Bernoulli **30**(4), 2024, 2676–2699
https://doi.org/10.3150/23-BEJ1690

Some rapidly mixing hit-and-run samplers for latent counts in linear inverse problems

MARTIN HAZELTON[1,a], MICHAEL MCVEAGH[1,b], CHRISTOPHER TUFFLEY[2,c] and BRUCE VAN BRUNT[2,d]

[1] *University of Otago, Dunedin, New Zealand,* [a]*martin.hazelton@otago.ac.nz,* [b]*mcveagh@gmail.com*
[2] *Massey University, Palmerston North, New Zealand,* [c]*C.Tuffley@massey.ac.nz,* [d]*B.vanBrunt@massey.ac.nz*

Linear inverse problems for count data arise in a myriad of settings. The latent counts lie on a fibre that is too large to enumerate in most practical problems, but inference can proceed by sampling the fibre. We examine the mixing properties of hit-and-run samplers in this context. In general convergence can be arbitrarily slow. However, there is a class of linear inverse problems for which rapid mixing for uniform fibre sampling is possible, using Markov sub-bases that are of minimum size but yet provide a sufficiently rich range of sampling directions to avoid the need for zig-zagging walks to ensure connectivity. Focussing on such problems, we study a particular class of bases that enjoy these properties under certain easily checkable conditions on the configuration matrix. We also examine the mixing properties of these bases when employing commonly used Poisson models. Our theoretical results provide practical guidance on optimizing these Markov sub-bases.

Keywords: Augmenting path; Eulerian matrix; fibre sampler; Markov basis; mixing time; random walk; second largest eigenvalue modulus

1. Introduction

Suppose that we are interested in a random process generating counts $\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0}$, but can observe only $\boldsymbol{y} \in \mathbb{Z}^n_{\geq 0}$ defined by the linear system

$$\boldsymbol{y} = A\boldsymbol{x} \tag{1}$$

where A is the $n \times r$ configuration matrix. Conducting inference for $\boldsymbol{x}$, and potentially for the parameters underlying its distribution, is a statistical linear inverse problem. Problems of this type arise frequently in science and engineering, and are known to be challenging (e.g. Hazelton, McVeagh and van Brunt, 2021, Kaipio and Somersalo, 2006).

We consider the common situation in which the entries of A are binary, and $r > n$. The resulting class of models covers a plethora of applications in which the ideal data $\boldsymbol{x}$ are observed only in summarized or otherwise corrupted form. For example, they arise when analysing mark-recapture data in ecology, where counts $\boldsymbol{y}$ of recorded histories of animal sightings are subject to identification errors and hence may differ from the true counts $\boldsymbol{x}$ (e.g. Link et al., 2010, Schofield and Bonner, 2015).

Example 1. Consider a mark-recapture experiment conducted over two observational windows, in which the identity of each animal spotted is compared to an existing catalogue. Let $\boldsymbol{x}_\omega$ denote the count of animals with history $\omega = (\omega_1, \omega_2)$, where $\omega_i = 0$ if the animal is not seen in window i; $\omega_i = 1$ if the animal is seen in window i and correctly identified; and $\omega_i = 2$ if the animal is seen in window i but misidentified as a new animal. The observed counts are collected as the vector $\boldsymbol{y} = (y_{01}, y_{10}, y_{11})^\mathsf{T}$, where the ith index is a binary indicator of animal sighting in the ith window. Following Schofield and Bonner (2015), these are related to the true counts $\boldsymbol{x} = (\boldsymbol{x}_{01}, \boldsymbol{x}_{10}, \boldsymbol{x}_{11}, \boldsymbol{x}_{02}, \boldsymbol{x}_{12}, \boldsymbol{x}_{20}, \boldsymbol{x}_{21}, \boldsymbol{x}_{22})^\mathsf{T}$ according

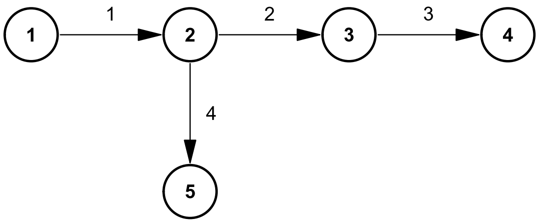

Figure 1. A simple example for network tomography.

to (1) with configuration matrix

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

One very important application of linear inverse models is network volume tomography, in which the aim is to make inferences about traffic volumes on routes through a network based on aggregate traffic counts collected at various sites. This is a standard practice in engineering, both for road systems and electronic communication networks. See for example Airoldi and Blocker (2013), Castro et al. (2004), Hazelton (2015), Tebaldi and West (1998), Vardi (1996) for some statistical contributions to the area.

Example 2. To illustrate, consider the simple network depicted in Figure 1. Let $\boldsymbol{x} = (x_{(1,2)}, x_{(1,3)}, x_{(1,4)}, x_{(2,5)}, x_{(3,4)}, x_{(2,3)}, x_{(1,5)}, x_{(2,4)})^\mathsf{T}$ denote the origin-destination traffic volumes, where $x_{(i,j)}$ is the count of traffic journeying between nodes i and j. (What may appear an idiosyncratic ordering of the origin-destination pairs is helpful in later examples.) Our information comes from observed traffic counts $\boldsymbol{y}$ on the network links, related to $\boldsymbol{x}$ through (1) where A is the link-path incidence matrix given by

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

Let $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ denote the distribution of $\boldsymbol{x}$, where $\boldsymbol{\theta}$ is a vector of parameters. Then the likelihood is

$$\begin{aligned} f(\boldsymbol{y} \mid \boldsymbol{\theta}) &= \sum_{\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0}} f(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{x} \in \mathcal{F}_{A,\boldsymbol{y}}} f(\boldsymbol{x} \mid \boldsymbol{\theta}) \end{aligned} \tag{2}$$

where $\mathcal{F}_{A,\boldsymbol{y}} = \{\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0} : \boldsymbol{y} = A\boldsymbol{x}\}$ is the $\boldsymbol{y}$-fibre of count vectors $\boldsymbol{x}$ consistent with the observed data. In practice this set will be far too large to enumerate in even modestly sized examples, so the likelihood cannot be computed directly. We must therefore resort to sampling-based methods of inference. In particular, Bayesian inference for $\boldsymbol{\theta}$ can be conducted by iterating between sampling from $f(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta})$ and $f(\boldsymbol{\theta} \mid \boldsymbol{x})$ (see Hazelton, McVeagh and van Brunt, 2021, for example).

