_2t_7$	$t_1t_5 - t_2t_8$	S-polynomial
t_1t_3	t_1t_4	t_1t_5	$S(t_1t_3 - t_2t_6, t_1t_4 - t_2t_7) = t_2t_3t_7 - t_2t_4t_6$
t_1t_3	t_1t_4	$-t_2t_8$	$S(t_1t_3 - t_2t_6, t_1t_4 - t_2t_7) = t_2t_3t_7 - t_2t_4t_6$
t_1t_3	$-t_2t_7$	t_1t_5	$S(t_1t_d - t_2t_6, -t_2t_8 + t_1t_5) = t_2t_3t_8 - t_2t_5t_6$
t_1t_3	$-t_2t_7$	$-t_2t_8$	$S(-t_2t_7 + t_1t_4, -t_2t_8 + t_1t_5) = t_1t_4t_8 - t_1t_5t_7$
$-t_2t_6$	t_1t_4	t_1t_5	$S(t_1t_4 - t_2t_7, t_1t_5 - t_2t_8) = t_2t_4t_8 - t_2t_5t_7$
$-t_2t_6$	t_1t_4	$-t_2t_8$	$S(-t_2t_6 + t_1t_3, -t_2t_8 + t_1t_5) = t_1t_3t_8 - t_1t_5t_6$
$-t_2t_6$	$-t_2t_7$	t_1t_5	$S(-t_2t_6 + t_1t_3, -t_2t_7 + t_1t_4) = t_1t_3t_7 - t_1t_4t_6$
$-t_2t_6$	$-t_2t_7$	$-t_2t_8$	$S(-t_2t_6 + t_1t_3, -t_2t_7 + t_1t_4) = t_1t_3t_7 - t_1t_4t_6$

We can see that no term in any of the given S-polynomials is divisible by the leading term of an element of F; and so no matter the term ordering, there is an S-polynomial whose terms are not divisible by the leading term of an element of F. Therefore F is not a Gröbner basis.

Chapter 4

Connectivity

4.1 Introduction

In this chapter we present some results on the main theme of this thesis, which is whether or not a column partition lattice basis is a Markov basis. Our main result is a condition on the matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}$$

that guarantees that U is a Markov basis. The statement and proof of this are in Section 4.4.4.

Other results in this chapter include a stronger condition on the matrix U that also guarantees that it is a Markov basis, which has other advantages and can be found in Section 4.2; and a simpler method for showing that a set of moves is a Markov basis, which can be found in Section 4.3. We also collect some results on potentially using collections of column partition lattice bases as a Markov basis, which can be found in Section 4.5.

In this chapter we assume an integral U matrix. This is guaranteed for unimodular configuration matrices. Non-unimodular configuration matrices may also have column partition lattice basis matrices that are integral — a condition that guarantees this for non-unimodular configuration matrices was given previously in Theorem 3.4.5. Having an integral U avoids the problem of parity errors described in Section 2.5.1.

The first result, presented in Section 4.2, gives a condition on the matrix U that guarantees that it is a Markov basis. This condition enjoys the advantage of translating simply into a condition on the column partition used, so it also provides a method of partitioning that produces such a U matrix. However, the availability of such a partition is not always guaranteed. This result generalises a result that was previously found by Schofield and Bonner [39].

Section 4.3 gives a simpler condition for a set of moves to be a Markov basis that can be used when some other Markov basis is known. From the definition of a Markov basis, a set $\mathcal{B}$ is a Markov basis for a configuration matrix A if $\mathcal{B}$ connects all $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ for all $\mathbf{y} \geq \mathbf{0}$. This is implied if $\mathcal{B}$ connects $\mathbf{z}^-$ and $\mathbf{z}^+$ for all $\mathbf{z} \in \mathcal{M}$, where $\mathcal{M}$ is some known Markov basis. We can think of this as being able to use the moves in $\mathcal{B}$ to simulate any move in $\mathcal{M}$. This will be the case if the geometry of the $\mathbb{Z}$-polytope is such that whenever it contains two points separated by $\mathbf{z}$, it is guaranteed that there is enough wiggle room to use the required moves in $\mathcal{B}$ to travel between these points.

If no known Markov basis is available, we may instead consider applying the idea to general elements of the integer kernel. The idea of simulating moves in a known Markov basis is revisited in Section 6.2.2.

Markov basis simulation is used in Section 4.4 to give a weaker condition on the matrix U matrix for a column partition lattice basis being a Markov basis. This result is based on the understanding of the geometry of projected $\mathbb{Z}$-polytopes representing fibres developed in Section 3.3. We conjecture that for unimodular configuration matrices, this condition on U is equivalent to U being a Markov basis.

Finally, Section 4.5 uses results from Section 3.5 concerning unimodular configuration matrices, circuits, and Graver bases to investigate how best to combine column partition lattice bases to get a Markov basis.

4.2 Column sums

One type of column partition lattice basis that is a Markov basis is that defined by some matrix

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}$$

where $-A_1^{-1}A_2 \leq 0$ and integral, and the inequality is componentwise. We state this as a theorem.

Theorem 4.2.1. *Let $A \in \{0,1\}^{n\times r}$ be a configuration matrix of rank n, and let $r > n$. Let the columns of A be a partitioned so that each of the $r-n$ columns of A_2 can be written as a non-negative combination of the n columns of A_1. Let U, the induced column partition lattice basis, be integral. Then U is a Markov basis.*

In this section we give a proof of this fact.

A condition for the entries of U being integral can be found in Theorem 3.4.5. The entries of U are non-negative if the column partition is such that each column in the A_2 part of A is a non-negative combination of columns of A_1. This result is a generalisation of a result of Schofield and Bonner [39], who proved the case where A_1 is the identity matrix.

Schofield and Bonner made use of their result in capture-recapture modelling. It can also often be applied in network tomography: for example, if on some traffic network, each link in the network is an allowed path, then the identity matrix is a maximal submatrix and the condition in Theorem 4.2.1 holds. This and other applications are discussed in Section 4.2.5.

We give here an example of this theorem in action using the link-path incidence matrix of the three-link linear network.

Example 4.2.2. Let

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

The columns of A are partitioned such that A_1 is the identity matrix. This induces the column partition lattice basis defined by the columns of

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

This matrix is integral and the $-A_1^{-1}A_2$ part is non-positive, so by Theorem 4.2.1, it defines a Markov basis.

Using the techniques of Diaconis and Sturmfels [19] from Section 2.4, a Gröbner basis for I_A under lex term ordering and $t_6 > t_5 > t_4 > t_3 > t_2 > t_1$ is given by

$$\{t_6 - t_1t_2t_3, t_5 - t_2t_3, t_4 - t_1t_2\}.$$

This is the monomial difference representation of U, above. This is also the Markov basis for A given by `4ti2` [44].

4.2.1 A non-negative $A_1^{-1}A_2$

The first step in the proof of Theorem 4.2.1 is showing that it if the condition on the columns of A_1 and A_2 holds, then A_1 is invertible and we can therefore use the partition to make a column partition lattice basis. This is stated in the following lemma.

Lemma 4.2.3. *Let A be an $n \times r$ matrix of full rank where $r > n$. Let the columns of A be partitioned such that each of the $r - n$ columns in A_2 can be written as a linear combination of the n columns of A_1. Then A_1 is invertible.*

Proof. The matrix A is $n \times r$ and of full rank, so $\text{rank}(A) = n$. The matrix A_1 is $n \times n$. The columns in A_2 are each a linear combination of columns in A_1, so columns of A_1 span $\text{CS}(A)$. This implies $\text{rank}(A_1) = \text{rank}(A) = n$, and so A_1 is invertible. □

The following example gives two partitions of a configuration matrix for which the condition does and does not hold.

Example 4.2.4. Let A be the link-path incidence matrix of the three-link linear network. Then

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

This matrix is of full rank. Choosing the partition so that

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

means that each column of A_2 can be written in terms of columns of A_1. In order,

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

The matrix A_1 is invertible.

Choosing instead the partition which switches the two parts, so that

$$A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

means that no column in A_2 can be written in terms of columns of A_1, and A_1 is not invertible.

If A is partitioned so that the conditions of Theorem 4.2.1 hold, then A_1 is invertible, and we can use the partition to create a column partition lattice basis for A. Theorem 4.2.1 requires that each column of A_2 is a non-negative combination of columns of A_1. The following lemma implies that if this condition is met, then the partition induces a column partition lattice basis such that the $-A_1^{-1}A_2$ part of U is non-positive.

Lemma 4.2.5. *Let $A_{\geq 0}^{n \times r}$ be of full rank, and let $r > n$. Let the columns of A be partititioned such that each column of A_2 is a non-negative combination of columns of A_1. Then $A_1^{-1}A_2$ is non-negative.*

Proof. Let $\mathbf{a}_i$ mean the ith column of A_2 under the column partition. Each such column of A_2 can be written as a non-negative combination of columns of A_1. We can therefore write $\mathbf{a}_i = A_1\mathbf{c}_i$, where $\mathbf{c}_i$ is a non-negative vector that creates the linear combination of columns of A_1 that sum to $\mathbf{a}_i$. The matrix $C = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_{r-n}\end{bmatrix}$ is therefore non-negative. We have

$$\begin{aligned} A_1C &= A_2 \\ A_1^{-1}A_1C &= A_1^{-1}A_2 \\ C &= A_1^{-1}A_2, \end{aligned}$$

and so all of the entries of $A_1^{-1}A_2$ are non-negative too. □

As an example, consider the link-path incidence matrix of the three-link linear network.

Example 4.2.6. We partition the columns of

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

so that A_1 is the identity matrix and

$$A_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Each column of A_2 is a non-negative combination of the columns of A_1. Clearly $A_1^{-1}A_2 = A_2$, which is non-negative.

This example was quite simple in that the A_1 part was the identity matrix; here is a non-trivial example.

Example 4.2.7. Let

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

This is the link-path incidence matrix for a six link linear network in which the first four nodes function as origins for traffic and the last five function as destinations (so that two nodes function as both). Partitioning the columns so that A_1 is made of the first six columns of A, we have

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and the remaining columns make up the A_2. Each of the columns of A_2 is a sum of columns in A_1. Then

$$\begin{aligned} A_1^{-1}A_2 &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

This matrix $A_1^{-1}A_2$ is non-negative.

The following lemma says that if some point $\mathbf{x} \in \mathcal{F}_\mathbf{y}$, then in the projected space all the points that lie between $\mathbf{x}$ and the origin are also in $\mathcal{F}_\mathbf{y}$.

Lemma 4.2.8. *Let A be a configuration matrix, and let the column partition into A_1 and A_2 be such that $A_1^{-1}A_2$ has all non-negative integer entries. Let $\mathbf{x} \in \mathcal{F}_\mathbf{y}$ for some $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, and let $\mathbf{x}$ be partitioned into $\mathbf{x}_1, \mathbf{x}_2$ according to the column partition of A. Let $\mathbf{m}_2 \in \mathbb{Z}^{r-n}_{\geq 0}$ be such that $\mathbf{0} \leq \mathbf{m}_2 \leq \mathbf{x}_2$, and let*

$$\mathbf{m} = \begin{bmatrix} \mathbf{x}_1 + A_1^{-1}A_2(\mathbf{x}_2 - \mathbf{m}_2) \\ \mathbf{m}_2 \end{bmatrix}.$$

Then $\mathbf{m} \in \mathcal{F}_\mathbf{y}$.

Proof. Let $\mathbf{m}_2$ be given, and set $\mathbf{m}_1 = \mathbf{x}_1 + A_1^{-1}A_2(\mathbf{x}_2 - \mathbf{m}_2)$. Each of $\mathbf{x}_1, A_1^{-1}A_2$, and

$\mathbf{x}_2 - \mathbf{m}_2$ are non-negative and integral, so $\mathbf{m}_1$ is non-negative and integral. We have

$$\begin{aligned} A\mathbf{m} &= A_1\mathbf{m}_1 + A_2\mathbf{m}_2 \\ &= A_1\left(\mathbf{x}_1 + A_1^{-1}A_2(\mathbf{x}_2 - \mathbf{m}_2)\right) + A_2\mathbf{m}_2 \\ &= A_1\mathbf{x}_1 + A_2(\mathbf{x}_2 - \mathbf{m}_2) + A_2\mathbf{m}_2 \\ &= A_1\mathbf{x}_1 + A_2\mathbf{x}_2 \\ &= A\mathbf{x} \\ &= \mathbf{y}. \end{aligned}$$

Therefore, $\mathbf{m} \in \mathcal{F}_{\mathbf{y}}$. □

In particular, if $\mathbf{m}_2 = \mathbf{0}$, then

$$\mathbf{m} = \begin{bmatrix} \mathbf{x}_1 + A_1^{-1}A_2\mathbf{x}_2 \\ \mathbf{0} \end{bmatrix} \in \mathcal{F}_{\mathbf{y}}.$$

We demonstrate with the following example.

Example 4.2.9. Consider again the three-link linear network, which has configuration matrix

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

and a column partition lattice basis

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

We set $\mathbf{y} = \begin{bmatrix} 2 & 3 & 2 \end{bmatrix}^\intercal$. We have

$$\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in \mathcal{F}_{\mathbf{y}}.$$

The last three co-ordinates give $\mathbf{x}_2 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\intercal$. The fibre $\mathcal{F}_{\mathbf{y}}$ also contains the vectors

$$\left\{ \begin{bmatrix} 2 \\ 3 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\},$$

which includes all vectors $\{\mathbf{m} \in \mathbb{Z}^6 : A\mathbf{m} = \mathbf{y}\}$ such that $\mathbf{0} \leq \mathbf{m}_2 \leq \mathbf{x}_2$.

4.2.2 Distance reducing proof of Theorem 4.2.1

Theorem 4.2.1 states that if the columns of a configuration matrix A are partitioned such that each of the columns of A_2 is a non-negative combination of columns of A_1, and U is integral, then U is a Markov basis. We present proofs for this using both distance reduction and ideal membership.

Here is the distance reducing proof.

Proof. Let a configuration matrix $A \in \{0,1\}^{n\times r}$ be partitioned such that each column of A_2 is a non-negative combination of columns of A_1. By Lemmata 4.2.3 and 4.2.5, this partition induces a column partition lattice basis

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix},$$

where $-A_1^{-1}A_2 \leq 0$. Let this U be integral. We will show that U is a Markov basis by taking a pair of arbitrary points $\mathbf{a}, \mathbf{b}$ in an arbitrary fibre and showing that moves in U can be used to reduce the distance from either point to another point $\mathbf{m}$, where $\mathbf{m}_2 = \min(\mathbf{a}_2, \mathbf{b}_2)$.

Let $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be given, and let $\mathbf{a}, \mathbf{b}$ be distinct points in $\mathcal{F}_\mathbf{y}$. By Lemma 4.2.8, the point

$$\mathbf{m} = \begin{bmatrix} \mathbf{a}_1 + A_1^{-1}A_2(\mathbf{a}_2 - \mathbf{m}_2) \\ \mathbf{m}_2 \end{bmatrix} \in \mathcal{F}_\mathbf{y}.$$

The moves required to get from $\mathbf{a}$ to $\mathbf{m}$ using U are given by

$$\mathbf{m} - \mathbf{a} = \sum_{i=1}^{r-n} (\mathbf{m} - \mathbf{a})_{n+i}\mathbf{u}_i.$$

For the distance measurement we use the L_1 norm in the projected co-ordinates,

$$\mathrm{d}(\mathbf{a}, \mathbf{b}) = \sum_{i=n+1}^{r} |\mathbf{a} - \mathbf{b}|_i .$$

This is the number of steps in a direct walk using between the two points using U. Because $\mathbf{m}_2 \leq \mathbf{a}_2$, the distance between $\mathbf{a}$ and $\mathbf{m}$ is given by

$$\mathrm{d}(\mathbf{a}, \mathbf{m}) = \sum_{i=n+1}^{r} (\mathbf{a} - \mathbf{m})_i.$$

We choose any integer $k \in \{(n+1), \ldots, r\}$ such that $(\mathbf{a} - \mathbf{m})_k \neq 0$ (i.e., $\mathbf{a}$ and $\mathbf{m}$ differ in co-ordinate k), and set $\mathbf{a}^\dagger = \mathbf{a} - \mathbf{u}_k$. By Lemma 4.2.8, $\mathbf{a}_2^\dagger \leq \mathbf{a}_2$ implies $\mathbf{a}^\dagger \in \mathcal{F}_\mathbf{y}$.

The distance between $\mathbf{a}^\dagger$ and $\mathbf{m}$ is

$$\begin{aligned} \mathrm{d}(\mathbf{a}^\dagger, \mathbf{m}) &= \sum_{i=n+1}^{r} (\mathbf{a}^\dagger - \mathbf{m})_i \\ &= \mathrm{d}(\mathbf{a}, \mathbf{m}) - 1 \end{aligned}$$

and we have reduced the distance from $\mathbf{a}$ to $\mathbf{m}$, proving that $\mathbf{a}$ is connected to $\mathbf{m}$.

The same is true for $\mathbf{b}$. Since both $\mathbf{a}$ and $\mathbf{b}$ are connected to $\mathbf{m}$, $\mathbf{a}$ and $\mathbf{b}$ are connected to each other by transitivity. The basis U is therefore a Markov basis. □

We illustrate this proof using again the example of the three-link linear network.

Example 4.2.10. Consider the configuration matrix

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

and the column partition lattice basis

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

We set $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$ and choose two points from $\mathcal{F}_\mathbf{y}$,

$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}.$$

We set $\mathbf{m}_2 = \min(\mathbf{a}_2, \mathbf{b}_2)$, so

$$\mathbf{m} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{m} &= \mathbf{a} - \mathbf{u}_2 - \mathbf{u}_3 \\ \mathbf{m} &= \mathbf{b} - 2\mathbf{u}_1 \end{aligned}$$

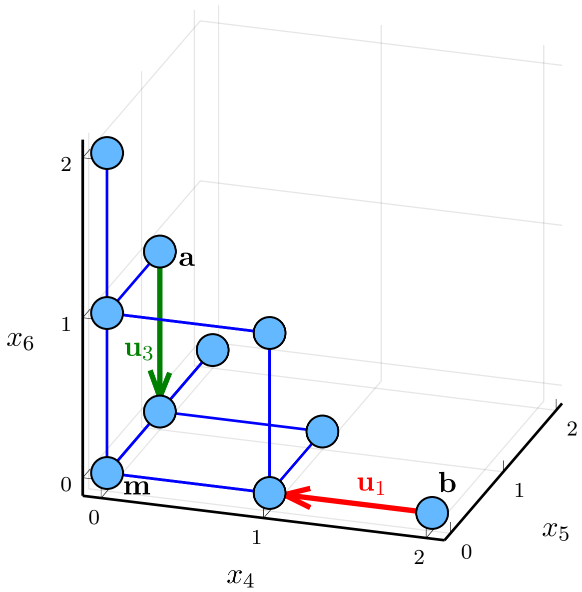


Figure 4.1: The $\mathbb{Z}$-polytope in Example 4.2.10 showing the distance reducing moves for $\mathbf{a}$ (green) and $\mathbf{b}$ (red).

and so

$$\mathrm{d}(\mathbf{a}, \mathbf{m}) = 2$$
$$\mathrm{d}(\mathbf{b}, \mathbf{m}) = 2.$$

Setting $\mathbf{a}^\dagger = \mathbf{a} - \mathbf{u}_3$ and $\mathbf{b}^\dagger = \mathbf{b} - \mathbf{u}_1$, we have

$$\mathbf{m} = \mathbf{a}^\dagger - \mathbf{u}_2$$
$$\mathbf{m} = \mathbf{b}^\dagger - \mathbf{u}_1,$$

so the distance from each point to $\mathbf{m}$ has been reduced to

$$\mathrm{d}(\mathbf{a}^\dagger, \mathbf{m}) = 1$$
$$\mathrm{d}(\mathbf{b}^\dagger, \mathbf{m}) = 1.$$

From Lemma 4.2.8, $\mathbf{0} \leq \mathbf{a}_2^\dagger \leq \mathbf{a}_2$ and $\mathbf{0} \leq \mathbf{b}_2^\dagger \leq \mathbf{b}_2$ means that $\mathbf{a}^\dagger, \mathbf{b}^\dagger \in \mathcal{F}_{\mathbf{y}}$, and we have used U to reduce the distance to from both $\mathbf{a}$ and $\mathbf{b}$ to $\mathbf{m}$. Both $\mathbf{a}$ and $\mathbf{b}$ are connected to $\mathbf{m}$ by U, and so they must be connected to each other.

4.2.3 Algebraic proof of Theorem 4.2.1

Theorem 4.2.1 states that if a column partition lattice basis U for a configuration matrix A is such that each of the colums of A_2 is a non-negative combination of columns of

A_1, and U is integral, then U is a Markov basis. A distance reducing proof was given in Section 4.2.2. Here we give an inductive proof using the Fundamental Theorem of Markov Bases (Theorem 2.4.5).

Proof. Let a column partition of a configuration matrix $A \in \{0,1\}^{n\times r}$ be such that each column of A_2 is a non-negative combination of columns of A_1. By Lemmata 4.2.3 and 4.2.5, this partition induces a column partition lattice basis

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix},$$

and $-A_1^{-1}A_2 \leq 0$. Let U be integral.

We split each vector in $\mathbb{Z}^r$ in accordance with the column partition so that

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix},$$

and we correspondingly split the collection of indeterminates $T = \{t_1, \ldots, t_r\}$ into $T_1 = \{t_1, \ldots, t_n\}$ and $T_2 = \{t_{n+1}, \ldots, t_r\}$, so that

$$T^{\mathbf{a}} = T_1^{\mathbf{a}_1} T_2^{\mathbf{a}_1}.$$

Splitting a vector $\mathbf{u}_i$ into its positive and negative parts $\mathbf{u}_i = \mathbf{u}_i^+ - \mathbf{u}_i^-$ produces

$$\mathbf{u}_i^+ = \begin{bmatrix} \mathbf{0} \\ \mathbf{e}_i \end{bmatrix} \quad \text{and} \quad \mathbf{u}_i^- = \begin{bmatrix} A_1^{-1}A_2\mathbf{e}_i \\ \mathbf{0} \end{bmatrix}.$$

The monomial difference form of this is

$$T^{\mathbf{u}_i^+} - T^{\mathbf{u}_i^-} = T_2^{\mathbf{e}_i} - T_1^{A_1^{-1}A_2\mathbf{e}_i}.$$

In these equations $\mathbf{u}_i$ denotes the ith column of U, and $\mathbf{e}_i$ is the ith standard basis vector.

The ideal I_U is generated by the monomial difference representations of U, and we can write

$$\begin{aligned} I_U &= \langle T^{\mathbf{u}^+} - T^{\mathbf{u}^-} : \mathbf{u} \in U \rangle \\ &= \langle T_2^{\mathbf{e}_i} - T_1^{A_1^{-1}A_2\mathbf{e}_i} : i = 1, \ldots, r-n \rangle. \end{aligned}$$

By the Fundamental Theorem of Markov Bases (Theorem 2.4.5), we need to show that for any $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, the monomial difference representation $T^{\mathbf{a}} - T^{\mathbf{b}}$ of any pair of points $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ is in the ideal I_U. We set $\mathbf{m}_2 = \mathbf{0}$, and

$$\mathbf{m} = \begin{bmatrix} \mathbf{a}_1 + A_1^{-1}A_2\mathbf{a}_2 \\ \mathbf{0} \end{bmatrix} \in \mathcal{F}_{\mathbf{y}}$$

by Lemma 4.2.8. We have $T^{\mathbf{a}} - T^{\mathbf{b}} = T^{\mathbf{a}} - T^{\mathbf{m}} + T^{\mathbf{m}} - T^{\mathbf{b}}$. The element $\mathbf{a}$ represents an arbitrary point in $\mathcal{F}_{\mathbf{y}}$, so if we can show that $T^{\mathbf{a}} - T^{\mathbf{m}} \in I_U$, then automatically $T^{\mathbf{b}} - T^{\mathbf{m}} \in I_U$ and so $T^{\mathbf{a}} - T^{\mathbf{b}} \in I_U$.

We have

$$\begin{aligned}T^{\mathbf{a}} - T^{\mathbf{m}} &= T_1^{\mathbf{a}_1} T_2^{\mathbf{a}_2} - T_1^{\mathbf{a}_1 + A_1^{-1}A_2\mathbf{a}_2} T_2^{\mathbf{0}} \\ &= T_1^{\mathbf{a}_1}(T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2}),\end{aligned}$$

so if we can show that $T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2} \in I_U$ for all $\mathbf{a}_2 \in \mathbb{Z}_{\geq 0}^{r-n}$, then $T^{\mathbf{a}} - T^{\mathbf{b}} \in I_U$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ for all $\mathbf{y}$, and U is a Markov basis.

We proceed via induction. For the base case, let $\mathbf{a}_2 = \mathbf{e}_k$ for some $k \in \{1, \ldots, r-n\}$. Clearly

$$\begin{aligned}T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2} &= T_2^{\mathbf{e}_k} - T_1^{A_1^{-1}A_2\mathbf{e}_k} \\ &\in \langle T_2^{\mathbf{e}_i} - T_1^{A_1^{-1}A_2\mathbf{e}_i} : i = 1, \ldots, r-n\rangle.\end{aligned}$$

For the induction, suppose that $T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2} \in I_U$. We need to show that

$$T_2^{(\mathbf{a}_2+\mathbf{e}_k)} - T_1^{A_1^{-1}A_2(\mathbf{a}_2+\mathbf{e}_k)} \in \langle T_2^{\mathbf{e}_i} - T_1^{A_1^{-1}A_2\mathbf{e}_i} : i = 1, \ldots, r-n\rangle.$$

For any $k \in \{1, \ldots, r-n\}$,

$$\begin{aligned}T_2^{(\mathbf{a}_2+\mathbf{e}_k)} - T_1^{A_1^{-1}A_2(\mathbf{a}_2+\mathbf{e}_k)} &= T_2^{(\mathbf{a}_2+\mathbf{e}_k)} - T_2^{\mathbf{a}_2}T_1^{A_1^{-1}A_2\mathbf{e}_k} + T_2^{\mathbf{a}_2}T_1^{A_1^{-1}A_2\mathbf{e}_k} - T_1^{A_1^{-1}A_2(\mathbf{a}_2+\mathbf{e}_k)} \\ &= T_2^{\mathbf{a}_2}(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1}A_2\mathbf{e}_k}) - T_1^{A_1^{-1}A_2\mathbf{e}_k}(T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2}) \\ &\in \langle T_2^{\mathbf{e}_i} - T_1^{A_1^{-1}A_2\mathbf{e}_i} : i = 1, \ldots, r-n\rangle,\end{aligned}$$

completing the proof. □

We demonstrate this proof with an example.

Example 4.2.11. Consider again the three-link linear network, which has the configuration matrix

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

and a column partition lattice basis

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

The monomial difference representation of U is given by

$$\{t_4 - t_1t_2, t_5 - t_2t_3, t_6 - t_1t_2t_3\}.$$

For our induction, we assume the monomial difference

$$T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a_2}} = t_4^{a_1}t_5^{a_2}t_6^{a_3} - t_1^{a_1+a_2}t_2^{a_2+a_3}t_3^{a_1+a_2+a_3}$$

lies in the ideal I_U. We will show that $T_2^{\mathbf{a}_2+\mathbf{e}_i} - T_1^{A_1^{-1}A_2(\mathbf{a_2}+\mathbf{e}_i)} \in I_U$ for $i = 1$. We have

$$\begin{aligned}
T_2^{\mathbf{a}_2+\mathbf{e}_1} - T_1^{A_1^{-1}A_2(\mathbf{a_2}+\mathbf{e}_1)} &= t_4^{(a_1+1)}t_5^{a_2}t_6^{a_3} - t_1^{(a_1+1)+a_2}t_2^{a_2+a_3}t_3^{(a_1+1)+a_2+a_3} \\
&= t_4^{(a_1+1)}t_5^{a_2}t_6^{a_3} - t_1t_2t_4^{a_1}t_5^{a_2}t_6^{a_3} + t_1t_2t_4^{a_1}t_5^{a_2}t_6^{a_3} \\
&\quad - t_1^{(a_1+1)+a_2}t_2^{a_2+a_3}t_3^{(a_1+1)+a_2+a_3} \\
&= t_4^{a_1}t_5^{a_2}t_6^{a_3}(t_4 - t_1t_2) + t_1t_2(t_4^{a_1}t_5^{a_2}t_6^{a_3} - t_1^{a_1+a_2}t_2^{a_2+a_3}t_3^{a_1+a_2+a_3}) \\
&= T_2^{\mathbf{a}_2}(T_2^{\mathbf{e}_1} - T_1^{A_1^{-1}A_2\mathbf{e}_1}) + T_1^{A_1^{-1}A_2\mathbf{e}_1}(T_2^{\mathbf{a}_2} - T_1^{A_1^{-1}A_2\mathbf{a}_2}).
\end{aligned}$$

We can demonstrate this for specific $\mathbf{a}$ too. We choose $\mathbf{a} = \begin{bmatrix}1 & 0 & 0 & 0 & 1 & 1\end{bmatrix}^\intercal$. We set $\mathbf{m}_2 = \mathbf{0}$, so

$$\begin{aligned}
\mathbf{m} &= \begin{bmatrix}\mathbf{a}_1 + A_1^{-1}A_2\mathbf{a}_2 \\ \mathbf{0}\end{bmatrix} \\
&= \begin{bmatrix}2\\2\\2\\0\\0\\0\end{bmatrix},
\end{aligned}$$

and we assume that

$$T^{\mathbf{a}} - T^{\mathbf{m}} = t_1t_5t_6 - t_1^2t_2^2t_2^2 \in I_U.$$

Choosing $k = 1$, we want to show that

$$T_2^{\mathbf{a}_2+\mathbf{e}_1} - T_1^{A_1^{-1}A_2(\mathbf{a_2}+\mathbf{e}_1)} = t_1t_4t_5t_6 - t_1^3t_2^3t_3^2 \in I_U.$$

We have

$$\begin{aligned}
T_2^{\mathbf{a}_2+\mathbf{e}_1} - T_1^{A_1^{-1}A_2(\mathbf{a_2}+\mathbf{e}_1)} &= t_1t_4t_5t_6 - t_1^3t_2^3t_3^2 \\
&= t_1t_4t_5t_6 - t_1^2t_2t_5t_6 + t_1^2t_2t_5t_6 - t_1^3t_2^3t_3^2 \\
&= t_1t_5t_6(t_4 - t_1t_2) + t_1t_2(t_1t_5t_6 - t_1^2t_2^2t_3^2) \\
&\in I_U,
\end{aligned}$$

and so U connects $\mathbf{a}$ to $\mathbf{m}$.

We can write down an expression for $T^{\mathbf{a}} - T^{\mathbf{b}}$ in terms of the monomial difference represenation of U. Such an expression is a telescoping series where the cancelling intermediate points define a path from $\mathbf{b}$ to $\mathbf{a}$ through $\mathcal{F}_{\mathbf{y}}$.

While walking from $\mathbf{m}$ to $\mathbf{a}$, each move $\mathbf{u}_i$ is used a_i times for each $a_i \in \mathbf{a}_2$. In a projection of a $\mathbb{Z}$-polytope where $-A_1^{-1}A_2$ is non-positive, these moves may be made in any order and the walk will still stay within $\mathcal{F}_{\mathbf{y}}$. In order to write $T^{\mathbf{a}} - T^{\mathbf{m}}$ in terms of elements of U, an ordering for the moves must be chosen. In our expression we perform the moves in the order $\mathbf{u}_1, \ldots, \mathbf{u}_{r-n}$.

If we are at a point $\mathbf{p}$ in the walk where we have completed all usages of $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$, and the first ℓ usages of $\mathbf{u}_k$, then

$$\begin{aligned}\mathbf{p} &= \mathbf{m} + a_1\mathbf{u}_1 + \cdots + a_{k-1}\mathbf{u}_{k-1} + \ell\mathbf{u}_k \\ &= \mathbf{m} + \sum_{i=1}^{k-1} a_i\mathbf{u}_i + \ell\mathbf{u}_k,\end{aligned}$$

where the a_i are elements of $\mathbf{a}_2$ and are indexed by their position therein. Recalling that

$$\mathbf{m} = \begin{bmatrix}\mathbf{a}_1 + A_1^{-1}A_2\mathbf{a}_2 \\ \mathbf{0}\end{bmatrix},$$

and so

$$T^{\mathbf{m}} = T_1^{\mathbf{a}_1} T_1^{A_1^{-1}A_2\mathbf{a}_2},$$

point $\mathbf{p}$ has the monomial representation

$$\begin{aligned}T^{\mathbf{p}} &= T^{\mathbf{m} + \sum_{i=1}^{k-1} a_i\mathbf{u}_i + \ell\mathbf{u}_k} \\ &= T_1^{\mathbf{a}_1} T_1^{A_1^{-1}A_2\mathbf{a}_2} T^{\sum_{i=1}^{k-1} a_i\mathbf{u}_i} T^{\ell\mathbf{u}_k} \\ &= T_1^{\mathbf{a}_1} T_1^{A_1^{-1}A_2\mathbf{a}_2} \frac{T_2^{\sum_{i=1}^{k-1} a_i\mathbf{e}_i}}{T_1^{\sum_{i=1}^{k-1} a_i A_1^{-1}A_2\mathbf{e}_i}} \frac{T_2^{\ell\mathbf{e}_k}}{T_1^{\ell A_1^{-1}A_2\mathbf{e}_k}} \\ &= T_1^{\mathbf{a}_1} T_1^{A_1^{-1}A_2\mathbf{a}_2} \left(\prod_{i=1}^{k-1} \frac{T_2^{a_i\mathbf{e}_i}}{T_1^{a_i A_1^{-1}A_2\mathbf{e}_i}}\right) \frac{T_2^{\ell\mathbf{e}_k}}{T_1^{\ell A_1^{-1}A_2\mathbf{e}_k}} \\ &= T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i\mathbf{e}_i}\right) T_2^{\ell\mathbf{e}_k} T_1^{(a_k-\ell)A_1^{-1}A_2\mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1}A_2\mathbf{e}_i}\right).\end{aligned}$$

The step in the walk between $\mathbf{m}$ and $\mathbf{a}$ that steps from $\mathbf{p}$ has the monomial difference representation

$$\begin{aligned}T^{\mathbf{p}+\mathbf{u}_k} - T^{\mathbf{p}} = {} & T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i\mathbf{e}_i}\right) T_2^{(\ell+1)\mathbf{e}_k} T_1^{(a_k-\ell-1)A_1^{-1}A_2\mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1}A_2\mathbf{e}_i}\right) \\ & - T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i\mathbf{e}_i}\right) T_2^{\ell\mathbf{e}_k} T_1^{(a_k-\ell)A_1^{-1}A_2\mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1}A_2\mathbf{e}_i}\right)\end{aligned}$$

$$= T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i \mathbf{e}_i} \right) T_2^{\ell \mathbf{e}_k} T_1^{(a_k - \ell - 1) A_1^{-1} A_2 \mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1} A_2 \mathbf{e}_i} \right) \times \left(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1} A_2 \mathbf{e}_k} \right).$$

The part of the walk that comprises of all steps $\mathbf{u}_k$ begins at the point

$$\mathbf{p} = \mathbf{m} + \sum_{i=1}^{k-1} a_i \mathbf{u}_i$$

and is described by

$$T^{\mathbf{p} + a_k \mathbf{u}_k} - T^{\mathbf{p}} = \sum_{\ell=0}^{a_k} \left(T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i \mathbf{e}_i} \right) T_2^{\ell \mathbf{e}_k} T_1^{(a_k - \ell - 1) A_1^{-1} A_2 \mathbf{e}_k} \times \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1} A_2 \mathbf{e}_i} \right) \left(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1} A_2 \mathbf{e}_k} \right) \right).$$

The path from $\mathbf{m}$ to $\mathbf{a}$ is taken by summing the above expression over all moves $\mathbf{u}_k$:

$$T^{\mathbf{m}} - T^{\mathbf{a}} = \sum_{k=1}^{r-n} \left(\sum_{\ell=0}^{a_k} \left(T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i \mathbf{e}_i} \right) T_2^{\ell \mathbf{e}_k} T_1^{(a_k - \ell - 1) A_1^{-1} A_2 \mathbf{e}_k} \times \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1} A_2 \mathbf{e}_i} \right) \left(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1} A_2 \mathbf{e}_k} \right) \right) \right).$$

Finally, a path connecting two $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ is given by

$$\begin{aligned} T^{\mathbf{b}} - T^{\mathbf{a}} &= (T^{\mathbf{m}} - T^{\mathbf{a}}) - (T^{\mathbf{m}} - T^{\mathbf{b}}) \\ &= \sum_{k=1}^{r-n} \left(\sum_{\ell=0}^{a_k} \left(T_1^{\mathbf{a}_1} \left(\prod_{i=1}^{k-1} T_2^{a_i \mathbf{e}_i} \right) T_2^{\ell \mathbf{e}_k} T_1^{(a_k - \ell - 1) A_1^{-1} A_2 \mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{a_i A_1^{-1} A_2 \mathbf{e}_i} \right) \left(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1} A_2 \mathbf{e}_k} \right) \right) \right) \\ &\quad - \sum_{k=1}^{r-n} \left(\sum_{\ell=0}^{b_k} \left(T_1^{\mathbf{b}_1} \left(\prod_{i=1}^{k-1} T_2^{b_i \mathbf{e}_i} \right) T_2^{\ell \mathbf{e}_k} T_1^{(b_k - \ell - 1) A_1^{-1} A_2 \mathbf{e}_k} \left(\prod_{i=k+1}^{r-n} T_1^{b_i A_1^{-1} A_2 \mathbf{e}_i} \right) \left(T_2^{\mathbf{e}_k} - T_1^{A_1^{-1} A_2 \mathbf{e}_k} \right) \right) \right). \end{aligned}$$

This path visits the origin, and is not necessarily the shortest path between $\mathbf{a}$ and $\mathbf{b}$.

4.2.4 A geometric interpretation

The non-negativity condition on $A_1^{-1}A_2$ in Theorem 4.2.1 has an interesting geometric interpretation. From Section 3.3 we know that the row vectors of U are the normal vectors to the bounding hyperplanes of the projected $\mathbb{Z}$-polytope. If these normal vectors are non-positive, then when their corresponding bounding hyperplane is in the non-negative orthant, they all point towards the origin. None of their bounding hyperplanes cuts any point in the projected $\mathbb{Z}$-polytope off from the origin. We will see this in the following example.

Example 4.2.12. Consider the configuration matrix of the three-link linear network

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

We partition A such that A_1 is the identity matrix. Clearly this partition meets the conditions of Theorem 4.2.1.

This partition induces the column partition lattice basis

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

The bounding hyperplanes whose normal vectors are the first three rows of U are shown in Figure 4.2. We can see that none of these bounding hyperplanes can be positioned so that in the non-negative orthant, a point on the positive side of the hyperplane is on the opposite side of the hyperplane to the origin.

Some example $\mathbb{Z}$-polytopes are shown in Figure 4.3. For each point in each $\mathbb{Z}$-polytope, it is possible to move towards the origin in any co-ordinate direction.

4.2.5 Applications

If a configuration matrix A contains an $n \times n$ identity matrix as a maximal submatrix, then this theorem implies that partitioning such that $A_1 = I$ induces a lattice basis that is a Markov basis. This has applications in capture-recapture models, as observed by Schofield and Bonner [39]. In network tomography, if a network is such that each edge is by itself an allowed path, then the identity matrix is a maximal submatrix and this theorem can also be applied.

Of course, a link-path incidence matrix that can be partitioned such that each column of A_2 is a positive sum of a selection of columns of A_1 need not have the identity matrix as a maximal submatrix. A one-way linear network such that the first k nodes can be origins, and the last l nodes can be destinations also has this property, a long as there is at least one vertex that is both an origin and a destination.