This provides a strong practical motivation to investigate the problem of sampling from the conditional distribution $f(\boldsymbol{x} \mid \boldsymbol{y}) = f_{\mathcal{F}_{A,\boldsymbol{y}}}(\boldsymbol{x})$, where $f_{\mathcal{F}}$ denotes a distribution truncated (and renormalized) to the set $\mathcal{F}$ and the dependence on $\boldsymbol{\theta}$ is suppressed for notational convenience. This is a challenging problem because we generally lack any convenient representation for the $\boldsymbol{y}$-fibre. It has been heavily studied in the context of contingency table resampling, where $\boldsymbol{x}$ is the vector of cell entries and $\boldsymbol{y}$ a vector of marginal totals corresponding to sufficient statistics for some log-linear model of interest. The most common approach is to conduct a random walk on $\mathcal{F}_{A,\boldsymbol{y}}$, but the difficulty is ensuring that a sufficiently rich array of sampling directions is available to ensure connectivity.

A breakthrough arrived when Diaconis and Sturmfels (1998) demonstrated the use of techniques from computational algebra to find a Markov basis corresponding to a given configuration matrix; that is, a set of sampling directions sufficient to guarantee connectivity of a random walk on $\mathcal{F}_{A,\boldsymbol{y}}$ for any $\boldsymbol{y} \geq \boldsymbol{0}$ (where such inequalities are assumed to apply componentwise henceforth). That article inspired a raft of follow-up articles, most of which focussed on the theory and computation of Markov bases for various log-linear models for contingency table data (e.g. Dobra, 2003, Rapallo, 2003, Takemura and Aoki, 2004). The monograph by Aoki, Hara and Takemura (2012) provides an overview. Computational algebra methods for finding Markov bases are conveniently implemented in the software `4ti2` (4ti2team, 2015).

Sampling can proceed using a simple random walk, in which the sampler at each iteration either takes a step defined by a randomly selected and randomly signed vector in the Markov basis, or holds fast. Such a sampler is adequate for the study of connectedness, but will mix increasingly slowly as the fibre size increases (Stanley and Windisch, 2018, Windisch, 2016). From a practical perspective a far more attractive proposition (and consequently a far more common choice by practitioners) is a hit-and-run sampler, which selects a new state from the lattice points on a ray through the fibre the direction of which is determined by a randomly selected vector from the Markov basis. See Section 2.

The mixing properties of hit-and-run samplers for fibres have received limited attention, despite the critical importance of this topic in determining the practicability of MCMC inference for the kinds of linear inverse problems described above. Examples of such hit-and-run samplers with very poor mixing properties in practice are readily available (e.g. Hazelton, McVeagh and van Brunt, 2021), but we lack theoretical results to properly explain this behaviour.

Stanley and Windisch (2018) recently made some progress in this direction when they examined uniform fibre sampling using hit-and-run samplers with Markov sub-bases; i.e. sufficient sets of sampling directions for pre-specified $\boldsymbol{y}$ (Chen, Dinwoodie and Sullivant, 2006). Using techniques develop by Sinclair (1992), they sought to find bounds on mixing times based on longest path length between points on a fibre. To obtain rapid mixing, the Markov sub-basis must contain a sufficiently rich selection of sampling directions so that zig-zagging sample paths are not required to ensure connectivity. Stanley and Windisch (2018) described such a basis as *augmenting*. It is also necessary to limit the size of the Markov sub-basis to avoid overly long paths even without zig-zags. Combining these ideas, Stanley and Windisch (2018) showed that rapid mixing is possible if we can find a sufficiently small augmenting Markov sub-basis: specifically, with cardinality not greater than $\dim(\ker_{\mathbb{Z}}(A))$. See Section 3.

While the existence of such small Markov sub-bases is by no means guaranteed for all linear inverse problems, nevertheless the class of problems for which they do exist warrants attention for two reasons. First, there are many important practical problems within this class including almost every reasonable model for mark-recapture data, and also network tomography problems for highways and transit systems. Second, the work of Stanley and Windisch (2018) provides reason to hope that we can design highly efficient samplers for application to problems in this class. We focus on such problems in this article.

We face a number of challenges. It can be difficult to ascertain whether a sufficiently small Markov sub-basis is available for any given problem. Moreover, even if we know that such bases exist, we may not be able to find them. Techniques from algebraic statistics will not necessarily help in this regard, since they are focussed on full Markov bases which may be (much) larger than a serviceable Markov sub-basis. Finally, it is unclear as to how well hit-and-run samplers based on these small bases will work for non-uniform target distributions. This is crucial, since for most practical linear inverse problems we will be focussed on non-uniform models: Possion models are the most common choice.

In response, we describe a type of lattice basis that will also be a minimally sized augmenting Markov sub-basis under certain, easily checkable, conditions on the configuration matrix. See Section 4. As a consequence, we can find hit-and-run samplers that mix rapidly for uniform fibre sampling in problems with such configuration matrices. We also analyze the mixing properties of these samplers when the components of $\boldsymbol{x}$ follow independent Poisson distributions. These depend critically on the relative magnitudes of the components of $\mathsf{E}[\boldsymbol{x}]$. Moves will necessarily be short in sampling directions necessitating changes in components of $\boldsymbol{x}$ with small mean, which can lead to bottlenecks in sample paths and arbitrarily poor mixing behaviour in extreme examples. See Section 5. Our theoretical analysis provides a highly useful guide to practical choice of basis, as we illustrate through a real data network tomography example in Section 6.

We assume throughout that the configuration matrix is of full rank. If this is not the case then we can delete redundant rows of A and corresponding entries of $\boldsymbol{y}$ without any loss of information. We also assume that each column of A contains at least one non-zero entry.

We will at times want to refer to the dimension of the $\boldsymbol{y}$-fibre. To that end, we make the following definition.

Definition 1.1 (Dimension of a $\boldsymbol{y}$-fibre). The dimension d of fibre $\mathcal{F}_{A,\boldsymbol{y}}$ is given by

$$d = \dim(\mathrm{span}\{\boldsymbol{x} - \boldsymbol{x}' \colon \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}\}).$$

We observe that if all entries of $\boldsymbol{y}$ are sufficiently large, then $\dim(\mathcal{F}_{A,\boldsymbol{y}}) = \dim(\ker_{\mathbb{Z}}(A)) = r - n$. However, if some of the entries of $\boldsymbol{y}$ are sufficiently small then $\dim(\mathcal{F}_{A,\boldsymbol{y}})$ may differ from the dimension of the solution space of (1) for continuous variables.

Example 3. Consider a linear inverse problem with configuration matrix

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

and observed data $\boldsymbol{y} = (1,1,1)^{\mathsf{T}}$. In the continuum the solution set can be written as $\{\boldsymbol{x} \in \mathbb{R}^5_{\geq 0} \colon x_1 \geq 0, x_2 \geq x_1, x_1 + x_2 \leq 1, x_3 = 1 - x_2, x_4 = 1 - x_1 - x_2, x_5 = x_2 - x_1\}$, which is a 2-dimensional polytope. However, for integer valued vectors the $\boldsymbol{y}$-fibre is $\mathcal{F}_{A,\boldsymbol{y}} = \{(0,0,1,1,0)^{\mathsf{T}}, (0,1,0,0,1)^T\}$ which is 1-dimensional. For $\boldsymbol{y} = (2,1,1)^T$ the fibre is $\mathcal{F}_{A,\boldsymbol{y}} = \{(1,1,0,0,0)^{\mathsf{T}}, (0,0,1,2,0)^{\mathsf{T}}, (0,1,0,1,1)^T\}$ which is 2-dimensional.

2. Hit-and-run samplers for latent counts

We consider the problem of sampling latent counts $\boldsymbol{x}$ from a distribution $f(\boldsymbol{x} \mid \boldsymbol{y})$ supported on the fibre $\mathcal{F}_{A,\boldsymbol{y}}$. Much of the theoretical study of Markov bases has focussed on random walk samplers in which candidates are drawn from the Markov basis with random allocation of sign. However, while that is sufficient for the analysis of connectedness, in practice we want the sampler to be able to take long steps when possible to facilitate fast mixing. We therefore focus on discrete versions of the hit-and-run sampler (Baumert et al., 2009, Smith, 1984).

Given a current vector $\boldsymbol{x} \in \mathcal{F}_{A,\boldsymbol{y}}$, the hit-and-run sampler works by randomly selecting a sampling direction $\boldsymbol{u}^\dagger$ from a set of basic moves $\mathcal{M} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$. For each $i = 1, \dots, N$, $\boldsymbol{u}_i$ is a vector with coprime entries, and $\boldsymbol{u}_i \in \ker_{\mathbb{Z}}(A)$. Then a candidate $\boldsymbol{x}^\dagger = \boldsymbol{x} + b\boldsymbol{u}^\dagger$ is sampled from the ray $\mathcal{R} \equiv \mathcal{R}_{\mathcal{F}_{A,\boldsymbol{y}},\boldsymbol{u}^\dagger}(\boldsymbol{x}) = \{\boldsymbol{x} + b\boldsymbol{u}^\dagger : b \in \mathbb{Z}\} \cap \mathcal{F}_{A,\boldsymbol{y}}$. For computation, the ray is conveniently expressed as $\mathcal{R}_{\mathcal{F}_{A,\boldsymbol{y}},\boldsymbol{u}^\dagger}(\boldsymbol{x}) = \{b_{\min}(\boldsymbol{x}, \boldsymbol{u}^\dagger), b_{\min}(\boldsymbol{x}, \boldsymbol{u}^\dagger) + 1, \dots, b_{\max}(\boldsymbol{x}, \boldsymbol{u}^\dagger)\}$ where

$$b_{\min}(\boldsymbol{x}, \boldsymbol{u}^\dagger) = -\left\lfloor \min_{\{i \in [r] :\, u_i^\dagger > 0\}} \{x_i / |u_i^\dagger|\} \right\rfloor \tag{3}$$

and

$$b_{\max}(\boldsymbol{x}, \boldsymbol{u}^\dagger) = \left\lfloor \min_{\{i \in [r] :\, u_i^\dagger < 0\}} \{x_i / |u_i^\dagger|\} \right\rfloor . \tag{4}$$

In cases where none of the entries of $\boldsymbol{y}$ are large it is feasible to implement a Gibbs sampling approach, in which $\boldsymbol{x}^\dagger$ is sampled from the conditional posterior distribution $f(\boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{R})$ and acceptance is guaranteed. More generally, $\boldsymbol{x}^\dagger$ can be generated by sampling from some proposal distribution $q_{\boldsymbol{u}^\dagger}(\cdot \mid \boldsymbol{x})$ supported on $\mathcal{R}_{\mathcal{F}_{A,\boldsymbol{y}},\boldsymbol{u}^\dagger}(\boldsymbol{x})$, with the sampler moving to this new state with probability determined by the Metropolis-Hastings acceptance probability

$$\begin{aligned}
\alpha &= \min\left\{1, \frac{f(\boldsymbol{x}^\dagger \mid \boldsymbol{y}) q_{\boldsymbol{u}^\dagger}(\boldsymbol{x}^\dagger \mid \boldsymbol{x})}{f(\boldsymbol{x} \mid \boldsymbol{y}) q_{\boldsymbol{u}^\dagger}(\boldsymbol{x} \mid \boldsymbol{x}^\dagger)}\right\} \\
&= \min\left\{1, \frac{f(\boldsymbol{x}^\dagger) q_{\boldsymbol{u}^\dagger}(\boldsymbol{x} \mid \boldsymbol{x}^\dagger)}{f(\boldsymbol{x}) q_{\boldsymbol{u}^\dagger}(\boldsymbol{x}^\dagger \mid \boldsymbol{x})}\right\} .
\end{aligned}$$

The sampler remains at $\boldsymbol{x}$ otherwise. See Algorithm 1.