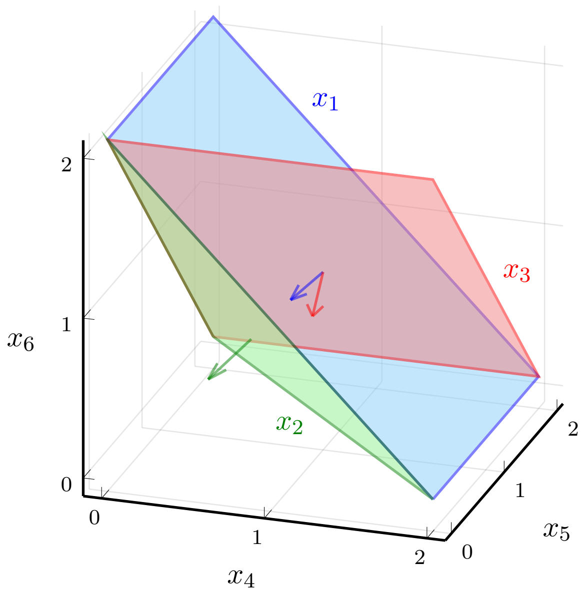
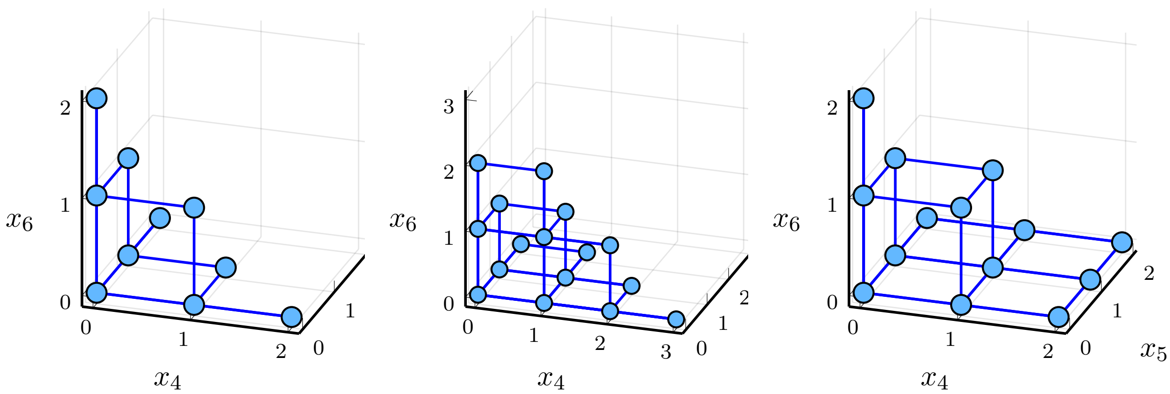


Figure 4.2: The bounding hyperplanes of the $\mathbb{Z}$-polytope in Example 4.2.12.



(a) For $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$. (b) For $\mathbf{y} = \begin{bmatrix} 4 & 4 & 4 \end{bmatrix}^\intercal$. (c) For $\mathbf{y} = \begin{bmatrix} 2 & 4 & 2 \end{bmatrix}^\intercal$.

Figure 4.3: The $\mathbb{Z}$-polytopes from Example 4.2.12.

Example 4.2.13. Consider this link-path incidence matrix for a five link network:

$$A = \left[\begin{array}{ccccc|cccccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right].$$

There are six nodes: the first three nodes are origins for traffic, then fourth node is both an origin and destination, and the last three are destinations. Under the given column ordering, the induced column partition lattice basis U is given by the columns of the matrix

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

Each move in U is equivalent to adding a car to some path that passes the node that is both an origin and a destination, and compensating by removing: one car that starts at the same origin node and stops at this central node; and one car that starts at the central node and stops at the same destination.

Suppose that the numbers of cars observed on links in this network is given by $\mathbf{y} = \begin{bmatrix} 3 & 4 & 5 & 5 & 2 \end{bmatrix}^{\mathsf{T}}$. Then the vectors

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 2 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \in \mathcal{F}_{\mathbf{y}},$$

and we will name them $\mathbf{a}$ and $\mathbf{b}$ respectively.

When traffic counts in $\mathbf{a}$ are observed, six cars are traversing paths in the A_2 part. Of these, only one of the two cars sixth path in A consisting of the first four links is common with $\mathbf{b}$. The other five are reduced down the paths in A_1 using $\mathbf{u}_1$ and $\mathbf{u}_5$ once each, and $\mathbf{u}_2$ and $\mathbf{u}_3$ twice each. This reduction is equivalent to a walking through the $\mathbb{Z}$-polytope to the point

$$\mathbf{m} = \mathbf{a} - \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 - \mathbf{u}_5.$$

Then $\mathbf{m} = \begin{bmatrix}2 & 1 & 1 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}^\intercal$. The walk then continues to $\mathbf{b}$ using the moves $\mathbf{u}_1, \mathbf{u}_4$ and $\mathbf{u}_6$ once each.

The point $\mathbf{m}$ that is visited is the same point $\mathbf{m}$ in the proof of Theorem 4.2.1 in Section 4.2.2: each of the A_2 co-ordinates of $\mathbf{m}$ is the minimum of the corresponding co-ordinates of $\mathbf{a}$ and $\mathbf{b}$.

A $\mathbb{Z}$-polytope walk constructed along the lines of the walk used in the algebraic proof in Section 4.2.3 would visit the point $\begin{bmatrix}3 & 1 & 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}^\intercal$, whose A_2 co-ordinates are all $\mathbf{0}$, and would require an additional use of $-\mathbf{u}_1$ and an additional use of $\mathbf{u}_1$ to construct.

4.3 Markov basis simulation

One way of showing that some set of moves in $\ker_{\mathbb{Z}}(A)$ is a Markov basis is to show that it is capable of simulating the moves in some other known Markov basis. Given a configuration matrix A and a vector $\mathbf{y}$, if a set $\mathcal{M}$ is a Markov basis for A then it can be used to construct a walk between any pair of points $\mathbf{x}_1, \mathbf{x}_k$ in the fibre $\mathcal{F}_\mathbf{y}$ that stays within $\mathcal{F}_\mathbf{y}$. This walk may be constructed by stepping between intermediate points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathcal{F}_\mathbf{y}$.

Suppose there is another set $\mathcal{B}$, and that $\mathcal{B}$ can be used to construct a walk between each pair of points $\mathbf{x}_i, \mathbf{x}_{i+1} \in \mathcal{F}_\mathbf{y}$. Then $\mathcal{B}$ connects $\mathbf{x}_1$ and $\mathbf{x}_k$, too. If $\mathcal{B}$ can do this for all pairs of sequential points in all walks using a Markov basis in all fibres for A, then $\mathcal{B}$ must also be a Markov basis for A.

Suppose we have a set $\mathcal{B}$, and a walk in a fibre which we wish to simulate. Let $\mathbf{x}_i, \mathbf{x}_{i+1} \in \mathcal{F}_\mathbf{y}$ be a sequential pair of points in this walk, and suppose that $\mathbf{z}$ is the move in $\mathcal{M}$ used to step between them. Then it must be the case that $\mathbf{z}^- \leq \mathbf{x}_i$ and $\mathbf{z}^+ \leq \mathbf{x}_{i+1}$.

If $\mathbf{z} = \mathbf{u}_1 + \cdots + \mathbf{u}_\ell$ where each $\mathbf{u}_j \in \mathcal{B}$, then we may wish to use these moves to simulate $\mathbf{z}$ in our walk. If $\mathbf{u}_1^- \leq \mathbf{x}_i$, then we can use $\mathbf{u}_1$ as the first step in our simulated walk — after taking the step $\mathbf{u}_1$ from $\mathbf{x}_i$, we are guaranteed to be at a point within the fibre. Then if $\mathbf{u}_1^- \leq \mathbf{z}^-$, we can use $\mathbf{u}_1$ as the first step in any walk that simulates the step $\mathbf{z}$.

The simulated walk is now at the point $\mathbf{x}_i + \mathbf{u}_1 \in \mathcal{F}_\mathbf{y}$. Similarly, if $\mathbf{u}_2^- \leq \mathbf{x}_i + \mathbf{u}_1$, then adding $\mathbf{u}_2$ will produce a point in $\mathcal{F}_\mathbf{y}$, and $\mathbf{u}_2$ may be used to continue the walk; and if $\mathbf{u}_2^- \leq \mathbf{z}^- + \mathbf{u}_1$, then it can be used as the second step in any walk that simulates the step $\mathbf{z}$ and begins with $\mathbf{u}_1$.

If $\mathbf{u}_m^- \leq \mathbf{z}^- + \sum_{j=1}^{m-1} \mathbf{u}_j$ for all $m = 1, \ldots, \ell$, then these moves can be used in the order $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ to walk not only from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ in our original walk, but between any pair of points that are separated by $\mathbf{z}$ in any fibre. If all moves in $\mathcal{M}$ can be simulated by $\mathcal{B}$

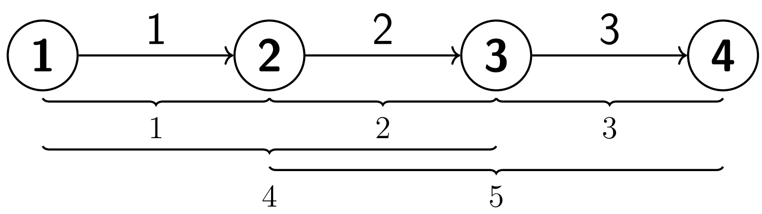


Figure 4.4: A three-link linear network. The underbraces show the allowed paths.

in this way, then $\mathcal{B}$ connects all pairs of points that $\mathcal{M}$ does, and so $\mathcal{B}$ must be a Markov basis.

On the other hand, if there is some $\mathbf{z} \in \mathcal{M}$ such that $\mathcal{B}$ cannot be used to construct a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$, then choosing $\mathbf{y} = A\mathbf{z}^-$ means that $\mathcal{F}_\mathbf{y}$ contains two points that are not connected by $\mathcal{B}$, so $\mathcal{B}$ is not a Markov basis.

We illustrate this with an example.

Example 4.3.1. Let A be the link-path incidence matrix of a three-link linear network where travel is allowed between any pair of nodes except for from the first to the last node, as in Figure 4.4. We form the column partition lattice basis U by taking the A_1 partition to be the identity matrix. Then

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad U = \begin{bmatrix} -1 & 0 \\ -1 & -1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The Graver basis is known to be a Markov basis. It consists of the two elements of $\mathcal{B}$ and $\mathbf{g} = -\mathbf{u}_1 + \mathbf{u}_2$. It is given by

$$\mathcal{G}_A = \left\{ \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

As in Theorem 3.3.3, the row vectors of the U matrix give the faces of the $\mathbb{Z}$-polytope and their positions are given by the $A_1^{-1}\mathbf{y}$ vector. We take $A_1^{-1}\mathbf{y} = \mathbf{y} = \begin{bmatrix} 6 & 8 & 4 \end{bmatrix}^t$. The projection of the $\mathbb{Z}$-polytope corresponding to this choice of column partition lattice basis is given in Figure 4.5.

Some of the elements of $\mathcal{F}_\mathbf{y}$ are given by

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 4 \\ 4 \end{bmatrix} \right\},$$

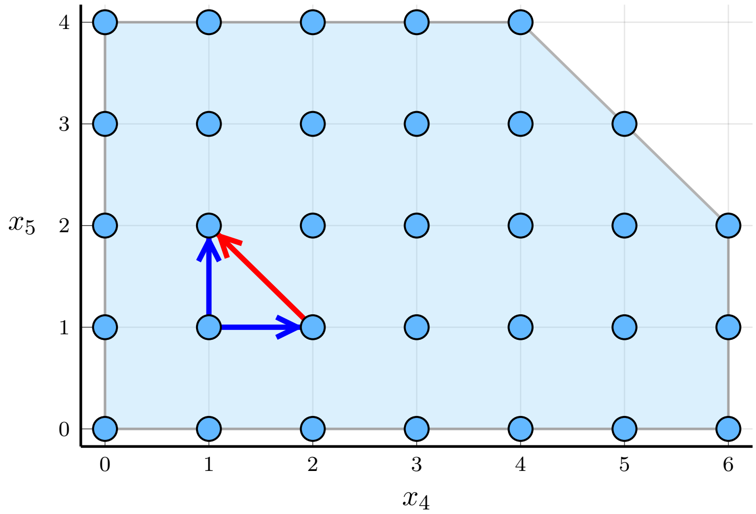


Figure 4.5: The $\mathbb{Z}$-polytope from Example 4.3.1. The lattice basis elements are shown in blue; the remaining Graver basis element is shown in red.

which we name $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and $\mathbf{x}_4$ respectively.

We can use $\mathcal{G}_A$ to walk from $\mathbf{x}_1$ to $\mathbf{x}_4$ via $\mathbf{x}_2$ and $\mathbf{x}_3$ using the sequence of moves $\mathbf{u}_2, \mathbf{g}, \mathbf{g}$. We can attempt to simulate this walk using only moves in U. We know we can perform $\mathbf{u}_2$ because it is in U. The two uses of $\mathbf{g}$ take the walk from

$$\begin{bmatrix}0\\0\\2\\6\\2\end{bmatrix} \text{ to } \begin{bmatrix}1\\0\\1\\5\\3\end{bmatrix}, \text{ and from } \begin{bmatrix}1\\0\\1\\5\\3\end{bmatrix} \text{ to } \begin{bmatrix}2\\0\\0\\4\\4\end{bmatrix}.$$

The moves in U required to simulate $\mathbf{g}$ are $-\mathbf{u}_1$ and $\mathbf{u}_2$. In both cases if we perform $-\mathbf{u}_1$ before $\mathbf{u}_2$, we are able to perform this sequence without leaving the $\mathbb{Z}$-polytope, and therefore we can simulate the entire walk without leaving the $\mathbb{Z}$-polytope. In this projection of the $\mathbb{Z}$-polytope, if the points $\mathbf{x}, \mathbf{x}-\mathbf{g} \in \mathcal{F}_\mathbf{y}$ for some $\mathbf{x}$, then $\mathbf{x}-\mathbf{u}_1 \in \mathcal{F}_\mathbf{y}$ too. In fact, the arrangement of the faces in this projection mean that we can always perform the sequence $-\mathbf{u}_1, \mathbf{u}_2$ whenever we can perform $\mathbf{g}$ in all $\mathbb{Z}$-polytopes associated with this configuration matrix. Whenever we perform $\mathbf{g}$, we must start from a point that is at least $\mathbf{g}^- = \begin{bmatrix}0 & 0 & 1 & 1 & 0\end{bmatrix}^\intercal$. The 1 in the 4th co-ordinate of $\mathbf{g}^-$ means that performing $-\mathbf{u}_1$ is always possible. This vector $\mathbf{g}$ is the only element of $\mathcal{G}_A$ that is not in U, and so $\mathcal{B}$ is a Markov basis.

We state this result as a theorem.

Theorem 4.3.2. *Let A be a configuration matrix and let $\mathcal{M}$ be a Markov basis for A. Let $\mathcal{B}$ be a set of moves in $\ker_\mathbb{Z}(A)$. Then $\mathcal{B}$ is a Markov basis for A if and only if $\mathcal{B}$ connects $\mathbf{z}^-$ and $\mathbf{z}^+$ for all $\mathbf{z} \in \mathcal{M}$.*

For the proof we use the Fundamental Theorem of Markov Bases.

Proof. Suppose that $\mathcal{B}$ connects $\mathbf{z}^-$ to $\mathbf{z}^+$ for each $\mathbf{z} \in \mathcal{M}$, so $T^{\mathbf{z}^-} - T^{\mathbf{z}^+} \in I_\mathcal{B}$ by Theorem 2.4.5. The set $\{T^{\mathbf{z}^-} - T^{\mathbf{z}^+} : \mathbf{z} \in \mathcal{M}\}$ generates $I_\mathcal{M}$, so $I_\mathcal{M} \subseteq I_\mathcal{B}$. The set $\mathcal{M}$ is a Markov basis for A, so we have $I_\mathcal{M} = I_A$, and $I_\mathcal{B} \subseteq I_A$ is always true, so $I_\mathcal{B} = I_A$. Therefore $\mathcal{B}$ is a Markov basis, as required.

Conversely, suppose there exists $\mathbf{z} \in \mathcal{M}$ such that the lattice basis $\mathcal{B}$ does not connect $\mathbf{z}^-$ to $\mathbf{z}^+$. Then by definition, $\mathcal{B}$ is not a Markov basis. □

We may not always have knowledge of a Markov basis that we can simulate: the same line of reasoning shows that in order to show that some $\mathcal{B}$ is a Markov basis, we need only show that $\mathcal{B}$ connects $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_\mathbf{y}$ when $\mathbf{x}_1$ and $\mathbf{x}_2$ have disjoint support.

Corollary 4.3.3. *Let A be a configuration matrix and let $\mathcal{B}$ be a set of moves in $\ker_\mathbb{Z}(A)$. Then $\mathcal{B}$ is a Markov basis for A if and only if $\mathcal{B}$ connects $\mathbf{z}^-$ and $\mathbf{z}^+$ for all $\mathbf{z} \in \ker_\mathbb{Z}(A)$.*

4.4 Connectivity of lattice bases

A column partition lattice basis for a configuration matrix A is given by the columns of a matrix U. In this section we give a sufficient condition on a matrix U that can determine if it is a Markov basis.

When checking if a matrix U meets this condition, we search it for a particular type of submatrix. If no such submatrix can be found, then U is a Markov basis. This is stated formally as Theorem 4.4.14. In this section we present and prove this condition.

The condition as we state and prove it requires all entries of U to be in $\{0, \pm 1\}$. It can therefore be applied to all column partition bases of unimodular configuration matrices. However, the idea does extend intuitively to matrices with larger integer entries.

The submatrices we are concerned with are non-zero Eulerian submatrices whose columns each contain entries that sum to zero, which we will call *zero column sum Eulerian*, or *c0-Eulerian* matrices. Recall from Definition 3.5.3 that a matrix $A \in \{0, \pm 1\}^{n \times r}$ is *Eulerian* if for each row and column, the sum of the entries is a multiple of two.

Definition 4.4.1 (Zero column sum Eulerian)**.** We say a non-null matrix M is *zero column sum Eulerian*, or *c0-Eulerian*, if:

- M is Eulerian,
- the entries within each column of M sum to zero.

Examples of c0-Eulerian matrices include

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

The matrix
$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix},$$
is also Eulerian, but the sums of the entries in each column are 2 and -2 respectively, so its presence as a submatrix of some matrix U does not affect whether or not it defines a Markov basis.

We also define *zero column and row sum Eulerian* matrices, or *cr*0*-Eulerian* matrices, for which the zero summing property holds for rows as well.

Definition 4.4.2 (Zero column and row sum Eulerian)**.** We say a matrix M is *zero column and row sum Eulerian*, or *cr*0*-Eulerian*, if:

- M is Eulerian,
- M does not have all entries equal to zero,
- the entries within each column of M sum to zero,
- the entries within each row of M sum to zero.

Clearly all cr0-Eulerian matrices are also c0-Eulerian.

The condition presented in this section is stated in Theorem 4.4.14, and claims that if a matrix U which defines a column partition lattice basis for a unimodular configuration matrix contains no c0-Eulerian submatrix, then its columns define a Markov basis.

We begin in Section 4.4.1 by looking at the geometric intuition behind this condition. This is based on Section 3.3 which discusses how the rows of U relate to the bounding hyperplanes of the corresponding projection of associated $\mathbb{Z}$-polytopes. Examples are presented in which connectivity problems (other than parity errors) appear when the matrix U contains a c0-Eulerian submatrix.

We then look at some examples of what happens when a matrix fails this condition. Two and three dimensional examples are presented in Section 4.4.2 in which we are given a U matrix that contains a c0-Eulerian submatrix, and we attempt to construct a walk between particular pairs of points. In each case, any walk must necessarily step outside of the $\mathbb{Z}$-polytope to a point with at least one negative co-ordinate. This interpretation of effect of the c0-Eulerian submatrices of U is more in line with how the proof of Theorem 4.4.14 works.

Section 4.4.4 gives a statement and proof of Theorem 4.4.14. Proving the validity of this condition requires some lemmata which are combinatorial in nature and are to do with how columns of U-style matrices can be ordered to meet certain conditions. The idea is that for any collection of moves used in any orientation, there exists an ordering such that a walk that uses that ordering involves visiting a sequence of points whose entries are non-decreasing, and then non-increasing, in each co-ordinate. This means that such a walk never visits a point with a negative co-ordinate. Since this is true for any collection of moves, for any pair of points in any $\mathbb{Z}$-polytope there is an ordering for the collection of moves that connects them, and so U is a Markov basis.

Section 4.4.5 discusses the theorem in terms of polynomial ideals. The Fundamental Theorem of Markov Bases (Theorem 2.4.5) gives a correspondence between a Markov basis for $\mathbb{Z}$-polytopes and a generating set of an ideal in a polynomial ring. Here, we show using induction that if the type of ordering of moves in Section 4.4.4 exists then the monomial difference representation of any combination of moves in U is in I_A, the ideal generated by the kernel of A. This is in line with the Fundamental Theorem.

Section 4.4.6 gives another geometric interpretation comparing the kinds of $\mathbb{Z}$-polytopes associated with U matrices that pass our condition, with the $\mathbb{Z}$-polytopes discussed in Section 4.2, in which $-A_1^{-1}A_2$ was non-positive. The $\mathbb{Z}$-polytopes from Section 4.2 can be thought of as a special case of those in this section.

In Section 4.4.7 we conjecture that the reverse of the implication in Theorem 4.4.14 is also true: that is, if U does contain a c0-Eulerian submatrix, then U is not a Markov basis.

Finally, Section 4.4.8 looks at how the theorem might be extended from $\{0, \pm 1\}$ matrices to matrices with entries in $\mathbb{Z}$ by looking at c0-Eulerian submatrices in U^ϵ, the matrix of signs of U.

4.4.1 Geometric intuition

In Section 2.5.1 we presented three potential reasons that a given column partition lattice basis might not be a Markov basis. We referred to them as parity errors, isolated spaces, and reduced dimension. If we are to show that some column partition lattice basis is a Markov basis, we must at a minimum show that these three problems cannot occur.

The first of these, parity errors, can be avoided if the matrix U which defines the column partition lattice basis contains only integers. Theorem 3.4.5 gives a condition on the column partition which guarantees this. The idea is that the I part of

$$U = \begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}$$

corresponds to the co-ordinates upon which the $\mathbb{Z}$-polytope is being projected and its non-zero entries are always 1. This means that unit sized steps in co-ordinate directions are always possible, provided the walk stays within the $\mathbb{Z}$-polytope's bounding hyperplanes. Our concern in this section is therefore with staying within these boundaries: that is, with avoiding the problems we labelled isolated spaces and reduced dimension.

Of these other problems, the simplest to address is reduced dimensionality. We will look again at the traffic network from Example 3.3.4.

Example 4.4.3. Consider the configuration matrix

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

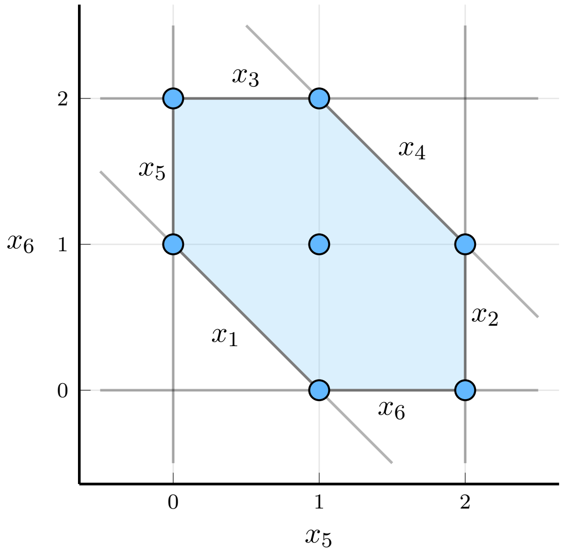


Figure 4.6: A $\mathbb{Z}$-polytope from Example 4.4.3 projected onto the x_5 and x_6 dimensions showing the bounding hyperplanes.

We choose the column partition such that

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then the column partition lattice basis is given by

$$U = \begin{bmatrix} -1 & -1 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which corresponds to a projection of the $\mathbb{Z}$-polytope onto the x_5 and x_6 co-ordinates. An example polytope for this projection with $\mathbf{y} = \begin{bmatrix} 3 & 3 & 2 & 2 \end{bmatrix}^\intercal$ is shown in Figure 4.6. The bounding hyperplanes for x_1 and x_2 oppose each other, and are not parallel to any axis. If $\mathbf{y}$ is such that these two bounding hyperplanes are in contact, then the underlying projected polytope is a diagonal line segment, and co-ordinate direction moves cannot connect the lattice points within.

The submatrix of U formed by taking the rows corresponding to these bounding hyperplanes is given by

$$M_1 = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}.$$

We note that the same effect would be achieved if the bounding hyperplanes were rotated ninety degrees — that is, if U contained the submatrix

$$M_2 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We note that both M_1 and M_2 are c0-Eulerian matrices.

A reduced dimension $\mathbb{Z}$-polytope may also arise when there are no bounding hyperplanes that directly oppose each other. The following is an example in three dimensions.

Example 4.4.4. Consider the configuration matrix

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

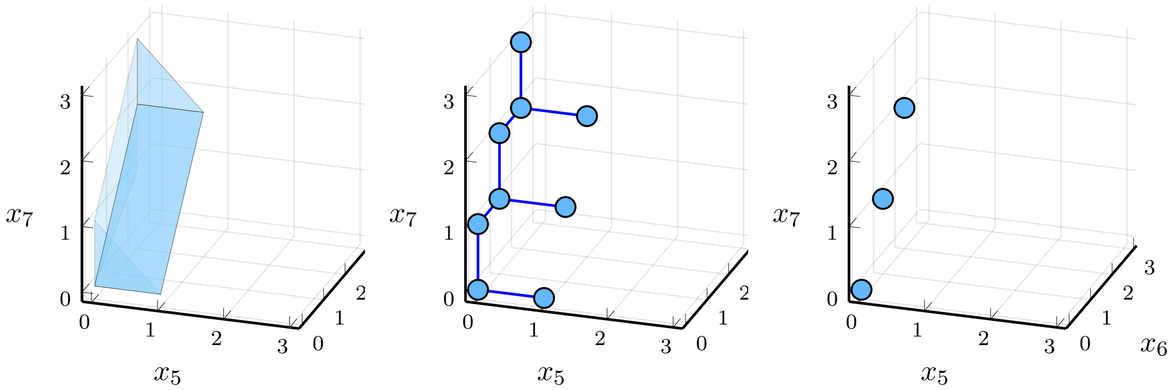
This matrix is a potential link-path incidence matrix for a traffic network on a four-link linearly connected directed tree. Choosing the column partition such that A_1 is made of the first four columns of A induces the basis

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

The $-A_1^{-1}A_2$ part of this matrix contains no pair of rows $\mathbf{u}_{i\cdot}, \mathbf{u}_{j\cdot}$ such that $\mathbf{u}_{i\cdot} = -\mathbf{u}_{j\cdot}$, and so none of the bounding hyperplanes directly opposes another in the corresponding projection of an associated $\mathbb{Z}$-polytope. The $\mathbb{Z}$-polytopes for this system for $\mathbf{y} = \begin{bmatrix} 1 & 3 & 2 & 1 \end{bmatrix}^\intercal$ and $\mathbf{y} = \begin{bmatrix} 0 & 2 & 2 & 1 \end{bmatrix}^\intercal$ are shown in Figure 4.7. Figure 4.7a shows that the $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix} 1 & 3 & 2 & 1 \end{bmatrix}^\intercal$ does not suffer from the problem of reduced dimension, and U is a Markov subbasis for this fibre.

The points in the $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix} 0 & 2 & 2 & 1 \end{bmatrix}^\intercal$ are given by

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 2 \\ 2 \end{bmatrix} \right\}.$$



(a) The polytope for $\mathbf{y} = \begin{bmatrix}1 & 3 & 2 & 1\end{bmatrix}^\intercal$. (b) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}1 & 3 & 2 & 1\end{bmatrix}^\intercal$. (c) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}0 & 2 & 2 & 1\end{bmatrix}^\intercal$.

Figure 4.7: The $\mathbb{Z}$-polytopes from Example 4.4.4

This $\mathbb{Z}$-polytope appears in Figure 4.7c, which shows that it is one dimensional and not aligned with the axis, and therefore that U is not a Markov basis.

The matrix U contains the submatrix

$$M = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

in the x_1 and x_2 co-ordinates of the columns corresponding to co-ordinate moves in the x_6 and x_7 directions. The x_1 and x_2 bounding hyperplanes do not directly oppose each other, but do oppose each other when considering only the x_6 and x_7 dimensions.

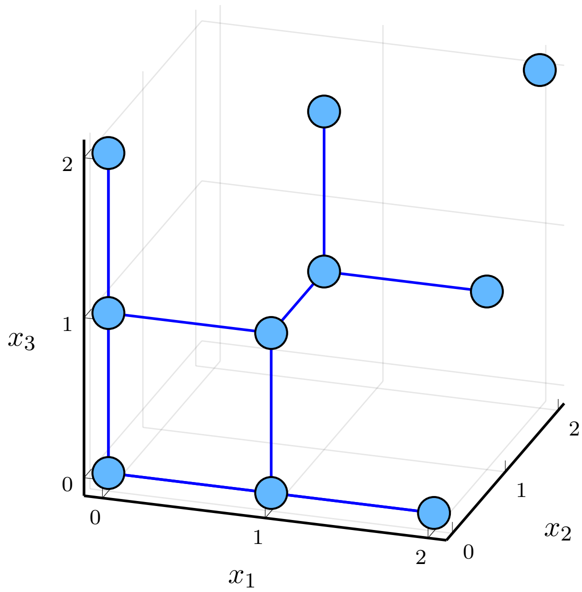
The third problem for column partition lattice bases' connectivity in Section 2.5.1 is the problem of isolated spaces. The example given is for a two-dimensional $\mathbb{Z}$-polytope. In order to obtain a two-dimensional $\mathbb{Z}$-polytope with an isolated vertex, we need a configuration matrix that was not unimodular. In three or more dimensions, unimodular configuration matrices can also produce $\mathbb{Z}$-polytopes with isolated vertices, as illustrated the following example.

Example 4.4.5. Consider the three-link linear network. The totally unimodular link-path incidence matrix for this network is

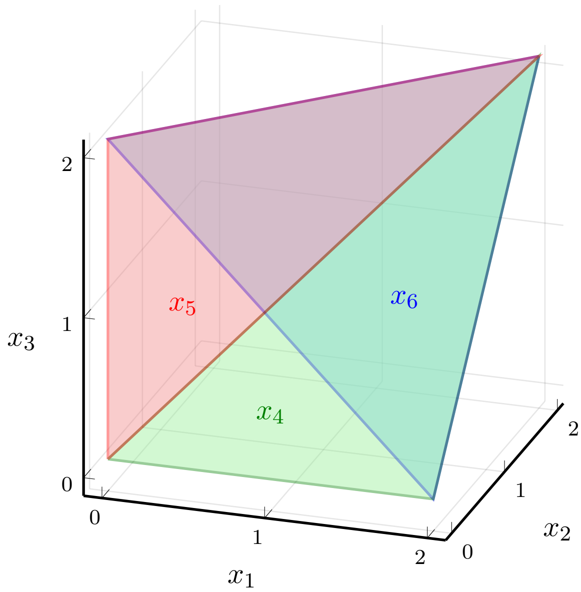
$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

We choose the partition such that

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$



(a) The isolated vertex.



(b) The bounding hyperplanes.

Figure 4.8: The $\mathbb{Z}$-polytope from Example 4.4.5.

and the induced column partition lattice basis U is given by the columns of the matrix

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

If two cars are observed on each link in the network, we have $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$, and the corresponding project of the $\mathbb{Z}$-polytope for $\mathcal{F}_\mathbf{y}$ is shown in Figure 4.8a. The point at $\begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$ in the projected space is not connected to the rest of the $\mathbb{Z}$-polytope. In a sense this is because the bounding hyperplanes representing x_5 and x_6 oppose each other when considering only the x_1 and x_2 dimensions; and because the x_4 and x_6 bounding hyperplanes are opposed to each other in the x_2 and x_3 dimensions. This can be seen in Figure 4.8b. The matrix U contains the submatrices

$$M_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

in the x_5 and x_6 co-ordinates of the first and second columns, and

$$M_2 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

in the x_4 and x_6 co-ordinates of the second and third columns.

In both Example 4.4.4 and 4.4.5, the column partition lattice basis chosen corresponds to a projection of the $\mathbb{Z}$-polytope where a pair of faces meet at awkward angles, denying access to part of the $\mathbb{Z}$-polytope. In each case, the submatrices of U that corresponds to the relevant bounding hyperplanes and projected space co-ordinates are c0-Eulerian matrices.

The idea that a column partition lattice bases that is not a Markov basis — despite not being afflicted by parity errors — corresponds to a projection of the $\mathbb{Z}$-polytope where the bounding hyperplanes meet at awkward angles, and with U containing a c0-Eulerian submatrix, extends to three dimensions too. This can be seen in the following example.

Example 4.4.6. Consider the link-path incidence matrix for a three-link linear network

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

We choose the partition such that

$$A_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

to get the column partition lattice basis

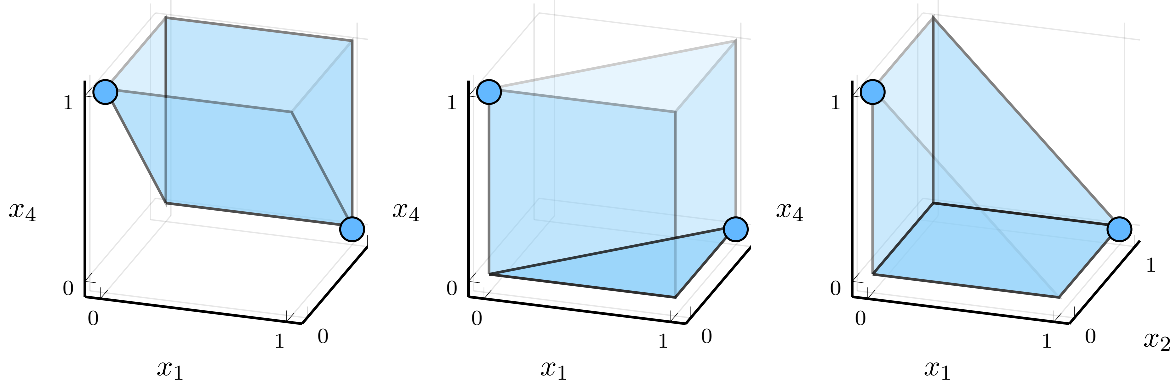
$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}.$$

This matrix contains the 3×3 c0-Eulerian submatrix

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

Choosing $\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^\intercal$ means that

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$



(a) The x_3 bounding hyperplane. (b) The x_5 bounding hyperplane. (c) The x_6 bounding hyperplane.

Figure 4.9: The action of each of the bounding hyperplanes on the projected $\mathbb{Z}$-polytope from Example 4.4.6.

This $\mathbb{Z}$-polytope lies entirely with the unit cube. Figure 4.9 shows the effect of each of the bounding hyperplanes and which points in the unit cube they remove from the $\mathbb{Z}$-polytope. Highlighted are the two vertices that are non-negative with respect to every bounding hyperplane. Their combined effect is to produce a $\mathbb{Z}$-polytope of reduced dimension that in this projection is not parallel to any of the axis. Therefore the moves in U, which are all in co-ordinate directions, cannot be used to move between them, and so U is not a Markov basis.

Together these examples illustrate the intuition behind Theorem 4.4.14. Examples 4.4.4 and 4.4.5 demonstrated that connectivity problems presented in Section 2.5.1 that column partition lattice bases might encounter other than parity errors might correlate with the presence of a c0-Eulerian submatrix in U. Example 4.4.6 suggested that this association might generalise to larger submatrices.

4.4.2 Eulerian submatrices

The presence of c0-Eulerian submatrices in column partition lattice bases has another interpretation, which we will explore in this section. In Section 4.4.1 we focussed on the rows of U and the interaction of the corresponding bounding hyperplanes. This section focuses more on the columns of the matrix U and their use as moves in a walk. This interpretation hints at the method of proof that we employ.

Suppose we have some matrix U that contains a c0-Eulerian submatrix. We are interested in what happens when we attempt use U to construct a walks in arbitrary $\mathbb{Z}$-polytopes, but fail. Potential endpoints of a walk that uses a collection of moves can be found by summing the required integer multiple of each of the columns of U.

Let the Eulerian submatrix of U be M. We claim that we can sum a ± 1 multiple of each of the columns of U that appear in M to a get a vector $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$ that has zeroes in the entries corresponding to the rows of U that are in M. Let $\mathcal{R}$ and $\mathcal{C}$ be respectively M's row and column indices in U. Then

$$\mathbf{z} = \sum_{i \in \mathcal{C}} \epsilon_i \mathbf{u}_i,$$

where each $\epsilon_i \in \pm 1$ and $\mathbf{z}_j = 0$ for each $j \in \mathcal{R}$. It is important that the ϵ_i are chosen so that the non-zero entries in each row cancel.

The vectors $\mathbf{z}^-, \mathbf{z}^+ \in \mathcal{F}_{\mathbf{y}}$ for $\mathbf{y} = A\mathbf{z}^+$. Constructing a path from $\mathbf{z}^-$ to $\mathbf{z}^+$ using U requires the use of the columns of U with indices in $\mathcal{C}$, and none of these moves can be applied because each $\mathbf{u}_i$ has a -1 where $\mathbf{z}^-$ has a zero. Applying any of them requires moving outside of the $\mathbb{Z}$-polytope.

Example 4.4.7. Consider the link-path incidence matrix and column partition lattice basis from Example 4.4.4. We had

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

and

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

The c0-Eulerian submatrix of concern was in the second and third columns, in rows one and two. The column multipliers necessary to make these entries cancel when we sum the columns are both 1. Setting $\mathbf{z} = \mathbf{u}_2 + \mathbf{u}_3$, we have

$$\mathbf{z} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Splitting $\mathbf{z}$ into positive and negative parts produces

$$\mathbf{z}^- = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{z}^+ = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

These are the endpoints of a potential walk in the $\mathbb{Z}$-polytope for $\mathbf{y} = A\mathbf{z}^- = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^\intercal$. The moves in U necessary for walking between $\mathbf{z}^-$ and $\mathbf{z}^+$ are $\mathbf{u}_2$ and $\mathbf{u}_3$. There are two possible orderings of these moves, shown in Table 4.1. We cannot walk from $\mathbf{z}^-$ to $\mathbf{z}^+$

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{u}_2$	$\mathbf{u}_3$	$-\mathbf{z}^+$
0	0	−1	1	0
0	0	1	−1	0
1	1	−1	0	0
0	0	0	0	0
0	0	0	0	0
−1	0	1	0	−1
−1	0	0	1	−1

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{u}_3$	$\mathbf{u}_2$	$-\mathbf{z}^+$
0	0	1	−1	0
0	0	−1	1	0
1	1	0	−1	0
0	0	0	0	0
0	0	0	0	0
−1	0	0	1	−1
−1	0	1	0	−1

Table 4.1: The two potential orderings of moves from the walk in Example 4.4.7.

using either of these orderings of moves while remaining within the $\mathbb{Z}$-polytope because of the -1s in $\mathbf{u}_2$ and $\mathbf{u}_3$ where $\mathbf{z}^-$ has a zero. The fact that in the c0-Eulerian submatrix M, each column contains entries that sum to zero ensures that each column contains a -1. The fact that it is Eulerian and each of the other columns also contains a -1 ensures that the signed sum of the entries in each row of M is zero. Together these conditions mean that we cannot use $\mathbf{u}_2$ and $\mathbf{u}_3$ to walk directly from $\mathbf{z}^-$ to $\mathbf{z}^+$.

This works for larger Eulerian submatrices as well.

Example 4.4.8. Recall the system in Example 4.4.6. The configuration matrix was

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

and the column partition lattice basis chosen was

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

The third, fifth, and sixth rows of U were made up of the 3×3 c0-Eulerian submatrix

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

When we assign the multipliers $\epsilon_1 = \epsilon_2 = 1, \epsilon_3 = -1$ to the columns of U and take the sum, the entries in these rows cancel and we have

$$\mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

Splitting $\mathbf{z}$ into its positive and negative parts produces two points in $\mathcal{F}_{\mathbf{y}}$ for $\mathbf{y} = A\mathbf{z}^-$ such that a walk connecting them using U requires the moves $\mathbf{u}_1, \mathbf{u}_2, -\mathbf{u}_3$, shown in Table 4.2. None of these moves can be applied to $\mathbf{z}^-$, so a direct walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ using U is not possible.

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{u}_3$	$-\mathbf{z}^+$
-1	0	1	0	0	-1
-1	0	0	1	0	-1
0	0	0	1	-1	0
1	1	0	0	-1	0
0	0	1	-1	0	0
0	0	-1	0	1	0

Table 4.2: The potential walk between $\mathbf{z}^-$ and $\mathbf{z}^+$ from Example 4.4.8.

4.4.3 Combinatorial Lemmata

In this section we state and prove several lemmata required for the proof of Theorem 4.4.14, which can be found in Section 4.4.4. Theorem 4.4.14 is the condition for whether a column partition lattice basis U is a Markov basis. The basic idea is that if U contains a c0-Eulerian submatrix, then any collection of columns of U where each column is multiplied by ± 1 can be ordered so that in each row, the 1s precede the -1s. If the collection of columns gives the moves required to move between two points in some $\mathbb{Z}$-polytope, then under such an ordering the walk given by the ordering never crosses any of the bounding hyperplanes. This means that it never visits a point with a negative co-ordinate. If U has only $\{0, \pm 1\}$ entries, then the walk visits only points with integer co-ordinates. Because this is true for any collection of moves, and therefore any pair of points in any fibre, U is a Markov basis.