We denote the hit-and-run sampler implemented using a set of basic moves $\mathcal{M}$ by $\mathcal{H}(\mathcal{M})$. The Markov chain generated by this sampler is reversible and aperiodic. Convergence to $f(\boldsymbol{x} \mid \boldsymbol{y})$ is therefore guaranteed if and only if the random walk is connected over $\mathcal{F}_{A,\boldsymbol{y}}$. In other words, irreducibility of $\mathcal{H}(\mathcal{M})$ requires that for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$ there exists a finite sequence $i_1, \dots, i_K \in \{1, \dots, N\}$ and

Algorithm 1 Hit-and-run fibre sampler

Require: Initial state $\boldsymbol{x}^0 \in \mathcal{F}_{A,\boldsymbol{y}}$; set of basic moves $\mathcal{M} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$; sample size N_{sim}.

```
for t = 1 : N_sim do
    Sample uniformly u† from M
    Sample x† from distribution q_{u†} with support R_{F_{A,y},u†}(x^{t-1})
    Sample w uniformly on [0,1]
    if w ≤ α then
        x^t = x†
    else
        x^t = x^{t-1}
    end if
end for
return x^0, ..., x^{N_sim}
```

integers $b_1, \dots, b_K$ such that

$$\boldsymbol{x}' = \boldsymbol{x} + \sum_{j=1}^{K} b_j \boldsymbol{u}_{i_j}, \text{ and } \boldsymbol{x} + \sum_{j=1}^{L} b_j \boldsymbol{u}_{i_j} \in \mathcal{F}_{A,\boldsymbol{y}}$$

for all $1 \le L \le K$. If $\mathcal{M}$ has this property then it is a Markov sub-basis for the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ (cf. Chen, Dinwoodie and Sullivant, 2006). The set $\mathcal{M}$ is a (full) Markov basis for A if it is a Markov sub-basis for $\mathcal{F}_{A,\boldsymbol{y}}$ for any $\boldsymbol{y} \in \mathbb{Z}^n_{\ge \boldsymbol{0}}$.

We will study the mixing properties of $\mathcal{H}(\mathcal{M})$. Our theoretical results on the rapidity of mixing hold for any set of proposal distributions so long as for all $\boldsymbol{u} \in \mathcal{M}$ the support of $q_{\boldsymbol{u}}$ is the ray $\mathcal{R}_{\mathcal{F}_{A,\boldsymbol{y}},\boldsymbol{u}^\dagger}(\boldsymbol{x})$. As a consequence, we will not dwell on the form of the proposal distribution. Our focus will be on the effect of the choice of basis (set of basic moves) $\mathcal{M}$ on mixing times.

The cardinality of the move set $\mathcal{M}$ is of importance in this analysis. It is standard to say that $\mathcal{M}$ is a *minimal* Markov basis if no proper subset of $\mathcal{M}$ is a Markov basis. However, for Markov sub-bases we will be interested in minimality in the stronger sense of minimum cardinality, motivating the following definition.

Definition 2.1 (c-minimal Markov sub-basis). A Markov sub-basis $\mathcal{M}$ for a fibre $\mathcal{F}$ is c-minimal (*cardinality*-minimal) if there is no Markov sub-basis for $\mathcal{F}$ of smaller cardinality.

Example 4. Consider a linear inverse problem with

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

Then for $\boldsymbol{y} = (3,1)^\top$ the fibre is $\mathcal{F}_{A,\boldsymbol{y}} = \{(x_1, 3 - x_1, x_3, 1 - x_3)^\top : x_1 = 0,1,2,3, x_3 = 0,1\}$. The move set $\mathcal{M} = \{(1,-1,0,0)^\top, (0,0,1,-1)^\top\}$ is evidently a Markov sub-basis, and is c-minimal since any Markov sub-basis must be of size at least $\dim(\mathcal{F}_{A,\boldsymbol{y}}) = 2$. The move set $\mathcal{M} = \{(1,-1,1,-1)^\top, (1,-1,-1,1)^T, (2,-2,1,-1)^\top\}$ is also an inclusion-minimal Markov sub-basis, but it is not c-minimal. We note that c-minimal Markov sub-bases are not unique. For instance, it is easily checked that $\mathcal{M} = \{(1,-1,0,0)^\top, (1,-1,1,-1)^\top\}$ is also a c-minimal Markov sub-basis for the fibre at hand.

3. Mixing properties

Let $f^t_{\boldsymbol{x}^0}$ denote the sampling distribution of $\mathcal{H}(\mathcal{M})$ at iteration t when initialized at $\boldsymbol{x}^0$. We are then interested in the rate at which $f^t_{\boldsymbol{x}^0}$ approaches the target $f = f(\boldsymbol{x} \mid \boldsymbol{y})$. The discrepancy between these distributions can be measured by the total variation distance $||f^t_{\boldsymbol{x}^0} - f||_{TV}$, which motivates defining the ϵ-mixing time as $\tau_{\boldsymbol{x}^0}(\epsilon) = \min\{t : ||f^s_{\boldsymbol{x}^0} - f||_{TV} \le \epsilon \text{ for all } s \ge t\}$. Bounds for this mixing time are well known. In particular, for all $\epsilon > 0$,

$$\log\left(\frac{1}{2\epsilon}\right)\left(\frac{\lambda}{1-\lambda}\right) \le \tau_{\boldsymbol{x}^0}(\epsilon) \le \log\left(\frac{1}{\epsilon f(\boldsymbol{x}^0 \mid \boldsymbol{y})}\right)\frac{1}{1-\lambda} \tag{5}$$

where λ is the second largest eigenvalue modulus of the transition matrix of $\mathcal{H}(\mathcal{M})$. See for example Levin, Peres and Wilmer (2009). It follows that the spectral gap $1 - \lambda$ is a critical measure of convergence. The further that λ is from one, the more rapid the mixing.

We shall be interested in the asymptotic behaviour of the spectral gap as the problem size becomes large, or whether λ can be bounded away from 1 for a class of target posterior distributions. In either case, standard arguments imply that there is little loss in focussing on the modulus of the second largest eigenvalue. Specifically, if this is smaller in modulus than the smallest eigenvalue, we may consider instead a sampler with a holding probability of 1/2 added to each state. The second largest eigenvalue modulus of this slowed sampler will be $(\lambda + 1)/2$ and so the spectral gap is reduced by a factor of 2 only.