We start by proving several lemmata concerning c0-Eulerian matrices. It is possible to assign ± 1 multipliers to the columns of a c0-Eulerian matrix to obtain a cr0-Eulerian matrix, so a matrix recording a collection of steps in a walk using U may have a cr0-Eulerian matrix as a maximal submatrix.

The lemmata look at how the columns of ± 1 multiples of c0-Eulerian matrices may be ordered to conform to certain conditions; this corresponds to orderings of steps in a walk in a $\mathbb{Z}$-polytope and the conditions guarantee that after each step, we are at a point in the $\mathbb{Z}$-polytope (that is, a point with non-negative integer co-ordinates).

The notation used in the proof is as follows: $\mathbf{m}_{i\cdot}$ denotes the ith row of the matrix M, and $\mathbf{m}_{\cdot j}$ denotes the jth column. The notation $m_{i,j}$ denotes the element in the ith row and jth column of M, or alternatively the ith element of the vector $\mathbf{m}_{\cdot j}$.

We begin with a lemma concerning cr0-Eulerian matrices.

Lemma 4.4.9. *A matrix $U \in \{0, \pm 1\}^{n \times r}$ that contains a cr0-Eulerian submatrix cannot have its columns reordered so that the 1s all precede the -1s in every row.*

Proof. The proof is by contradiction. Let U be a matrix containing a cr0-Eulerian submatrix, V. Suppose that we can reorder the columns of U such that 1s all precede the -1s in every row.

Under this reordering, let $\mathbf{v}_{\cdot j}$ be the first column of V that contains a non-zero entry. The entries of $\mathbf{v}_{\cdot j}$ sum to zero, so it must contain at least one 1 and one -1.

Let i be an index such that $v_{i,j} = -1$. The entries of the ith row also sum to zero, so the ith row must also contain a 1.

But $\mathbf{v}_{\cdot j}$ is the first column of V that contains non-zero entries, and so the 1 in the ith row must be in a column that succeeds the jth column.

Therefore, the ith row of V has a -1 that precedes a 1. Since V is a submatrix of U, it also has a row with a -1 that precedes a 1. This contradicts our hypothesis, and so U cannot be ordered so the the 1s precede the -1s in every row. □

We now prove the converse.

Lemma 4.4.10. *A matrix $U \in \{0, \pm 1\}^{n \times r}$ that cannot have its columns reordered so that the 1s all precede the -1s in every row must contain a cr0-Eulerian submatrix.*

Proof. Let $<$ over $\{0, \pm 1\}^r$ define a binary relation such that $\mathbf{u} < \mathbf{v}$ if $u_k = 1$ and $v_k = -1$ for some $k \in \{1, \ldots, n\}$. Let $\prec$ be the transitive closure of $<$. It is therefore a binary relation over $\{0, \pm 1\}^r$ where $\mathbf{u} \prec \mathbf{v}$ if $\mathbf{u} < \mathbf{v}$ or if $\mathbf{u} < \cdots < \mathbf{v}$.

If $\prec$ is a strict partial order over $\mathcal{U}$, the set of columns of U, then any reordering of $\mathcal{U}$ that conforms to $\prec$ satisfies the condition that no -1 precedes a 1 in any row. Conversely, if $\prec$ is not a strict partial order over $\mathcal{U}$, then no such reordering of columns is possible. Recall that a strict partial order $\prec$ is a binary relation over a set $\mathcal{S}$ with the following properties:

1. $\forall s \in \mathcal{S} : \neg(s \prec s)$ (*irreflexivity*)
2. $\forall r, s, t \in \mathcal{S} : (r \prec s) \wedge (s \prec t) \Rightarrow (r \prec t)$ (*transitivity*)

Suppose that $\prec$ does not define a strict partial ordering over $\mathcal{U}$. Then the irreflexivity property must have been violated: transitivity cannot have been violated because it is part of how we defined $\prec$. If irreflexivity has been violated and some $\mathbf{u}_{\cdot j_1} \prec \mathbf{u}_{\cdot j_1}$, then either $\mathbf{u}_{\cdot j_1} < \mathbf{u}_{\cdot j_1}$, or $\mathbf{u}_{\cdot j_1} < \cdots < \mathbf{u}_{\cdot j_m} < \mathbf{u}_{\cdot j_1}$ for some $m \in \mathbb{Z}^+$. It cannot be $\mathbf{u}_{\cdot j_1} < \mathbf{u}_{\cdot j_1}$, since that would imply that there is an index k such that $u_{k,j_1} = 1$ and $u_{k,j_1} = -1$, which is a contradiction. Therefore the second condition must be true, and we choose $\{\mathbf{u}_{\cdot j_1}, \ldots, \mathbf{u}_{\cdot j_m}\}$ such that m is the minimum over all such sets. In what follows we consider the indices on the j_k modulo m, so that $j_{m+1} = j_1$.

We claim that each of the j_k are distinct. If the j_k were not distinct and $j_p = j_q$ for some $p < q$, then we would have $\mathbf{u}_{\cdot j_1} < \cdots < \mathbf{u}_{\cdot j_p} < \mathbf{u}_{\cdot j_{q+1}} < \cdots < \mathbf{u}_{\cdot j_m} < \mathbf{u}_{\cdot j_1}$ and m would not be minimal.

From the definition of $<$, for every $k \in \{1, \ldots, m\}$ there is i_k such that $u_{i_k,j_k} = 1$ and $u_{i_k,j_{k+1}} = -1$. We claim that each of the i_k are distinct. If the i_k were not distinct and $i_p = i_q$ for some $p < q$, then we would have $u_{i_p,j_{p-1}} = 1$ and $u_{i_p,j_p} = -1$, and $u_{i_q,j_{q-1}} = 1$, and $u_{i_q,j_q} = -1$. Then we would have $\mathbf{u}_{\cdot j_1} < \cdots < \mathbf{u}_{\cdot j_{p-1}} < \mathbf{u}_{\cdot j_q} < \cdots < \mathbf{u}_{\cdot j_m} < \mathbf{u}_{\cdot j_1}$ and m would not be minimal.

We construct a submatrix V of U by taking the i_kth rows and j_kth columns of U for $k \in \{1, \ldots, m\}$. We claim that the entries of this matrix not already defined by $v_{i_k,j_k} = 1$ and $v_{i_k,j_{k+1}} = -1$ are all 0. To the contrary, if one of these entries $v_{i_q,j_q} = 1$ where $p \neq q$ and $p + 1 \neq q$, then $\mathbf{u}_{\cdot j_q} < \mathbf{u}_{\cdot j_{p+1}}$, which if $q < p$ produces $\mathbf{u}_{\cdot j_1} < \cdots < \mathbf{u}_{\cdot j_q} < \mathbf{u}_{\cdot j_{p+1}} < \cdots < \mathbf{u}_{\cdot i_m} < \mathbf{u}_{\cdot i_1}$; or if $p < q$ produces $\mathbf{u}_{\cdot j_q} < \cdots < \mathbf{u}_{\cdot j_{p+1}} < \mathbf{u}_{\cdot j_q}$ and m was not minimal. On the other hand, if one of these entries $v_{i_q,j_q} = -1$ where $p \neq q$ and $p + 1 \neq q$, then $\mathbf{u}_{\cdot j_p} < \mathbf{u}_{\cdot j_q}$, which if $p < q$ produces $\mathbf{u}_{\cdot j_1} < \cdots < \mathbf{u}_{\cdot j_p} < \cdots < \mathbf{u}_{\cdot j_q} < \cdots < \mathbf{u}_{\cdot j_m} < \mathbf{u}_{\cdot j_1}$, and m was not minimal; or if $q < p$ produces $\mathbf{u}_{\cdot j_q} < \cdots < \mathbf{u}_{\cdot j_p} < \mathbf{u}_{\cdot j_q}$, and again m was not minimal.

Then the matrix V contains rows and columns that each contain one 1 and one -1, with all other entries equal to 0, and so the sum of each row or column of V is 0. This means V is a cr0-Eulerian matrix. This matrix V is a submatrix of U, so U contains a cr0-Eulerian submatrix. $\square$

Lemmata 4.4.9 and 4.4.10 are combined into the following theorem.

Theorem 4.4.11. *A matrix with $\{0, \pm 1\}$ entries cannot have its columns reordered so that the* 1*s all precede the* -1*s in every row if and only if it contains a cr*0*-Eulerian submatrix.*

Having established this result, we now establish which column partition lattice basis-defining matrices it is applicable to.

Theorem 4.4.12. *Let $U \in \{0, \pm 1\}^{n \times r}$ be a matrix that contains no c*0*-Eulerian submatrix. Let U^σ be a matrix obtained from U by independently multiplying each column of U by ± 1. Then U^σ contains no cr*0*-Eulerian submatrix.*

Proof. Suppose that the theorem is false, and we have a matrix U^σ that contains a cr0-Eulerian submatrix V^σ. All entries of V^σ are in $\{0, \pm 1\}$, so in each row the count of 1s must equal the count of -1s. Therefore, each row of V^σ has an even number of non-zero entries.

Since U^σ was constructed by multiplying columns of U by ± 1, the original U can be found by performing the same multiplications on U^σ, and U contains a submatrix V corresponding to V^σ.

The sum of each column of V^σ is 0. Multiplying any particular column of V^σ by ± 1 does not change this, so each column of V sums to 0.

Each row of V^σ contains an even number of non-zero entries, so each row of V contains an even number of non-zero entries. The count of 1s and the count of -1s in any row of U_M must be either both even, or both odd. Therefore the sum of the entries of each row of V must be a multiple of 2.

The matrix U must therefore contain a non-zero submatrix V whose columns sum to 0 and whose rows sum to a multiple of 2. This submatrix is therefore c0-Eulerian, and we have a contradiction. □

4.4.4 Main theorem

We can now give a condition on a column partition lattice basis $U \in \{0, \pm 1\}^{r \times (r-n)}$ of a unimodular configuration matrix that guarantees it will be a Markov basis.

Theorem 4.4.13. *Let U be a column partition lattice basis for a unimodular configuration matrix $A_{n \times r}$, and let U contain no c0-Eulerian submatrix. Then U is a Markov basis.*

Proof. By Corollary 4.3.2, we need only show that U connects the positive and negative parts $\mathbf{z}^-$ and $\mathbf{z}^+$ of each $\mathbf{z}$ in some known Markov basis for A. We choose $\mathcal{G}_A$, the Graver basis of A. For a unimodular A, each $\mathbf{z} \in \mathcal{G}_A$ has all entries in $\{0, \pm 1\}$.

Choose any $\mathbf{z} \in \mathcal{G}_A$. Since U is a column partition lattice basis for $\ker_{\mathbb{Z}}(A)$ we can write $\mathbf{z}$ as a combination of $\{0, \pm 1\}$ multiples of a subset of columns of U, or

$$\mathbf{z} = \sum_{i=1}^{r-n} \epsilon_i \mathbf{u}_i,$$

where $\epsilon_i \in \{0, \pm 1\}$ gives either the required multiplier.

We can construct a new matrix U^ϵ by multiplying the ith column of U by the corresponding sign ϵ_i and concatenating these columns. We exclude columns with a multiplier of zero. If this matrix has k columns and $\mathbf{u}_i^\epsilon$ is the ith column of U^ϵ, then we can write

$$\mathbf{z} = \sum_{i=1}^{k} \mathbf{u}_i^\epsilon$$

without requiring signs. The matrix U contains no c0-Eulerian submatrix, so by Theorem 4.4.12, U^ϵ contains no cr0-Eulerian submatrix. By Theorem 4.4.11, we can reorder the columns of U^ϵ such that no -1 precedes a 1 in any row.

This ordering of the columns of U^ϵ gives the order in which the moves in U should be applied to get from $\mathbf{z}^-$ to $\mathbf{z}^+$ without leaving the fibre. After each step we are at an integer point, and using this ordering of the columns means that in every co-ordinate, any moves that decrease the entry come after the moves that increase the entry. If after some step we are at position $\hat{\mathbf{x}}$, then the ith entry of $\hat{\mathbf{x}}$ is an integer greater than or equal to $\min(z_i^-, z_i^+) \geq 0$. We can do this for any $\mathbf{z} \in \mathcal{G}_A$, a known Markov basis, and therefore U is also Markov basis. □

The idea extends naturally to column partition lattice bases $U \in \{0, \pm 1\}^{n\times r}$ of non-unimodular matrices, using Corollary 4.3.3.

Theorem 4.4.14. *Let $U \in \{0, \pm1\}^{r\times r-n}$ be a column partition lattice basis for a configuration matrix $A_{n\times r}$, and let U contain no c0-Eulerian submatrix. Then U is a Markov basis.*

Proof. By Corollary 4.3.3, we need only show that U connects the positive and negative parts $\mathbf{z}^-$ and $\mathbf{z}^+$ of each $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$.

Choose any such $\mathbf{z}$. Since U is a lattice basis for $\ker_{\mathbb{Z}}(A)$ we can write

$$\mathbf{z} = \sum_{i=1}^{r-n} \epsilon_i a_i \mathbf{u}_i,$$

where $\epsilon_i \in \{\pm 1\}$ gives the sign and $\mathbf{a} \in \mathbb{Z}_+^{r-n}$ gives the number of copies of $\mathbf{u}_i$ required.

We can construct a new matrix U^ϵ by multiplying the ith column of U by the corresponding sign ϵ_i. Then we can write

$$\mathbf{z} = \sum_{i=1}^{r-n} a_i \mathbf{u}_i^\epsilon$$

without a sign. From Theorem 4.4.12, this matrix U^ϵ contains no cr0-Eulerian submatrix. By Theorem 4.4.11, we can reorder the columns of U^ϵ such that no -1 precedes a 1 in any row.

This ordering of the columns of U^ϵ gives the order in which the moves in U should be applied to get from $\mathbf{z}^-$ to $\mathbf{z}^+$ without leaving the fibre: after each step we are at an integer point, and the ordering of the columns means that in every co-ordinate, any moves that decrease the entry come after the moves that increase the entry. If after some step we are at position $\hat{\mathbf{x}}$, then the ith entry of $\hat{\mathbf{x}}$ is an integer greater than or equal to $\min(z_i^-, z_i^+) \geq 0$. We can do this for any $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$, therefore U is a Markov basis. □

4.4.5 Ideal membership

The Fundamental Theorem of Markov Bases (Theorem 2.4.5) says that a set $\mathcal{B}$ connects two points $\mathbf{a}, \mathbf{b} \in \mathcal{F}_\mathbf{y}$ if and only if

$$T^\mathbf{a} - T^\mathbf{b} \in I_\mathcal{B},$$

where $I_{\mathcal{B}}$ is the ideal generated by the monomial difference representations of the elements of $\mathcal{B}$.

According to Theorem 4.4.14, if U is a column partition lattice basis containing only $\{0, \pm 1\}$ entries, and it contains no c0-Eulerian submatrix, then U is a Markov basis. We can use the property of ordering columns of c0-Eulerian matrices in Theorem 4.4.11 to show that if $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$, then $T^{\mathbf{a}} - T^{\mathbf{b}} \in I_U$ when U has this property. By Corollary 4.3.3, we need only show this when $\mathbf{a}$ and $\mathbf{b}$ have disjoint support, so we will write $\mathbf{a} = \mathbf{z}^-$ and $\mathbf{b} = \mathbf{z}^+$, and $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$.

Let $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$ be given, and write

$$\mathbf{z} = \sum_{i=1}^{r-n} \epsilon_i a_i \mathbf{u}_i.$$

We write U^ϵ for the matrix made by concatenating the columns $\epsilon_i \mathbf{u}_i$ such that $a_i \neq 0$. This matrix U^ϵ contains ± 1 multiples of columns of U, which is a matrix with no c0-Eulerian submatrix. By Theorems 4.4.11 and 4.4.12, the columns of U^ϵ can be reordered so that the 1s precede the -1s in every row. Let V be the matrix containing the columns of U^ϵ in such an order.

We correspondingly reorder the elements of $\mathbf{a}$ so that

$$\begin{aligned}\mathbf{z} &= \sum_{i=1}^{r-n} a_i \mathbf{v}_i \\ &= V\mathbf{a}.\end{aligned}$$

Using V as a set of moves means using a subset of the moves in U. We can now show ideal membership using induction on the vector $\mathbf{a}$. Let

$$I_V = \langle T^{\mathbf{v}_i^+} - T^{\mathbf{v}_i^-} : i = 1, \ldots, r-n \rangle$$

be the ideal generated by the monomial difference representations of the columns of V. Then $I_V \subseteq I_U$. First the base case: suppose that $\mathbf{a} = \mathbf{e}_k$, the vector with a 1 at the kth co-ordinate and 0 elsewhere, so that $\mathbf{z} = V\mathbf{e}_k = \mathbf{v}_k$. Then

$$\mathbf{z}^+ = \mathbf{v}_k^+ \text{ and } \mathbf{z}^- = \mathbf{v}_k^-$$

and we can see that

$$T^{\mathbf{z}^+} - T^{\mathbf{z}^-} = T^{\mathbf{v}_k^+} - T^{\mathbf{v}_k^-} \in I_V.$$

Now the induction: suppose that for some $\mathbf{z} = V\mathbf{a}$ with $\mathbf{a} \in \mathbb{Z}^r_{\geq 0}$ we know that

$$T^{\mathbf{z}^+} - T^{\mathbf{z}^-} \in I_V.$$

Consider $\mathbf{x} = V(\mathbf{a} + \mathbf{e}_i) = \mathbf{z} + \mathbf{v}_i$ for some i such that $i \geq j$ for all $j \in \text{supp}(\mathbf{a})$. This restriction on i together with the ordering of the columns of V means that $V\mathbf{e}_1$ cannot contain a positive value where $V\mathbf{a}$ contains a negative value, which is to say

$\text{supp}(\mathbf{v}_1^+) \cap \text{supp}(\mathbf{z}^-) = \varnothing$. If $\mathbf{z}$ contains a negative entry $\mathbf{z}^{(k)}$, then at least one of $\mathbf{u}_j^{(k)}$ for some $j \leq i$ must have been negative, and so $\mathbf{u}_i^{(k)}$ must not be positive.

We need to show that $T^{\mathbf{x}^+} - T^{\mathbf{x}^-} \in I_V$ too. Given that $\mathbf{x}^+ - \mathbf{v}_i = \mathbf{x}^- + \mathbf{z}$, we have

$$\begin{aligned} T^{\mathbf{x}^+} - T^{\mathbf{x}^-} &= T^{\mathbf{x}^+} - T^{\mathbf{x}^+ - \mathbf{v}_i} + T^{\mathbf{x}^+ - \mathbf{v}_i} - T^{\mathbf{x}^-} \\ &= T^{\mathbf{x}^+} - T^{\mathbf{x}^+ - \mathbf{v}_i} + T^{\mathbf{x}^- + \mathbf{z}} - T^{\mathbf{x}^-} \\ &= T^{\mathbf{x}^+ - \mathbf{v}_i^+ + \mathbf{v}_i^+} - T^{\mathbf{x}^+ - \mathbf{v}_i^+ + \mathbf{v}_i^-} + T^{\mathbf{x}^- - \mathbf{z}^- + \mathbf{z}^+} - T^{\mathbf{x}^- - \mathbf{z}^- + \mathbf{z}^-} \\ &= T^{\mathbf{x}^+ - \mathbf{v}_i^+}(T^{\mathbf{v}_i^+} - T^{\mathbf{v}_i^-}) + T^{\mathbf{x}^- - \mathbf{z}^-}(T^{\mathbf{z}^+} - T^{\mathbf{z}^-}) \end{aligned}$$

which is in the ideal I_V if $\mathbf{x}^+ - \mathbf{v}_i^+$ and $\mathbf{x}^- - \mathbf{z}^-$ are both non-negative.

Taking the positive part of $\mathbf{x}$ shows

$$\begin{aligned} \mathbf{x} &= \mathbf{z} + \mathbf{v}_i \\ &= \mathbf{z}^+ - \mathbf{z}^- + \mathbf{v}_i^+ - \mathbf{v}_i^- \\ \mathbf{x}^+ &= (\mathbf{z}^+ - \mathbf{z}^- + \mathbf{v}_i^+ - \mathbf{v}_i^-)^+. \end{aligned}$$

The fact that $\mathbf{z}^-$ has disjoint support with both positive components $\mathbf{z}^+$ and $\mathbf{v}_i^+$ mean that it contributes nothing to the positive part of the vector and can be ignored.

$$\mathbf{x}^+ = (\mathbf{z}^+ + \mathbf{v}_i^+ - \mathbf{v}_i^-)^+.$$

The vector $\mathbf{v}_i^+$ has disjoint support with $\mathbf{v}_i^-$, so it can be taken outside the brackets.

$$\begin{aligned} \mathbf{x}^+ &= \mathbf{v}_i^+ + (\mathbf{z}^+ - \mathbf{v}_i^-)^+ \\ \mathbf{x}^+ - \mathbf{v}_i^+ &= (\mathbf{z}^+ - \mathbf{v}_i^-)^+ \\ &\geq \mathbf{0}. \end{aligned}$$

Similarly for the negative part of $\mathbf{x}$,

$$\begin{aligned} \mathbf{x}^- &= (\mathbf{z}^+ - \mathbf{z}^- + \mathbf{v}_i^+ - \mathbf{v}_i^-)^- \\ &= (\mathbf{z}^+ - \mathbf{z}^- - \mathbf{v}_i^-)^- \\ &= \mathbf{z}^- + (\mathbf{z}^+ - \mathbf{v}_i^-)^- \\ \mathbf{x}^- - \mathbf{z}^- &= (\mathbf{z}^+ - \mathbf{v}_i^-)^- \\ &\geq \mathbf{0}, \end{aligned}$$

and so both of these vectors are non-negative. Then

$$T^{\mathbf{x}^+} - T^{\mathbf{x}^-} \in I_V$$

which means that $\mathbf{x}^+$ and $\mathbf{x}^-$ are connected by U by induction. This is true of all vectors $\mathbf{x}^+, \mathbf{x}^-$ with disjoint support such that $A\mathbf{x}^+ = A\mathbf{x}^-$, and so by Theorem 4.4.11, U is a Markov basis.

We will demonstrate this with the three-link linear network.

Example 4.4.15. Consider the three-link linear network, whose configuration matrix is given by

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

and a column partition lattice basis is given by

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

This matrix contains no c0-Eulerian submatrix. The monomial difference representation of U is

$$\{t_1t_2 - t_4, t_5 - t_2t_3, t_6 - t_3t_4\}.$$

We choose arbitrary column multipliers ϵ_i so that $\epsilon = \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}^\intercal$, and

$$U^\epsilon = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix has a -1 preceeding a 1 in the second row, so we reorder these columns to get

$$V = \begin{bmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where the 1s preceed the -1s in every row. This matrix V defines the same set of moves that U does.

We choose an arbitrary non-negative integer vector $\mathbf{a} = \begin{bmatrix} 2 & 3 & 0 \end{bmatrix}^\intercal$, which generates the kernel element

$$\begin{aligned} \mathbf{z} &= V\mathbf{a} \\ &= \begin{bmatrix} -3 & -1 & 2 & 3 & -2 & 0 \end{bmatrix}^\intercal, \end{aligned}$$

which has the monomial difference representation $T^{\mathbf{z}^+} - T^{\mathbf{z}^-} = t_3^2t_4^3 - t_1^3t_2t_5^2$.

For our induction, we need to choose a vector $\mathbf{e}_i$ such that $i \geq j$ for all $j \in \text{supp}(\mathbf{a})$: both $\mathbf{e}_2$ and $\mathbf{e}_3$ meet this condition. We will assume that $T^{\mathbf{z}^+} - T^{\mathbf{z}^+} \in I_U$ and show that both $T^{(\mathbf{z}+\mathbf{v}_2)^+} - T^{(\mathbf{z}+\mathbf{v}_2)^-} \in I_U$ and $T^{(\mathbf{z}+\mathbf{v}_3)^+} - T^{(\mathbf{z}+\mathbf{v}_3)^-} \in I_U$. For $\mathbf{e}_2$, we have

$$\begin{aligned} T^{(\mathbf{z}+\mathbf{v}_2)^+} - T^{(\mathbf{z}+\mathbf{v}_2)^-} &= t_3^2t_4^4 - t_1^4t_2^2t_5^2 \\ &= t_3^2t_4^4 - t_1t_2t_3^2t_4^3 + t_1t_2t_3^2t_4^3 - t_1^4t_2^2t_5^2 \\ &= t_3^2t_4^3(t_4 - t_1t_2) + t_1t_2(t_3^2t_4^3 - t_1^3t_2t_5^2) \\ &= t_3^2t_4^3(T^{\mathbf{v}_2^+} - T^{\mathbf{v}_2^-}) + t_1t_2(T^{\mathbf{z}^+} - T^{\mathbf{z}^-}), \end{aligned}$$

and so $T^{(\mathbf{z}+\mathbf{v}_2)^+} - T^{(\mathbf{z}+\mathbf{v}_2)^-} \in I_U$.

For $\mathbf{e}_3$,

$$\begin{aligned} T^{(\mathbf{z}+\mathbf{v}_3)^+} - T^{(\mathbf{z}+\mathbf{v}_3)^-} &= t_3^3t_4^4 - t_1^3t_2t_5^2t_6 \\ &= t_3^3t_4^4 - t_3^2t_4^3t_6 + t_3^2t_4^3t_6 - t_1^3t_2t_5^2t_6 \\ &= t_3^2t_4^3(t_3t_4 - t_6) + t_6(t_3^2t_4^3 - t_1^3t_2t_5^2) \\ &= t_3^2t_4^3(T^{\mathbf{v}_3^+} - T^{\mathbf{v}_3^-}) + t_6(T^{\mathbf{z}^+} - T^{\mathbf{z}^-}), \end{aligned}$$

and so $T^{(\mathbf{z}+\mathbf{v}_3)^+} - T^{(\mathbf{z}+\mathbf{v}_3)^-} \in I_U$ too.

4.4.6 A geometric interpretation

Theorem 4.4.14 can also be as a generalisation of the idea presented in Section 4.2. In that section, Theorem 4.2.1 stated that for a configuration matrix A and partition of columns of A, if the entries of $A_1^{-1}A_2$ are non-negative, then U is a Markov basis. One way to understand this is to see that from any point in any associated fibre, it is always possible to use U to move in any dimension through the projected space towards the origin.

One example of this is the three-link linear network.

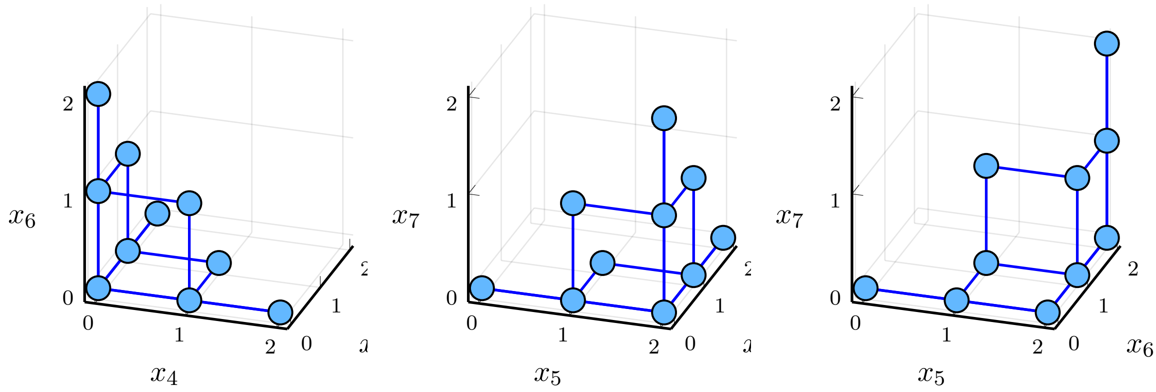
Example 4.4.16. Consider the configuration matrix

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

and the column partition lattice basis

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

Then $A_1^{-1}A_2$ is non-negative, and by Theorem 4.2.1 U is a Markov basis. The $\mathbb{Z}$-polytope for this system for $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^\intercal$ is shown in Figure 4.10a. We can see that from any



(a) The x_3 bounding hyperplane.

(b) The x_5 bounding hyperplane.

(c) The x_6 bounding hyperplane.

Figure 4.10: Three projected $\mathbb{Z}$-polytopes for which the corresponding column partition lattice basis is a Markov basis.

point in the $\mathbb{Z}$-polytope, we can always use co-ordinate direction moves to move towards the origin.

The same idea should work if the origin is not necessarily the privileged point, as shown by the following example.

Example 4.4.17. In Figure 4.10b, we see a rotation of the $\mathbb{Z}$-polytope from Example 4.4.16. In this projection of a $\mathbb{Z}$-polytope, from any point in this $\mathbb{Z}$-polytope, we can use co-ordinate direction steps to move towards the point at $\begin{bmatrix}2 & 0 & 0\end{bmatrix}^\intercal$ in the projected space. The column partition lattice basis corresponding to this projection should also be a Markov basis. This projected $\mathbb{Z}$-polytope may potentially arise for the configuration matrix

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

with the vector $\mathbf{y} = \begin{bmatrix}2 & 2 & 2 & 2\end{bmatrix}^\intercal$, when we choose the column partition lattice basis

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

In both Examples 4.4.16 and 4.4.17, the order of the moves used did not matter. However, the restrictions required on U matrices in order to be a Markov basis are weaker if we allow projections of $\mathbb{Z}$-polytopes where the order of the moves used to reach the privileged point does matter.

Example 4.4.18. In Figure 4.10c, we see a projection of a $\mathbb{Z}$-polytope where from any point we can always move towards, for example, the origin of the projected space using co-ordinate direction moves moves, provided they are ordered correctly. In this example, choosing x_7, then x_6, then x_5 directed moves will do. This $\mathbb{Z}$-polytope may appear when studying the configuration matrix

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

and the vector $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^\intercal$, and choosing the column partition lattice basis

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

We note that none of the U matrices in Examples 4.4.16, 4.4.17, or 4.4.18 contains a c0-Eulerian submatrix.

4.4.7 Reverse implication

Theorem 4.4.13 says that if a column partition lattice basis for a unimodular configuration matrix contains no c0-Eulerian submatrices, then it is a Markov basis. We conjecture that the reverse implication is also true.

Conjecture 4.4.19. *Let U be a column partition lattice basis for a unimodular configuration matrix, and suppose U contains a c0-Eulerian submatrix. Then U is not a Markov basis.*

This clearly holds when the c0-Eulerian submatrix is maximal.

Example 4.4.20. Consider the configuration matrix

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

which has a column partition lattice basis

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

Setting $\mathbf{y} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^\intercal$ produces the fibre

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

These vectors are respectively the negative and positive parts of $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$, so to walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ using U, we require the moves $\mathbf{u}_1$ and $\mathbf{u}_2$. Neither of these moves can be applied to $\mathbf{z}^-$, so U does not connect $\mathbf{z}^-$ to $\mathbf{z}^+$ and is not a Markov basis.

It is not so clear that the presence of a c0-Eulerian submatrix precludes U from being a Markov basis when the c0-Eulerian submatrix is not maximal. If some pair of points in some $\mathcal{F}_{\mathbf{y}}$ is not connected by U with a direct walk, it may still be possible to walk between them by performing a detour into some part of $\mathcal{F}_{\mathbf{y}}$ with more space to perform the required moves. Suppose that in Example 4.4.20, the basis contains another move $\mathbf{c}$ that had a 1 in the first entry. It may then be possible to conjugate the sequence of moves $(\mathbf{u}_1, \mathbf{u}_2)$ by $\mathbf{c} \in U$ in order to walk from $\mathbf{z}^-$ to $\mathbf{z}^+$. Then $\mathbf{y} = A\mathbf{z}^-$ would not provide a counterexample to U being a Markov basis.

In this section we will consider bases containing c0-Eulerian submatrices and attempt to find a conjugating move or moves. We refer to these moves with the vector $\mathbf{c}$, or $\mathbf{c}_i$, to distinguish from the required moves $\mathbf{u}_i$, although all $\mathbf{c}_i$ and $\mathbf{u}_i$ are columns of U, and therefore elements of the column partition lattice basis under consideration. We will assume:

1. U will define a potential column partition lattice basis for a unimodular configuration matrix.

2. U will contain a 2×2 cr0-Eulerian submatrix in $\mathbf{u}_1$ and $\mathbf{u}_2$.

3. The positive and negative parts of $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$ will be connected by U.

These conditions imply that U will be totally unimodular; that $\mathbf{z}$ and each column of U will each contain at least one 1 and one -1; and U and $\mathbf{z}$ will have all $\{0, \pm 1\}$ entries.

The matrix U defines a column partition lattice basis, so any conjugating move such as $\mathbf{c}$ has a 1 in the A_2 co-ordinates where every other move has a 0. This means that the moves $\mathbf{c}$ and $-\mathbf{c}$ must be used in that order, in order to avoid visiting a point that is negative in that co-ordinate. When initially listing the permutation of the moves used, the necessary moves $\mathbf{u}_i$ are interchangeable and we will adopt the convention that the lowest indexed move is used first. Each move $\mathbf{u}_i$ could also in fact be $-\mathbf{u}_i$: we use $\mathbf{u}_i$ by convention. The same is true for conjugating moves $\mathbf{c}_i$.

We begin with a three column U matrix. The non-maximal c0-Eulerian submatrix takes up exactly two columns, leaving only one column for conjugation.

Example 4.4.21. If U has three columns, the moves must be performed in the order $(\mathbf{c}, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c})$; any other ordering requires that either $\mathbf{u}_1$ or $\mathbf{u}_2$ is applied to $\mathbf{z}^-$ or $\mathbf{z}^+$, which we have claimed is impossible.

We have begun to fill in the columns of U in Table 4.3. We have included $-\mathbf{z}, \mathbf{z}^-$, and $-\mathbf{z}^+$ in this table: this makes it easier to see when the walk is in danger of visiting a point that is negative in some co-ordinate. The vector $\mathbf{z}$ is shown as its negation $-\mathbf{z}$ so that its entries match $\mathbf{z}^-$ and $-\mathbf{z}^+$. Then each row (excluding $-\mathbf{z}$) must sum to zero, and since $\mathbf{z}^-$ provides the starting point for the walk, the partial sums of the columns (excluding $-\mathbf{z}$) must always be non-negative. Therefore each row must have a 1 as its first non-zero entry, and a -1 as its last non-zero entry. We first place the cr0-Eulerian submatrix in

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{c}$	$-\mathbf{z}^+$
0	0		1	-1		0
0	0		-1	1		0
		-1			1	

Table 4.3: An attempted detour using a three column U matrix.

rows 1 and 2 of columns $\mathbf{u}_1$ and $\mathbf{u}_2$. We have $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$, so this forces the values of $\mathbf{z}, \mathbf{z}^-$, and $-\mathbf{z}^+$ in this co-ordinate. The vector $\mathbf{c}$ requires a -1, so we place this in row three. It may be that this -1 appears in the same row as the cr0-Eulerian submatrix: in this case, row three will end up being identical to row one or row two.

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{c}$	$-\mathbf{z}^+$
0	0		1	-1		0
0	0	1	-1	1	-1	0
1	1	-1			1	0

Table 4.4: An attempted detour using a three column U matrix.

Table 4.4 shows a more complete U. The second entry of $\mathbf{c}$ must be 1 to avoid a negative partial sum in this co-ordinate, and we can also add a corresponding -1 to $-\mathbf{c}$. The third entry of $\mathbf{z}^-$ be 1 must too, which forces $-\mathbf{z}^+$ to be zero. This means the third row ends with a 1, and as the row sums to zero this means the penultimate partial sum must be -1 in this co-ordinate. The sequence of moves $\mathbf{c}, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}$ does not connect the points $\mathbf{z}^-$ to $\mathbf{z}^+$ for $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$, and this was the only sequence of moves that potentially might connect them. Therefore U is not a Markov basis.

Furthermore, the cr0-Eulerian submatrix in U is the only non-maximal cr0-Eulerian submatrix of a three column matrix, and so Conjecture 4.4.19 holds for three column U matrices.

We will attempt to fill in four column U matrices that contain 2×2 cr0-Eulerian submatrices. With four columns, there are more possibilities for ordering the moves. There are 720 permutations for the moves $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}_1, -\mathbf{c}_2\}$. Many of these violate our convention that $\mathbf{u}_1$ precedes $\mathbf{u}_2$. Many are impossible, for example those that use $\mathbf{u}_1$ as the first move or use $-\mathbf{c}_1$ before $\mathbf{c}_1$. Removing all these leaves five permutations:

$$\begin{aligned}
P_1 &= (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}_2, -\mathbf{c}_1),\\
P_2 &= (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, -\mathbf{c}_2, \mathbf{u}_2, -\mathbf{c}_1),\\
P_3 &= (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, -\mathbf{c}_1, \mathbf{u}_2, -\mathbf{c}_2),\\
P_4 &= (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}_1, -\mathbf{c}_2),\\
P_5 &= (\mathbf{c}_1, \mathbf{u}_1, \mathbf{c}_2, -\mathbf{c}_1, \mathbf{u}_2, -\mathbf{c}_2).
\end{aligned}$$

Example 4.4.22. Consider the sequence of moves $P_1 = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}_2, -\mathbf{c}_1)$. In Table 4.5 we have begun to fill in some of the rows that may be necessary in any U matrix for this ordering of the moves. We have filled in the cr0-Eulerian submatrix and

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{c}_1$	$-\mathbf{z}^+$
0	0			1	-1			0
0	0			-1	1			0
			-1			1		
1	1	-1					1	0

Table 4.5: An attempted detour P_1 using a four column U matrix.

the required -1s in $\mathbf{c}_1$ and $\mathbf{c}_2$. The -1 in $\mathbf{c}_1$ forces a 1 in $\mathbf{z}^-$, which must have come from a -1 in $\mathbf{u}_1$ or $\mathbf{u}_2$.

This is sufficient to show that this ordering of moves cannot connect $\mathbf{z}^-$ and $\mathbf{z}^+$. In row four, the 1 in $-\mathbf{c}_1$ means that before taking this step, the partial sum must have been -1 in this co-ordinate. This sequence of moves cannot connect $\mathbf{z}^+$ and $\mathbf{z}^-$.

Example 4.4.23. Consider the sequence of moves $P_2 = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, -\mathbf{c}_2, \mathbf{u}_2, -\mathbf{c}_1)$. Table 4.6 shows that this ordering of moves suffers from the same problem as P_1. We have

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$-\mathbf{c}_2$	$\mathbf{u}_2$	$-\mathbf{c}_1$	$-\mathbf{z}^+$
0	0			1		-1		0
0	0			-1		1		0
1	1	-1					1	0

Table 4.6: An attempted detour P_2 using a four column U matrix.

filled in the cr0-Eulerian submatrix and the required -1s in $\mathbf{c}_1$. Again, this is sufficient to show that this ordering of moves cannot connect $\mathbf{z}^-$ and $\mathbf{z}^+$. In row three, the 1 in $-\mathbf{c}_1$ means that before taking this step, the partial sum must have been -1 in this co-ordinate. This sequence of moves cannot connect $\mathbf{z}^+$ and $\mathbf{z}^-$.

Example 4.4.24. Consider the sequence of moves $P_3 = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, -\mathbf{c}_1, \mathbf{u}_2, -\mathbf{c}_2)$. In Table 4.7 we have begun to fill in the potential entries of U.

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$-\mathbf{c}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0			1		−1		0
0	0			−1		1		0
			−1				1	
		−1			1			

Table 4.7: Two potential detours P_3 using a four column U matrix.

The -1 in $\mathbf{c}_1$ and the 1 in $-\mathbf{c}_2$ force a 1 in $\mathbf{z}^-$ and a -1 in $-\mathbf{z}^+$ respectively. These entries force the remaining entries of $\mathbf{z}^-$ and $\mathbf{z}^+$. Similarly, the third entry of $\mathbf{c}_1$ must be 1 and the second entry of $-\mathbf{c}_2$ must be -1, and the respective entries in $-\mathbf{c}_1$ and $\mathbf{c}_2$ can be filled as well. The potential U matrix as it now stands is shown in Table 4.8.

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$-\mathbf{c}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0			1		−1		0
0	0		1	−1		1	−1	0
−1	0	1	−1		−1		1	−1
1	1	−1			1			0

Table 4.8: Two potential detours P_3 using a four column U matrix.

The second entry of $\mathbf{c}_1$ must be 0: the row cannot start with a -1, and setting this entry to 1 would mean a non-unimodular submatrix. The first entry of $-\mathbf{c}_2$ cannot be ± 1 for similar reasons. The third entry of $\mathbf{u}_1$ must be 1 to avoid a negative entry in the partial sum. The third entry of $\mathbf{u}_2$ must therefore be 0, maintaining the value of $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$. All forced entries are now shown: the current state of U is shown in Table 4.9.

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$-\mathbf{c}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0		0	1		−1	0	0
0	0	0	1	−1	0	1	−1	0
−1	0	1	−1	1	−1	0	1	−1
1	1	−1			1			0

Table 4.9: Two potential detours P_3 using a four column U matrix.