Intuitively, the rate of mixing depends on the number of transitions required for the sampler to move between any two vectors in the fibre. The work of Sinclair (1992) makes this idea more concrete. Consider the underlying graph describing possible transitions of a sampler for some set $\mathcal{F}$. Let $\gamma_{\boldsymbol{x},\boldsymbol{x}'}$ denote a path through this graph connecting $\boldsymbol{x}$ and $\boldsymbol{x}'$, and let $\Gamma = \{\gamma_{\boldsymbol{x},\boldsymbol{x}'} : \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}\}$ denote a set of canonical paths containing a single transitional path for each pair $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}$. Let Q denote the transition matrix for the sampler. Then the second largest eigenvalue λ is bounded above by

$$\lambda \leq 1 - \frac{1}{\rho \ell} \tag{6}$$

where ℓ is the maximum canonical path length, and

$$\rho \equiv \rho(\Gamma) = \max_{e \in E} \frac{1}{Q(e)} \sum_{\gamma_{\boldsymbol{x},\boldsymbol{x}'} \ni e} f(\boldsymbol{x} \mid \boldsymbol{y}) f(\boldsymbol{x}' \mid \boldsymbol{y}), \tag{7}$$

in which E indexes all pairs of vectors $e = (\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{F}^2$ for which $Q(e) = Q(\boldsymbol{v}, \boldsymbol{w}) > 0$. The term ρ captures a sense of the ease with which each transition can be made relative to the need for that transition during sampling. Sinclair (1992) used the notion of flows of probability over the graph, when the term ρ is a measure of the maximum probability loading of any edge e in the graph relative to capacity.

Consider the sampler $\mathcal{H}(\mathcal{M})$ applied to a sequence of fibres for fixed A, but with strictly increasing fibre size. In such circumstances it is quite possible that the number of transitions required for a walk between some pair of points on the fibre (i.e. the path length) may grow without limit. We then know from (6) that the mixing time for $\mathcal{H}(\mathcal{M})$ will become arbitrarily large. This kind of behaviour will certainly occur for a sampler with bounded step size b. Moreover, even though hit-and-run samplers provide the facility for arbitrarily long moves, a poor choice of sampling directions can nevertheless lead to long mixing times in even low-dimension problems.

Example 5. Consider the network tomography problem from Example 2. Suppose that the observed counts are $\boldsymbol{y} = (1, M, M, 1)^\mathsf{T}$ for some non-negative integer M. It is straightforward to show that

$$\begin{aligned}\mathcal{M} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\} = \{&(0,1,-1,0,1,0,0,0)^\mathsf{T}, (1,-1,0,0,0,1,0,0)^\mathsf{T},\\ &(-1,0,0,-1,0,0,1,0)^\mathsf{T}, (1,0,-1,0,0,0,0,1)^\mathsf{T}\}\end{aligned}$$

is a Markov sub-basis. The observed counts imply the constraints $0 \leq x_i \leq 1$ for $i = 1,2,3,4,7$. Each of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_4$ has non-zero entries of differing signs in these coordinates, so no move using the basis $\mathcal{M}$ can have length greater than one. Observe that $-\boldsymbol{u}_2 - \boldsymbol{u}_1 + \boldsymbol{u}_4 = (0,0,0,0,-1,-1,0,1)^\mathsf{T}$, and that the moves in that implied sequence can be applied to any $\boldsymbol{x} \geq (1,0,0,0,1,1,0,0)^T$ without leaving the corresponding fibre. The shortest walk from $\boldsymbol{x} = (1,0,0,0,M,M,0,0)^\mathsf{T}$ to $\boldsymbol{x}' = (1,0,0,0,0,0,0,M)^\mathsf{T}$ uses that sequence of moves M times, and so is of length $3M$. It follows that the spectral gap for the sampler $\mathcal{H}(\mathcal{M})$ can be arbitrarily small for sufficiently large M, and hence the mixing time can be arbitrarily large.

Evidently the difficulties experienced in this example stem from the zig-zagging nature of the path. Such problems can be avoided if we can find a Markov sub-basis with a more advantageous set of sampling directions, such that there is no need for repeated use of any basic move on a path between any pair of points on the fibre. Stanley and Windisch (2018) referred to such a path as *augmenting*. They used the same term to describe a set of moves sufficient to provide an augmenting walk between any two points on the fibre. We formalize the definition as follows so as to respect the standard distinction between Markov bases and sub-bases.

Definition 3.1 (Augmenting Markov sub-basis). A Markov sub-basis $\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$ for a fibre $\mathcal{F}_{A,\boldsymbol{y}}$ is augmenting if for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$, there exist integers $b_1, \dots, b_K$ and distinct indices $i_1, \dots, i_K \in \{1, \dots, N\}$ such that $\boldsymbol{x}' = \boldsymbol{x} + \sum_{j=1}^{K} b_j \boldsymbol{u}_{i_j}$ and $\boldsymbol{x} + \sum_{j=1}^{L} b_j \boldsymbol{u}_{i_j} \in \mathcal{F}_{A,\boldsymbol{y}}$ for all $1 \leq L \leq K \leq N$. The path defined by the sequence of coefficient-index pairs $((b_j, i_j))_{j=1,\dots,K}$ is an augmenting path.

In theory, augmenting Markov sub-bases are readily available. For instance, the set of all pairwise differences $\mathcal{M} = \{\boldsymbol{x} - \boldsymbol{x}' : \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}\}$ is an augmenting Markov sub-basis for the fibre $\mathcal{F}_{A,\boldsymbol{y}}$. A less trivial example is provided by the Graver basis for A, which generates an augmenting Markov sub-basis for the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ for any observed counts $\boldsymbol{y} \geq \boldsymbol{0}$ (De Loera, Hemmecke and Lee, 2015). However, we will want to employ small augmenting Markov sub-bases; if possible, a c-minimal augmenting Markov sub-basis (i.e., the smallest possible augmenting Markov sub-basis for the fibre at hand). From a theoretical perspective this is advantageous since we can use the cardinality of an augmenting Markov sub-basis as a bound on the maximum canonical path length, ℓ. In addition, there is no point in including moves in $\mathcal{M}$ that are impossible for the fibre at hand. For example, if some entries of $\boldsymbol{y}$ are zero then this will constrain various coordinates of $\boldsymbol{x}$ to be zero, and hence render infeasible vectors in the Graver basis which are non-zero in those coordinates.

The work of Stanley and Windisch (2018) illustrates the advantage in using a small augmenting Markov sub-basis. For uniform sampling on $\mathcal{F}_{A,\boldsymbol{y}}$, they showed that if we can find an augmenting Markov sub-basis with no more than $\dim(\ker A_{\mathbb{Z}})$ elements, then it is possible to uniformly bound the second largest eigenvalue of $\mathcal{H}(\mathcal{M})$ away from one when sampling from the infinite sequence of increasing fibres $\{\mathcal{F}_{A,\boldsymbol{y}} : \boldsymbol{y} = i\boldsymbol{y}_0, \boldsymbol{y}_0 \in \mathbb{Z}_0^n, i \in \mathbb{N}\}$ (Stanley and Windisch, 2018, Corollary 5.11). A hit-and-run sampler using such a basis therefore mixes rapidly for uniform sampling. Indeed, the underlying sequence of transition graphs is an expander.

While results for uniform sampling on $\mathcal{F}_{A,\boldsymbol{y}}$ have their value, we will almost always be focussed on non-uniform distributions when working with linear inverse models in practice. We will therefore be interested in the mixing properties of $\mathcal{H}(\mathcal{M})$ over classes of probability model applied to fibres of fixed sizes. To that end, we characterize the notion of rapid mixing for families of distributions on a fixed fibre through the following definition.

Definition 3.2 (Uniformly rapid mixing for a family of stationary distributions). Let $\mathcal{D}_\Theta = \{f(\boldsymbol{x} \mid \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ be a family of stationary distributions, each supported on some given and fixed set $\mathcal{F}$. A sampler mixes uniformly rapidly for $\mathcal{D}_\Theta$ if there exists $\epsilon > 0$ such that $\lambda(\boldsymbol{\theta}) < 1 - \epsilon$ for all $\theta \in \Theta$, where $\lambda(\boldsymbol{\theta})$ denotes the second largest eigenvalue modulus when the sampler is applied to $f(\boldsymbol{x} \mid \boldsymbol{\theta})$.

The hit-and-run sampler with a c-minimal augmenting Markov sub-basis mixes well for uniform sampling because the set of available sampling directions is propitious for the fibre at hand, and we can take long steps. For non-uniform target distributions, we may encounter circumstances in which the probability of taking long steps is arbitrarily small for members of the family $\mathcal{D}_\Theta$. The maximum loading ρ may then be unbounded in this family, and so mixing may not be uniformly rapid.

Example 6. Two-way contingency tables arise in network tomography when data are available on counts of traffic exiting (row totals) and entering (column totals) each zone of the network. In such circumstances a Poisson model is typically applied, with any pattern of mean values across the individual cells of the table.

Consider the 2×3 table

$$\begin{array}{c|ccc} & y_3 & y_4 & y_5 \\ \hline y_1 & x_1 & x_2 & x_3 \\ y_2 & x_4 & x_5 & x_6 \end{array}.$$

The configuration matrix matrix is

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

where the redundant fifth row (corresponding to y_5) has been omitted so that A is of full rank. Suppose that $\boldsymbol{y} = (1,2,1,1)^{\mathsf{T}}$, when the $\boldsymbol{y}$-fibre is

$$\mathcal{F}_{A,\boldsymbol{y}} = \{\boldsymbol{x}_1 = (0,0,1,1,1,0)^{\mathsf{T}}, \boldsymbol{x}_2 = (0,1,0,1,0,1)^{\mathsf{T}}, \boldsymbol{x}_3 = (1,0,0,0,1,1)^{\mathsf{T}}\}.$$

It is straightforward to see that

$$\mathcal{M} = \{\boldsymbol{u}_1 = (-1,0,1,1,0,-1)^{\mathsf{T}}, \boldsymbol{u}_2 = (-1,1,0,1,-1,0)^{\mathsf{T}}\}$$

is an augmenting Markov sub-basis for this fibre. For uniform sampling the transition matrix for $\mathcal{H}(\mathcal{M})$ is

$$Q = \frac{1}{4} \begin{bmatrix} 3\,0\,1 \\ 0\,3\,1 \\ 1\,1\,2 \end{bmatrix}$$

with second largest eigenvalue modulus $\lambda = 3/4$.

Now consider Gibbs sampling using a hit-and-run sampler when $x_1, \dots, x_6$ are independent Poisson random variables with $\mathsf{E}[\boldsymbol{x}] = (\delta, 1, 1, 1, 1, 1)^{\mathsf{T}}$. The transition matrix is

$$Q = \frac{1}{2+\delta} \begin{bmatrix} 2+\delta & 0 & \delta \\ 0 & 2+\delta & \delta \\ 1 & 1 & 2\delta \end{bmatrix}$$

which has second largest eigenvalue modulus $\lambda = (2+\delta)/(2+2\delta)$. This is not bounded uniformly away from one on $\delta > 0$, and so the sampler does not mix uniformly rapidly when $\mathcal{D}_\Theta = \{p(\boldsymbol{x} \mid \boldsymbol{\theta}) \colon \boldsymbol{\theta} > \boldsymbol{0}\}$, where $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ denotes a product of independent Poisson distributions with mean vector $\boldsymbol{\theta}$.

4. Some simple augmenting Markov sub-bases

In this section we describe a simple class of c-minimal augmenting Markov sub-bases that can be employed on any fibre when the configuration matrix A has certain helpful properties. Our first requirement is that we can form a lattice basis containing only integer-valued vectors using the Hermite normal form of A, (cf. Airoldi and Haas, 2011). To that end, let $A = [A_1 \mid A_2]$ be a partition of A such that A_1 is invertible, and define

$$U = \begin{pmatrix} -A_1^{-1}A_2 \\ I_{r-n} \end{pmatrix}. \tag{8}$$

Note that $AU = 0$, so the columns of U form a basis for ker(A). It is particularly convenient when the choice of partition of A leads to a matrix U with integer-valued entries, motivating the following definition.