We now branch on which of $\mathbf{u}_1$ and $\mathbf{u}_2$ are -1 and 0 in the fourth co-ordinate. The two cases are shown in Table 4.10. In each case, the remaining entries are forced by the need to avoid negative partial sums and non-unimodular matrices. This produces the

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$-\mathbf{c}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0	0	0	1	0	−1	0	0
0	0	0	1	−1	0	1	−1	0
−1	0	1	−1	1	−1	0	1	−1
1	1	−1	0	0	1	−1	0	0
0	0	0	0	1	0	−1	0	0
0	0	0	1	−1	0	1	−1	0
−1	0	1	−1	1	−1	0	1	−1
1	1	−1	1	−1	1	0	−1	0

Table 4.10: Two potential detours P_3 using a four column U matrix.

following two matrices, which may be submatrices of a column partition lattice basis:

$$U_1 = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & 1 & 0 \\ -1 & 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \end{bmatrix}.$$

For each of these bases, the points $\mathbf{z}^-$ and $\mathbf{z}^+$ are connected via an indirect path. However, in each case the conditions force the creation of more c0-Eulerian submatrices. Setting $\mathbf{z}_2 = \mathbf{c}_1 + \mathbf{c}_2 + \mathbf{u}_1 + \mathbf{u}_2$ produces a $\mathbf{z}_2^-$ with 0 in each of these four entries. Then $\mathbf{z}_2^+$ is another point in the same fibre, and U cannot connect this new pair of points without conjugating by some other moves. If the four columns shown are the only columns in U, then U is not a Markov basis, although $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$ does not provide the counterexample.

Example 4.4.25. Consider the sequence of moves $P_4 = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{u}_1, \mathbf{u}_2, -\mathbf{c}_1, -\mathbf{c}_2)$. In Table 4.11 we have begun to fill in the potential entries of U. In the third row we must

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{c}_1$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0			1	−1			0
0	0			−1	1			0
−1	0	1	−1			1	1	−1
1	1	−1	1			1	−1	0

Table 4.11: Two potential detours P_4 using a four column U matrix.

choose which of $\mathbf{u}_1$ and $\mathbf{u}_2$ contains a 0 and which contains a 1. This choice forces the rest of the entries, as shown in Table 4.12. This produces the following two matrices, which may be submatrices of a column partition lattice basis:

$$U_1 = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \end{bmatrix}.$$

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{u}_1$	$\mathbf{u}_2$	$-\mathbf{c}_1$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0	0	0	1	−1	0	0	0
0	0	1	0	−1	1	−1	0	0
−1	0	1	−1	0	1	1	1	−1
1	1	−1	1	0	−1	1	−1	0
0	0	0	0	1	−1	0	0	0
0	0	0	1	−1	1	0	−1	0
−1	0	1	−1	1	0	1	1	−1
1	1	−1	1	−1	0	1	−1	0

Table 4.12: Two potential detours P_4 using a four column U matrix.

For each of these bases, the points $\mathbf{z}^-$ and $\mathbf{z}^+$ are connected via an indirect path. As in Example 4.4.24, the conditions force the creation of more c0-Eulerian submatrices.

For U_1, we set $\mathbf{z}_1 = \mathbf{c}_2 + \mathbf{u}_1 + \mathbf{u}_2$. This produces $\mathbf{z}_1^-$ with 0 in each of these four entries. Neither a direct path, nor an indirect path conjugating by $\mathbf{c}_1$ can connect $\mathbf{z}_1^-$ and $\mathbf{z}_1^+$. The same is true for U_2 and $\mathbf{z}_2 = -\mathbf{c}_1 + \mathbf{u}_1 + \mathbf{u}_2$. Again, if the four columns shown are the only columns in U, then U can not be Markov basis, although $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2$ does not directly provide the counterexample.

Example 4.4.26. Consider the sequence of moves $P_5 = (\mathbf{c}_1, \mathbf{u}_1, \mathbf{c}_2, -\mathbf{c}_1, \mathbf{u}_2, -\mathbf{c}_2)$. All entries of these moves are forced. These are shown in Table 4.13. This means we are

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}_1$	$\mathbf{u}_1$	$\mathbf{c}_2$	$-\mathbf{c}_1$	$\mathbf{u}_2$	$-\mathbf{c}_2$	$-\mathbf{z}^+$
0	0	0	1	0	0	−1	0	0
0	0	1	−1	1	−1	1	−1	0
−1	0	0	1	−1	0	0	1	−1
1	1	−1	0	0	1	−1	0	0

Table 4.13: A potential detour P_5 using a four column U matrix.

dealing with a column partition lattice basis containing the submatrix

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

Setting $\mathbf{z} = -\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{u}_1 + \mathbf{u}_2$ provides us with points $\mathbf{z}^-$ and $\mathbf{z}^+$ that U does not connect: $\mathbf{z}^-$ has zeroes in each of the four indices shown, and so none of the moves shown in U can be applied to $\mathbf{z}^-$. If there are only four moves in the basis, then it is not a Markov basis.

We have shown that if a c0-Eulerian submatrix spans two columns of a four column basis U, then U cannot be a Markov basis. If U contains a maximal c0-Eulerian submatrix then it cannot be a Markov basis. In order to prove Conjecture 4.4.19 for four column U matrices, we need to show that it holds for three column c0-Eulerian submatrices.

If a c0-Eulerian submatrix spans three of the four columns in a column partition lattice basis, then there is only one column left to conjugate by. This suffers from the same problem as P_1 and P_2 in Examples 4.4.22 and 4.4.23 above. There is only one permutation of moves to consider: $(\mathbf{c}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, -\mathbf{c})$. There are different permutations of columns of the cr0-Eulerian submatrix involved, but we can skip these as only the -1 in $\mathbf{c}$ requirement is necessary to show that in this case, U does not connect $\mathbf{z}^-$ and $\mathbf{z}^+$.

Example 4.4.27. Consider the sequence of moves $P = (\mathbf{c}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, -\mathbf{c})$. Table 4.14 shows that under P, U does not connect $\mathbf{z}^-$ and $\mathbf{z}^+$ for $\mathbf{z} = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$. The 1 in $-\mathbf{c}$

$-\mathbf{z}$	$\mathbf{z}^-$	$\mathbf{c}$	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$-\mathbf{c}$	$-\mathbf{z}^+$
1	1	-1				1	

Table 4.14: An attempted detour P using a five column U matrix.

means $-\mathbf{z}^+$ must be -1 in this co-ordinate. However the -1 in $\mathbf{c}$ means we have already assigned a non-zero value to $\mathbf{z}^-$, and the parts of $\mathbf{z}$ must have disjoint support. This sequence of moves cannot connect $\mathbf{z}^+$ and $\mathbf{z}^-$.

With these examples we have proved Conjecture 4.4.19 for column partition lattice bases with up to four elements.

4.4.8 Extension to $\mathbb{Z}$

The idea of ordering columns so that the positive entries precede the negative entries in every row generalises to U matrices in $\mathbb{Z}^{r\times(r-n)}$ too. The requirement that U have integer entries is to avoid the problem of parity errors described in Section 2.5.1.

Remark 4.4.28. Let $U \in \mathbb{Z}^{r\times(r-n)}$ be column partition lattice basis. Let $U^\sigma \in \{0, \pm 1\}^{n\times r}$ be the matrix of signs of U, so that

$$u^{\sigma}{}_{i,j} = \begin{cases} u_{i,j} = 0 & 0, \\ u_{i,j} < 0 & -1, \\ u_{i,j} > 0 & 1. \end{cases}$$

If U^σ contains no c0-Eulerian submatrix, then U is a Markov basis.

The matrix U gives a lattice basis, given $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, the difference between any pair of points $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ can be written in the form

$$\mathbf{b} - \mathbf{a} = \sum_{i=1}^{r-n} a_i \epsilon_i \mathbf{u}_i.$$

A collection of U moves required to walk from $\mathbf{a}$ to $\mathbf{b}$ are given by this sum: each move $\mathbf{u}_i$ must be used in the ϵ_i orientation a_i times. This collection of moves can be ordered so that in each row, the positive entries precede the negative entries exactly if ± 1 multiples of the moves in U^σ can: that is, if U^σ contains c0-Eulerian submatrix. If this condition holds, then U is a Markov basis.

We give an example, but provide no configuration matrix.

Example 4.4.29. Consider the column partition basis U and the matrix U^σ, given by

$$U = \begin{bmatrix} -2 & -4 & 0 & -1 \\ 1 & 0 & 0 & -3 \\ 0 & -5 & -3 & 0 \\ -1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad U^\sigma = \begin{bmatrix} -1 & -1 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The matrix U^σ contains no c0-Eulerian submatrix so by Remark 4.4.28, U is a Markov basis.

Suppose that for some $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, the pair of points $\mathbf{a}, \mathbf{b} \in \mathcal{F}_{\mathbf{y}}$ has the difference

$$\mathbf{b} - \mathbf{a} = -2\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3.$$

The collection of moves required to walk from $\mathbf{a}$ to $\mathbf{b}$ is given by

$$\left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ -5 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \\ -2 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Writing $\mathbf{z} = -2\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3$, we have $\mathbf{a} \geq \mathbf{z}^-$ and $\mathbf{b} \geq \mathbf{z}^+$. We can order the moves as in Table 4.15 to get a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that visits only points with non-negative integer co-ordinates. Since $\mathbf{a} \geq \mathbf{z}^-$ and $\mathbf{b} \geq \mathbf{z}^+$, the walk from $\mathbf{a}$ to $\mathbf{b}$ that uses this ordering of moves stays within $\mathcal{F}_{\mathbf{y}}$.

If $U \in \mathbb{Z}^{r \times (r-n)}$ contains a c0-Eulerian submatrix, then there exists a pair of points in some fibre between which a direct walk is not possible.

Example 4.4.30. Consider the potential column partition lattice basis

$$U = \begin{bmatrix} 2 & 3 \\ -3 & -1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$-\mathbf{z}$	$\mathbf{z}^-$	$-\mathbf{u}_1$	$-\mathbf{u}_1$	$\mathbf{u}_3$	$-\mathbf{u}_2$	$-\mathbf{z}^+$
0	0	2	2	0	−4	0
2	2	−1	−1	0	0	0
2	2	0	0	3	−5	0
0	0	1	1	−2	0	0
2	2	−1	−1	0	0	0
−1	0	0	0	0	1	−1
1	1	0	0	−1	0	0
0	0	0	0	0	0	0

Table 4.15: A walk from **a** to **b** using an integral U as a Markov basis.

Then

$$U^\sigma = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This matrix has a c0-Eulerian submatrix in the first two rows. We can combine the two columns such that there is cancellation in both rows: $\mathbf{z} = \mathbf{u}_1 - \mathbf{u}_2$. Then $\mathbf{z} = \begin{bmatrix} -1 & -2 & -1 & 1 & 1 & -1 \end{bmatrix}^\intercal$ can be split into

$$\mathbf{z}^- = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{z}^+ = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Neither $\mathbf{u}_1$ nor $-\mathbf{u}_2$ can be applied to $\mathbf{z}^-$ without ending at a point with a negative co-ordinate, and so a direct walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that stays within the fibre is impossible.

4.5 Collections of lattice bases

We have seen in Section 3.4.3 that the union of the integer valued column partition lattice bases of some configuration matrix A is the set of circuits $\mathcal{C}_A$, and in Section 3.5.2 that if A is unimodular then $\mathcal{C}_A$ is equal to the Graver basis $\mathcal{G}_A$. This implies that the union of the lattice bases is a Markov basis.

Section 4.2 showed that there exist lattice bases that are Markov bases. An interesting question is given a configuration matrix, how many and which lattice bases do we need to combine to get a Markov basis, or the Graver basis?

We will first consider an example.

Example 4.5.1. Let A be the link-path incidence matrix of the three-link linear network, so

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

There are $\binom{6}{3} = 20$ possible partition choices of which 16 lead to invertible A_1 parts, and therefore 16 distinct column partition lattice bases. Each one contains three elements. There are seven distinct Graver basis elements:

$$\mathcal{G}_A = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Corollary 3.5.16 gives an upper limit on the size of the $\mathcal{G}_A$,

$$|\mathcal{G}_A| \leq \sum_{i=1}^{m} \binom{r-n}{i}$$

where $m = \min(r-n, n+1)$. Here, $m = \min(3,4) = 3$, so

$$\begin{aligned} |\mathcal{G}_A| &\leq \sum_{i=1}^{m} \binom{r-n}{i} \\ &\leq \sum_{i=1}^{3} \binom{3}{i} \\ &\leq 3+3+1 \\ &\leq 7 \end{aligned}$$

and the size of $\mathcal{G}_A$ is exactly this upper bound.

Each lattice basis has three elements, so to get the entire Graver basis we need to combine at least three lattice bases. It turns out that three lattice bases are sufficient. The two column partitions formed by taking the last three columns of A, and then the first three columns of A, as the A_1 part respectively produce

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 0 & -1 \\ -1 & -1 & -1 \\ 0 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which together give us the first six elements of $\mathcal{G}_A$ as listed above. The last element can be had by choosing any column partition lattice basis that contains it.

For this configuration matrix, only one lattice basis is required to get a Markov basis: Section 4.2 shows that the second of the above matrices is a Markov basis.

Suppose that for some reason we had access to the first of these two lattice bases but not to the second. We want an upper bound on the number of other lattice bases we need to combine with this basis get a Markov basis. Call this lattice basis $\mathcal{B}$. We know from Section 4.3 that some elements of $\mathcal{G}_A$ that are not in $\mathcal{B}$ may not be necessary for a Markov basis because we can simulate them with elements of $\mathcal{B}$. From Theorem 4.4.14 any element of $\mathcal{G}_A$ that cannot be simulated by $\mathcal{B}$ is a sum of columns of U that contain an Eulerian submatrix whose columns sum to zero. Therefore to get a Markov basis, we need only to combine these Graver basis elements with $\mathcal{B}$, and not necessarily the full Graver basis.

Chapter 5

Network tomography

5.1 Introduction

One application for our sampling methods is volume network tomography [25, 26, 45]. Volume network tomography is the art of estimating traffic flow volume on various potential journeys on a traffic network, given observed traffic volume at different locations on the network. In this chapter we will look at what properties the kinds of configuration matrices that arise in network tomography might have, particularly those that might affect column partition lattice bases.

We represent a real world network with a *digraph*, or directed graph, and a collection of *paths* on the graph.

Definition 5.1.1 (Digraph [31])**.** A digraph $\mathcal{D}$ consists of a non-empty finite set V together with a set E of ordered pairs of distinct elements of V.

The set V is the set of *vertices*, or *nodes*, of the digraph, and E is the set of *directed edges*, or *links*. This definition implies that a digraph has no loops (self-connected vertices), and that there is at most one directed link from any node to any other node. How the vertices and links of a digraph correspond to elements of the transport network being modelled varies, but generally the vertices represent cities, intersections, or stations; while the links represent the roads or rails connecting them.

Particular journeys in a network are represented by a kind of *walk* in the digraph called a *path*.

Definition 5.1.2 (Walk [31])**.** A *walk* in a digraph is a finite sequence of vertices $v_0, v_1, \ldots, v_k$, where for each $i = 0, 1, \ldots, k-1$, the pair (v_i, v_{i+1}) is a link.

Definition 5.1.3 (Path [31])**.** A *path* is a walk in which all the visited vertices are distinct.

This is often referred to as an acyclic path in the transport literature. We will in general be considering only networks on connected graphs.

The configuration matrices of interest here are the link-path incidence matrices.

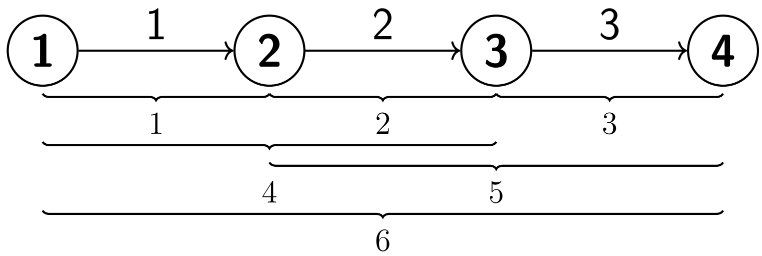


Figure 5.1: A representation of the three-link linear network, a traffic network with four nodes connected linearly by three links. The underbraces show the six allowed paths.

Definition 5.1.4 (Link-path incidence matrix)**.** The *link-path incidence matrix* of a traffic network with n links and r paths is a matrix $A \in \{0,1\}^{n \times r}$ such that $a_{ij} = 1$ if the ith link is in the jth path, and $a_{ij} = 0$ otherwise.

Different labellings of the network's links and paths correspond to reorderings of the configuration matrix's rows and columns. The structure of a link-path incidence matrix is partly dependent on the routing policy of the network, which is to say the choice of routes (or set of allowed paths). This is essentially arbitrary. This is unlike other matrices derived from graphs, such as vertex-edge incidence matrices, which is uniquely determined by the network topology. As the chapter progresses, we will at times look at link-path incidence matrices that arise given certain assumptions on the choice of routes. Such a set of assumptions will be termed a *routing policy.*

Example 5.1.5. Consider the three-link linear network, shown in Figure 5.1. This network previously appeared in Examples 2.4.6 and 3.2.1. Traffic on the network must follow the orientations of the links, so only eastward travel is possible.

This network could use nodes to represent cities and links to represent the highways connecting them. Suppose that one car is observed on each link in the network. We represent this with the vector $\mathbf{y} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$, where there are y_i cars observed on the ith link. There are multiple possible combinations of path car counts that could cause this. For example, there could be one car that drives the entire length of the network, or there could be three cars that drive a path consisting of only one link.

For each combination, we collect the number of cars driving each link into the vectors $\mathbf{x}$, where x_i records how many cars drive the ith path. The set $\mathcal{F}_\mathbf{y}$ gives the possible path traffic count vectors $\mathbf{x}$ that could result in the link traffic count vector $\mathbf{y}$. For $\mathbf{y} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$, we have

$$\mathcal{F}_\mathbf{y} = \left\{ \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\\0\\0\\0\end{bmatrix} \right\}.$$

Then each $\mathbf{x} \in \mathcal{F}_{\mathbf{y}}$ is related to $\mathbf{y}$ by the equation $A\mathbf{x} = \mathbf{y}$, where

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

is the link-path incidence matrix for the network.

We are interested in sampling from this set $\mathcal{F}_{\mathbf{y}}$ of path traffic counts given $\mathbf{y}$, the link traffic counts. This is a critical step in performing Bayesian network tomography — see for example Tebaldi and West [45] and Hazelton [25].

In Section 5.2 we consider what properties link-path incidence matrices might have in relation to unimodularity. We show using examples that they may or may not be unimodular or totally unimodular, and may or may not have unimodular partitions.

In Section 5.3 we look at traffic networks on a particular type of graph called a *polytree*. Polytrees include many important types of graph including linear networks and star networks. We find that all link-path incidence matrices of polytrees are *network matrices*, which are known to be totally unimodular. We give two new proofs that link-path incidence matrices of polytrees are totally unimodular. We also provide some observations concerning paths and circuits on polytrees that are of use in Chapter 6.

In Section 5.4 we consider a type of traffic network we call *symmetric directed network*. We begin with a type of symmetric directed network we call a peripheral bidirectional tree network, which can be thought of as a star network with some interior structure. Symmetric directed tree networks include the Monroe network [45], shown in Figure 5.15. We show how symmetric directed trees may be combined to form symmetric digraphs. We prove that the link-path incidence matrices of symmetric directed networks contain unimodular maximal submatrices.

5.2 Unimodularity

In Section 2.5.1 we discussed a type of problem for connectivity of fibres under column partition lattice bases that we called parity errors. We found that parity errors can be avoided if the determinant of the A_1 matrix divides the determinant of the other maximal submatrices; this is guaranteed if A_1 is unimodular. Moreover, no column partition lattice basis can suffer from parity errors if A as a whole is unimodular, meaning that every maximal invertible submatrix is unimodular. We are therefore interested in which traffic networks are unimodular, or have unimodular maximal submatrices.

Unimodularity and even total unimodularity seem to be common properties of link-path incidence matrices. Airoldi and Haas [2] conjectured that most reasonable link-path incidence matrices encountered in practice would be totally unimodular, and noted that they had not found a counterexample. Airoldi and Blocker [1] noted that all the link-path incidence matrices they encountered in the literature were unimodular.

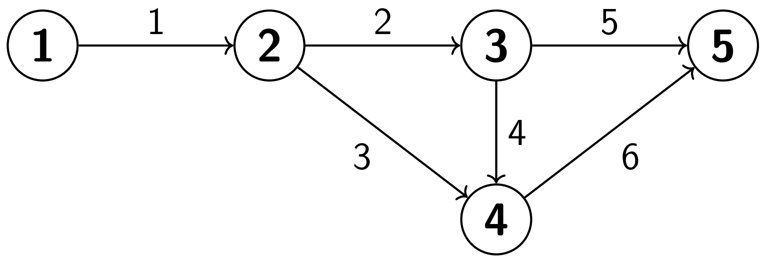


Figure 5.2: The graph of the network in Example 5.2.1.

Hazelton and Bilton [26] considered several networks that do not have unimodular link-path incidence matrices. Of those they looked at, the highest percentage of non-unimodular invertible maximal submatrices of any network was 3.7%. However, they were able to construct a network with a link-path matrix with no unimodular invertible maximal submatrices.

Example 5.2.1. Consider a traffic network on the graph shown in Figure 5.2. The link-path incidence matrix is given by

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

This matrix has seven maximal submatrices of which three are invertible: all three have determinant -2. This matrix has no unimodular partitions. Although A is not unimodular, its maximal invertible submatrices do have determinants that are equal in absolute value. This is, for the purposes of constructing a column partition lattice basis, as good as unimodularity. By Theorem 3.5.5, any column partition lattice basis will not only be integral but the U matrix that defines it will be totally unimodular. The kernel of A is one dimensional, so A has one circuit that comprises every column partition lattice basis for A. This is $\mathbf{c} = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 & -1 \end{bmatrix}^{\intercal}$.

Hazelton and Bilton note that the routing scheme in the network in Example 5.2.1 is quite perverse. For example, travel is permitted from node 1 to node 4 via node 2, and from node 2 to node 5 via node 4; so clearly node 2 may function as an origin for traffic that uses link 3, and node 4 may function as a destination for traffic that arrives via link 3. But traffic is not permitted to travel from node 2 to node 4 along link 3. Here is an example of a network with a more sensible routing scheme whose link-path incidence matrix has no unimodular maximal submatrices.

Example 5.2.2. Consider a traffic network on the digraph in Figure 5.3. The routing policy is as follows: the nodes marked with A, B, and C may function as both an origin

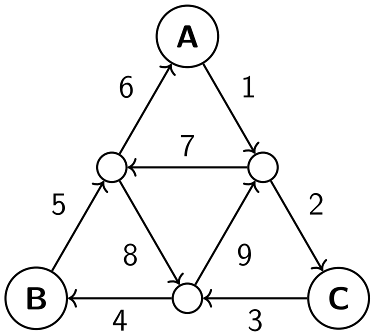


Figure 5.3: A triangular network whose configuration matrix contains no unimodular maximal submatrices.

or a destination for traffic. Any path from an origin to a destination that follows the directions on the links is permitted.

For traffic starting at node A, there are three possible paths: one path to each of nodes B and C that involve clockwise travel around the perimeter of the network, and one that zigzags through the centre of the network to node B. These three paths make up the first three columns of the configuration matrix. By symmetry, there are nine paths in the full network and the full configuration matrix is

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

This matrix has rank six, and so a maximal square submatrix after deleting dependent rows is 6×6. Checking the determinants of the invertible 6×6 submatrices with a computer shows that 800 have determinant ± 2 and 50 have determinant ± 4. There is no maximal unimodular submatrix.

There is no upper limit on the absolute value of the determinant of a maximal invertible submatrix of link-path incidence matrices.

Observation 5.2.3. The determinants of maximal submatrices of link-path incidence matrices may be arbitrarily large in absolute value.

Consider a traffic network on the directed cycle graph with n links C_n, where all paths of length $n-1$ are permitted routes. A maximal submatrix of a link-path incidence matrix

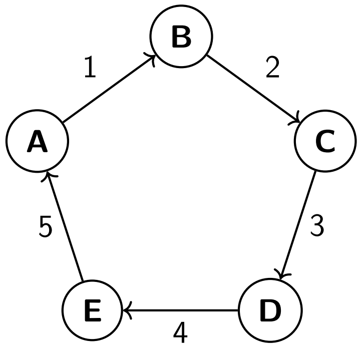


Figure 5.4: The cyclic graph C_5.

for a network on this graph is given by the $n \times n$ matrix

$$M = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 & 1 \\ 1 & 0 & 1 & \dots & 1 & 1 \\ 1 & 1 & 0 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{bmatrix}.$$

Matrices of this type have determinant $\det(M) = (-1)^{(n-1)}(n-1)$. For example, the cycle graph C_5 is shown in Figure 5.4. There are five paths of length four, one originating at each node. The part of the link-path incidence matrix corresponding to these paths is given by

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

This matrix has $\det(A) = 4$.

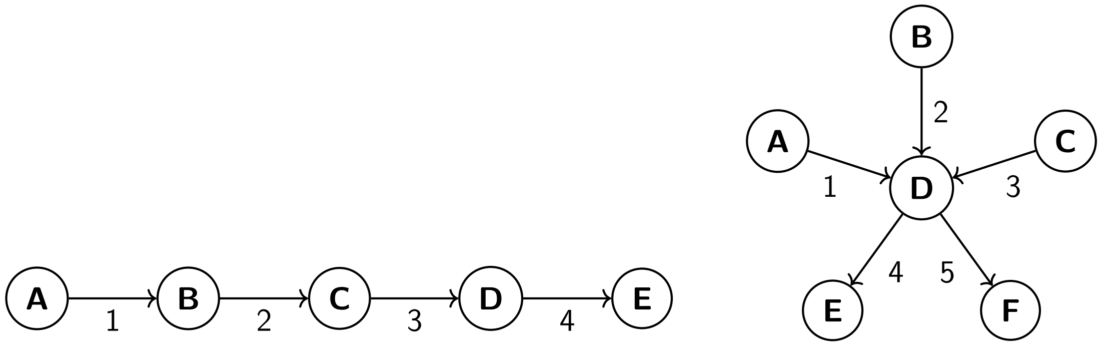
Hazelton and Bilton [26] did find two types of traffic network for which the link-path incidence matrices are guaranteed to be totally unimodular.

Proposition 5.2.4 (Hazelton and Bilton's Proposition 1)**.** *The link-path incidence matrix A is totally unimodular in the following cases:*

1. *The network is a linear highway in which we observe traffic counts on a sequence of unidirectional links connected in series.*

2. *The network has star topology, with routes connecting every pair of peripheral nodes via the central node.*

Some example linear and star networks are shown in Figure 5.5.



(a) A linear network.

(b) A star network.

Figure 5.5: Some examples of polytrees.

5.3 Polytrees

One type of graph we will consider is called a *polytree*. A polytree is a tree whose links are directed; it differs from a *directed tree* in that all of the links in a directed tree must be directed either towards or away from a particular vertex, called the root. Any directed tree is also a polytree. The graph in Figure 5.5b is a polytree but is not a directed tree.

Definition 5.3.1 (Polytree [14])**.** A polytree is a directed acyclic graph with the property that ignoring the directions on links yields a graph with no undirected cycles.

Polytrees therefore include linear networks and star networks, shown in Figure 5.5. Link-path incidence matrices for polytrees are examples of *network matrices*.

Definition 5.3.2 (Network matrix [40])**.** Let V be a collection of vertices, and let E and P be collections of directed links on V such that $\mathcal{T} = (V, E)$ is a polytree and $\mathcal{D} = (V, P)$ is a digraph. Digraph $\mathcal{D}$ is not necessarily connected. Let A be the matrix defined by, for $p = (u, v) \in P$ and $e \in E$:

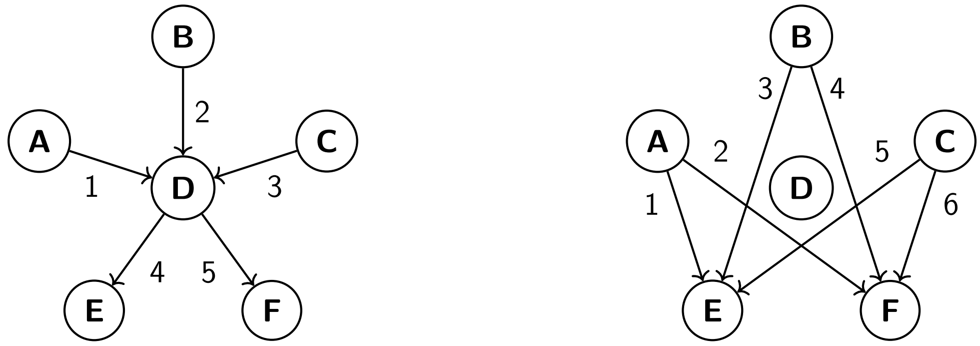
$A_{e,p} = 1$ if the unique u—v path in $\mathcal{T}$ passes through e in the direction of its orientation;

$A_{e,p} = -1$ if the unique u—v path in $\mathcal{T}$ passes through e in the opposite direction to its orientation;

$A_{e,p} = 0$ if the unique u—v path in $\mathcal{T}$ does not pass through e.

Matrices arising in this way are called *network matrices*.

A network matrix is a variation of a link-path matrix for a polytree where we allow paths to traverse edges in the opposite direction to their orientation, and we denote this with a -1 entry in the matrix. This implies the following fact.



(a) The polytree $\mathcal{P}$. (b) The graph $\mathcal{D}$.

Figure 5.6: The graphs from Example 5.3.4.

Observation 5.3.3. A link-path incidence matrix for a polytree corresponds to a network matrix where every link in $\mathcal{D}$ has a corresponding path in $\mathcal{P}$ whose direction matches the orientation of every link.

Example 5.3.4. Consider the network shown in Figure 5.6. The network matrix for $\mathcal{P}$ and $\mathcal{D}$ is given by

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

The paths in $\mathcal{P}$ corresponding to the links in $\mathcal{D}$ follow the orientations of the links in $\mathcal{P}$, so A is non-negative. Therefore A is also the link-path incidence of a traffic network on T where nodes A, B and C function as origins for traffic, and E and F function as destinations.

Tutte [46] has shown that network matrices are totally unimodular, so link-path incidence matrices of polytrees are too. We give two independently found proofs that the link-path incidence matrices of polytrees are totally unimodular.

Theorem 5.3.5. *Any link-path incidence matrix for a polytree is totally unimodular.*

The first proof uses the following property of totally unimodular matrices, given in Section 3.5.1:

Theorem 5.3.6. *A matrix* $A \in \{0, \pm 1\}^{n \times r}$ *is totally unimodular if and only if every square Eulerian submatrix is singular.*

Proof of Theorem 5.3.5 using Theorem 5.3.6. Let A be a link-path incidence matrix for a polytree $\mathcal{P}$. If we delete a column from A, the resulting matrix is a valid link-path incidence matrix for $\mathcal{P}$. If we delete a row from A, the resulting matrix is a valid link-path incidence matrix for the minor of $\mathcal{P}$ found by contracting the corresponding link of $\mathcal{P}$. Any minor of $\mathcal{P}$ constructed in this way must also be a polytree.

If A has no Eulerian submatrices, then we are done. If not, take any Eulerian submatrix of A and call it M, and let $\mathcal{P}_M$ be the minor of $\mathcal{P}$ constructed by contracting the links in M corresponding to rows of A not included in M. We label some link L of $\mathcal{P}_M$ with a 1. Then at the two vertices adjoining L, we label any other adjoining links with a 1 if the link has the same orientation as L relative to their shared vertex (either towards, or away from), or with a -1 if it has the opposite orientation. We propagate this labelling system through $\mathcal{P}_M$ — this is always possible because $\mathcal{P}_M$ is a tree, and so no contradiction in how to label any link can occur. Then each path through $\mathcal{P}_M$ passes through a series of links whose labels alternate between 1 and -1.

Construct the row vector $\mathbf{z}$ where z_i takes the value of the label of the link corresponding to row i of M. Because M is Eulerian, every path contains an even number of links. The sum of the labels of any path is therefore 0.

The vector $\mathbf{z} \neq \mathbf{0}$ is therefore in the left nullspace of M, and so $\det(M) = 0$. Because we can follow this procedure with any square Eulerian submatrix of A, every square Eulerian submatrix of A is singular, and so by Theorem 5.3.6 A is totally unimodular. □

Our second proof of Theorem 5.3.5 uses a different property of totally unimodular matrices due to Ghouila-Houri [23]:

Theorem 5.3.7. *A matrix A with n rows (columns) is totally unimodular if and only if for every subset $\mathcal{A}$ of the columns (rows) of A, each column (row) $\mathbf{a} \in \mathcal{A}$ can be assigned a multiplier $\epsilon_{\mathbf{a}} \in \{\pm 1\}$ such that*

$$\sum_{\mathbf{a} \in \mathcal{A}} \epsilon_{\mathbf{a}} \mathbf{a} \in \{0, \pm 1\}^n.$$

The proof of Theorem 5.3.5 using Theorem 5.3.7 is as follows:

Alternative proof of Theorem 5.3.5 using Theorem 5.3.7. Let A be a link-path incidence matrix for a polytree $\mathcal{P}$. If we delete a row from A, the resulting matrix is a valid link-path incidence matrix for the minor of $\mathcal{P}$ found by contracting the corresponding link of $\mathcal{P}$. Any minor of $\mathcal{P}$ constructed in this way must also be a polytree.

Take any submatrix of A formed by choosing a subset of the rows, and call it M. Let $\mathcal{P}_M$ be the minor of $\mathcal{P}$ constructed by contracting the links in $\mathcal{P}$ corresponding to rows of A not included in M. We label some link L of $\mathcal{P}_M$ with a 1. Then at the two vertices adjoining L, we label any other adjoining links with a 1 if the link has the same orientation as L relative to their shared vertex (either towards, or away from), or with a -1 if it has the opposite orientation, and we propagate this labelling system through $\mathcal{P}_M$ — this is always possible because $\mathcal{P}_M$ is a tree, and so no contradiction in how to label any link can occur.

Each path passes through a series of links such that the labels alternate between 1 and -1. The sum of the labels is 0 if the path contains an even number of links, or ± 1 otherwise.

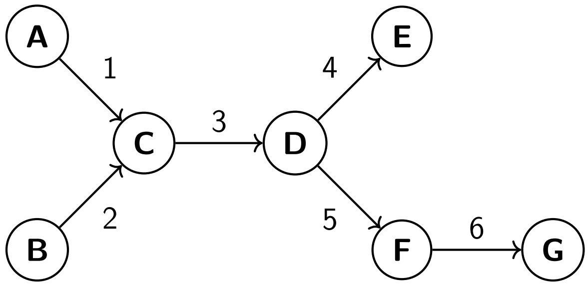


Figure 5.7: The polytree $\mathcal{P}$ from Example 5.3.8.

These labels therefore assign a multiplier to each row of M such that

$$\sum_{\mathbf{m}\in\mathcal{M}} \epsilon_{\mathbf{m}}\mathbf{m} \in \{0, \pm 1\}^r.$$

We can follow this procedure with any subset of the rows A, so A is totally unimodular. □

We illustrate both proofs of Theorem 5.3.5 with the following example.

Example 5.3.8. Consider the polytree network $\mathcal{P}$ shown in Figure 5.7. The link-path incidence matrix is given by

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

First proof: Consider the highlighted Eulerian submatrix,

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

We construct the polytree $\mathcal{P}_M$ which has link-path incidence matrix M by contracting links 3 and 6. This has the effect of merging nodes C and D, and nodes F and G. Link 1 is labelled with a 1, and the other links are labelled as prescribed in the proof. The result is shown in Figure 5.8. We collect these labels into the vector $\mathbf{z} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^\intercal$. Each column of M refers to a path in $\mathcal{P}_M$, and sums an alternating sequence of ± 1s. Because M is Eulerian, each sequence has an even

(a) Forming $\mathcal{P}_M$ by contracting links. (b) The polytree $\mathcal{P}_M$ with its links labelled.

Figure 5.8: The polytree $\mathcal{P}_M$ from Example 5.3.8.

number of elements and so it sums to 0, and we have

$$\begin{aligned} \mathbf{z}^\intercal M &= \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

so $\mathbf{z}$ is in the left nullspace of M, and M is singular.

Second proof: Consider the submatrix comprising the first four rows of A,

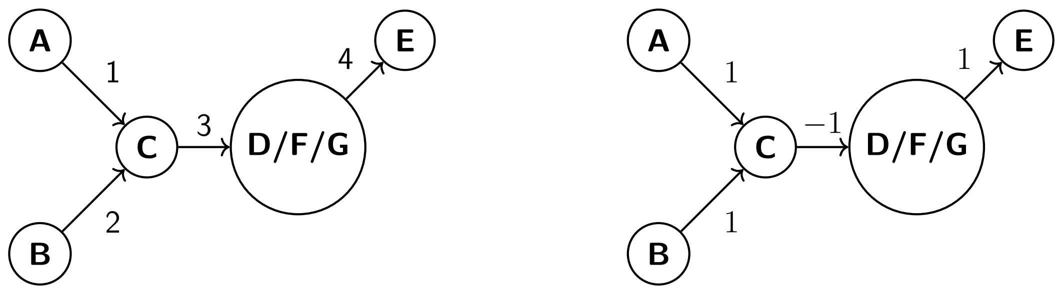
$$A_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

This is the link-incidence matrix of a traffic network on the minor of $\mathcal{P}$ formed by contracting links 5 and 6, which we will call $\mathcal{P}_4$. Contracting these links means merging nodes D, F, and G. Link 1 is again labelled with a 1, and the other links are labelled as prescribed in the proof. The result is show in Figure 5.9. The rows of A_4 are collected into the set $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$. Each row in this collection is assigned as a multiplier the label of the corresponding link in $\mathcal{P}_4$. The signed sum in the alternative proof of Theorem 5.3.5 is given by

$$\begin{aligned} \sum_{\mathbf{a} \in \mathcal{A}} \epsilon_{\mathbf{a}} \mathbf{a} &= \mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 + \mathbf{a}_4 \\ &= \begin{bmatrix} 1 & 1 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 \end{bmatrix}, \end{aligned}$$

and every entry in this vector is 0 or ± 1.

The converse of Theorem 5.3.5 is not necessarily true: given a totally unimodular matrix A, there is not necessarily a polytree for which A is a link-path incidence matrix.



(a) Forming $\mathcal{P}_4$ by contracting links. (b) The polytree $\mathcal{P}_4$ with its links labelled.

Figure 5.9: The polytree $\mathcal{P}_4$ from Example 5.3.8.

We claim that there is no polytree for which the following totally unimodular matrix is a valid link-path incidence matrix:

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

Let $\mathcal{P}$ be a polytree for which A is a valid link-path incidence matrix. We will demonstrate that $\mathcal{P}$ cannot exist. The fifth column of A says that there is a path in $\mathcal{P}$ that includes all four links. The four links must therefore be connected linearly, and there must be two links at the ends that are adjacent to only one of the other three links.

The first four columns of A each describe two link paths in $\mathcal{P}$, so each column gives a pair of adjacent links. But each link appears twice in these columns, each time paired with a different link. This contradicts the linearity of $\mathcal{P}$, so $\mathcal{P}$ cannot exist.

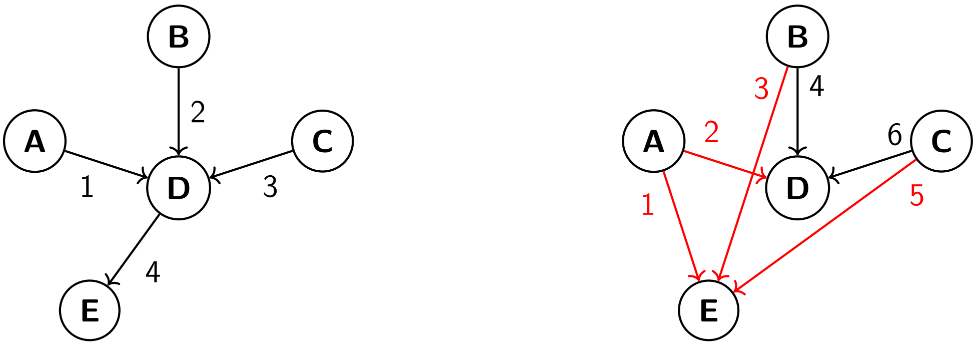
5.3.1 Reduced echelon normal form

It can be shown that the reduced row echelon form of a network matrix is also a network matrix. Schrijver [40, Section 19.3, (36)] notes that two network matrices A and $A_1^{-1}A$ are related in the following way:

Theorem 5.3.9. *If A is a network matrix of full row rank, and A_1 is a basis for the column space of A, then $A_1^{-1}A$ is a network matrix again. If A is represented by the polytree $\mathcal{P} = (V, E)$ and the digraph $\mathcal{D} = (V, P)$, then the columns in A_1 correspond to the links E_1 of a spanning polytree, say $\mathcal{P}_1$, in $\mathcal{D}$, and $A_1^{-1}A$ is represented by $\mathcal{P}_1 = (V, E_1)$ and $\mathcal{D}$.*

Example 5.3.10. Consider the traffic network from Example 5.3.4 which had network matrix

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



(a) The polytree $\mathcal{P}$. (b) The graph $\mathcal{D}$ showing the spanning tree $\mathcal{P}_1$.

Figure 5.10: The graphs from Example 5.3.10.

The fifth row is omitted to obtain a matrix of full row rank, and we redefine A to be

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

This is the network matrix represented by the polytree $\mathcal{P}$ and digraph $\mathcal{D}$, show in Figure 5.10. We choose the spanning tree $\mathcal{P}_1$ consisting of links 1, 2, 3, and 5, shown highlighted in red in Figure 5.10. We partition A accordingly, and we have

$$A_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Then the network matrix represented by the polytree $\mathcal{P}_1$ and the digraph $\mathcal{D}$ is given by

$$A_1^{-1}A = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Each column in $A_1^{-1}A$ gives the corresponding link in $\mathcal{D}$ in terms of the spanning polytree $\mathcal{P}_1$. Links 1, 2, 3, and 5 are each in $\mathcal{P}_1$, so their columns in $A_1^{-1}A$ are standard basis vectors. Link 4 in G connects node B to D, so column 4 in $A_1^{-1}A$ gives the unique $B \to D$ path in $\mathcal{P}_1$, which comprises links 3, 1, and 2, where link 1 is traversed in the opposite direction to its orientation, so its entry in the fourth column of $A_1^{-1}A$ is negative.

5.3.2 Polytree circuits

An element of the integer kernel of a link-path incidence matrix specifies a collection of paths, each with an orientation and a traversal count, such that, when each path is traversed the specified number of times in the specified direction, the signed sum of the traversals of each link is zero. If the traffic network is on a polytree, and the integer kernel element is a circuit, then the paths can be placed nose to tail to form a closed walk. We will prove this property for network matrices, which are a superset of polytree link-path incidence matrices.

Theorem 5.3.11. *Let $A \in \{0, \pm 1\}^{n \times r}$ be a network matrix represented by polytree $\mathcal{P}$ and digraph $\mathcal{D}$, and let $\mathbf{c}$ be a circuit of A. Form the matrix $A_{\mathbf{c}} = \begin{bmatrix}-c_i \mathbf{a}_i\end{bmatrix}$ for $i \in \operatorname{supp}(\mathbf{c})$. Then the oriented paths in $\mathcal{P}$ specified by $A_{\mathbf{c}}$ can be joined nose to tail to form a closed walk.*

Proof. Let P and D be the incidence matrices of $\mathcal{P}$ and $\mathcal{D}$ respectively. We claim that $\ker(D) = \ker(A)$. Since $PA = D$ and we have $\ker(PA) = \ker(D)$. The graph $\mathcal{P}$ is a tree, so $\ker(P) = \{\mathbf{0}\}$. Suppose $\mathbf{z} \in \ker(D) = \ker(PA)$. Then we have

$$PA\mathbf{z} = \mathbf{0} \iff A\mathbf{z} \in \ker(P) \iff A\mathbf{z} = \mathbf{0} \iff \mathbf{z} \in \ker(A).$$

The vector $\mathbf{c}$ is a circuit of A, so it is a circuit of D too. Since A is totally unimodular, $\mathbf{c}$ has all entries in $\{0, \pm 1\}$. We form $D_{\mathbf{c}} = \begin{bmatrix}-c_i \mathbf{d}_i\end{bmatrix}$, and from this we form $\mathcal{D}_{\mathbf{c}}$, the subgraph of $\mathcal{D}$ containing only the links in $D_{\mathbf{c}}$, where each ith link is reoriented according to c_i. The entries in each row of $D_{\mathbf{c}}$ sum to 0, so the indegree of each node in $\mathcal{D}_{\mathbf{c}}$ is equal to the outdegree and the digraph is balanced. The link set can therefore be expressed as a disjoint union of directed cycles, and each connected component of $\mathcal{D}_{\mathbf{c}}$ has an Eulerian walk. We claim there is only one connected component in $\mathcal{D}_{\mathbf{c}}$: if there were more than one, then the support of $\mathbf{c}$ would not minimal by inclusion, and so $\mathbf{c}$ would not be a circuit.

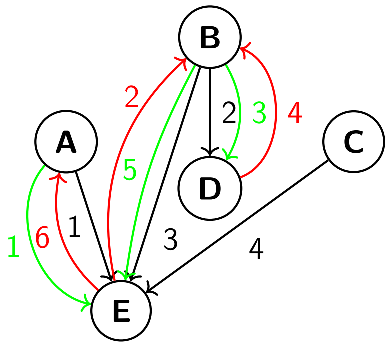
The links in $\mathcal{D}_{\mathbf{c}}$ correspond to paths in $\mathcal{P}$, so the Eulerian walk on $\mathcal{D}_{\mathbf{c}}$ specifies an order in which these paths in $\mathcal{P}$ can joined nose to tail to form a closed walk. □

Example 5.3.12. Consider the network matrix from Example 5.3.10,

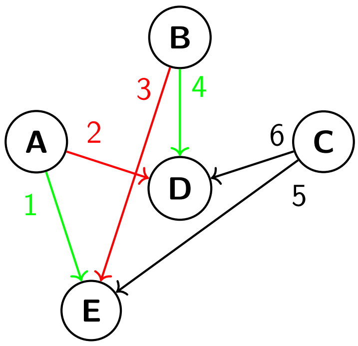
$$A_1^{-1}A = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This matrix is represented by the polytree $\mathcal{P}_1$ and digraph $\mathcal{D}$, show in Figure 5.11. The vector $\mathbf{c} = \begin{bmatrix}1 & -1 & -1 & 1 & 0 & 0\end{bmatrix}^\intercal$ is a circuit of $A_1^{-1}A$. On $\mathcal{D}$, $\mathbf{c}$ corresponds to the closed path $A \to E \to B \to D \to A$. On $\mathcal{P}_1$, $\mathbf{c}$ corresponds to the closed walk $A \to E \to B \to D \to (B) \to (E) \to A$, where nodes in parenthesis indicate that a path passes through that node but does not terminate there. These two walks are shown in Figure 5.11.

Theorem 5.3.11 does not hold for general networks on digraphs, as demonstrated by the following example.



(a) The polytree $\mathcal{P}_1$. Numbers on the coloured arrows give the order of their traversal when laid nose to tail.



(b) The graph $\mathcal{D}$.

Figure 5.11: The graphs from Example 5.3.12 with walks corresponding to the circuit **c**. Green arrows show links that are traversed in the direction of their orientation; red arrows show links that are traversed against the direction of their orientation.

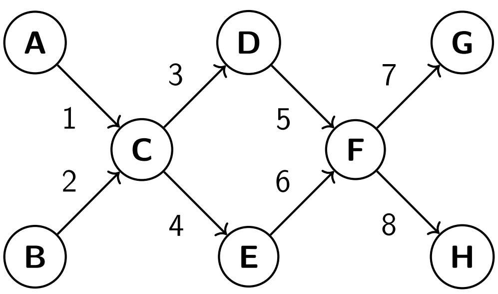


Figure 5.12: The graph of the network in Example 5.3.13.

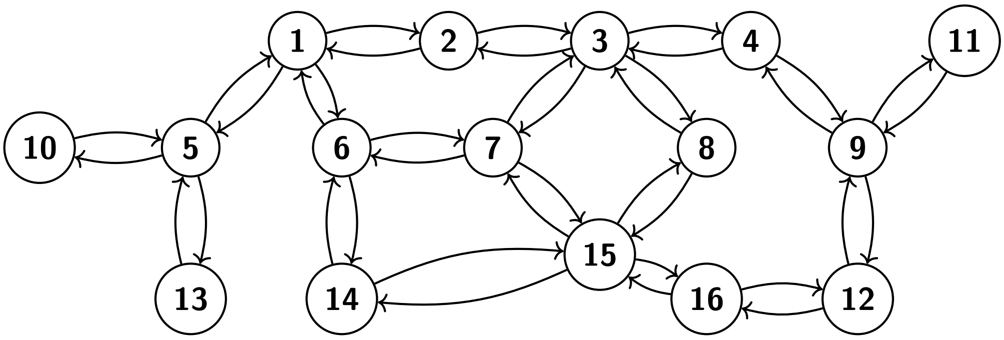


Figure 5.13: An example of a symmetric digraph.

Example 5.3.13. Consider a transport network on the graph in Figure 5.12. Part of the link-path incidence matrix is given by the columns of

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

The paths in this matrix consist of two different paths from node A to node G, and two different paths from node B to node H. The vector $\mathbf{z} = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}^\mathsf{T}$ is in the kernel of A, and would correspond to a circuit of the full link-path incidence matrix. Unlike the polytree network in Example 5.3.12, there is no closed walk that uses all of the paths, and neither of the two closed walks corresponding to $\mathbf{z}$ corresponds to an element of the kernel of A.

5.4 Symmetric directed networks

We turn now to networks on symmetric directed graphs, or symmetric digraphs.

Definition 5.4.1 (Symmetric digraph [31])**.** A *symmetric digraph* $\mathcal{D}$ is a digraph such that (u, v) is a link in $\mathcal{D}$ whenever (v, u) is.

The notation (u, v) means a directed link from a node u to a node v. In this case, (u, v) and (v, u) are called a *symmetric pair of links*. A symmetric digraph is therefore the graph obtained by taking a simple undirected graph and replacing each link with a symmetric pair of links. The graph $\mathcal{G}$ from which the symmetric digraph $\mathcal{D}$ is obtained is called the *underlying graph* of $\mathcal{D}$. An example of a symmetric digraph is shown in Figure 5.13.

We are interested in networks in which every node is designated as either a *station* or a *junction*. Stations will function as both origins and destinations for traffic, and junctions will function as neither. The routing policy we use will be as in the triangular network in Example 5.2.2, above: travel is permitted between any pair of stations along any path that visits each node at most once. We will call traffic networks with this routing policy *symmetric directed networks*.

Definition 5.4.2 (Symmetric directed network)**.** A *symmetric directed network* is a traffic network on a symmetric directed graph that has the following properties:

1. Each peripheral node is designated as a station.

2. All other nodes are designated as either a station or a junction.

3. Travel between any distinct pair of stations along any path is permitted.

4. Every cycle on the graph that uses at most one of each pair of symmetric edges visits at least two stations.

Our aim in this section is to investigate whether traffic networks on symmetric digraphs have link-path incidence matrices that contain maximal unimodular submatrices. This implies the existence of column partitions for A for which A_1 is unimodular, and hence for which the parity problems described in Section 2.5.1 cannot occur.

We will begin in Section 5.4.1 with star networks that can considered as symmetric digraphs. Star networks can also be considered as polytrees, so their link-path incidence matrices are totally unimodular by Theorem 5.3.5.

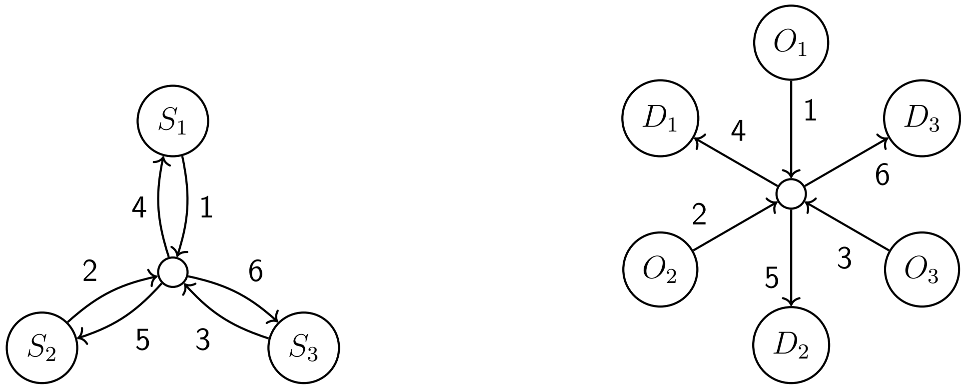
In Section 5.4.2 we will add internal structure to the graphs of star networks to form *bidirectional trees*. Traffic networks on bidirectional trees whose stations are all peripheral stations we will call *peripheral bidirectional tree networks*. We will show that any link-path incidence matrix of a peripheral bidirectional tree network has at least one maximal unimodular submatrix. We will additionally show that there is good reason to think there are very many maximal unimodular submatrices.

In Section 5.4.3 we combine collections of peripheral bidirectional tree networks to create bidirectional tree networks. These can be thought of as peripheral bidirectional tree networks, but where internal nodes may also be designated as stations. We show that such networks also have maximal unimodular submatrices.

In Section 5.4.4 we look at general symmetric directed networks. These are networks on symmetric digraphs which differ from bidirectional tree networks in that their underlying graphs may contain cycles. Again, these may be formed by combining bidirectional tree networks. We prove that the link-path incidence matrices of symmetric directed networks also contain maximal unimodular submatrices.

5.4.1 Star networks

We first consider symmetric digraphs whose underlying graph is that of a star network.



(a) The symmetric directed star network. (b) The polytree star network.

Figure 5.14: The equivalence between the symmetric directed and polytree forms of a star network.

Definition 5.4.3 (Star network)**.** A *star network* is a symmetric directed network on a symmetric digraph consisting of one central junction connected to n peripheral stations.

The routes on this network are the paths that connect each pair of stations.

The link-path incidence matrix for this is shown by Hazelton [26] to be totally unimodular. It is also totally unimodular by Theorem 5.3.5, since it is a valid link-path matrix for the polytree constructed by splitting each peripheral node in two: one origin node with a link pointing towards the central node; and one destination node with one link pointing away from the central node.

Figure 5.14 shows this equivalence. On the symmetric directed star network (Figure 5.14a), travel between any pair of peripheral nodes is allowed. On the polytree star network (Figure 5.14b), travel between any origin and destination pair (O_i, D_j) is allowed except when $i = j$. Then the link-path incidence matrices for both networks are given by

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

The rank of a link-path incidence matrix is given by the following theorem.

Theorem 5.4.4. *The link-path incidence matrix of a star network with s peripheral nodes has rank $2s - 1$.*

Proof. A star network with s peripheral nodes has $2s$ links. The central node is a junction, and so no path can begin or end there. Therefore the sum of the traffic counts on the paths directed towards the central node is equal to the traffic counts directed away from the central node, and so $\text{rank}(A) < 2s$.

To complete the proof we will show that there are $2s-1$ independent columns. We label the s links of the graph that are directed towards the central node $e_1, \ldots, e_s$, and the s links of the graph that are directed away from the central node $f_1, \ldots, f_s$. We order the rows of A so that they correspond to the ordering of the links $e_1, \ldots, e_s, f_1, \ldots, f_{s-1}$, and exclude the row f_s because of the rank deficiency. We take the submatrix of A formed by collecting the columns corresponding to the following paths:

1. The paths (e_i, f_s) for $i = 1, \ldots, s-1$. With the row corresponding to f_s excluded, these columns are the standard basis vectors $\mathbf{e}_i$ for $i = 1, \ldots, s-1$.

2. The path (e_2, f_1). This column is the sum of the two standard basis vectors $\mathbf{e}_2 + \mathbf{e}_{s+1}$.

3. The paths (e_s, f_i) for $i = 1, \ldots, s-1$. Each of these columns is the sum of $\mathbf{e}_s + \mathbf{e}_{s+i}$, two standard basis vectors.

The matrix is

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

Subtracting the $(s+1)$th column from the sth column does not change the determinant, and produces the following upper triangular matrix:

$$A' = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 1 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & -1 & 1 & 1 & \ldots & 1 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

Then $\det(A) = \det(A') = -1$, so there are $2s-1$ independent columns, completing the proof. $\square$

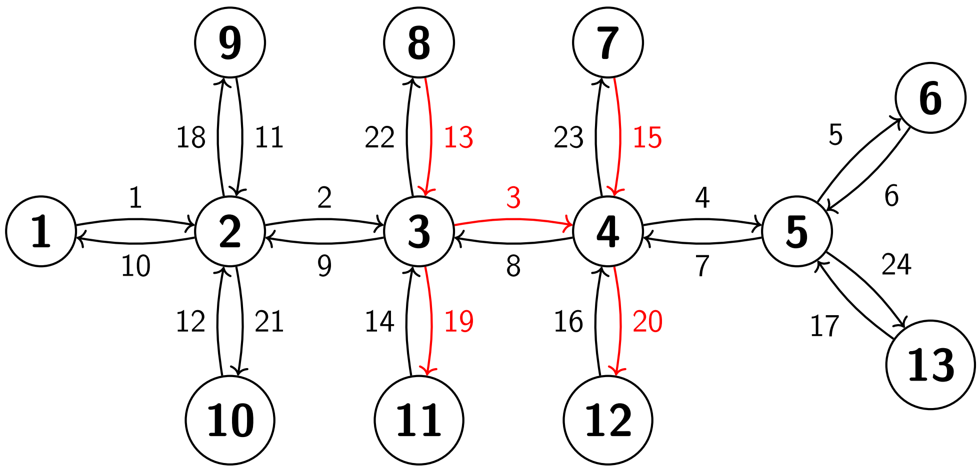


Figure 5.15: The graph from the Monroe network, used in Tebaldi and West's seminal paper on Bayesian network tomography [45]. The red links correspond to rows of a submatrix of the link-path incidence matrix with determinant −2.

5.4.2 Peripheral bidirectional tree networks

We turn now to a type of network we call a *bidirectional tree network*. These can be thought of as a star network with some internal structure. We will call the graph for this type of network a *bidirectional tree*.

Definition 5.4.5 (Bidirectional tree)**.** A *bidirectional tree* is a symmetric digraph whose underlying graph is a tree.

In accordance with the routing policy, the leaves of the underlying tree are designated as stations. We do not designate any internal stations.

Definition 5.4.6 (Peripheral bidirectional tree network)**.** A *peripheral bidirectional tree network* is a network on a bidirectional tree where the stations are the peripheral nodes.

We omit trees containing nodes of degree 2: the routing policy states that no such nodes can be a station. Therefore on the induced symmetric digraph, the traffic counts on either of the two pairs of symmetric links can be determined from the other, and so one pair is redundant.

Example 5.4.7 (The Monroe network [45])**.** The graph from the Monroe network is an example of a bidirectional tree. The Monroe network with traffic demand as specified by Tebaldi and West [45] excludes some pairs of peripheral nodes as origin/destination pairs and so does not follow our routing policy and is not a peripheral bidirectional tree network. The Monroe network is shown in Figure 5.15.

Matrices for peripheral bidirectional tree networks fail to be totally unimodular when they contain more than one junction. To see this, consider the submatrix M of the link-path incidence matrix containing the rows corresponding to links 3, 13, 15, 19, and 20;

and the columns corresponding to paths $8 \to 7$, $11 \to 12$, $7 \to 12$, $7 \to 11$; and $8 \to 11$. This produces the matrix

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

This submatrix has determinant -2, so any link-path incidence matrix that contains it is not totally unimodular. This or a similar submatrix exists whenever there is at least one internal bidirectional link and two stations connected (perhaps via other links) to each end. The matrix M, or a reordering of it, therefore appears in any link-path incidence matrix for a peripheral bidirectional tree network.

Peripheral bidirectional tree networks do not in general have unimodular link-path incidence matrices either, as demonstrated by the following example.

Example 5.4.8. Let A be the link-path incidence matrix of the 10 link peripheral bidirectional tree network in Figure 5.16, given by

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

This matrix contains eight 5×5 and eight 7×7 non-singular Eulerian submatrices so it is not totally unimodular. The rank of the matrix is $\text{rank}(A) = 8$, so the potential A_1 partitions are the 8×8 non-singular submatrices. Checking all 8000 such submatrices with a computer we find that only 64 are non-unimodular — all 64 have determinant ± 2.

We first determine the rank of the link-path incidence matrix of a peripheral bidirectional tree network. The rank determines the size of a maximal invertible submatrix.

Theorem 5.4.9. *The link-path incidence matrix of a peripheral bidirectional tree network with s stations and j junctions is of rank $2s + j - 2$.*

Proof. Given a peripheral bidirectional tree with s stations and $j \geq 2$ junctions, we can merge two adjacent junctions to form a smaller peripheral bidirectional tree with s stations and $j - 1$ junctions. The proof is by induction on the number of junctions.

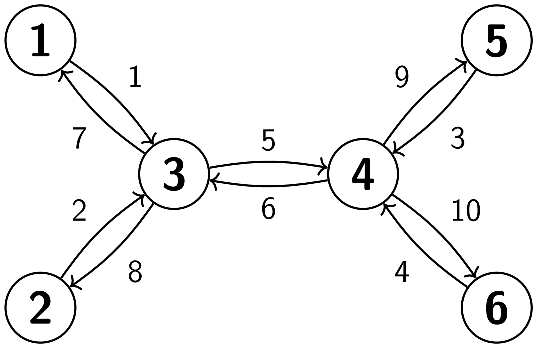


Figure 5.16: The graph for the network in Example 5.4.8.

For the base case we consider a star network, which is a peripheral bidirectional tree network with s stations and $j = 1$ junction. By Theorem 5.4.4 we have

$$\begin{aligned}\operatorname{rank}(A) &= 2s - 1 \\ &= 2s + j - 2.\end{aligned}$$

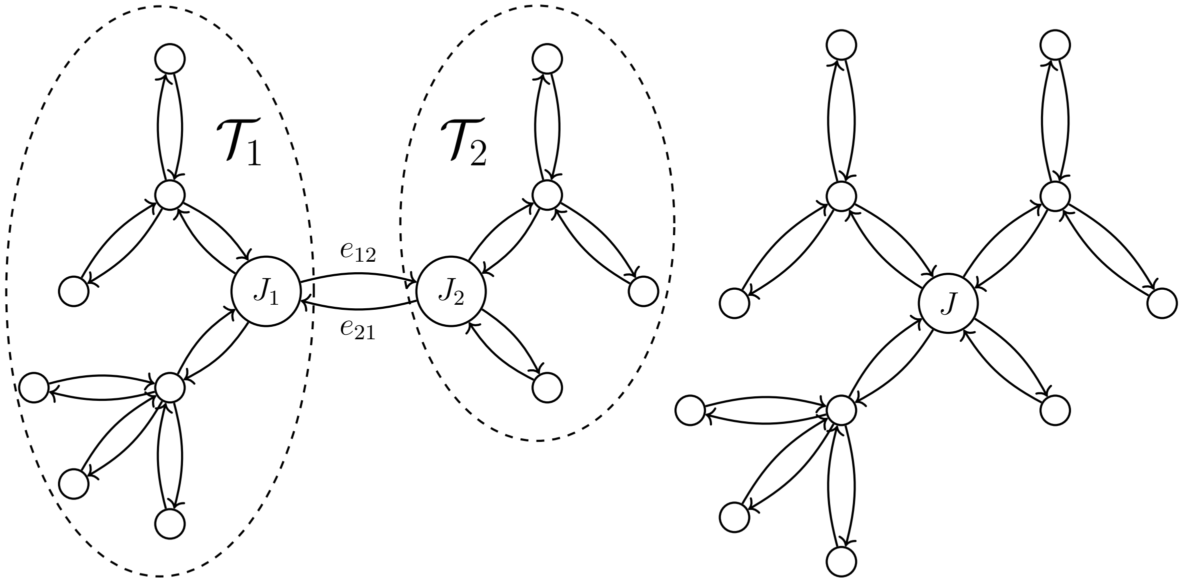
For the induction, consider a peripheral bidirectional tree network on a bidirectional tree $\mathcal{T}$ containing $j > 2$ junctions and s stations. A pair of adjacent junctions J_1 and J_2 are connected by a pair of links e_{12} and e_{21}, which run from J_1 to J_2 and from J_2 to J_1 respectively. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be the parts of $\mathcal{T}$ connected to J_1 and to J_2 respectively when links e_{12} and e_{21} are removed. Links e_{12} and e_{21} can be contracted to form another peripheral bidirectional tree $\mathcal{T}'$. There are still s stations in $\mathcal{T}'$, so the origin/destination pairs in the network remain unchanged. There are $j - 1$ junctions in $\mathcal{T}'$ since J_1 and J_2 have merged to form the junction J. The relationship between $\mathcal{T}$ and $\mathcal{T}'$ is shown in Figure 5.17. If $\mathcal{T}'$ has link-path incidence matrix A', then $\mathcal{T}$ has link-path incidence matrix

$$A = \begin{bmatrix} A' \\ \mathbf{r}_{12} \\ \mathbf{r}_{21} \end{bmatrix},$$

where $\mathbf{r}_{12}$ and $\mathbf{r}_{12}$ are rows corresponding to links e_{12} and e_{21} respectively. The row vector $\mathbf{r}_{12}$ contains 1s in the columns corresponding to paths that originate in $\mathcal{T}_1$ and whose destination lies in $\mathcal{T}_2$, and 0s elsewhere. Similarly, $\mathbf{r}_{21}$ contains 1s in the columns corresponding to paths that originate in $\mathcal{T}_2$ and whose destination lies and $\mathcal{T}_1$, and 0s elsewhere. Appending these two rows to A' either leaves the rank of the matrix unchanged; or increased it by one or by two. We need to show that the rank of the matrix has increased by one, which we will do by showing there is a dependency involving these rows, and by showing that we have removed a linear dependence amongst the columns.

Let $\mathcal{R}_\alpha$ be the set of row vectors in A corresponding to the links in $\mathcal{T}_1$ that enter J_1, and let $\mathcal{R}_\omega$ denote the set of row vectors in A corresponding to the links in $\mathcal{T}_1$ that exit J_1. The net flow through J_1 is given by

$$\sum_{\mathbf{r} \in \mathcal{R}_\alpha} \mathbf{r} + \mathbf{r}_{21} = \sum_{\mathbf{r} \in \mathcal{R}_\omega} \mathbf{r} + \mathbf{r}_{12},$$



(a) The graph $\mathcal{T}$.

(b) The graph $\mathcal{T}'$.

Figure 5.17: Contracting links e_{12} and e_{21} of $\mathcal{T}$ to form $\mathcal{T}'$, as described in the proof of Theorem 5.4.9.

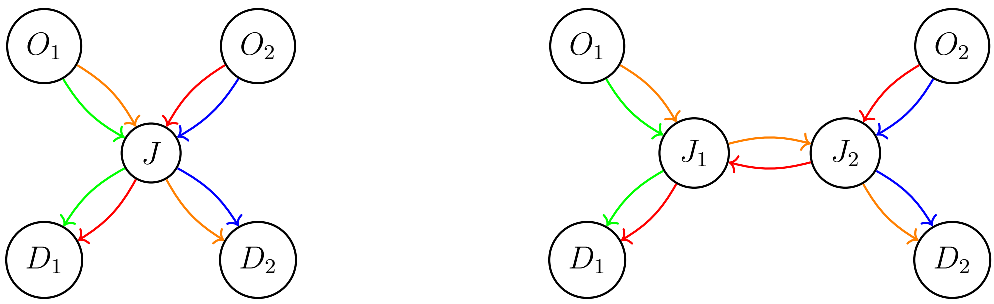
which shows a linear dependence involving the newly appended rows.

We now need to show that by appending $\mathbf{r}_{12}$ and $\mathbf{r}_{21}$ to A' we have removed a linear dependence amongst the columns. Choose one origin O_1 and one destination D_1 from the designated stations in $\mathcal{T}_1$ such that the path from O_1 to D_1 passes through J_1, and similarly choose O_2 and D_2 from $\mathcal{T}_2$ so that the connecting path passes through J_2. A representation of these paths on $\mathcal{T}'$ and $\mathcal{T}$ are shown in Figure 5.18.

If $\mathbf{a}'_{ij}$ is the column of A' corresponding to the path from O_i to D_j, then clearly

$$\mathbf{a}'_{11} + \mathbf{a}'_{22} = \mathbf{a}'_{12} + \mathbf{a}'_{21}.$$

In the corresponding columns of A, only $\mathbf{a}_{12}$ has a 1 in the entry corresponding to e_{12};



(a) The paths on $\mathcal{T}'$.

(b) The paths on $\mathcal{T}$.

Figure 5.18: The paths showing the eliminated linear dependence.

the other columns have 0. Appending these rows therefore removes a linear dependence amongst the columns of A', so $\text{rank}(A) \geq \text{rank}(A')+1$. By hypothesis we have $\text{rank}(A') = 2s + (j-1) - 2$, so we have $\text{rank}(A) = 2s + j - 2$ as required. □

We now use Theorem 5.4.9 to prove that there is a unimodular partition.

Theorem 5.4.10. *Let A be the link-path incidence matrix of a peripheral bidirectional tree network. Then A has a maximal unimodular submatrix.*

Proof. The proof will work as follows: we take an arbitrary peripheral bidirectional tree network $\mathcal{T}$ and its link-path incidence matrix A. We delete links from $\mathcal{T}$ to form a polytree $\mathcal{P}$, and form the full-rank submatrix A_{S_1} of A by deleting the corresponding rows of A.. We select paths in $\mathcal{T}$ to construct a maximal invertible submatrix M of A_{S_1}. We then show that we can transform M to a link-path incidence matrix M' for $\mathcal{P}$ without altering the determinant. Then since M' is totally unimodular, we have $\det(M) = \pm 1$.

Consider a peripheral bidirectional tree network on a bidirectional tree $\mathcal{T}$ which has link-path incidence matrix A. The network has s stations and j junctions and therefore $2(s+j-1)$ links and $s(s-1)$ paths, so A has $2s+2j-1$ rows and $s(s-1)$ columns. By Theorem 5.4.9, we have $\text{rank}(A) = 2s + j - 2$. Following the proof of Theorem 5.4.9, a full rank submatrix of A can be found by removing rows corresponding to one of each symmetric pair of links connecting pairs of junctions, and one row corresponding to a link connecting to any station.

We designate the stations S_i and the junctions J_i. The link from node N_i to N_k is N_iN_k. Each station is connected to only one junction, so we will use S_iJ and JS_i to mean the links connecting S_i to its junction.

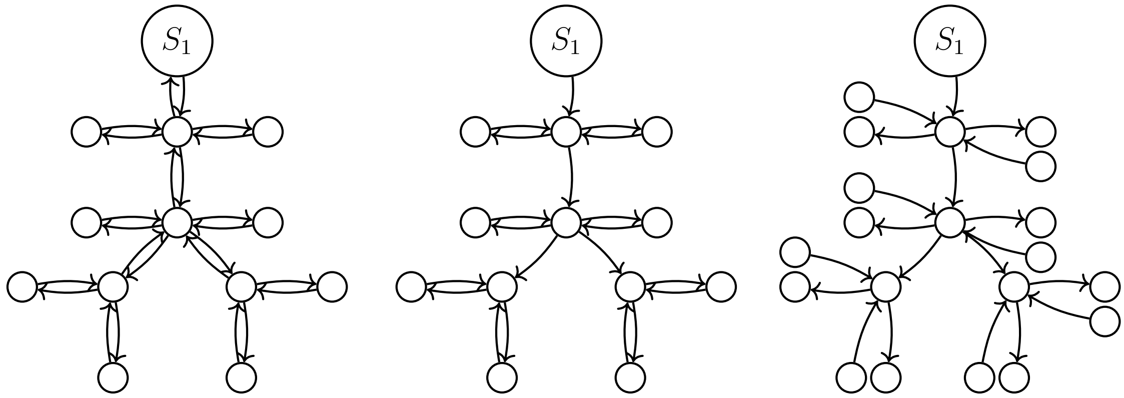
We designate one station, say S_1, as the root station. We say a node N_i is an *ancestor* of another node N_k if N_i lies between N_k and S_1. We will extend this definition for stations, and say that S_i is an ancestor of S_k if the junction that S_i is connected to is an ancestor of S_k. Note that it is possible for two stations to be each other's ancestors if they are connected to the same junction.

We form the subgraph $\mathcal{T}_{S_1}$ of $\mathcal{T}$ by deleting from $\mathcal{T}$ the link JS_1, and all links J_iJ_k such that J_k is an ancestor of J_i. An example of the formation of $\mathcal{T}_{S_1}$ is shown in Figure 5.19b. The junction-to-junction links and the link JS_1 form a directed tree rooted at S_1 that spans S_1 and the junctions.

We form the polytree $\mathcal{P}_{S_1}$ from $\mathcal{T}_{S_1}$ by splitting each station S_i (except for S_1) into two nodes, one connected by the link S_iJ and the other by the link JS_i. An example of the formation of $\mathcal{T}_{S_1}$ from $\mathcal{T}_{S_1}$ is show in Figure 5.19c. Any link path incidence matrix of $\mathcal{T}_{S_1}$ is also a link path incidence matrix for $\mathcal{P}_{S_1}$.

The paths in the network on $\mathcal{T}$ are represented on $\mathcal{T}_{S_1}$ in the following ways.

- Any path $S_i \to S_k$ such that S_i is an ancestor of S_k is also a path on $\mathcal{T}_{S_1}$.
- Any path $S_i \to S_1$ is represented on $\mathcal{T}_{S_1}$ by the path comprising only the link S_iJ.



(a) The network $\mathcal{T}$. (b) The graph $\mathcal{T}_{S_1}$. (c) The polytree $\mathcal{P}_{S_1}$.

Figure 5.19: Formation of the graphs used in the proof of Theorem 5.4.10.

- Any other path $S_i \to S_k$ such that S_i is not an ancestor of S_k is represented by two disjoint paths: one comprising only the link S_iJ, and the other being the path $\text{LCA}(S_i, S_k) \to S_k$, where LCA returns the least common ancestor of its arguments.

We form the submatrix A_{S_1} of A by including only the rows corresponding to links that are included in $\mathcal{T}_{S_1}$. Following the proof of Theorem 5.4.9, this matrix A_{S_1} is of full rank. We choose a full rank maximal submatrix M of A_{S_1} such that if M contains the column corresponding to the path $S_i \to S_k$ where S_i is not an ancestor of S_k, then the column corresponding to the path $S_i \to S_1$ is also in M. This is guaranteed if all $(s-1)$ paths terminating at S_1 are included in the $(2s+j-2)$ paths in M. The column of A_{S_1} corresponding to the path $S_i \to S_1$ has one non-zero entry, which is a 1 in the row corresponding to the link S_iJ. This link is unique among paths terminating at S_1, so the columns are all linearly independent.

We claim that this matrix has $\det(M) = \pm 1$. We have chosen M to be full rank, so we now show that it can be transformed to a totally unimodular matrix M' without changing the determinant.

Each column in M corresponds to a path in $\mathcal{T}$. We form M' from M by subtracting from each column corresponding to a path $S_i \to S_k$, where S_i is not an ancestor of S_k, the column corresponding to the path $S_i \to S_1$. This column now corresponds to a path $\text{LCA}(S_i, S_k) \to S_k$ on $\mathcal{T}_{S_1}$. These operations do not affect the determinant, so we have $\det(M) = \det(M')$.

Each column in M' corresponds to a path in $\mathcal{T}_{S_1}$, or equivalently in $\mathcal{P}_{S_1}$. Then M' is a link-path incidence matrix for a polytree, so by Theorem 5.3.5 M' is totally unimodular. We have $\det(M) = \det(M')$, and M is of full rank, so $\det(M) = \pm 1$ and A contains a maximal unimodular submatrix. □

In practice there will be many maximal unimodular submatrices. We expect that different choices of station as root, forming $\mathcal{T}_{S_1}$ by including junction links directed towards rather than away from the root, and including a different selection of paths in M will all result in different M matrices.

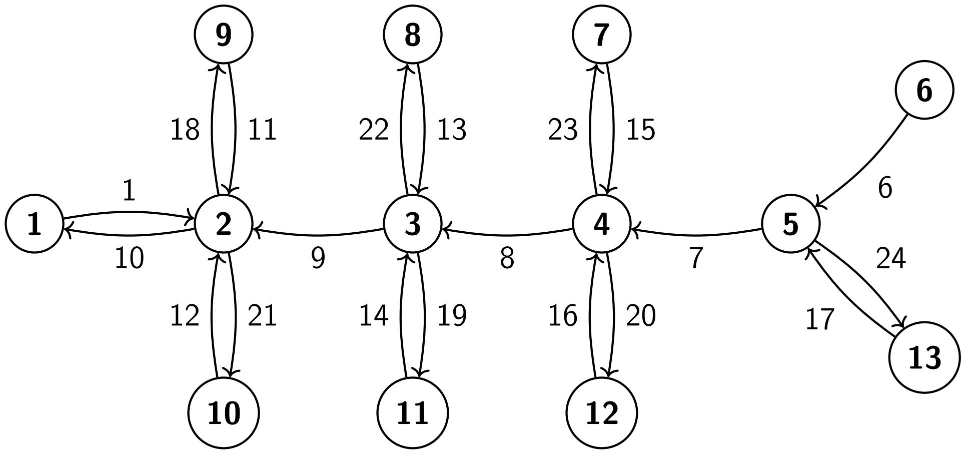
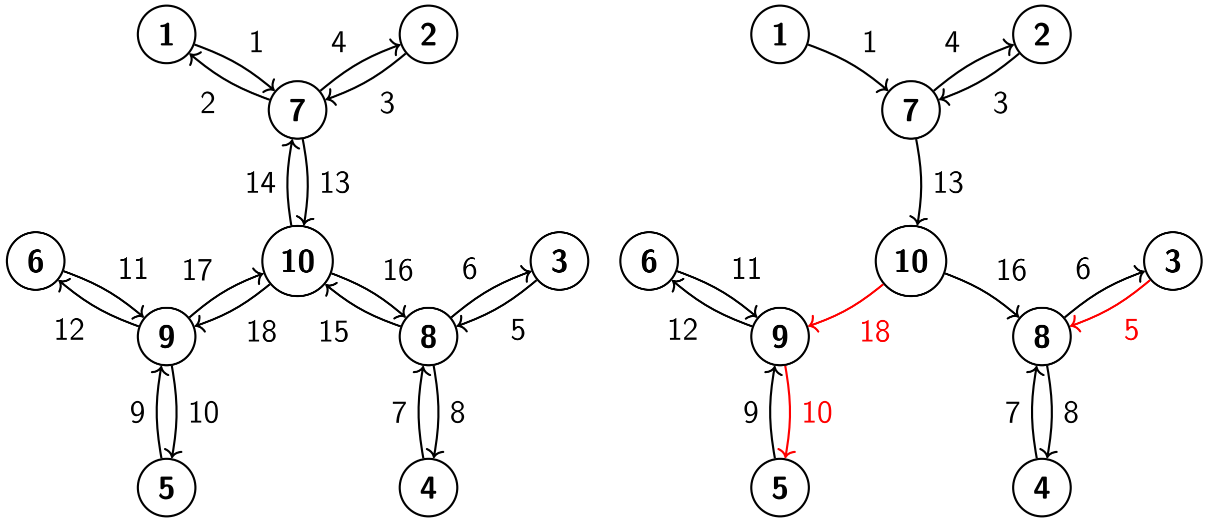


Figure 5.20: The graph $\mathcal{T}_6$ from Example 5.4.11. Node 13 is an ancestor of every other node, so any path in $\mathcal{T}$ originating at node 13 is also a valid path on $\mathcal{T}_6$.

Example 5.4.11 (The Monroe network revisited)**.** Consider the peripheral bidirectional tree network on the graph of the Monroe network, introduced in Example 5.4.7, and whose graph is shown in Figure 5.15. This network has nine stations and four junctions, with 24 links and 72 paths. The full link-path incidence matrix has rank 20. We choose node 6 as the root and form the polytree $\mathcal{T}_6$ by deleting links $2, 3, 4$ and 5, shown in Figure 5.20. There are 42 paths that originate at an ancestor of the path's destination, and seven paths (other than $13 \to 6$, since 13 is an ancestor of 6) terminating at node 6, making 49 paths on $\mathcal{T}$ that are also valid paths on $\mathcal{T}_6$. Collecting the columns corresponding to these paths produces a totally unimodular 20×49 submatrix of full rank, so any full rank maximal submatrix has determinant ± 1.

If not every junction is connected to a station, then collecting the columns of A_{S_1} corresponding to the paths that are also valid paths on $\mathcal{T}_{S_1}$ will not produce a matrix of full row rank. These are the paths that originate at an ancestor of the destination, and the paths that terminate at the root. It is then necessary to include other paths from $\mathcal{T}$ in the submatrix M. The remaining choices are paths in $\mathcal{T}$ are those that do not originate at an ancestor of the destination, and do not terminate at the root. These paths are not paths in $\mathcal{T}_{S_1}$ — they include links that were deleted during the formation of $\mathcal{T}_{S_1}$, and so on $\mathcal{T}_{S_1}$ the remaining included links make up a disjoint union of paths. We illustrate this with the following example.

Example 5.4.12. Consider the bidirectional tree network $\mathcal{T}$ shown in Figure 5.21a. There are 18 links and 30 paths. The full link-path incidence matrix is of rank 14. Choosing node 1 as the root, we form $\mathcal{T}_1$ by deleting links $2, 14, 15$ and 17. The submatrix of A formed by including only rows and columns corresponding to links and paths in $\mathcal{T}_1$ is



(a) The network $\mathcal{T}$. The central junction 10 has no station attached.