Definition 4.1 (Partition Lattice Basis (PLB)). Let $A = [A_1 \mid A_2]$ be a partition with A_1 invertible. Let $\mathcal{M}_U$ be the set of columns of U defined by (8). Then we call $\mathcal{M}_U$ a partition lattice basis if all entries of U are integers.

We can find a PLB for a given configuration matrix if there is some choice of partition for which U is integer-valued. A sufficient condition for this is that A has at least one unimodular maximal square submatrix (i.e. submatrix of size $n \times n$). This property is very common in practical linear inverse problems in which the entries of A are binary. We can then reorder the columns of the configuration matrix as necessary to create the partition $A = [A_1 \mid A_2]$ in which A_1 is unimodular, so that U is an integer matrix and hence $\mathcal{M}_U$ is a PLB. Nevertheless, that condition is not necessary. As a trivial example, if $A = [A_1 \mid A_1]$ for any invertible matrix A_1 then $U = [-I \mid I]^T$ and so $\mathcal{M}_U$ is a PLB.

Example 7. Consider the one-way circuit network depicted in Figure 2, in which travel is possible on acyclic paths between any pair of nodes. The link-path incidence matrix (i.e. the configuration matrix for network tomography) is

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

We apply equation (8) to A; to a version of A with columns reordered so that the partition A_1 comprises columns 1, 2, 4 of the original matrix; and to a version of A with columns reordered so that the partition A_1 comprises columns 4, 5, 6 of the original matrix. The corresponding matrices computed

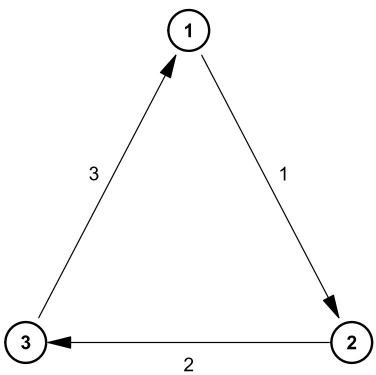

Figure 2. A one-way circuit network.

from (8), and with rows reordered to match the original column ordering of A, are given below:

$$U_1 = \frac{1}{2}\begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad U_2 = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & -1 \\ 1 & 0 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad U_3 = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

With the original partition $\det(A_1) = 2$, and the entries of U_1 are not all integer. The columns of U_1 do not form a PLB in this case. For both the second and third partitions A_1 is unimodular, and so the columns of both U_2 and U_3 form PLBs. The latter contains only entries in $\{0, \pm 1\}$ which is very common for PLBs derived from configuration matrix partitions encountered in practice.

Geometrically, the form of a $\boldsymbol{y}$-fibre is a $\mathbb{Z}$-polytope (i.e. the points on the integer lattice within a polytope). If the entries of $\boldsymbol{x}$ are reordered to match the column reordering implied by the partition $A = [A_1 \mid A_2]$ used to form U, then the fibre may be represented in terms of the values taken by the final $r - n$ coordinates of $\boldsymbol{x}$. The corresponding PLB corresponds to unit vector moves with respect to those coordinates.

In many applications A will have multiple partitions for which U is an integer matrix, and hence there will be many different PLBs available. For example, Hazelton and Bilton (2017) examined the configuration matrices arising in a wide range of realistic network tomography examples. In each case they found that for those partitions where A_1 was invertible, it was also unimodular more than 85% of the time. The abundance of PLBs is helpful because it provides wide scope to find some with beneficial characteristics.

Consider a random walk conducted by a hit-and-run sampler employing a PLB. If two points are to be connected by such a walk on the fibre, then we must be able to find a sequence of moves such that at each step of the walk the state vector $\boldsymbol{x}$ is non-negative in every coordinate. In general there is no guarantee that we will be able to do so. Moreover, even if we can find such a walk, its path need not be augmenting.

A straightforward way of forming an augmenting path would be to order application of moves from $\mathcal{M}_U$ so that all positive changes in each coordinate of $\boldsymbol{x}$ precede negative changes. We shall refer to such an augmenting path as 'up-and-down'. Whether or not an up-and-down augmenting path is available depends solely on the signs of the entries of basis vectors, and not their magnitudes. (See the proof of Theorem A.4 for details.) It is therefore convenient to work with the matrix of signs, $\mathsf{sgn}(U)$, the i, jth entry of which is defined to be $u_{i,j}/|u_{i,j}|$ when $u_{i,j} \neq 0$ and 0 otherwise. In practice it is rather common to find that $\mathsf{sgn}(U) = U$.

As we show in Theorem 4.4, the existence of certain types of Eulerian submatrices in $\mathsf{sgn}(U)$ is critical in determining whether a given lattice basis will be capable of generating augmenting paths of this sort for $\mathcal{H}(\mathcal{M}_U)$. Recall that a matrix is Eulerian if all its column and row sums are even. We define some more restrictive types of Eulerian matrix as follows.

Definition 4.2 (c0-Eulerian matrix). A non-null square matrix is c0-Eulerian if it is Eulerian and all its column sums are zero.

Definition 4.3 (cr0-Eulerian matrix). A matrix is cr0-Eulerian if it is c0-Eulerian and all its row sums are zero.

Some examples of c0-Eulerian matrices are

$$E_1 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \text{and} \quad E_3 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}.$$

Of these, E_2 is also cr0-Eulerian. Moreover, both E_1 and E_3 become cr0-Eulerian if the signs of their final columns are changed. Viewing either matrix as a submatrix of $\mathsf{sgn}(U)$, such a sign change corresponds to an irrelevant switching of sign in a member of the partition lattice basis. The cr0-Eulerian property is important, because it is impossible to reorder the columns of this type of matrix so that all the ones precede the minus ones in each row (a result proved in Lemma A.1 in the Appendix). The presence of a cr0-Eulerian submatrix of $\mathsf{sgn}(U)$ means that we cannot always find an up-and-down augmenting path, and can result in the need for zig-zagging behaviour during a walk using the basis $\mathcal{M}_U$.