(b) The graph $\mathcal{T}_1$. The links in the path $3 \to 5$ are highlighted in red.

.

Figure 5.21: The graphs from Example 5.4.12.

given by the 14×18 matrix

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

This matrix is rank 13, so in order to make a full rank maximal submatrix we need to include paths in $\mathcal{T}$ that are not paths in $\mathcal{T}_1$: these are the paths $S_i \to S_k$ with $k \neq 1$ such that S_i is not an ancestor of S_k. A maximal submatrix M that includes such paths will still be unimodular if for $i, k \neq 1$, the column corresponding to the path $S_i \to S_1$ is in M whenever the column corresponding to the path $S_i \to S_k$. Using a computer to check the determinants of all maximal submatrices of A_{S_1} that include all of the columns corresponding to paths terminating at the root, we find that $2, 190, 110$ are singular and

119, 536 have determinant ± 1. In particular, the submatrix

$$M = \begin{bmatrix}
1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

has determinant $\det(M) = -1$, and includes as column 14 the path $3 \to 5$. Node 3 is not an ancestor of node 5. This path includes links $5, 15, 18$ and 10, although row 15 is not included in A_{S_1}. This path is highlighted in Figure 5.21. Column 7 corresponds to the path $3 \to 1$, and so in this submatrix includes only link 5. Subtracting column 9 from column 14 produces a link-path incidence matrix for the graph $\mathcal{T}_1$, which is also a link-path incidence matrix for the polytree formed by splitting the stations into an origin and a destination node, so it is totally unimodular.

5.4.3 Bidirectional tree networks

Real networks on bidirectional trees may include internal stations.

Definition 5.4.13 (Bidirectional tree network)**.** A *bidirectional tree network* is a network on a bidirectional tree where all peripheral nodes are stations, and internal nodes may be designated as stations.

All peripheral nodes are designated stations — if they are not, then because they have no through-traffic, they are superfluous and may be omitted from the network. Any bidirectional tree network $\mathcal{T}$ with an internal station S could be decomposed into two smaller networks $\mathcal{T}_1, \mathcal{T}_2$ by splitting S into two stations. The set of allowed paths on one of the new networks $\mathcal{T}_i$ is the subset of the paths on $\mathcal{T}$ that use only links in $\mathcal{T}_i$.

The two smaller networks will themselves be either bidirectional tree networks, or *bilateral networks*.

Definition 5.4.14. A *bilateral network* is a network consisting of two stations joined by a symmetric pair of links.

Observation 5.4.15. The link-path incidence matrix of a bilateral network is

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

This matrix is unimodular.

If $\mathcal{T}_1$ or $\mathcal{T}_2$ also has an internal station, it can also be decomposed into two smaller networks. Any bidirectional tree can therefore be recursively decomposed into a collection of peripheral bidirectional trees networks and bilateral networks.

Theorem 5.4.16. *The link-path incidence matrix of a bidirectional tree network has at least one unimodular maximal invertible submatrix.*

Proof. Suppose the bidirectional tree network $\mathcal{T}$ is constructed from a collection of k peripheral bidirectional tree networks and bilateral networks $\{\mathcal{T}_1, \ldots, \mathcal{T}_k\}$, where $\mathcal{T}_i$ has link-path incidence matrix $A^{\mathcal{T}_i}$. Then the link-path incidence matrix A of $\mathcal{T}$ is

$$A = \left[\begin{array}{cccc} A^{\mathcal{T}_1} & 0 & \cdots & 0 \\ 0 & A^{\mathcal{T}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A^{\mathcal{T}_k} \end{array} \quad A^{\mathcal{T}_{1\cdots k}} \right]$$

where the columns of $A^{\mathcal{T}_{1\cdots k}}$ correspond to paths that span multiple subnetworks $\mathcal{T}_i$. Each $A^{\mathcal{T}_i}$ has a unimodular maximal submatrix by Theorem 5.4.10 or by Observation 5.4.15. We partition each $A^{\mathcal{T}_i}$ into $A_1^{\mathcal{T}_i}$ and $A_2^{\mathcal{T}_i}$ such that $A_1^{\mathcal{T}_i}$ is unimodular, and include the corresponding columns of A in the A_1 partition. This produces the partition

$$A = \left[\begin{array}{cccc|cccc} A_1^{\mathcal{T}_1} & 0 & \cdots & 0 & A_2^{\mathcal{T}_1} & 0 & \cdots & 0 \\ 0 & A_1^{\mathcal{T}_2} & \cdots & 0 & 0 & A_2^{\mathcal{T}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_1^{\mathcal{T}_k} & 0 & 0 & \cdots & A_2^{\mathcal{T}_k} \end{array} \quad A^{\mathcal{T}_{1\cdots k}} \right].$$

The A_1 part of this matrix is block diagonal, where each block has determinant ± 1, so it also has determinant $\det(A_1) = \pm 1$. □

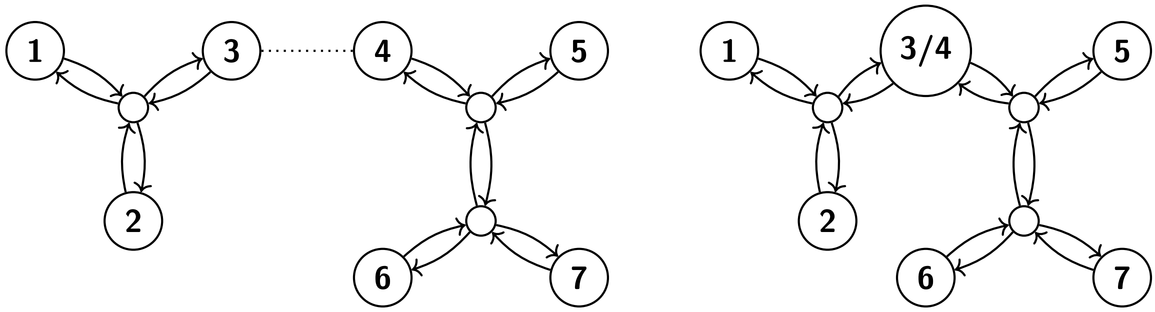
The construction of such a unimodular maximal submatrix is illustrated by the following example.

Example 5.4.17. Consider the peripheral bidirectional tree $\mathcal{T}$ shown in Figure 5.22b. The bidirectional tree network $\mathcal{T}$ can be formed by identifying station 3 in $\mathcal{T}_1$ with station 4 in $\mathcal{T}_2$.

The networks $\mathcal{T}_1$ and $\mathcal{T}_2$ are both peripheral bidirectional tree networks, so their respective link-path incidence matrices $A^{\mathcal{T}_1}$ and $A^{\mathcal{T}_2}$ have at least one unimodular maximal submatrix. These become $A_1^{\mathcal{T}_1}$ and $A_2^{\mathcal{T}_1}$. Then the matrix

$$A_1 = \begin{bmatrix} A_1^{\mathcal{T}_1} & 0 \\ 0 & A_1^{\mathcal{T}_2} \end{bmatrix}$$

is a maximal submatrix of A, the link-path incidence matrix of $\mathcal{T}$, and $\det(A_1) = \pm 1$.



(a) The star network $\mathcal{T}_1$ and peripheral bidirectional tree network $\mathcal{T}_2$.

(b) The bidirectional tree network $\mathcal{T}$, which includes node 3/4 as an internal station.

.

Figure 5.22: The graphs from Example 5.4.17.

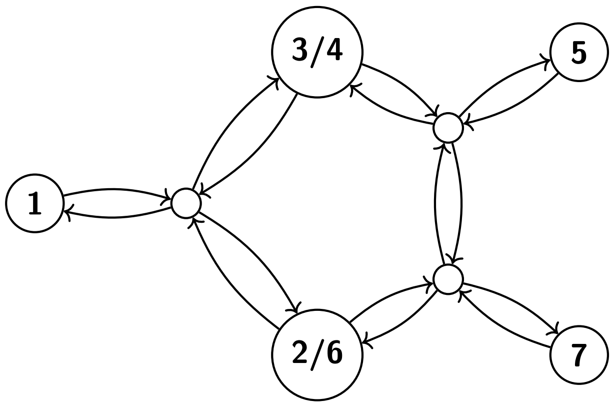


Figure 5.23: The symmetric directed network $\mathcal{D}$ from Section 5.4.4.

.

5.4.4 Symmetric directed networks

We now give a brief discussion of how the techniques of this section might be extended to more general symmetric directed networks. As an example, consider again the bidirectional tree networks $\mathcal{T}_1$ and $\mathcal{T}_2$ from Example 5.4.17 in Figure 5.22. As well as identifying stations 3 and 4, we can also identify stations 2 and 6 to construct a symmetric directed network, $\mathcal{D}$, whose underlying graph is not a tree. The network $\mathcal{D}$ is shown in Figure 5.23. As before, the set of paths on $\mathcal{T}_1$ and $\mathcal{T}_2$ are also allowed on $\mathcal{D}$. For this to work, it must be the case that every cycle in the underlying graph includes at least two stations — this guarantees that no station has been identified with another on its own network. If a symmetric digraph $\mathcal{D}$ is constructed by identifying some station S_1 in a bidirectional tree network $\mathcal{T}_1$ with another station S_2 in $\mathcal{T}$, then the path $S_1 \to S_2$ in $\mathcal{T}$ is not an allowed path in $\mathcal{D}$. Then a version of Theorem 5.4.16 for symmetric directed networks would not guarantee that the link-path incidence matrix of $\mathcal{D}$ has a unimodular maximal submatrix.

Chapter 6

Chapter minus one

6.1 Introduction

If a lattice basis is not a Markov basis it can still be used to construct a walk between any two points in any associated fibre, provided the walk is allowed to visit points with negative co-ordinates. For such a walk to be useful for sampling, a lower bound on the co-ordinates of visited points must be known. Using lower bounds on co-ordinates that are less than zero enables the walk to visit more points — some of these points may then function as stepping stones between fibre elements that are not connected by a walk with a lower bound of zero.

For any particular fibre, a lower bound must exist because there is a finite distance between any pair of points. We may be able to compute this lower bound, and sampling from the fibre may commence. However, knowing a priori that a lower bound for all fibres exists, and what it is, saves us the computational work for the fibre in question, and can provide the foundations for development of new samplers.

When using MCMC sampling methods, the connectedness of walks on the fibre being sampled are generally unknown. For walks that use a lower bound less than zero to be useful, we must have a lower bound that we know is uniform across all potential fibres.

In this chapter we discuss such lower bounds.

Example 6.1.1. Consider the configuration matrix

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

This is a link-path incidence matrix for the three-link linear network that appears in the example in Section 1.1, where we have removed the path consisting of just the third link. The network is shown in Figure 6.1.

Partitioning the columns of A such that

$$A_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

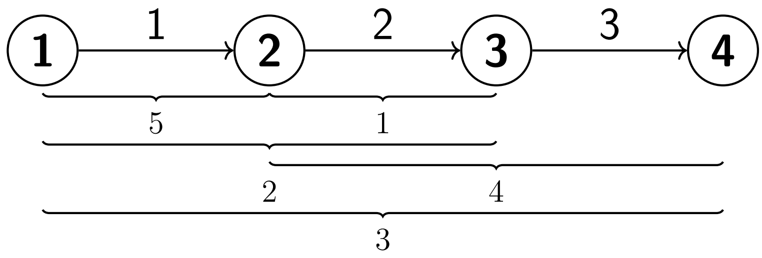


Figure 6.1: The three-link linear network from Example 6.1.1. Note that the third link is not itself an allowed path.

induces the column partition lattice basis

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

Setting $\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\intercal$, we have

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

If we were to construct a direct walk between these two points using U, we would need to use each move exactly once. Such a walk necessarily leaves the fibre because both of these two points have zeroes in the first and second co-ordinates, and each move in U has a -1 at one of them. However, if we were to allow walks that visit a point such as $\begin{bmatrix} -1 & 1 & 0 & 1 & 0 \end{bmatrix}^\intercal$ or $\begin{bmatrix} 1 & -1 & 1 & 0 & 1 \end{bmatrix}^\intercal$, both of which have a -1 in some co-ordinate, then such a walk would connect the two elements of $\mathcal{F}_{\mathbf{y}}$.

In Example 6.1.1 the two points are connected by a walk that is allowed to visit points that have -1 in either of the first two co-ordinates. In fact, for this configuration matrix this condition is sufficient to connect all pairs of points in all fibres using moves in U.

Two of the standard basis vectors are $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^\intercal$ and $\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}^\intercal$. We will say that U is both an $\mathbf{e}_1$-*Markov basis* and an $\mathbf{e}_2$-*Markov basis*, because both $-\mathbf{e}_1$ and $-\mathbf{e}_2$ are lower bounds on visited points that guarantee for every fibre, a walk using U can visit every point in the fibre.

More generally, we make the following definition:

Definition 6.1.2 (An $\mathbf{m}$-Markov basis)**.** Let $A_{n\times r}$ be a configuration matrix, let $\mathcal{B}$ be a set of moves, and let $\mathbf{m} \in \mathbb{Z}^r_{\geq 0}$ be a non-negative integer vector. If for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, the basis $\mathcal{B}$ can be used to construct a walk between any pair of points in $\mathcal{F}_{\mathbf{y}}$ such that the walk visits only points that are elementwise at least $-\mathbf{m}$, then $\mathcal{B}$ is an $\mathbf{m}$-*Markov basis.*

Often, vectors $\mathbf{m}$ will have all entries the same, which prompts the following definition.

Definition 6.1.3 (An m-Markov basis)**.** Let $\mathcal{B}$ be an $\mathbf{m}$-Markov basis, and suppose that $\mathbf{m}$ has all entries equal to m. Then we will say $\mathcal{B}$ is an *m-Markov basis.*

By these definitions, a true Markov basis is both a $\mathbf{0}$-Markov basis and a 0-Markov basis. In Example 6.1.1, the basis $\mathcal{B}$ is also an 1-Markov basis, because $\mathbf{m}_1 \leq \mathbf{1}$, or because $\mathbf{m}_2 \leq \mathbf{1}$.

We can also define analogous versions of Markov sub-bases.

Definition 6.1.4 (An $\mathbf{m}$-Markov sub-basis)**.** Let $A_{n\times r}$ be a configuration matrix, let $\mathcal{B}$ be a set of moves, and let $\mathbf{m} \in \mathbb{Z}^r_{\geq 0}$ and $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be non-negative integer vectors. If $\mathcal{B}$ can be used to construct a walk between any pair of points in $\mathcal{F}_{\mathbf{y}}$ such that the walk visits only points that are elementwise at least $-\mathbf{m}$, then $\mathcal{B}$ is an $\mathbf{m}$-*Markov sub-basis* for $\mathbf{y}$.

The scalar version is an m-Markov sub-basis.

Definition 6.1.5 (An m-Markov sub-basis)**.** Let $\mathcal{B}$ be an $\mathbf{m}$-Markov sub-basis for some configuration matrix A and vector $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, and suppose that $\mathbf{m}$ has all entries equal to m. Then we will say $\mathcal{B}$ is an *m-Markov sub-basis* for $\mathbf{y}$.

The focus of this chapter is $\mathbf{m}$-Markov bases and m-Markov bases of unimodular matrices. The main topic of this chapter is the Minus One Conjecture (Conjecture 6.3.2), found in Section 6.3. The Minus One Conjecture claims that for unimodular configuration matrices, any column partition lattice basis is a 1-Markov basis. In fact it is slightly stronger — it says that only the entries corresponding to the A_1 part of the column partition need drop down to -1. That is, for any fibre and any column partition lattice basis of any unimodular configuration matrix, a random walk that can visit points whose co-ordinates are all at least -1 in their A_1 co-ordinates (and at least 0 in their A_2 co-ordinates) is guaranteed to be able to visit, and therefore sample, any point in the fibre.

We begin in Section 6.2 by looking at $\mathbf{m}$-Markov bases generally. In Section 6.2.1 we discuss two interpretations of the idea of an $\mathbf{m}$-Markov basis. These are:

- For any fibre for the given configuration matrix, the basis allows construction of a walk that can visit every point if the walk may visit points outside the fibre that are at least $-\mathbf{m}$.
- For any fibre for the given configuration matrix, the basis allows construction of a walk that can visit every point in the fibre that is greater than or equal to $\mathbf{m}$, without ever leaving the fibre.

Theorem 4.3 says that in order to show that a set of moves $\mathcal{B}$ is a Markov basis, we need only show that $\mathcal{B}$ connects the positive and negative parts of each element of some other known Markov basis. In Section 6.2.2 we extend this principle to $\mathbf{m}$-Markov bases.

Section 6.3 concerns the Minus One Conjecture, which is stated in Conjecture 6.3.2. We then give a justification for why we are only concerned with the A_1 co-ordinates.

The interpretations of an $\mathbf{m}$-Markov basis given in Section 6.2.1 are specialised for unimodular configuration matrices in Section 6.3.1. One interpretation says that the points in any fibre that are at least $\mathbf{m}$ are connected by any $\mathbf{m}$-Markov basis. The Minus One Conjecture claims that a column partition lattice basis of a unimodular configuration matrix is an $\mathbf{m}_{(1,0)}$-Markov basis. The vector $\mathbf{m}_{(1,0)}$ is defined in Definition 6.3.1. With this choice of $\mathbf{m}$, the points in the fibre that are at least $\mathbf{m}$ become the points in the $\mathbb{Z}$-polytope that do not lie on a bounding hyperplane that corresponds to an A_1 co-ordinate.

In Section 6.3.2 we prove a theorem that reduces the problem of proving that a particular move can be simulated using a column partition lattice basis to a problem of ordering columns of a submatrix of $A_1^{-1}A$. This has a few advantages over using the definition of a Markov basis directly.

These advantages will be put to use in Section 6.4, where we prove some specific cases of the Minus One Conjecture. We begin by looking at a known 1-Markov basis from Chen, Dinwoodie and Yoshida [12]. Their Proposition 0.2.1 states that if the monomial difference representations of a set of moves $\mathcal{B}$ generates a radical ideal, then $\mathcal{B}$ is a 1-Markov basis. We provide a more detailed proof of this in Section 6.4.1. We then specialise the proposition for column partition lattice bases. If it were shown that any column partition lattice basis of a unimodular configuration matrix generates a radical ideal, then the Minus One Conjecture would have been proved.

In Section 6.4.2, we prove the Minus One Conjecture for the case of traffic networks on polytrees (Theorem 6.4.5). Polytrees are a kind of graph that were described in more detail in Section 5.3. The proof is valid for all network matrices, which are a generalisation of polytree link-path incidence matrices. Apart from being directly useful for analysis of traffic on polytree networks, many other kinds of problems may have network matrices as configuration matrices. These include configuration matrices of two-way contingency tables. Schrijver [40, Section 20.1] notes the existence of several algorithms for recognising network matrices that work in polynomial time. This is useful because the Minus One Conjecture automatically holds for any configuration matrix or column partition lattice basis matrix recognised as a network matrix by one of these algorithms.

In Section 6.4.3, we give a weaker lower bound for column partition lattice bases of unimodular configuration matrices that depends on the size and nullity of the configuration matrix. This is based on the work in Theorem 3.5.15 and Section 4.3.

In Section 6.4.4 we show that if the number of column partition lattice basis elements required to simulate some move is sufficiently few, and if every ordering of these moves necessarily breaks the minus one lower bound, then the matrix containing this collection of moves can not be totally unimodular. A limit on the number of moves necessary to simulate a Graver basis element based on the number of rows in the configuration matrix is given in Theorem 3.5.15. Section 6.4.4 can therefore be taken as a proof of the Minus One Conjecture for small configuration matrices as measured by number of rows.

Section 6.4.2 contains the proof of the Minus One Conjecture for network matrices. As well as being useful in its own right for configuration matrices that are network matrices, such as link-path incidence matrices of polytrees, the material in Section 6.4.5 may also provide a useful stepping stone towards a proof of the full Minus One Conjecture. Seymour [41] showed that all totally unimodular matrices arise from network matrices and certain 5×5 totally unimodular matrices using a certain set of operations. If it were shown that these operations preserve the minus one property, this would constitute a proof the Minus One Conjecture.

6.2 Alternative lower bounds

In this section we look at $\mathbf{m}$-Markov bases generally, before moving on to the specific case of the Minus One Conjecture. Recall from Definition 6.1.2 that an $\mathbf{m}$-Markov basis $\mathcal{M}$ is a set of moves such that for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, $\mathcal{M}$ can be used to construct a walk between any pair of points in $\mathcal{F}_{\mathbf{y}}$ that only visits points that are elementwise at least $-\mathbf{m}$.

6.2.1 Interpretations

There are two ways to interpret some set $\mathcal{B}$ being an $\mathbf{m}$-Markov basis for some $\mathbf{m} \in \mathbb{Z}^r_{\geq 0}$. The first is as above, which says that for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, $\mathcal{B}$ can be used to construct a walk that may visit points whose co-ordinates are at least $-\mathbf{m}$ that connects a set that includes $\mathcal{F}_{\mathbf{y}}$. This is how for example Yoshida's Proposition 1 [50], discussed in Section 6.4.1, is phrased.

The second is given by the Theorem below. It claims that $\mathcal{B}$ connects points that are in some sense internal to the polytope, with respect to co-ordinates whose entry in $\mathbf{m}$ is not zero.

Theorem 6.2.1. *Let A be a configuration matrix, and let the vector $\mathbf{m} \in \mathbb{Z}^r_{\geq 0}$. A set of moves $\mathcal{B}$ is an $\mathbf{m}$-Markov basis for A if and only if for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, $\mathcal{B}$ connects all pairs of points in $\mathcal{F}_{\mathbf{y}}$ that are at least $\mathbf{m}$ with a walk that remains within $\mathcal{F}_{\mathbf{y}}$.*

Proof. We first show that if $\mathcal{B}$ is an $\mathbf{m}$-Markov basis, then it connects all pairs of points that are at least $\mathbf{m}$. Let $\mathbf{y}$ be such that $\mathcal{F}_{\mathbf{y}}$ has at least two points that are at least $\mathbf{m}$, and write them as $\mathbf{x}_0 + \mathbf{m}$ and $\mathbf{x}_k + \mathbf{m}$, where $\mathbf{x}_0, \mathbf{x}_k \in \mathbb{Z}^r_{\geq 0}$. We need to show that there exists a walk in $\mathcal{F}_{\mathbf{y}}$ using $\mathcal{B}$ that connects these two points.

Let $\mathbf{y}_0 = A\mathbf{x}_0 = A\mathbf{x}_k$, and construct the fibre $\mathcal{F}_{\mathbf{y}_0}$. The points $\mathbf{x}_0$ and $\mathbf{x}_k$ are both in this $\mathcal{F}_{\mathbf{y}_0}$.

Because $\mathcal{B}$ is an $\mathbf{m}$-Markov basis, there is a walk using $\mathcal{B}$ connecting $\mathbf{x}_0$ to $\mathbf{x}_k$ that only visits points that are each at least $-\mathbf{m}$. This walk can be written

$$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k.$$

Then the walk

$$\mathbf{x}_0 + \mathbf{m}, \mathbf{x}_1 + \mathbf{m}, \mathbf{x}_2 + \mathbf{m}, \ldots, \mathbf{x}_k + \mathbf{m}$$

in $\mathcal{F}_{\mathbf{y}}$ travels from $\mathbf{x}_0 + \mathbf{m}$ to $\mathbf{x}_k + \mathbf{m}$ using only moves in $\mathcal{B}$. Each point is at least $\mathbf{0}$ because each $\mathbf{x}_i \geq -\mathbf{m}$. The points $\mathbf{x}_0$ and $\mathbf{x}_k$ can be any that are at least $\mathbf{m}$, so all such points are connected, as required.

Suppose now that in any fibre, a set $\mathcal{B}$ connects all pairs of points that are at least $\mathbf{m}$. We need to show that $\mathcal{B}$ is an $\mathbf{m}$-Markov basis.

Let $\mathbf{y}$ be given, and let $\mathbf{x}_0$ and $\mathbf{x}_k$ be any two points in $\mathcal{F}_{\mathbf{y}}$. Set $\mathbf{y}_m = A(\mathbf{x}_0 + \mathbf{m})$, and construct the fibre $\mathcal{F}_{\mathbf{y}_m}$. The points $\mathbf{x}_0 + \mathbf{m}$ and $\mathbf{x}_k + \mathbf{m}$ are both in $\mathcal{F}_{\mathbf{y}_m}$.

In any fibre, the set $\mathcal{B}$ connects every pair of points that are each at least $\mathbf{m}$, so there is a walk from $\mathbf{x}_0 + \mathbf{m}$ and $\mathbf{x}_k + \mathbf{m}$ which we can write as

$$\mathbf{x}_0 + \mathbf{m}, \mathbf{x}_1 + \mathbf{m}, \mathbf{x}_2 + \mathbf{m}, \ldots, \mathbf{x}_k + \mathbf{m},$$

where each $\mathbf{x}_i$ is at least $-\mathbf{m}$. Then there is a walk from $\mathbf{x}_0$ to $\mathbf{x}_k$ that visits only points that are at least $-\mathbf{m}$, which we can write as

$$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k.$$

This walk visits only points that are at least $-\mathbf{m}$, and because for any fibre, $\mathbf{x}_0$ and $\mathbf{x}_k$ can be any points in the fibre, $\mathcal{B}$ is an $\mathbf{m}$-Markov basis as required. □

We illustrate this with the following example.

Example 6.2.2. Consider the configuration matrix

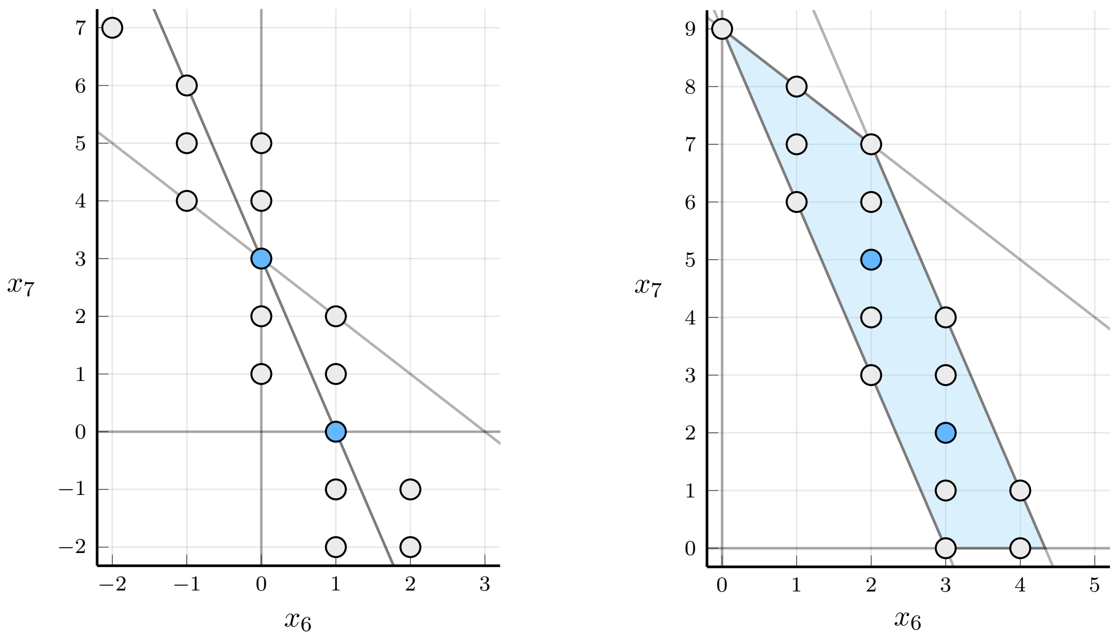
$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix},$$

and the vector $\mathbf{y}_1 = \begin{bmatrix} 2 & 0 & 3 & 1 & 1 \end{bmatrix}^\intercal$, which produces the fibre

$$\mathcal{F}_{\mathbf{y}_1} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 3 \end{bmatrix} \right\}.$$

We select the column partition lattice basis defined by

$$U = \begin{bmatrix} 3 & 1 \\ -3 & -1 \\ -1 & -1 \\ -1 & 0 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$



(a) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}2 & 0 & 3 & 1 & 1\end{bmatrix}^\intercal$. (b) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}10 & 4 & 9 & 5 & 5\end{bmatrix}^\intercal$.

Figure 6.2: The $\mathbb{Z}$-polytopes from Example 6.2.2 showing the two interpretations of a 2-Markov sub-basis.

This matrix U is a 2-Markov sub-basis for $\mathbf{y}$. The first interpretation of this is that U connects $\mathcal{F}_{\mathbf{y}_1}$ with walks that are allowed to visit points that are at least -2 in each co-ordinate. The projected $\mathbb{Z}$-polytope of $\mathcal{F}_\mathbf{y}$ under U is shown in Figure 6.2a. We can see that using U to construct a direct walk between the elements of $\mathcal{F}_\mathbf{y}$ (in blue) requires that we visit one of the points with projected co-ordinates $\begin{bmatrix}0 & 1\end{bmatrix}^\intercal$ or $\begin{bmatrix}1 & 2\end{bmatrix}^\intercal$. These points, and the others with all co-ordinates at least -2 which we may also visit, are shown in grey. The full co-ordinates of these points are $\begin{bmatrix}-2 & 2 & 2 & 1 & 1 & 0 & 1\end{bmatrix}^\intercal$ and $\begin{bmatrix}2 & -2 & 0 & 0 & 0 & 1 & 2\end{bmatrix}^\intercal$, so each has a -2 in some co-ordinate.

The second interpretation of an m-Markov sub-basis is that it connects all of the points in the fibre that are at least m with a walk that remains within the fibre. To illustrate, consider

$$\begin{aligned}\mathbf{y}_2 &= \mathbf{y}_1 + A\begin{bmatrix}2 & 2 & 2 & 2 & 2 & 2 & 2\end{bmatrix}^\intercal \\ &= \begin{bmatrix}10 & 4 & 9 & 5 & 5\end{bmatrix}^\intercal.\end{aligned}$$

The matrix U is also a 2-Markov sub-basis for this fibre, which is given by

$$\mathcal{F}_{\mathbf{y}_2} = \left\{ \begin{bmatrix}0\\4\\0\\5\\5\\0\\9\end{bmatrix} \begin{bmatrix}0\\4\\2\\4\\4\\1\\6\end{bmatrix} \begin{bmatrix}1\\3\\1\\4\\4\\1\\7\end{bmatrix} \begin{bmatrix}2\\2\\0\\4\\4\\1\\8\end{bmatrix} \begin{bmatrix}0\\4\\4\\3\\3\\2\\3\end{bmatrix} \begin{bmatrix}1\\3\\3\\3\\3\\2\\4\end{bmatrix} \begin{bmatrix}2\\2\\2\\3\\3\\2\\5\end{bmatrix} \begin{bmatrix}3\\1\\1\\3\\3\\2\\6\end{bmatrix} \begin{bmatrix}4\\0\\0\\3\\3\\2\\7\end{bmatrix} \begin{bmatrix}0\\4\\6\\2\\2\\3\\0\end{bmatrix} \begin{bmatrix}1\\3\\5\\2\\2\\3\\1\end{bmatrix} \begin{bmatrix}2\\2\\4\\2\\2\\3\\2\end{bmatrix} \begin{bmatrix}3\\1\\3\\2\\2\\3\\3\end{bmatrix} \begin{bmatrix}4\\0\\2\\2\\2\\3\\4\end{bmatrix} \begin{bmatrix}3\\1\\5\\1\\1\\4\\0\end{bmatrix} \begin{bmatrix}4\\0\\4\\1\\1\\4\\1\end{bmatrix} \right\}.$$

Only two of these elements are at least 2 in every co-ordinate:

$$\begin{bmatrix}2\\2\\2\\3\\3\\2\\5\end{bmatrix}, \quad \text{and} \quad \begin{bmatrix}2\\2\\4\\2\\2\\3\\2\end{bmatrix}.$$

These are the two elements of $\mathcal{F}_{\mathbf{y}_1}$ with two added in every co-ordinate. Figure 6.2b shows these two elements in blue, and the other elements of $\mathcal{F}_{\mathbf{y}_2}$ in grey, and comparing the two panels shows the obvious correspondence between these two interpretations of an $\mathbf{m}$-Markov sub-basis.

6.2.2 m-Markov basis simulation

In Section 4.3 we saw that for a set to be a Markov basis, we need only show that it connects $\mathbf{z}^-$ to $\mathbf{z}^+$ for each element $\mathbf{z}$ of some other known Markov basis, such as the Graver basis. The same argument applies when determining whether, given some $\mathbf{m} \geq \mathbf{0}$, a set $\mathcal{B}$ is an $\mathbf{m}$-Markov basis. If we can show that for each element $\mathbf{z}$ of some known Markov basis $\mathcal{M}$, we can use $\mathcal{B}$ to construct a walk between $\mathbf{z}^-$ and $\mathbf{z}^+$ that only visits points that are each at least $-\mathbf{m}$, then $\mathcal{B}$ is an $\mathbf{m}$-Markov basis.

Theorem 6.2.3. *Let A be a configuration matrix and let $\mathcal{M}$ be a Markov basis for A, and let $\mathbf{m} \geq \mathbf{0}$. Let $\mathcal{B}$ be a set of moves in $\ker_{\mathbb{Z}}(A)$. Then $\mathcal{B}$ is an $\mathbf{m}$-Markov basis if and only if for all $\mathbf{z} \in \mathcal{M}$, moves in $\mathcal{B}$ can be used to construct a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that visits only points that are at least $-\mathbf{m}$.*

Proof. Suppose first that for each $\mathbf{z} \in \mathcal{M}$, $\mathcal{B}$ can be used to construct a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that only visits points that are at least $-\mathbf{m}$. By the proof of the Fundamental Theorem of Markov Bases (Theorem 2.4.2),

$$T^{\mathbf{m}+\mathbf{z}^-} - T^{\mathbf{m}+\mathbf{z}^+} \in I_{\mathcal{B}}.$$

By Theorem 6.2.1, in order to show that $\mathcal{B}$ is an $\mathbf{m}$-Markov basis we must show that in all fibres, $\mathcal{B}$ connects all pairs of points $\mathbf{m} + \mathbf{x}_1$, $\mathbf{m} + \mathbf{x}_2$ where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{Z}^r_{\geq 0}$. Given such a pair, the set $\mathcal{M}$ is a Markov basis, so we have $T^{\mathbf{x}_1} - T^{\mathbf{x}_2} \in I_{\mathcal{M}}$, and we can factor out common factors, leaving the positive and negative parts of a kernel element $\mathbf{z}$:

$$T^{\mathbf{x}_1} - T^{\mathbf{x}_2} = \sum_{\mathbf{z}\in\mathcal{M}} T^{\mathbf{x}_0}(T^{\mathbf{z}^-} - T^{\mathbf{z}^+})$$

for some $\mathbf{x}_0 \in \mathbb{Z}^r_{\geq 0}$. Multiplying by $T^{\mathbf{m}}$, we have

$$T^{\mathbf{m}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) = \sum_{\mathbf{z}\in\mathcal{M}} T^{\mathbf{x}_0}T^{\mathbf{m}}(T^{\mathbf{z}^-} - T^{\mathbf{z}^+}) \in I_{\mathcal{B}},$$

since $T^{\mathbf{m}}(T^{\mathbf{z}^-} - T^{\mathbf{z}^+}) \in I_{\mathcal{B}}$ for each $\mathbf{z} \in \mathcal{M}$. Therefore,

$$T^{\mathbf{m}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) \in I_{\mathcal{B}},$$

so $\mathcal{B}$ is an $\mathbf{m}$-Markov basis as required.

Suppose now that there exists $\mathbf{z} \in \mathcal{M}$ such that $\mathcal{B}$ cannot connect $\mathbf{z}^-$ to $\mathbf{z}^+$ with a walk that visits points that are at least $-\mathbf{m}$. Then there exists $\mathbf{y} = A\mathbf{z}^+ = A\mathbf{z}^-$ that contains two points $\mathbf{z}^-, \mathbf{z}^+ \in \mathcal{F}_{\mathbf{y}}$ that are not connect by a walk using $\mathcal{B}$ that only visits points that are at least $-\mathbf{m}$. Then by definition $\mathcal{B}$ is not an $\mathbf{m}$-Markov basis. □

Using this theorem requires that we have knowledge of some Markov basis. Failing this, we can instead prove connectivity via simulating elements of the integer kernel.

Theorem 6.2.4. *Let A be a configuration matrix, let $\mathbf{m} \in \mathbb{Z}^n_{\geq 0}$, and let $\mathcal{B}$ be a set of moves in $\ker_{\mathbb{Z}}(A)$. Then $\mathcal{B}$ is an $\mathbf{m}$-Markov basis if and only if for all $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$, moves in $\mathcal{B}$ can be used to construct a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that only visits points that are at least $-\mathbf{m}$.*

Proof. Suppose first that $\mathcal{B}$ is an $\mathbf{m}$-Markov basis. For any $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$, we can set $\mathbf{y} = A\mathbf{z}^+ = A\mathbf{z}^-$, and $\mathbf{z}^-, \mathbf{z}^+ \in \mathcal{F}_{\mathbf{y}}$. From the definition of an $\mathbf{m}$-Markov basis, $\mathcal{B}$ connects $\mathbf{z}^-$ to $\mathbf{z}^+$ with a walk that visits only points that are at least $-\mathbf{m}$.

Suppose now that for any $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$, $\mathcal{B}$ can be used to construct a walk from $\mathbf{z}^-$ to $\mathbf{z}^+$ that only visits points that are at least $-\mathbf{m}$. By the Fundamental Theorem of Markov Bases 2.4.5,

$$T^{\mathbf{m}}(T^{\mathbf{z}^+} - T^{\mathbf{z}^-}) \in \langle T^{\mathbf{u}^+} - T^{\mathbf{u}^-} : \mathbf{u} \in \mathcal{B}\rangle = I_{\mathcal{B}}.$$

We need to show that $\mathcal{B}$ is an $\mathbf{m}$-Markov basis. Let $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$ be given. If $\mathcal{F}_{\mathbf{y}}$ is empty or contains only one element, we are done. If not, let $\mathbf{x}_1, \mathbf{x}_2$ be any pair of points in $\mathcal{F}_{\mathbf{y}}$. We can write $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$, where $\mathbf{z} \in \ker_{\mathbb{Z}}(A)$. We set $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{z}^+ = \mathbf{x}_2 - \mathbf{z}^-$. Then

$$T^{\mathbf{m}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) = T^{\mathbf{x}_0}T^{\mathbf{m}}(T^{\mathbf{z}^-} - T^{\mathbf{z}^+}) \in I_{\mathcal{B}},$$

so $\mathbf{x}_1$ and $\mathbf{x}_2$ are connected by a walk using $\mathcal{B}$ that only visits points that are at least $-\mathbf{m}$. Therefore $\mathcal{B}$ is an $\mathbf{m}$-Markov basis, as required. □

6.3 The Minus One Conjecture

Following Definition 6.1.3, a 1-Markov basis for a configuration matrix A is a collection of moves that, given any $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, can be used to construct a random walk that visits every point in $\mathcal{F}_{\mathbf{y}}$ if we allow the walk to step outside $\mathcal{F}_{\mathbf{y}}$ to also visit points with a -1 in some co-ordinate.

We conjecture that any column partition lattice bases of a unimodular configuration matrix is a 1-Markov basis. In fact, due to the nature of column partition lattice bases, we can strengthen this conjecture. Suppose that a column partition lattice basis U of a configuration matrix A is a 1-Markov basis, and suppose that in order to walk between any pair of points we can construct the walk such that there are no moves that are later undone. Then we know that during the walk, each of the co-ordinates corresponding to columns of A in the A_2 partition move from either 1 to 0, or 0 to 1, if they move at all. We never visit a point whose value in this co-ordinate is anything other than zero or one. For the purposes of lower bounds we are therefore only concerned with the co-ordinates corresponding to the columns of A_1. Depending on the dimensions of the configuration matrix, this can be a very small proportion of the total number of co-ordinates.

The Minus One Conjecture requires the following definition:

Definition 6.3.1 (The vector $\mathbf{m}_{(1,0)}$)**.** Let A be a configuration matrix. Given a partition of the columns of A into A_1 and A_2, the vector $\mathbf{m}_{(1,0)}$ is the vector containing 1 in the entries corresponding to columns of A in the A_1 partition, and 0 in the entries corresponding to columns in the A_2 partition.

Conjecture 6.3.2 (The Minus One Conjecture)**.** *Let $A \in \{0,1\}^{r\times n}$ be a unimodular configuration matrix and let U be a column partition lattice basis for A. Then U is an $\mathbf{m}_{(1,0)}$-Markov basis.*

Example 6.3.3. Consider again the unimodular configuration matrix and column partition lattice basis from Example 6.1.1, where

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

and

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

Then the vector $\mathbf{m}_{(1,0)}$ is given by $\mathbf{m}_{(1,0)} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \end{bmatrix}^\intercal$. This column partition lattice basis is an $\mathbf{e}_1$-Markov basis. Since $\mathbf{m}_{(1,0)} \geq \mathbf{e}_1$, this basis is also an $\mathbf{m}_{(1,0)}$-Markov basis.

For larger matrices, the order in which the moves are used may be important, as illustrated by the following example.

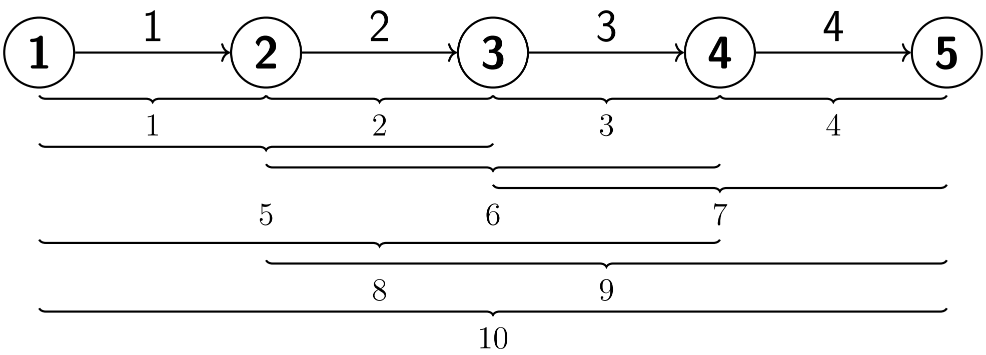


Figure 6.3: The four-link linear network in Example 6.3.4. Underbraces show the allowed paths.

Example 6.3.4. Let A be the link-path incidence matrix of the four-link linear traffic network shown in Figure 6.3. Then

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

and A is totally unimodular. The matrix

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

defines a column partition lattice basis. Setting $\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 0 \end{bmatrix}^\intercal$, we have

$$\mathcal{F}_{\mathbf{y}} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

We will call the first two elements of $\mathcal{F}_{\mathbf{y}}$ $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. We wish to see if U connects $\mathbf{x}_1$ to $\mathbf{x}_2$ with a walk with a lower bound of -1.

The signed moves in U required to walk from $\mathbf{x}_1$ to $\mathbf{x}_2$ are given by the columns of the matrix

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

If the moves are used in the order given, then after two steps the walk is at a point with -2 in the tenth co-ordinate. Switching either of the first two columns with either of the last two columns gives an ordering of moves such that the walk visits only points with at least -1 in all co-ordinates.

The following example shows that the U matrix having all entries in $\{0, \pm 1\}$ does not guarantee that the column partition lattice basis it defines is an $\mathbf{m}_{(1,0)}$-Markov basis.

Example 6.3.5. Consider the matrix

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

which has a column partition lattice basis

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

Setting $\mathbf{y} = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \end{bmatrix}^{\intercal}$, we have

$$\begin{aligned} \mathcal{F}_{\mathbf{y}} = \{ & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^{\intercal}, \\ & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}^{\intercal} \}. \end{aligned}$$

Every walk between the two elements of $\mathcal{F}_{\mathbf{y}}$ using only moves in U must visit a point with a -2 in one of the first six co-ordinates. Therefore U is not an $\mathbf{m}_{(1,0)}$-Markov basis. The matrix U is non totally unimodular, so it does not constitute a counterexample to the Minus One Conjecture.

6.3.1 The internal points

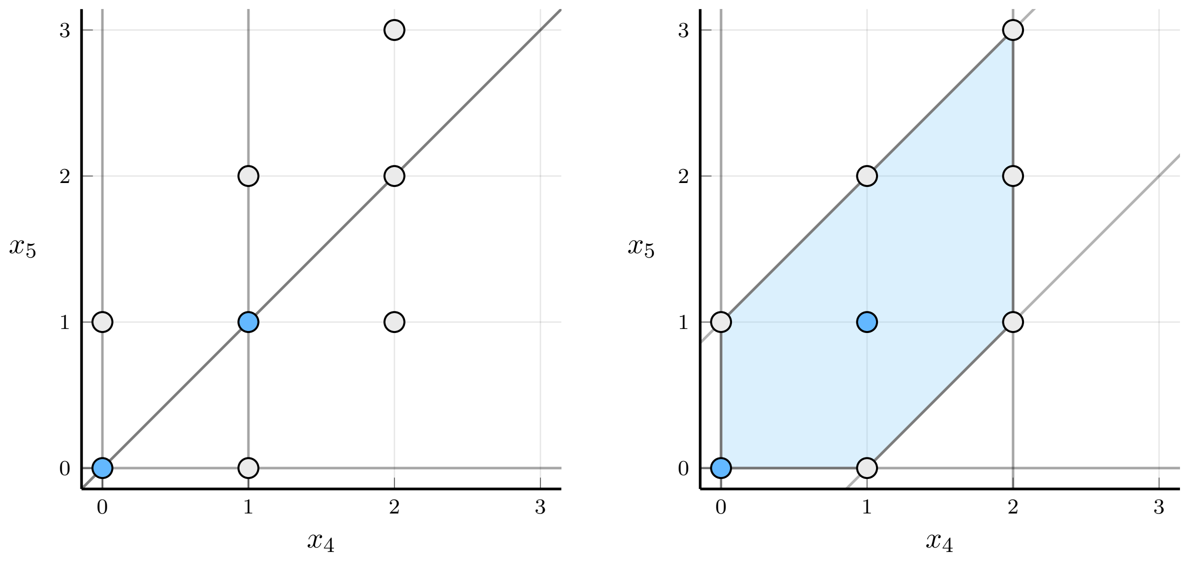
In Section 6.2.1 we discussed two possible interpretations of an $\mathbf{m}$-Markov basis. The second is that in any given fibre, $\mathcal{B}$ connects all of the points that are at least $\mathbf{m}$ to each other with a walk that does not leave the fibre. These points are in some sense internal to the polytope. In terms of column partition lattice bases that are $\mathbf{m}_{(1,0)}$-Markov bases, this means all of the points in $\mathcal{F}_{\mathbf{y}}$ that do not have a zero in any of the A_1 co-ordinates. Geometrically, these are the points that do not lie on one of the A_1 bounding hyperplanes.

We demonstrate these two interpretations with an example.

Example 6.3.6. Consider again the configuration matrix from Example 6.1.1,

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

This is a link-path incidence matrix for the three-link linear network, where we have removed the path consisting of just the third link. The column partition lattice basis in



(a) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$. (b) The $\mathbb{Z}$-polytope for $\mathbf{y} = \begin{bmatrix}3 & 4 & 2\end{bmatrix}^\intercal$.

Figure 6.4: The $\mathbb{Z}$-polytopes from Example 6.3.6 showing the two interpretations of a $\mathbf{m}_{(1,0)}$-Markov sub-basis.

that example is given by

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

Setting $\mathbf{y}_1 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$ produces the fibre

$$\mathcal{F}_{\mathbf{y}_1} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

This fibre is shown Figure 6.4a. The elements of $\mathcal{F}_{\mathbf{y}_1}$ are shown in blue; the points in grey are the solutions to $A\mathbf{x} = \mathbf{y}$ for $\mathbf{y} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal$ that are at least -1 in the A_1 co-ordinates. A walk that may also visit the grey points can connect the elements of $\mathcal{F}_{\mathbf{y}_1}$.

In Figure 6.4b, we see the $\mathbb{Z}$-polytope for

$$\begin{aligned} \mathbf{y}_2 &= \mathbf{y}_1 + A_1 \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\intercal \\ &= \begin{bmatrix}3 & 4 & 2\end{bmatrix}^\intercal. \end{aligned}$$

The points in $\mathcal{F}_{\mathbf{y}_2}$ that are at least 1 in the A_1 co-ordinates are shown in blue. A random walk through the fibre can connect these points. Note the correspondence between the points that are guaranteed to be connected and the points used to connect them under each interpretation.

6.3.2 Column orderings

Theorem 6.2.3 states that we can prove that a collection of moves is an $\mathbf{m}$-Markov basis by showing that it can simulate any move in some known Markov basis $\mathcal{M}$ with a walk that visits only points that are at least $-\mathbf{m}$. If we have a column partition lattice basis and wish to prove that it is an $\mathbf{m}$-Markov basis, we may need to show that we can simulate particular Markov basis moves. In this section we look at simulating an integer kernel element $\mathbf{z}$ using moves from a column partition lattice basis U, and present a condition for $\mathbf{z}^+$ and $\mathbf{z}^-$ to be connected with a walk that uses U and only visits points that are at least $\mathbf{m}$.

We first give a general theorem for column partition lattice bases with all integer entries. A condition on the configuration matrix and choice of partition that guarantees that U has all integer entries can be found in Theorem 3.4.5.

Theorem 6.3.7. *Let A be an $n \times r$ configuration matrix and let π be a partition of columns of A. Let $U \in \mathbb{Z}^{r\times(r-n)}$ be the column partition lattice basis for A induced by π.*

Let $\mathbf{m} \in \mathbb{Z}^n_{\geq 0}$ be a vector, and construct the vector

$$\mathbf{m}_0 = \begin{bmatrix} \mathbf{m} \\ \mathbf{0} \end{bmatrix} \in \mathbb{Z}^r_{\geq 0}.$$

Let $\mathbf{z}$ be an element of $\ker_{\mathbb{Z}}(A)$.

Construct the matrix $(A_1^{-1}A)_{\mathbf{z}}$ by collecting $|z_i|$ copies of the ith column of $A_1^{-1}A$ for $i = 1, \dots, r$, where each column is signed according to the sign of $-z_i$. If we can place the s columns of $(A_1^{-1}A)_{\mathbf{z}}$ in an order $\mathbf{a}_1, \dots, \mathbf{a}_s$ (reindexed to reflect this order) such that each of the sums $\sum_{i=1}^{j} \mathbf{a}_i$ satisfies

$$\sum_{i=1}^{j} \mathbf{a}_i \geq -\mathbf{m} \text{ for } j = 1, \dots, s,$$

then U can simulate $\mathbf{z}$ with a walk that only visits points that are at least $-\mathbf{m}_0$.

Proof. We need to show that $\mathbf{z}^-$ and $\mathbf{z}^+$ are connected with a walk that visits points that are at least $\mathbf{m}_0$. The matrix U is a column partition lattice basis, so the moves in U required to simulate $\mathbf{z}$ are given by $\mathbf{z}_2$, the co-ordinates of $\mathbf{z}$ correspond to the A_2 part of A. The moves required are $z_i\mathbf{u}_i$ for each $z_i \in \mathbf{z}_2$, so we need to show that these moves can be ordered $\mathbf{u}_{i_1}, \dots, \mathbf{u}_{i_k}$ such that for each $j = 1, \dots, k$, each of the partial sums $\mathbf{z}^- + \sum_{m=1}^{j} z_{i_m}\mathbf{u}_{i_m}$ satisfies

$$\mathbf{z}^- + \sum_{i=1}^{j} z_i\mathbf{u}_i \geq -\mathbf{m}.$$

By hypothesis, for each $j = 1, \ldots, s$, we have

$$\sum_{i=1}^{j} \mathbf{a}_i \geq -\mathbf{m}.$$

If some $\mathbf{a}_p \geq \mathbf{0}$, then for $j = 1, \ldots, p-1$ we have

$$\mathbf{a}_p + \sum_{i=1}^{j} \mathbf{a}_i \geq -\mathbf{m}.$$

In other words, we can move all of the non-negative $\mathbf{a}_i$ to the start of the sequence without breaking the inequalities.

We can move the non-positive entries to the end of the sequence, too. If for some q we have $\mathbf{a}_q \leq \mathbf{0}$, then for $j = q+1, \ldots, s$ we have

$$\sum_{i=1}^{q-1} \mathbf{a}_i + \sum_{q+1}^{j} \mathbf{a}_i \geq -\mathbf{m}.$$

The columns of $A_1^{-1} A_{\mathbf{z}}$ that derive from the A_1 part of A all have one non-zero entry, so they all meet one of these two conditions. We move the non-negative columns that derive from the A_1 part of A to the beginning of the sequence and combine them by summing, and the non-positive ones to the end and sum them. The sequence is now of the form

$$\left(\sum_{\mathbf{e}_i \in A_1^{-1} A_{\mathbf{z}}} \mathbf{e}_i \right), \mathbf{a}_p, \ldots, \mathbf{a}_q, \left(\sum_{-\mathbf{e}_i \in A_1^{-1} A_{\mathbf{z}}} -\mathbf{e}_i \right)$$

for some indices p and q. Each vector $\mathbf{a}_i$ is equal to the first n entries of a vector $\mathbf{u} \in U$. The sum of this sequence is $\mathbf{0}$, and the sum of those elements $\mathbf{a}_i$ is equal to the first n entries of $\mathbf{z}$, so

$$\sum_{\mathbf{e}_i \in A_1^{-1} A_{\mathbf{z}}} \mathbf{e}_i = \hat{\mathbf{z}}^- \quad \text{and} \quad \sum_{-\mathbf{e}_i \in A_1^{-1} A_{\mathbf{z}}} \mathbf{e}_i = \hat{\mathbf{z}}^+,$$

where the caret means only the first n entries. Rewriting the sequence in these terms gives

$$\hat{\mathbf{z}}^-, \hat{\mathbf{u}}_p, \ldots, \hat{\mathbf{u}}_q, \hat{\mathbf{z}}^+,$$

and each of the partial sums is at least $-\mathbf{m}$.

When proving that a column partition lattice basis is an $\mathbf{m}_0$-Markov basis, we are only interested in the first n terms. This means that there is a sequence

$$\mathbf{z}^-, \mathbf{u}_p, \ldots, \mathbf{u}_q, \mathbf{z}^+,$$

such that each of the partial sums is a least $-\mathbf{m}$, so U is an $\mathbf{m}_0$-Markov basis. □

This theorem can be simplified when applied to proving Conjecture 6.3.2. If the configuration matrix is unimodular, we can choose the Graver basis as our Markov basis. Each $\mathbf{g} \in \mathcal{G}_A$ has all entries either zero or ± 1, and so at most one copy of each column of $A_1^{-1}A$ is needed to construct $(A_1^{-1}A)_{\mathbf{g}}$.

Corollary 6.3.8. *Let A be an $n \times r$ unimodular configuration matrix, let π be a partition of the columns of A, and let $\mathbf{m}_{(1,0)}$ be defined as in Definition 6.3.1. Let $\mathbf{g}$ be an element of $\mathcal{G}_A$, and let U be the column partition lattice basis for A induced by π. Construct the matrix $(A_1^{-1}A)_{\mathbf{g}}$ by collecting the columns $-g_i\mathbf{a}_i$ for each $i \in \operatorname{supp}(\mathbf{g})$, where $\mathbf{a}_i$ is the ith column of $A_1^{-1}A$. If we can place the columns of $(A_1^{-1}A)_{\mathbf{g}}$ in an order $\mathbf{a}_1, \dots, \mathbf{a}_s$ such that for $j = 1, \dots, s$, each of the sums $\sum_{i=1}^{j} \mathbf{a}_i$ is at least -1, then U connects $\mathbf{g}^-$ to $\mathbf{g}^+$ with a walk that only visits points whose co-ordinates are at least $-\mathbf{m}_{(1,0)}$.*

There are a few advantages of using this condition over the usual condition of finding an ordering of the necessary moves $\mathbf{u}_i \in U_{\mathbf{g}}$ in U such that each of the partial sums $\mathbf{g}^- + \sum_{i=1}^{j} \mathbf{u}_i$ satisfies

$$\mathbf{g}^- + \sum_{i=1}^{j} \mathbf{u}_i \geq \mathbf{m}_{(1,0)} \text{ for } j = 1, \dots, k.$$

Some of these are:

1. The matrix $(A_1^{-1}A)_{\mathbf{g}}$ is totally unimodular.

2. The sum of the entries in each row of $(A_1^{-1}A)_{\mathbf{g}}$ is zero.

3. If A is the link-path incidence matrix of a polytree, then the columns of $A_1^{-1}A$ and $(A_1^{-1}A)_{\mathbf{g}}$ correspond to paths on another related polytree (see Section 5.3.1).

The first property is true of the matrix $U_{\mathbf{g}}$, but not the augmented matrix $\begin{bmatrix}\mathbf{g}^- & U_{\mathbf{g}} & \mathbf{g}^+\end{bmatrix}$; and the second is true of $\begin{bmatrix}\mathbf{g}^- & U_{\mathbf{g}} & \mathbf{g}^+\end{bmatrix}$ but not $U_{\mathbf{g}}$. In the proofs in Section 6.4.4, these two properties mean that operating on columns of $(A_1^{-1}A)_{\mathbf{g}}$ instead of $U_{\mathbf{g}}$ or $\begin{bmatrix}\mathbf{g}^- & U_{\mathbf{g}} & \mathbf{g}^+\end{bmatrix}$ significantly reduces the number of calculations required.

The third property is used in the proof of Theorem 6.4.5.

We illustrate the correspondence between ordering columns of $U_{\mathbf{g}}$ and columns of $(A_1^{-1}A)_{\mathbf{g}}$ with the following example.

Example 6.3.9. Consider the four-link linear network from Example 6.3.4. The configuration matrix is

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

and a column partition lattice basis is given by

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

Suppose we wish to simulate the move $\mathbf{g} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}^\intercal \in \mathcal{G}_A$.

The moves in U required are $\mathbf{u}_2$ and $\mathbf{u}_3$. If we want to simulate $\mathbf{g}$ using U, we need to place the moves

$$\epsilon_{i_1}\mathbf{u}_{i_1}, \epsilon_{i_2}\mathbf{u}_{i_2}$$

in an order such that the partial sums

$$\begin{aligned} &\mathbf{g}^- \\ &\mathbf{g}^- + \epsilon_{i_1}\mathbf{u}_{i_1} \\ &\mathbf{g}^- + \epsilon_{i_1}\mathbf{u}_{i_1} + \epsilon_{i_2}\mathbf{u}_{i_2} \end{aligned}$$

never drop below -1 in any entry. The ϵ_i represent the signs of the moves.

Together with $\mathbf{g}^-$ and $\mathbf{g}^+$, the list of columns to be ordered is as shown in Table 6.1. When the moves are ordered $(\mathbf{u}_2, \mathbf{u}_3)$, the partial sums never go below -1 in any co-

$-\mathbf{g}$	$\mathbf{g}^-$	$\mathbf{u}_2$	$\mathbf{u}_3$	$-\mathbf{g}^+$
0	0	0	0	0
−1	0	1	0	−1
−1	0	0	1	−1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	1	−1	0
1	1	0	−1	0
1	1	−1	0	0
−1	0	0	1	−1

Table 6.1: The moves required to simulate $\mathbf{g}$ in Example 6.3.9.

ordinate. The matrix $A_1^{-1}A$ is given by

$$A_1^{-1}A = \begin{bmatrix} 0 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

so the columns of $A_1^{-1}A$ that $\mathbf{g}$ specifies sum to $\mathbf{0}$ are given by

$$(A_1^{-1}A)_{\mathbf{g}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

Under this ordering of columns, no partial sum goes below -1.

Clearly, we can then move any column $\mathbf{e}_i$ to the beginning and $-\mathbf{e}_i$ to the end of the sequence. Doing so can only increase or leave unchanged the partial sums. Keeping the order of the two remaining columns produces

$$\begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Again, the sequence of partial sums of this ordering of columns never goes below -1.

There are two columns $\mathbf{e}_i$ at the beginning of the sequence; summing these produces the A_1 part of $\mathbf{g}^-$. The A_1 part of $-\mathbf{g}^+$ is given by the fifth column. This ordering then corresponds to using the move $\mathbf{u}_2$ first and the move $\mathbf{u}_3$ second to simulate $\mathbf{g}$ with U.

6.4 Proofs

In this section we give some proofs of specific cases of the Minus One Conjecture, and in Section 6.4.3 we give a proof of a weaker lower bound for unimodular configuration matrices.

6.4.1 Radical ideals

One condition that guarantees that a lattice basis is a 1-Markov basis is given by Chen et al. [12]. Chen et al. were focused solely on resampling contingency tables, but their result applies to configuration matrices for general statistical linear inverse problems. The condition requires the following definition.

Definition 6.4.1 (Radical ideal [13])**.** An ideal I is *radical* if $f^m \in I$ for some integer $m \geq 1$ implies $f \in I$.

Proposition 6.4.2 (Proposition 0.2.1 in Chen et al. [12])**.** *Suppose $I_{\mathcal{B}}$ is a radical ideal, and suppose the moves in $\mathcal{B}$ form a lattice basis. Then for any $\mathbf{y} \geq \mathbf{0}$ the Markov chain using the moves in $\mathcal{B}$ that allows entries to drop down to -1 connects a set that includes the set $\mathcal{F}_{\mathbf{y}}$.*

Any column partition lattice basis that generates a radical ideal is therefore a 1-Markov basis. As we will see, because it is a column partition lattice bases, it is also a $\mathbf{m}_{(1,0)}$-Markov basis. If a proof could be found that all column partition lattice basis of unimodular configuration matrices generate radical ideals, this would constitute a proof of the Minus One Conjecture.

To prove Proposition 6.4.2 we require the following lemma, which is a standard fact of factor ring multiplication.

Lemma 6.4.3. *Let I be an ideal, let $X - Y \in I$, and let $a \in \mathbb{N}$. Then $X^a - Y^a \in I$.*

Proof. Because $X - Y \in I$, we have

$$\begin{aligned} X^a - Y^a &= (X-Y)(X^{a-1} + X^{a-2}Y + X^{a-3}Y^2 + \cdots + X^2Y^{a-3} + XY^{a-2} + Y^{a-1}) \\ &\in I \end{aligned}$$

as required. □

We can now give the proof.

Proof of Proposition 6.4.2. Using the proof of Theorem 2.4.5, we need to show that $T^{\mathbf{1}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) \in I_{\mathcal{B}}$ whenever $A\mathbf{x}_1 = A\mathbf{x}_2$. First we will show that a monomial multiple of $T^{\mathbf{x}_1 - \mathbf{x}_2}$ lies in $I_{\mathcal{B}}$.

The moves in $\mathcal{B}$ form a lattice basis, so for any two points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$ we have

$$\mathbf{x}_2 - \mathbf{x}_1 = \sum_{i=1}^{k} a_i \mathbf{u}_i$$

for $a_i \in \mathbb{Z}$ and $\mathbf{u}_i \in \mathcal{B}$. We will assume without loss of generality that the lattice basis elements $\mathbf{u}_i$ are oriented in the direction we wish to use them, so that each $a_i \in \mathbb{N}$. Then

$$\begin{aligned} \mathbf{x}_1 - \mathbf{x}_2 &= \sum_{i=1}^{k} a_i(\mathbf{u}_i^+ - \mathbf{u}_i^-) \\ T^{\mathbf{x}_1 - \mathbf{x}_2} &= \prod_{i=1}^{k} T^{a_i(\mathbf{u}_i^+ - \mathbf{u}_i^-)} \\ \frac{T^{\mathbf{x}_1}}{T^{\mathbf{x}_2}} &= \prod_{i=1}^{k} \frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}} \\ T^{\mathbf{x}_1 - \mathbf{x}_2} - 1 &= \prod_{i=1}^{k} \frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}} - 1 \\ T^{\mathbf{x}_1} - T^{\mathbf{x}_2} &= T^{\mathbf{x}_2}\left(\prod_{i=1}^{k} \frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}} - 1\right). \end{aligned}$$

So we need to show that there is a monomial $T^{\mathbf{m_k}}$ such that

$$T^{\mathbf{m_k}}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\in I_{\mathcal{B}}.$$

We will proceed using induction. For the base case, let $k=1$ and choose $T^{\mathbf{m}_1}=T^{a_1\mathbf{u}_1^-}$. Then

$$\begin{aligned}T^{\mathbf{m}_1}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)&=T^{a_1\mathbf{u}_1^-}T^{\mathbf{x}_2}\left(\frac{T^{a_1\mathbf{u}_1^+}}{T^{a_1\mathbf{u}_1^-}}-1\right)\\&=T^{\mathbf{x}_2}\left(T^{a_1\mathbf{u}_1^+}-T^{a_1\mathbf{u}_1^-}\right)\\&\in I_{\mathcal{B}}\end{aligned}$$

by Lemma 6.4.3, since $T^{\mathbf{u}_1^+}-T^{\mathbf{u}_1^-}\in I_{\mathcal{B}}$.

For the induction, we assume that there is $T^{\mathbf{m}_{k-1}}$ such that

$$T^{\mathbf{m}_{k-1}}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\in I_{\mathcal{B}}.$$

We now need to show the same is true for k, and we choose $T^{\mathbf{m}_k}=T^{\mathbf{m}_{k-1}}\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}$. Then

$$\begin{aligned}T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)&=T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}+\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\\&=T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(T^{a_k\mathbf{u}_k^+}\frac{\prod_{i=1}^{k-1}T^{a_i\mathbf{u}_i^+}}{\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}}-\frac{T^{a_k\mathbf{u}_k^-}}{T^{a_k\mathbf{u}_k^-}}\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}+\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\\&=T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(T^{a_k\mathbf{u}_k^+}\frac{\prod_{i=1}^{k-1}T^{a_i\mathbf{u}_i^+}}{\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}}-T^{a_k\mathbf{u}_k^-}\frac{\prod_{i=1}^{k-1}T^{a_i\mathbf{u}_i^+}}{\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}}+\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\\&=T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(\frac{\prod_{i=1}^{k-1}T^{a_i\mathbf{u}_i^+}}{\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}}\left(T^{a_k\mathbf{u}_k^+}-T^{a_k\mathbf{u}_k^-}\right)+\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\\&=T^{\mathbf{x}_2}T^{\mathbf{m}_{k-1}}\left(\prod_{i=1}^{k-1}T^{a_i\mathbf{u}_i^+}\right)\left(T^{a_k\mathbf{u}_k^+}-T^{a_k\mathbf{u}_k^-}\right)\\&\quad+\left(\prod_{i=1}^{k}T^{a_i\mathbf{u}_i^-}\right)T^{\mathbf{x}_2}T^{\mathbf{m}_{k-1}}\left(\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right).\end{aligned}$$

Because both $T^{a_k\mathbf{u}_k^+}-T^{a_k\mathbf{u}_k^-}\in I_{\mathcal{B}}$ by Lemma 6.4.3 and $T^{\mathbf{m}_{k-1}}\left(\prod_{i=1}^{k-1}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\in I_{\mathcal{B}}$, we have

$$T^{\mathbf{m}_k}T^{\mathbf{x}_2}\left(\prod_{i=1}^{k}\frac{T^{a_i\mathbf{u}_i^+}}{T^{a_i\mathbf{u}_i^-}}-1\right)\in I_{\mathcal{B}}$$

as required.

We have shown that for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_{\mathbf{y}}$, there is a monomial $T^{\mathbf{m}}$ such that

$$T^{\mathbf{m}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) \in I_{\mathcal{B}}.$$

We now need only to show that if $I_{\mathcal{B}}$ is radical, then this implies that

$$T^{\mathbf{1}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) \in I_{\mathcal{B}}.$$

Let $n = \max_{m \in \mathbf{m}} m$, and let $\mathbf{n}$ be the vector whose entries are all n. Then

$$\begin{aligned} T^{\mathbf{n}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) &\in I_{\mathcal{B}} \\ T^{\mathbf{n}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2})^n &\in I_{\mathcal{B}} \\ (T^{\mathbf{1}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}))^n &\in I_{\mathcal{B}}. \end{aligned}$$

But $I_{\mathcal{B}}$ is radical, so $T^{\mathbf{1}}(T^{\mathbf{x}_1} - T^{\mathbf{x}_2}) \in I_{\mathcal{B}}$ as required. □

We can apply this result to prove the Minus One Conjecture in cases where the column partition lattice basis generates a radical ideal.

Theorem 6.4.4 (The Minus One Theorem for column partition lattice bases that generate a radical ideal)**.** *Let $A \in \{0,1\}^{r \times n}$ be a unimodular configuration matrix and let U be a column partition lattice basis for A. Suppose that the ideal I_U is radical. Then for all $\mathbf{y} \in \mathbb{Z}^n_{\geq 0}$, the Markov chain using the moves in U that allows entries to drop down to -1 in the first n co-ordinates connects a set that includes the set $\mathcal{F}_{\mathbf{y}}$.*

Proof. Following the proof of Proposition 6.4.2, we can see that each move $\mathbf{u}_i$ is required exactly a_i times, and no other moves are required. This means that there is a path between $\mathbf{x}_1$ and $\mathbf{x}_2$ that requires no detours, or moves that are later undone.

By Proposition 6.4.2, none of the entries in any co-ordinate drops below -1, so it remains to check that none of the co-ordinates $n+1, \ldots, r$ drops below 0. In a path that uses a column partition lattice basis, for $i = n+1, \ldots, r$ the ith co-ordinate is affected only by the basis element $\mathbf{u}_{n-i}$. When following such a path, the entries in the ith co-ordinate make up either a non-increasing or non-decreasing sequence from $\mathbf{x}_1^i$ to $\mathbf{x}_2^i$ using steps of size $\mathbf{u}_{n-i}^i$. Because $\mathbf{x}_1^i, \mathbf{x}_2^i \geq 0$, the ith entry never goes below zero, as required. □

6.4.2 Network matrices

One of our key motivating applications in studying Markov bases comes from network tomography. Link-path incidence matrices, the configuration matrices of interest in network tomography, were the subject of Chapter 5. Section 5.3 covered traffic networks on a type of graph called a polytree. The link-path incidence matrices of polytrees are non-negative network matrices. Theorem 5.3.5 shows that they are totally unimodular. In this section we prove the Minus One Conjecture for the particular case of network matrices (Definition 5.3.2), and therefore for the link-path incidence matrices of polytrees.

Schrijver [40, Section 20.1] notes the existence of polynomial time algorithms for recognising network matrices, designed by Auslander and Trent [4, 5], Gould [24], Tutte [47, 46, 48], and Bixby and Cunningham [7]. As well as being useful for analysis of traffic on polytrees, this theorem also shows that the Minus One Conjecture holds for any configuration matrix or U matrix that these algorithms recognise as a network matrix.

Theorem 6.4.5. *Let A be a network matrix, and let U be a column partition lattice basis for A. Then U is an $\mathbf{m}_{(1,0)}$-Markov basis, where $\mathbf{m}_{(1,0)}$ is as defined in Definition 6.3.1.*

Proof. Let $\mathcal{P}_A$ be a polytree with link-path incidence matrix A. Then A is also a network matrix represented by the polytree $\mathcal{P}_A$, and G, the graph whose edges show the origin/destination pairs of the allowed paths in the network. By Theorem 5.3.5, A is totally unimodular. Let U be a column partition lattice basis for A under the column partition π. By Theorem 5.3.9, $A = A_1^{-1}A = \begin{bmatrix} I & C \end{bmatrix}$ is a network matrix represented by another polytree $\mathcal{P}_{A_1^{-1}A}$ and G.

Let $\mathbf{g} \in \mathcal{G}_A$ be given — the unimodularity of A means $\mathbf{g} \in \{0, \pm 1\}^r$. This $\mathbf{g}$ takes a subset of columns of $A_1^{-1}A$ and assigns them a multiplier of ± 1 such that the sum is $\mathbf{0}$.

Construct the matrix $(A_1^{-1}A)_{\mathbf{g}}$ by taking this subset of signed columns of $A_1^{-1}A$. By Theorem 6.3.7, we must show that the k columns of $(A_1^{-1}A)_{\mathbf{g}}$ can be ordered

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}$$

so that the sequence of partial sums $\sum_{j=1}^{m} \mathbf{a}_{i_j}$ satisfies

$$\sum_{j=1}^{m} \mathbf{a}_{i_j} \geq -\mathbf{1}$$

for $m = 1, \ldots, k$, where $\mathbf{1}$ is the vector with all entries equal to 1.

By Corollary 5.3.11, we can order the columns of $(A_1^{-1}A)_{\mathbf{g}}$ so that the corresponding paths on $\mathcal{P}_{A_1^{-1}A}$ are joined nose to tail. If we follow the journey they define, then we must traverse any particular edge in alternate directions each time we encounter it. In any row of $(A_1^{-1}A)_{\mathbf{g}}$, the non-zero entries record in which paths the corresponding edge is traversed, and the sign gives the direction. Therefore in every row the non-zero entries alternate in sign.

Such an ordering of the columns of $(A_1^{-1}A)_{\mathbf{g}}$ gives an ordering of the columns of U such that in the sequence of partial sums, all entries alternate between either 0 and 1 or 0 and -1. □

We illustrate this proof with the following example.

Example 6.4.6. Consider the four-link linear network and column partition lattice basis from Example 6.3.9. Then the configuration matrix A is given by

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

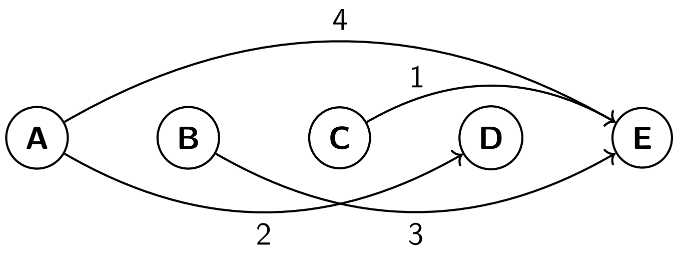


Figure 6.5: The polytree $\mathcal{P}_{A_1^{-1}A}$ from Example 6.4.6.

and the matrix $A_1^{-1}A$ is given by

$$A_1^{-1}A = \begin{bmatrix} 0 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The matrix $A_1^{-1}A$ is a network matrix represented by the polytree $\mathcal{P}_{A_1^{-1}A}$ and the graph G, shown in Figure 6.5.

Suppose we wish to simulate the move $\mathbf{g} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}^\intercal \in \mathcal{G}_A$. Then the matrix $(A_1^{-1}A)_\mathbf{g}$ is given by

$$(A_1^{-1}A)_\mathbf{g} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & -1 & 1 & 1 \end{bmatrix}.$$

The partial sums of these columns in the order given includes a -2 in the fourth coordinate.

The columns of $(A_1^{-1}A)_\mathbf{g}$ refer to a closed walk on $\mathcal{P}_{A_1^{-1}A}$. Respectively, they refer to the paths:

- $(B) \to (A)$
- $(E) \to (D)$
- $(D) \to (B)$
- $(A) \to (E)$

These paths can be placed nose to tail in the following order:

$$(A) \to (E) \to (D) \to (B) \to (A),$$

which requires use of the edges in $\mathcal{P}_{A_1^{-1}A}$ in the order $(4, -4, 2, -2, 4, -3, 3, -4)$, where a negative sign indicates the edge is traversed in the opposite direction to its orientation.

Each edge in $\mathcal{P}_{A_1^{-1}A}$ is traversed in alternating directions each time it is encountered. Ordering the columns of $(A_1^{-1}A)_{\mathbf{g}}$ accordingly produces the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & -1 & 1 & -1 \end{bmatrix},$$

where in each row the non-zero entries alternate in sign. In each row, the sequence of partial sums therefore alternates between 0 and 1, or 0 and -1, and therefore never drops below -1.

6.4.3 A weaker lower bound

In this section we give a lower bound on the co-ordinates of points we are required to visit in a walk connecting any two points in any given fibre using a column partition lattice basis for a unimodular configuration matrix. The lower bound comes from the upper limit on the number of column partition lattice basis steps required to simulate any Graver basis element given in Theorem 3.5.15.

Theorem 6.4.7. *Let A be a unimodular configuration matrix and let U be a column partition lattice basis for* $\ker_{\mathbb{Z}}(A)$*. Then U is an $\mathbf{m}_0$-Markov basis, where $\mathbf{m}_0$ is the vector made up of m in the first n entries and 0 in the other entries, where*

$$m = \min\left(\left\lfloor \frac{r-n}{2} \right\rfloor, \left\lfloor \frac{n+1}{2} \right\rfloor\right).$$

Proof. The matrix U is a column partition lattice basis, so in co-ordinates $n+1, \ldots, r$, the moves simulating a direct walk (that is, one with no moves that are later undone) all have the same sign. In these co-ordinates the walk travels from $\mathbf{x}_1$ to $\mathbf{x}_2$ in a non-decreasing or non-increasing sequence, so they never go below 0. We are therefore only concerned with co-ordinates $1, \ldots, n$. By Theorem 6.2.3, we only need to worry about simulating elements of a known Markov basis. We choose the Graver basis, $\mathcal{G}_A$.

The matrix A is unimodular, so by Theorem 3.5.9 any $\mathbf{g} \in \mathcal{G}_A$ is a circuit, and it has all entries in $\{0, \pm 1\}$. Therefore each step in a walk using a column partition lattice basis alters each co-ordinate by at most ± 1. By Theorem 3.5.15, there are at most $k = \min(r-n, n+1)$ steps in any direct walk that simulates a Graver basis element.

Suppose to the contrary that there is a Graver basis element $\mathbf{g}$ such that a direct walk W simulating $\mathbf{g}$ visits a point $\mathbf{x}$ with ith co-ordinate less than $-m$. Since $\mathbf{g}_i^-, \mathbf{g}_i^+ \geq 0$; $\mathbf{x}_i \leq -m-1$; and the ith co-ordinate can change by at most 1 at each step, there must be at least $m+1$ steps from $\mathbf{g}^-$ to $\mathbf{x}$ and then at least $m+1$ steps from $\mathbf{x}$ to $\mathbf{g}^+$. But then W is at least $2m+2 > k$ steps long, contradicting Theorem 3.5.15. □

Example 6.4.8. Consider the four-link linear network which has unimodular configuration matrix

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

and a column partition lattice basis

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

For this configuration matrix we have $n = 4$ and $r = 10$, so by Theorem 6.4.7 a direct walk between any pair of points can visit at worst a point with some co-ordinate $-m$, where

$$\begin{aligned} m &= \min\left(\left\lfloor \frac{r-n}{2} \right\rfloor, \left\lfloor \frac{n+1}{2} \right\rfloor\right) \\ &= \min\left(\left\lfloor \frac{6}{2} \right\rfloor, \left\lfloor \frac{5}{2} \right\rfloor\right) \\ &= \min(3, 2) \\ &= 2. \end{aligned}$$

We may wish to simulate $\mathbf{g} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & 0 \end{bmatrix}^{\intercal} \in \mathcal{G}_A$. Then the points $\mathbf{g}^-$ and $\mathbf{g}^+$, and the moves in U required to connect them, are given in Table 6.2.

$-\mathbf{g}$	$\mathbf{g}^-$	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$-\mathbf{u}_4$	$-\mathbf{u}_5$	$-\mathbf{g}^+$
0	0	−1	0	0	1	0	0
0	0	−1	−1	0	1	1	0
0	0	0	−1	−1	1	1	0
0	0	0	0	−1	0	1	0
−1	0	1	0	0	0	0	−1
−1	0	0	1	0	0	0	−1
−1	0	0	0	1	0	0	−1
1	1	0	0	0	−1	0	0
1	1	0	0	0	0	−1	0
0	0	0	0	0	0	0	0

Table 6.2: The moves required to simulate $\mathbf{g}$ in Example 6.4.8.

With this ordering of moves the walk visits points with -2 in the x_2 and x_3 co-ordinates. It is not possible to reorder the columns such that the walk visits a point with -3 in some co-ordinate.

Of course, a walk between $\mathbf{g}^-$ and $\mathbf{g}^+$ need not step outside the fibre at all, since U is a Markov basis by Theorem 4.2.1.

6.4.4 Proofs for small matrices

In this section we give a proof that for any unimodular configuration matrix A, for column partition lattice basis U, if some Graver basis element $\mathbf{g}$ has $\text{supp}(\mathbf{g}) \leq 6$, then it can be simulated using U with a walk that only visits points that are at least $\mathbf{m}_{(1,0)}$. That is, it can be simulated under the conditions of the Minus One Conjecture. For unimodular matrices, any Graver basis element is also a column partition lattice basis element and so by Theorem 3.5.10 it has support of size at most $n+1$, where n is the number of rows of A. Therefore, the Minus One Conjecture holds for all unimodular configuration matrices with five or fewer rows.

In Example 6.3.5 we gave an example of a column partition lattice basis $U \in \{0, \pm 1\}^{r \times (r-n)}$ that is not an $\mathbf{m}_{(1,0)}$-Markov basis. The first six rows of the matrix U were made up of the submatrix

$$R = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

The move $\mathbf{z} = \sum_{i=1}^{4} \mathbf{u}_i$ is a counterexample to U being a $\mathbf{m}_{(1,0)}$-Markov basis. This is because the moves required to simulate $\mathbf{z}$ with U are $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ and $\mathbf{u}_4$, and any ordering of these moves means that in one of the co-ordinates in R, the -1 lower bound is broken. The sum of the entries in each row of R is equal to 0, so in co-ordinates $i = 1, \ldots, 6$, we have $z_i^- = z_i^+ = 0$. For any permutation π of the columns of U, there is a row $\mathbf{r}$ in R such that $\pi(\mathbf{r}) = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$, and the walk corresponding to the ordering π breaks the -1 lower bound in that co-ordinate.

Every row in R is required to be in U for the walk simulating $\mathbf{z}$ to necessarily break the -1 lower bound. If some row $\mathbf{r}$ in R were not in U, then the column permutation π for which $\pi(\mathbf{r}) = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$ would no longer break the -1 lower bound. Moreover, when using a column partition lattice basis $U \in \{0, \pm 1\}^{r \times 4}$ to simulate a move $\mathbf{z}$ that requires all four moves from U, $\begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$ is the only potential row of a reordered U that breaks the -1 lower bound.

Observation 6.4.9. If a column partition lattice basis $U \in \{0, \pm 1\}^{r \times 4}$ is not an $\mathbf{m}_{(1,0)}$-Markov basis, it must have R as a submatrix.

The matrix R is not totally unimodular, and so neither is U. For any row $\mathbf{r}$ in R, there is only one other row that when paired with $\mathbf{r}$ does *not* yield a non-totally unimodular submatrix. For any given U, we need only check, for example, that for the column permutations $\pi_1 = 1234$ and $\pi_2 = 1324$ there exist rows $\mathbf{r}_1$ and $\mathbf{r}_2$ of U such that $\pi_1(\mathbf{r}_1) = \pi_2(\mathbf{r}_2) = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$ before we find a non-totally unimodular matrix.

In this section we aim to generalise this approach to show that column partition lattice bases meeting certain conditions are $\mathbf{m}_{(1,0)}$-Markov basis. We demonstrate that for a unimodular configuration matrix A and column partition lattice basis U, we can simulate any Graver basis element $\mathbf{g} \in \mathcal{G}_A$ with a walk using U that never visits points less than $\mathbf{m}_{(1,0)}$ if $\mathbf{g}$ has sufficiently small support. We use k to mean the largest possible size of the support of $\mathbf{g}$, so it is equal to the number of columns of $(A_1^{-1}A)_{\mathbf{g}}$. We demonstrate this by contradiction: we assume that some k column matrix $(A_1^{-1}A)_{\mathbf{g}}$ breaks the $\mathbf{m}_{(1,0)}$ lower limit for every ordering of its columns, and show by enumeration of cases that $(A_1^{-1}A)_{\mathbf{g}}$ cannot be totally unimodular.

The Graver basis is known to be a Markov basis. By Theorem 6.2.3, if we can show that U is capable of simulating all of the moves in the Graver basis with a walk that only visits points that are at least $-\mathbf{m}_{(1,0)}$, then U is an $\mathbf{m}_{(1,0)}$-Markov basis. The matrix A is unimodular, so by Theorem 3.5.10, $\mathbf{g}$ is also an element of some column partition lattice basis of A and therefore $k \leq n+1$. A proof that all Graver basis elements $\mathbf{g}$ with at most k non-zero entries can be simulated in this way is therefore a proof of the Minus One Conjecture for configuration matrices where the number of rows n satisfies $n \leq k-1$.

By Corollary 6.3.8, we can simulate $\mathbf{g}$ if we can place the k columns of $(A_1^{-1}A)_{\mathbf{g}}$ in an order

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}$$

such that for $m = 1, \ldots, k$ the partial sums $\sum_{j=1}^{m} \mathbf{a}_{i_j}$ satisfy the inequality

$$\sum_{j=1}^{m} \mathbf{a}_{i_j} \geq -\mathbf{1}. \tag{6.4.1}$$

For a given k, we will take an arbitrary $(A_1^{-1}A)_{\mathbf{g}}$ matrix and assume that it cannot have its columns ordered in such a way that inequality 6.4.1 is satisfied. If this is the case, then for every ordering of the columns there must be at least one row in $(A_1^{-1}A)_{\mathbf{g}}$ that breaks the inequality. For each k, there is a finite number of rows that do this; for every ordering, at least one of these must appear. We can enumerate the potential rows and collect them into the matrix R_k. For example, if $k = 4$, then to break the inequality every ordering of the columns of $(A_1^{-1}A)_{\mathbf{g}}$ must contain the row

$$R_4 = \begin{bmatrix} -1 & -1 & 1 & 1 \end{bmatrix},$$

and if $k = 5$, every ordering of columns of $(A_1^{-1}A)_{\mathbf{g}}$ must contain at least one row of the matrix

$$R_5 = \begin{bmatrix} -1 & -1 & 1 & 1 & 0 \\ -1 & -1 & 1 & 0 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & 0 & -1 & 1 & 1 \\ 0 & -1 & -1 & 1 & 1 \end{bmatrix}.$$

We will show that for $k = 4, 5$, and 6, for every combination of such rows, there must be a submatrix with a determinant that is not 0 or ± 1, and so $(A_1^{-1}A)_{\mathbf{g}}$ cannot be totally unimodular and therefore A cannot be unimodular.

Algorithm 1: Proofs for small matrices algorithm

Data: Π_k, R_k, M_k.
$\mathcal{S} \leftarrow \{M_k\}$;
for $\pi \in \Pi_k$ **do**
  $\mathcal{S}' \leftarrow \varnothing$;
  for $M \in \mathcal{S}$ **do**
    if $\mathrm{rows}(\pi(M)) \cap \mathrm{rows}(R_k) \neq \varnothing$ **then**
      $\mathcal{S}' \leftarrow \mathcal{S}' \cup \{M\}$;
    else
      for $\mathbf{r} \in \mathrm{rows}(R_k)$ **do**
        if $\begin{bmatrix} \pi(M) \\ \mathbf{r} \end{bmatrix}$ *is totally unimodular* **then**
          $\mathcal{S}' \leftarrow \mathcal{S}' \cup \left\{ \begin{bmatrix} M \\ \pi^{-1}(\mathbf{r}) \end{bmatrix} \right\}$;
        end
      end
    end
  end
  $\mathcal{S} \leftarrow \mathcal{S}'$;
end
return $\mathcal{S}$;

Algorithm 1 performs this check. The algorithm is initialised with the data Π_k, R_k, and M_k. Here, Π_k is the symmetric group on k elements; it acts on matrices with k

columns by permuting the columns. The matrix R_k comprises all of the potential rows $\mathbf{r} \in \{0, \pm 1\}^k$ that violate inequality 6.4.1; and M_k is the empty matrix with k columns to which we will append potential rows. This matrix M_k is stored in the collection $\mathcal{S}$.

The algorithm then iterates through every column permutation of a k column matrix, forming from the elements of $\mathcal{S}$ a new generation of matrices $\mathcal{S}'$. For the column permutation π, the new generation $\mathcal{S}'$ is formed by taking each matrix M currently in $\mathcal{S}$ and ensuring that $\pi(M)$ violates inequality 6.4.1. This is achieved by first checking if it already violates inequality 6.4.1, and carrying it forward to the next generation $\mathcal{S}'$ if it does; or by forming new matrices by appending each row in R_k in turn. The inverse permutations of matrices generated in this way that are totally unimodular are included in $\mathcal{S}'$, and the iteration continues with the next column permutation.

When the iteration through column permutations is complete, the collection of surviving matrices $\mathcal{S}$ is returned. Any k column $(A_1^{-1}A)_{\mathbf{g}}$ matrix that violates inequality 6.4.1 must contain some matrix in $\mathcal{S}$ as a submatrix. If, however, the returned collection $\mathcal{S}$ is empty, then the conditions of the Minus One Conjecture are met by all potential k column $(A_1^{-1}A)_{\mathbf{g}}$ matrices.

Algorithm 1 can potentially generate matrices $(A_1^{-1}A)_{\mathbf{g}}$ with all-zero columns. Practically, if the returned collection $\mathcal{S}$ includes such an $(A_1^{-1}A)_{\mathbf{g}}$, then the c all-zero columns can be dropped, leaving a c column matrix that is totally unimodular that violates inequality 6.4.1 for every ordering of its columns. This means that if Algorithm 1 terminates for some k with an empty $\mathcal{S}$, then for any unimodular configuration matrix A and any column partition lattice basis U, any Graver basis element $\mathbf{g} \in \mathcal{G}_A$ with support size $|\text{supp}(\mathbf{g})| \leq k$ can be simulated with a walk using U under the conditions of the Minus One Conjecture.

Theorem 3.5.10 says that for a unimodular configuration matrix, any Graver basis element $\mathbf{g}$ is also an element of some column partition lattice basis. If the configuration matrix has n rows, then $\mathbf{g}$ has support of size at most $n + 1$. Therefore, if Algorithm 1 terminates for some k with an empty $\mathcal{S}$, then the Minus One Conjecture holds for configuration matrices with $n \leq k - 1$ rows.

Four column matrices

Theorem 6.4.10. *Suppose A is a configuration matrix, U is a column partition lattice basis, and $\mathbf{g}$ is an element of the Graver basis of A such that $|\text{supp}(g)| \leq 4$. Then $\mathbf{g}$ can be simulated by a walk using U that satisfies inequality 6.4.1.*

Proof. We run Algorithm 1 with $k = 4$. The initial data is given by

$$\begin{aligned} \Pi_4 &= (1234, 1243, 1324, \ldots, 4321)\,, \\ R_4 &= \begin{bmatrix} -1 & -1 & 1 & 1 \end{bmatrix}, \\ M_4 &= \begin{bmatrix} & & & \end{bmatrix}, \quad \text{an empty matrix with four columns.} \end{aligned}$$

The permutations in Π_4 are ordered lexicographically. We initialise $\mathcal{S} = \{M_4\}$.

The first column permutation is the identity permutation, $\pi_1 = 1234$. There is only one element in $\mathcal{S}$, which is the empty matrix M_4. The matrix $\pi_1(M)$ has no rows in common with R_4, so we append the only row of R_4 to $\pi_1(M_4)$ and apply the permutation π_1^{-1} to form $\begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$. This matrix is totally unimodular, so it is added to $\mathcal{S}'$. We have completed iteration through the elements of $\mathcal{S}$ for this permutation, so we set $\mathcal{S}$ to the new generation $\mathcal{S}'$ with $\mathcal{S} \leftarrow \mathcal{S}'$, and move on to the next column permutation.

The second column permutation is $\pi_2 = 1243$. We set $\mathcal{S}' \leftarrow \varnothing$. There is one element in $\mathcal{S}$, which is $M = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$. The matrix $\pi_2(M)$ contains only the row $\begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$, which violates inequality 6.4.1. This M is added to $\mathcal{S}'$, and iteration through $\mathcal{S}$ is complete, and $\mathcal{S}'$ is moved to $\mathcal{S}$.

The next column permutation is 1324. Again, the collection $\mathcal{S}$ contains one element, $M = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$. We form $\pi_3(M) = \begin{bmatrix}-1 & 1 & -1 & 1\end{bmatrix}$ and find $\text{rows}(\pi_3(M)) \cap \text{rows}(R_4) = \varnothing$. We append $\mathbf{r} = \begin{bmatrix}-1 & -1 & 1 & 1\end{bmatrix}$ to form the matrix

$$\begin{bmatrix}\pi(M)\\ \mathbf{r}\end{bmatrix} = \begin{bmatrix}-1 & 1 & -1 & 1\\ -1 & -1 & 1 & 1\end{bmatrix}.$$

The highlighted submatrix has determinant 2, so this matrix is not totally unimodular and is discarded. This concludes iteration through $\mathcal{S}$ for π_3 and we are left with $\mathcal{S}' = \varnothing$.

Because $\mathcal{S}$ is now empty, no other permutations have any effect on $\mathcal{S}$ and the algorithm terminates with $\mathcal{S} = \varnothing$. This proves the case for $k = 4$. □

Five column matrices

In Section 6.4.4 we ran Algorithm 1 with $k = 4$. The second iteration through column permutations Π_4 used the permutation $\pi_2 = 1243$. This iteration left the collection $\mathcal{S}$ unchanged. Algorithm 1 with $k = 4$ would have terminated more quickly had we selected $\pi_3 = 1324$ for the second iteration. We may be able demonstrate that the algorithm terminates with $\mathcal{S} = \varnothing$ more efficiently through judicious choice of column permutation ordering.

Theorem 6.4.11. *Suppose A is a configuration matrix, U is a column partition lattice basis, and $\mathbf{g}$ is an element of the Graver basis of A such that $|\text{supp}(g)| \leq 5$. Then $\mathbf{g}$ can be simulated by a walk using U that satisfies inequality 6.4.1.*

Proof. We will prove that for $k = 5$, Algorithm 1 terminates with $\mathcal{S} = \varnothing$. The potential rows of M that violate inequality 6.4.1 are given by

$$R_5 = \begin{bmatrix}-1 & -1 & 1 & 1 & 0\\ -1 & -1 & 1 & 0 & 1\\ -1 & -1 & 0 & 1 & 1\\ -1 & 0 & -1 & 1 & 1\\ 0 & -1 & -1 & 1 & 1\end{bmatrix}.$$

Each row of R_5 contains the same elements in a different order, so any one can be permuted to form any of the others, so any M must contain a row that can be permuted to form

$\begin{bmatrix}-1 & -1 & 0 & 1 & 1\end{bmatrix}$. Without loss of generality, we initialise M_5 to be this row, rather than the empty matrix. The collection $\mathcal{S}$ is initialised as $\left\{\begin{bmatrix}-1 & -1 & 0 & 1 & 1\end{bmatrix}\right\}$.

We begin iterating over Π_5. We choose $\pi_1 = 14325$ and form

$$\pi_2(M) = \begin{bmatrix}-1 & 1 & 0 & -1 & 1\end{bmatrix}.$$

Appending each row of R_5 in turn produces the matrices

$$\begin{bmatrix}-1 & 1 & 0 & -1 & 1\\ -1 & -1 & 1 & 1 & 0\end{bmatrix}, \begin{bmatrix}-1 & 1 & 0 & -1 & 1\\ -1 & -1 & 1 & 0 & 1\end{bmatrix}, \begin{bmatrix}-1 & 1 & 0 & -1 & 1\\ -1 & -1 & 0 & 1 & 1\end{bmatrix},$$
$$\begin{bmatrix}-1 & 1 & 0 & -1 & 1\\ -1 & 0 & -1 & 1 & 1\end{bmatrix}, \text{ and } \begin{bmatrix}-1 & 1 & 0 & -1 & 1\\ 0 & -1 & -1 & 1 & 1\end{bmatrix}.$$

The highlighted submatrices each have determinant ± 2, so none of the generated matrices is totally unimodular, and so Algorithm 1 terminates with $\mathcal{S} = \varnothing$. This proves the case for $k = 5$. □

Six column matrices

Theorem 6.4.12. *Suppose A is a configuration matrix, U is a column partition lattice basis, and $\mathbf{g}$ is an element of the Graver basis of A such that $|\mathrm{supp}(g)| \leq 6$. Then $\mathbf{g}$ can be simulated by a walk using U that satisfies inequality 6.4.1.*

Proof. We will prove that Algorithm 1 with $k = 6$ terminates with $\mathcal{S} = \varnothing$. The potential rows of M that violate inequality 6.4.1 are given by

$$R_6 = \begin{bmatrix}
-1 & -1 & 1 & 1 & 0 & 0\\
-1 & -1 & 1 & 0 & 1 & 0\\
-1 & -1 & 1 & 0 & 0 & 1\\
-1 & -1 & 0 & 1 & 1 & 0\\
-1 & -1 & 0 & 1 & 0 & 1\\
-1 & -1 & 0 & 0 & 1 & 1\\
-1 & 0 & -1 & 1 & 1 & 0\\
-1 & 0 & -1 & 1 & 0 & 1\\
-1 & 0 & -1 & 0 & 1 & 1\\
-1 & 0 & 0 & -1 & 1 & 1\\
0 & -1 & -1 & 1 & 1 & 0\\
0 & -1 & -1 & 1 & 0 & 1\\
0 & -1 & -1 & 0 & 1 & 1\\
0 & -1 & 0 & -1 & 1 & 1\\
0 & 0 & -1 & -1 & 1 & 1\\
1 & -1 & -1 & -1 & 1 & 1\\
-1 & 1 & -1 & -1 & 1 & 1\\
-1 & -1 & 1 & 1 & 1 & -1\\
-1 & -1 & 1 & 1 & -1 & 1\\
-1 & -1 & 1 & -1 & 1 & 1\\
-1 & -1 & -1 & 1 & 1 & 1
\end{bmatrix}.$$

We initialise

$$\mathcal{S} = \left\{ \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix} \right\}$$

for similar reasons to those given in the proof of Theorem 6.4.11.

We specify that the initial row chosen contains the minimum number of zeroes over all rows of $(A_1^{-1}A)_{\mathbf{g}}$. Then we run the algorithm in two parts: once with $\mathcal{S} = \{\begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}\}$ and R_6 as above; and once with $\mathcal{S} = \{\begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}\}$, and use

$$R_6' = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}.$$

We first address the case where $\mathcal{S} = \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}$. We choose as the first permutation $\pi = 153426$, and form $\pi(M) = \begin{bmatrix} -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix}$. Appending each row of R_6' to $\pi(M)$ in turn forms only non-totally unimodular matrices. To see this, note that $\pi(M)$ begins and ends with $\begin{bmatrix} -1 & 1 \end{bmatrix}$, while each row of R_6' either begins with $\begin{bmatrix} -1 & -1 \end{bmatrix}$ or ends with $\begin{bmatrix} 1 & 1 \end{bmatrix}$. These combine to form a non-totally unimodular submatrix. For example, appending the first row of R_6' produces

$$\begin{bmatrix} -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}.$$

The highlighted submatrix has determinant 2, so this matrix is not totally unimodular.

We now turn to $\mathcal{S} = \{\begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}\}$. We note that almost every row of R_6 either begins or ends with $\begin{bmatrix} \pm 1 & \pm 1 \end{bmatrix}$. Suppose that there is a column permutation π such that for all $M \in \mathcal{S}$, $\pi(M)$ begins and ends with $\begin{bmatrix} \pm 1 & \mp 1 \end{bmatrix}$. Then appending a row of R_6 that either begins or ends with $\begin{bmatrix} \pm 1 & \pm 1 \end{bmatrix}$ or ends with $\begin{bmatrix} 1 & 1 \end{bmatrix}$ produces a non-totally unimodular matrix. Under such a column permutation we need only concern ourselves with rows of R_6 that do not begin or end with $\begin{bmatrix} \pm 1 & \pm 1 \end{bmatrix}$, which are

$$\begin{bmatrix} -1 & 0 & -1 & 1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 & 1 \\ 0 & -1 & -1 & 1 & 1 & 0 \\ 0 & -1 & -1 & 1 & 0 & 1 \end{bmatrix}.$$

We choose the permutation $\pi_1 = 153426$ and form $\pi_1(M) = \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix}$.

Appending each of the four rows of R_6' in turn produces the following matrices:

$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & -1 & 1 & 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & -1 & 1 & 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & -1 & 1 & 0 & 1 \end{bmatrix}.$$

Submatrices with determinant ± 2 are shown in two of these matrices. By Theorem 3.5.4, these matrices are not totally unimodular. We set $\mathcal{S}'$ to contain the other two matrices with their original column orderings:

$$\mathcal{S} = \left\{ \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 & -1 & 0 \end{bmatrix} \right\}.$$

Then $\mathcal{S}'$ is stored in $\mathcal{S}$.

Next we choose permutation $\pi_2 = 154326$. Again, this permutation of each matrix in $\mathcal{S}$ has $\begin{bmatrix} -1 & 1 \end{bmatrix}$ at the beginning and end of each row, so appending any row of R_6 that begins or ends with $\pm \begin{bmatrix} 1 & 1 \end{bmatrix}$ will result in a matrix that is not totally unimodular; consequently, these rows can be ignored. Appending each of the other rows of R_6' to each of the permuted matrices in $\mathcal{S}$ produces the matrices:

$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \\ -1 & 0 & -1 & 1 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 & 1 & 0 \end{bmatrix},$$
$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix},$$
$$\begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & -1 & -1 & 1 & 1 & 0 \end{bmatrix}, \text{ and } \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & -1 & -1 & 1 & 0 & 1 \end{bmatrix}.$$

In each case there is a highlighted submatrix with determinant ± 2, so none of these matrices is totally unimodular and the algorithm terminates with $\mathcal{S} = \varnothing$. This proves the case $k = 6$. □

We have shown for a unimodular configuration matrix A, column partition lattice basis U, and Graver basis element $\mathbf{g}$ that if $|\text{supp}(\mathbf{g})| \leq 6$, then $\mathbf{g}$ can be simulated with U with a walk that only visits points that are at least $\mathbf{m}_{(1,0)}$. If A has at most five rows, then $|\text{supp}(\mathbf{g})| \leq 6$ is guaranteed. Therefore, any column partition lattice basis of a unimodular configuration matrix with five or fewer rows is an $\mathbf{m}_{(1,0)}$-Markov basis.

6.5 A possible way forward

In Section 6.4.2 we proved the Minus One Conjecture for network matrices. Although as we showed in Section 5.3 network matrices do not account for all totally unimodular matrices, they do act as important building blocks. This suggests another possible method by which the full Minus One Conjecture might be proved.

Hoffman [29] and Bixby [6] gave the two matrices

$$F_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 \\ -1 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ -1 & 0 & 0 & -1 & 1 \end{bmatrix}$$

and

$$F_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

as examples of totally unimodular matrices such that neither they nor their transposes are network matrices. Seymour [41] showed that these two matrices together with network matrices can be combined using a certain collection of operations and compositions to produce all totally unimodular matrices. These are listed in Schrijver [40]. The operations are:

1. permuting rows or columns;
2. taking the transpose;
3. multiplying a row or column by -1;
4. pivoting, i.e. replacing

$$\begin{bmatrix} \epsilon & \mathbf{c}^\intercal \\ \mathbf{b} & D \end{bmatrix} \quad \text{by} \quad \begin{bmatrix} -\epsilon & \epsilon\mathbf{c}^\intercal \\ \epsilon\mathbf{b} & D - \epsilon\mathbf{b}\mathbf{c}^\intercal \end{bmatrix},$$

   where $\epsilon = \pm 1$, $\mathbf{b}$ is a column vector, $\mathbf{c}^\intercal$ is a row vector, and D is a matrix;
5. appending an all-zero row or column; or appending a row or column with one non-zero, being ± 1;
6. repeating a row or column.

The compositions are:

1. 1-sum:
$$A \oplus_1 B := \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

2. 2-sum:
$$\begin{bmatrix} A & \mathbf{a} \end{bmatrix} \oplus_2 \begin{bmatrix} \mathbf{b}^\intercal \\ B \end{bmatrix} := \begin{bmatrix} A & \mathbf{a}\mathbf{b}^\intercal \\ 0 & B \end{bmatrix}$$

3. 3-sum:
$$\begin{bmatrix} A & \mathbf{a} & \mathbf{a} \\ \mathbf{c}^\intercal & 0 & 1 \end{bmatrix} \oplus_3 \begin{bmatrix} 1 & 0 & \mathbf{b}^\intercal \\ \mathbf{d} & \mathbf{d} & B \end{bmatrix} := \begin{bmatrix} A & \mathbf{a}\mathbf{b}^\intercal \\ \mathbf{d}\mathbf{c}^\intercal & B \end{bmatrix}$$

where A and B are matrices, $\mathbf{a}$ and $\mathbf{d}$ are column vectors, and $\mathbf{b}^\intercal$ and $\mathbf{c}^\intercal$ are row vectors. These compositions are only applied if for both A and B, the number of rows plus the number of columns is at least four.

The matrices F_1 and F_2 have at most one row that contains more than one non-zero entry, so it seems likely that they would meet the conditions of the Minus One Conjecture. The Minus One Theorem for network matrices (Theorem 6.4.5) may make up another important building block towards proof of the full Minus One Conjecture based on Seymour's work. We leave this as a possible future research direction.

Chapter 7

Discussion

We conclude the thesis by summarising the main contributions.

7.1 $\mathbb{Z}$-polytope samplers

In this thesis we looked at random walk fibre sampling for statistical inverse problems using Markov bases. The way this problem is usually approached is algebraic and based on the Fundamental Theorem of Markov Bases. This involves a generating set for an ideal in a polynomial ring. The kernel of the configuration matrix corresponds to the ideal, and Markov bases correspond to generating sets of this ideal. Finding Markov bases with this approach involves finding such a generating set, such as a Gröbner basis. This is computationally expensive, often prohibitively so.

In contrast, our aim was to determine whether representing the fibre geometrically as a $\mathbb{Z}$-polytope can provide insight into whether or not any particular collection of moves is a Markov basis. We studied a particular type of potential Markov basis that we called a column partition lattice basis. Column partition lattice bases have a simple geometric interpretation, which is that they provide co-ordinate direction moves when the $\mathbb{Z}$-polytope is projected onto a subset of the co-ordinates. Constructing a column partition lattice basis involves one matrix inversion and one matrix multiplication, which on medium to large problems should make it much more computationally efficient to find a column partition lattice basis than a Gröbner bases.

7.1.1 Markov basis results

Our aim was to find a general method for determining when a column partition lattice basis is a Markov basis, and to investigate ways of constructing lattice bases that guarantee that they are Markov bases. We identified several problems that can prevent a column partition from being a Markov basis.

The first problem we discussed was one we called parity errors (Section 2.5.1). This occurs when the elements of the basis are non-integral, and the $\mathbb{Z}$-polytope is split into two or more cliques that are not connected to each other. We proved that non-integral column partition lattice basis matrices are avoided when the column partition is such that the determinant of A_1 divides the determinants of the other maximal submatrices (Theorem 3.4.1). We also proved that if the determinants of each of the maximal invertible submatrices are equal in absolute value, then every column partition lattice basis is integral, and the matrix that defines the column partition lattice basis is totally unimodular.

If the column partition lattice basis is integral, then every lattice point in the projected space that is within the bounding hyperplanes is an element of the fibre. The question of connectivity then comes down to the geometry of the projected polytope. We found that the choice of the column partition determines the angles between the faces of the projected polytope, as discussed in Section 3.3. The angles between the faces of the projected polytope are important in determining whether a walk constructed with coordinate direction moves will be able to access all the points in the projected polytope: faces coming together at acute angles can result in some vertices or cliques of vertices being inaccessible.

We found that there are conditions that guarantee that we can construct a column partition lattice basis that is a Markov basis. Section 4.2 gives one condition, concerning the case where we can partition the columns of the configuration matrix A such that each column of the A_2 partition is a positive sum of columns of the A_1 partition. This is applicable for example when the configuration matrix has the identity matrix as a maximal submatrix, a situation common in capture-recapture models. This situation also appears in link-path incidence matrices in network tomography when each edge is an allowed path, or more generally each of the A_2 paths can be constructed by linking together paths in A_1.

Generalising this result produced a test that can be used to determine whether a column partition lattice basis is a Markov basis. This is detailed in Theorem 4.4.14. If U is a column partition lattice basis with all entries in $\{0, \pm 1\}$, then this theorem states:

$$\begin{gathered}U \text{ contains no non-zero Eulerian submatrix whose columns each sum to zero} \\ \Longrightarrow \\ U \text{ is a Markov basis.}\end{gathered}$$

We conjectured that this implication is bidirectional.

We extended this idea to general integral U matrices: in that case, we must take U^σ, the matrix of signs of the entries of U. The theorem becomes:

$$\begin{gathered}U^\sigma \text{ contains no non-zero Eulerian submatrix whose columns each sum to zero} \\ \Longrightarrow \\ U \text{ is a Markov basis.}\end{gathered}$$

7.1.2 Walks with other lower bounds

In Chapter 6 we discussed the possibility of using polytope samplers that may visit points that are not in the polytope, although such points are not included in the obtained sample. The use of such a sampler requires knowledge of a lower bound on the coordinates of points the sampler needs to visit in order to be able to visit every point in the ℤ-polytope. For column partition lattice bases, it is known that the lower bound for the $r - n$ co-ordinates in the A_2 part is 0.

We gave a lower bound for column partition lattice bases where the non-zero entries of the U matrix are all ± 1, which is guaranteed if the configuration matrix is unimodular. For an $n \times r$ configuration matrix, this lower bound is

$$-\min\left(\left\lfloor \frac{r-n}{2} \right\rfloor, \left\lfloor \frac{n+1}{2} \right\rfloor\right)$$

for the n co-ordinates in the A_1 part. We proved this lower bound in Section 6.4.3.

We gave a conjecture we called the Minus One Conjecture (Conjecture 6.3.2). This conjectures that for unimodular configuration matrices, this lower bound can be strengthened to -1 in the n A_1 co-ordinates and 0 in the $r - n$ A_2 co-ordinates. This is known to be the case for two-way contingency tables, whose configuration matrices are totally unimodular.

In Section 6.4 we proved the Minus One Conjecture in a few specific cases. It is known that a lower bound of -1 applies for any basis which contains a lattice basis if the ideal generated by the monomial difference representations of the basis elements is radical. In Section 6.4.1, we apply this result to column partition lattice bases. In the case of column partition lattice bases, the lower bound of -1 would only apply to the n co-ordinates in the A_1 part; the lower bound for the $r - n$ co-ordinates in the A_2 part is 0.

In Section 6.4.4 we prove the Minus One Conjecture for configuration matrices with $n \leq 5$ rows. For such matrices each circuit (or Graver basis element) is a sum of at most six columns of the reduced row echelon form of the configuration matrix. By enumerating the possible combinations of rows of this matrix we showed that if the lower bound of minus one does not hold, then this matrix must not be totally unimodular and hence the configuration matrix is not unimodular.

We also proved a lower bound of -1 for the link-path incidence matrices of polytrees, which are totally unimodular. The proof is given in Section 6.4.2. It is in fact a proof of the -1 lower bound for network matrices. Polynomial time algorithms exist for recognising network matrices, so this result is applicable not only to polytrees in network tomography, but also to any configuration matrix recognised as a network matrix by these algorithms.

The proof for network matrices also provides a possible research avenue for a proof of the Minus One Conjecture in full. Network matrices are an important building block for all totally unimodular matrices. Together with two certain 5×5 matrices, they can be combined using a certain set of operations to produce any totally unimodular matrix. If it could be shown that these operations preserve properties that imply the -1 lower bound, it could lead to a proof of the Minus One Conjecture.

7.1.3 Other results

In Section 3.5.2 we prove that for a configuration matrix A, if the determinants of all of the invertible maximal submatrices are equal in absolute value, the union of all column partition lattice bases of A is equal to the set of circuits of A, and hence the Graver basis of A. One application of this is in the adaptive lattice basis sampler of Hazelton et al. [27]. This theorem guarantees that a $\mathbb{Z}$-polytope sampler that dynamically changes which column partition lattice basis it uses, but is capable of accessing all column partition lattice bases, will be capable of sampling every point in the $\mathbb{Z}$-polytope and hence generate an irreducible Markov chain.

7.2 Applications

The Fundamental Theorem of Markov Bases was originally developed for work on contingency tables [19]. Much subsequent work has focussed on these, and they are frequently used as examples in the literature. There are other applications: Markov Chain Monte Carlo is applicable to a range of statistical linear inverse problems, which includes capture-recapture models in ecology and network tomography.

This thesis has a particular focus on volume network tomography, where the configuration matrices are the link-path incidence matrices of the network. We found several conditions on traffic networks that guarantee favourable properties in the configuration matrix. One important result is that the link-path incidence matrices for a network on a polytree, defined and discussed in Section 5.3, is totally unimodular. We gave two new proofs of this fact. We also demonstrated that link-path incidence matrices for polytrees are a subset of network matrices, which are known to be totally unimodular.

In Section 5.4 we looked at networks on symmetric digraphs, a type of graph that closely models real world traffic networks, which are not in general totally unimodular. We showed that when the underlying graph is a tree and the collection of allowed paths conforms to a certain set of rules, there exist many unimodular maximal submatrices. We speculated that this is true for all symmetric digraphs.

7.3 Avenues for future research

In this thesis we have made some conjectures which may provide avenues for future research. We collect them here.

Theorem 4.4.14 gives a condition that guarantees that a column partition lattice basis for a unimodular configuration matrix is a Markov basis. It states that if the matrix U that defines the basis contains no Eulerian submatrices M such that for each column $\mathbf{m}_i$ of M, the sum of the entries in $\mathbf{m}_i$ is zero, then U is a Markov basis. Conjecture 4.4.19 claims that the implication is bidirectional.

In Section 5.4 we looked at link-path incidence matrices of traffic networks on symmetric digraphs that have a certain routing policy. We found that if the underlying graph of the network is a tree, there are many different maximal unimodular submatrices. In Section 5.4.4, we speculate that this result extends to traffic networks on any symmetric digraph that follow the routing policy.

In Chapter 6 we gave the Minus One Conjecture (Conjecture 6.3.2). This conjecture claims that all column partition lattice bases of unimodular configuration matrices are what we called $\mathbf{m}_{(1,0)}$-Markov bases. This means that for any $\mathbf{y}$-fibre, they can be used to construct a walk between any pair of points such that the walk only visits points that are at least -1 in the first n co-ordinates, and non-negative elsewhere. If true, the Minus One Conjecture has consequences both for constructing walks that connect a set that includes the fibre of interest; and for approximate samplers, which are only concerned with visiting points that are internal to the polytope, which for many models is where the vast bulk of the probability might lie.

Bibliography

[1] E. Airoldi and A. Blocker. "Estimating Latent Processes on a Network from Indirect Measurements". *Journal of the American Statistical Association* 108 (2013), pp. 149–164.

[2] E. Airoldi and B. Haas. "Polytope Samplers for Inference in Ill-posed Inverse Problems". *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 2011, pp. 110–118.

[3] S. Aoki, H. Hara, and A. Takemura. *Markov Bases in Algebraic Statistics*. Springer New York, 2012.

[4] L. Auslander and H.M. Trent. "Incidence Matrices and Linear Graphs". *Journal of Mathematics and Mechanics* 8.5 (1959), pp. 827–835.

[5] L. Auslander and H.M. Trent. "On the Realization of a Linear Graph Given its Algebraic Specification". *The Journal of the Acoustical Society of America* 33.9 (1961), pp. 1183–1192.

[6] R.E. Bixby. "Kuratowski's and Wagner's Theorems for Matroids". *Journal of Combinatorial Theory, Series B* 22.1 (1977), pp. 31–53.

[7] R.E. Bixby and W.H. Cunningham. "Converting Linear Programs to Network Problems". *Mathematics of Operations Research* 5.3 (1980), pp. 321–357.

[8] S. Bonner et al. "Extending the Latent Multinomial Model with Complex Error Processes and Dynamic Markov Bases". *The Annals of Applied Statistics* 10.1 (2016), pp. 246–263.

[9] S. Borgwardt, T. Stephen, and T. Yusun. "On the Circuit Diameter Conjecture". *Discrete & Computational Geometry* 60.3 (2018), pp. 558–587.

[10] B. Buchberger. "A Theoretical Basis for the Reduction of Polynomials to Canonical Forms". *ACM SIGSAM Bulletin* 10.3 (1976), pp. 19–29.

[11] P. Camion. "Characterization of Totally Unimodular Matrices". *Proceedings of the American Mathematical Society* 16.5 (1965), pp. 1068–1073.

[12] Y. Chen, I. Dinwoodie, and R. Yoshida. "Markov Chains, Quotient Ideals, and Connectivity with Positive Margins". *Algebraic and Geometric Methods in Statistics* (2008), pp. 99–110.

[13] D.A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer International Publishing, 2015.

[14] S. Dasgupta. "Learning Polytrees". *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999, pp. 134–141.

[15] J. De Loera, R. Hemmecke, and M. Koeppe. *Algebraic and Geometric Ideas in the Theory of Discrete Optimization*. Society for Industrial and Applied Mathematics, 2012.

[16] J. De Loera and S. Onn. "Markov Bases of Three-way Tables are Arbitrarily Complicated". *Journal of Symbolic Computation* 41.2 (2006), pp. 173–181.

[17] L. Denby et al. "Statistical Aspects of the Analysis of Data Networks". *Technometrics* 49.3 (2007), pp. 318–334.

[18] P. Diaconis, D. Eisenbud, and B. Sturmfels. "Lattice Walks and Primary Decomposition". *Mathematical Essays in Honor of Gian-Carlo Rota*. Birkhauser Boston, 1998, pp. 173–193.

[19] P. Diaconis and B. Sturmfels. "Algebraic Algorithms for Sampling from Conditional Distributions". *The Annals of Statistics* 26.1 (1998), pp. 363–397.

[20] I. Dinwoodie. "The Diaconis-Sturmfels Algorithm and Rules of Succession". *Bernoulli* 4.3 (1998), pp. 401–410.

[21] I. Dinwoodie and Y. Chen. "Sampling Large Tables with Constraints". *Statistica Sinica* 21.4 (2011), pp. 1591–1609.

[22] A. Dobra. "Dynamic Markov Bases". *Journal of Computational and Graphical Statistics* 21.2 (2012), pp. 496–517.

[23] A. Ghouila-Houri. "Caractérisation des Matrices Totalement Unimodulaires". *Comptes Redus Hebdomadaires des Séances de l'Académie des Sciences* 254 (1962), pp. 1192–1194.

[24] R. Gould. "Graphs and Vector Spaces". *Journal of Mathematics and Physics* 37.1-4 (1958), pp. 193–214.

[25] M. Hazelton. "Network Tomography for Integer-valued Traffic". *The Annals of Applied Statistics* 9.1 (2015), pp. 474–506.

[26] M. Hazelton and T. Bilton. "Polytope Samplers for Network Tomography". *Australian & New Zealand Journal of Statistics* 59.4 (2017), pp. 495–511.

[27] M. Hazelton, M. McVeagh, and B. van Brunt. "Geometrically Aware Dynamic Markov Bases for Statistical Linear Inverse Problems". *Biometrika* (Oct. 2020).

[28] R. Hemmecke and P.N. Malkin. "Computing Generating Sets of Lattice Ideals and Markov Bases of Lattices". *Journal of Symbolic Computation* 44.10 (2009), pp. 1463–1476. ISSN: 0747-7171.

[29] A.J. Hoffman. "Some Recent Applications of the Theory of Linear Inequalities to Extremal Combinatorial Analysis". *Combinatorial Analysis, AMS, Providence* (1960), pp. 113–128.

[30] T. Kahle and J. Rauh. *The Markov Bases Database*. 2017. URL: `http://www.markov-bases.de` (visited on 05/24/2018).

[31] M. Koh et al. *Graph Theory*. World Scientific, 2015.

[32] E. Lawrence et al. "Network Tomography: A Review and Recent Developments". *Frontiers in Statistics*. Imperial College Press, 2012, pp. 345–366.

[33] A. Lee. "Generating Synthetic Microdata from Published Marginal Tables and Confidentialised Files". *Official Statistics Research Series* 5.123 (2009).

[34] A. Lee. "Generating Synthetic Unit-record-data from Published Marginal Tables". *Official Statistics Research Series* 1 (2007).

[35] B. Li. "Bayesian Inference for Origin-Destination Matrices of Transport Networks Using the EM Algorithm". *Technometrics* 47.4 (2005), pp. 399–408.

[36] W. Link et al. "Uncovering a Latent Multinomial: Analysis of Mark-Recapture Data with Misidentification". *Biometrics* 66.1 (2010), pp. 178–185.

[37] A. Martin del Campo, S. Cepeda, and C. Uhler. "Exact Goodness-of-Fit Testing for the Ising Model". *Scandinavian Journal of Statistics* 44.2 (2016), pp. 285–306.

[38] F. Rapallo and R. Yoshida. "Markov Bases and Subbases for Bounded Contingency Tables". *Annals of the Institute of Statistical Mathematics* 62.4 (2010), pp. 785–805.

[39] M. Schofield and S. Bonner. "Connecting the Latent Multinomial". *Biometrics* 71.4 (2015), pp. 1070–1080.

[40] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

[41] P.D. Seymour. "Decomposition of Regular Matroids". *Journal of Combinatorial Theory, Series B* 28.3 (1980), pp. 305–359.

[42] B. Sturmfels. *Groebner Bases and Complex Polytopes*. American Mathematical Society, 1996.

[43] S. Sullivant. *Algebraic Statistics*. American Mathematical Society, 2018. URL: `https://www4.ncsu.edu/~smsulli2/Pubs/publications.html`.

[44] 4ti2 team. *4ti2—A Software Package for Algebraic, Geometric and Combinatorial Problems on Linear Spaces*. Available at www.4ti2.de.

[45] C. Tebaldi and M. West. "Bayesian Inference on Network Traffic Using Link Count Data". *Journal of the American Statistical Association* 93.442 (1998), pp. 557–573.

[46] W. Tutte. "Lectures on Matroids". *Journal of Research of the National Bureau of Standards* 69.1 (1965), pp. 1–47.

[47] W.T. Tutte. "An Algorithm for Determining Whether a Given Binary Matroid is Graphic". *Proceedings of the American Mathematical Society* 11.6 (1960), pp. 905–917.

[48] W.T. Tutte. “On Even Matroids”. *J. Res. Nat. Bur. Standards Sect. B* 71 (1967), pp. 213–214.

[49] Y. Vardi. “Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data”. *Journal of the American Statistical Association* 91.433 (1996), pp. 365–377.

[50] R. Yoshida. “Open Problems on Connectivity of Fibers with Positive Margins in Multi-dimensional Contingency Tables”. *Journal of Algebraic Statistics* 1.1 (2010), pp. 13–26.