Example 8. The Markov sub-basis $\mathcal{M} = \mathcal{M}_U$ presented in Example 5 is the partition lattice basis for the configuration matrix in Example 2 without any column reordering. In this case,

$$\mathsf{sgn}(U) = U = \begin{bmatrix} 0 & 1 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which has E_3 as a submatrix formed using the first three rows of columns 1, 2 and 4. Recall that the sequence $-\boldsymbol{u}_2 - \boldsymbol{u}_1 + \boldsymbol{u}_4$ had to be used repeatedly in the walk described in Example 5. Indeed, the need to zig-zag with this basis is so bad that $\mathcal{H}(\mathcal{M})$ is not connected on thin fibres. For instance, $\boldsymbol{x} = (0,0,0,0,1,1,0,0)^\mathsf{T}$ and $\boldsymbol{x}' = (0,0,0,0,0,0,0,1)^\mathsf{T}$ are not connected by any walk on the fibre with $\boldsymbol{y} = (0,1,1,0)^\mathsf{T}$, and so $\mathcal{M}$ is not a full Markov basis.

It transpires that absence of c0-Eulerian submatrices in $\mathsf{sgn}(U)$ is sufficient to ensure that its columns form a Markov basis that is also a c-minimal augmenting Markov sub-basis when the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ is of full dimension, $r - n$.

Theorem 4.4. *Let A be a configuration matrix for which $\mathcal{M}_U$ is a partition lattice basis. If the matrix of signs $\mathsf{sgn}(U)$ contains no c0-Eulerian submatrix, then*

(i) *$\mathcal{M}_U$ is a Markov basis for A.*
(ii) *$\mathcal{M}_U$ is a c-minimal augmenting Markov sub-basis for any fibre $\mathcal{F}_{A,\boldsymbol{y}}$ of dimension $r - n$.*

The general idea of the proof is as follows. Consider walking between $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$. Since the columns of U form a lattice basis for $\ker_{\mathbb{Z}}(A)$ we can write

$$\boldsymbol{x}' - \boldsymbol{x} = \sum_{i=1}^{r-n} \epsilon_i a_i \boldsymbol{u}_i,$$

where $\epsilon_i \in \{\pm 1\}$ gives the sign and $a_i \in \mathbb{Z}_{\geq \mathbf{0}}$ gives the number of copies of $\boldsymbol{u}_i$ required. Let U^ϵ be a version of U in which the ith column is multiplied by ϵ_i. Using each of the moves $\{a_i \boldsymbol{u}_i^\epsilon : i = 1, \ldots, n\}$ once will take us from $\boldsymbol{x}$ to $\boldsymbol{x}'$, but we must apply these moves in an appropriate order. The condition that $\mathsf{sgn}(U)$ contains no c0-Eulerian submatrix is sufficient to ensure that we can reorder the columns of U^ϵ so that no strictly negative entry precedes a strictly positive one in any row. Applying the moves in that implied order ensures that all positive changes in a coordinate are applied before negative ones. It follows that at every intermediate vector in the walk all coordinates are guaranteed to be non-negative, and hence the walk remains on the fibre. This demonstrates that M_U is a Markov basis. What is more, the walk just described utilizes an augmenting path with no more than $r - n$ steps. Full proofs (of this result and subsequent ones) are collected together in the Appendix.

Example 9. We return to the network tomography problem presented in Example 2, and select columns 1, 5, 6, 7 to form partition A_1. The corresponding PLB $\mathcal{M}_U$ is then formed from the columns of

$$U = \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{9}$$

where the row ordering matches the column ordering for the original A matrix. The matrix $\mathsf{sgn}(U) = U$ contains no c0-Eulerian submatrices, and so $\mathcal{M}_U$ is a Markov basis for A. It is also a c-minimal augmenting Markov sub-basis for any fibre $\mathcal{F}_{A,\boldsymbol{y}}$ with $\boldsymbol{y} \geq (1,1,1,1)^\mathsf{T}$.

Theorem 4.4 provides a checkable condition for a PLB $\mathcal{M}_U$ to be a c-minimal augmenting Markov sub-basis. Such checking is computationally feasible for the matrices that typically arise in practical problems of moderate size, in part because of the structure and sparseness of U. For example, c0-Eulerian submatrices can only exist in the n rows of U corresponding to the columns of A_1. Nevertheless, checking the condition in very large problems will typically be impractical. It is therefore interesting and important to be able to identify types of linear inverse model for which the configuration matrix contains no c0-Eulerian submatrices. The following corollary is an example of the kind of result that can be helpful in so doing.

Corollary 4.5. *Suppose that a configuration matrix can be partitioned as $A = [A_1 \mid A_2]$ such that each column of A_2 can be written as a linear combination of columns of A_1 with non-negative integer coefficients. Let $\mathcal{M}_U$ be the set of columns of U computed from (8). Then*

(i) *$\mathcal{M}_U$ is a Markov basis for A.*
(ii) *$\mathcal{M}_U$ is a c-minimal augmenting Markov sub-basis for any fibre $\mathcal{F}_{A,\boldsymbol{y}}$ of dimension $r - n$.*

The sum condition for Corollary 4.5 is trivially satisfied if A_1 is the identity matrix. This will hold in almost all models for mark-recapture data when A_1 is formed from the columns corresponding to no animal identification errors. The matrix A presented in Example 1 is a case in point.

Similarly, consider a network in which every node is both an origin and destination of travel, and traffic counts are observed on all links. Then the configuration matrix will contain the identity matrix as a maximal submatrix, corresponding to trips between neighbouring nodes. We may then order the columns of A so that A_1 is the identity and Corollary 4.5 holds. This result will hold in more general network tomography problems in which every path corresponding to the columns of the submatrix A_2 of the link-path incidence matrix can be formed by joining end on end the paths corresponding to the columns in A_1. Examples of this type occur frequently when studying highway systems, as illustrated by the data analysis presented in Section 6.

Theorem 4.4 works because the absence of a c0-Eulerian submatrix in $\mathsf{sgn}(U)$ discounts the possibility that a cr0-Eulerian submatrix will arise after some set of sign changes to the columns of $\mathsf{sgn}(U)$, so we can always find an up-and-down augmenting path between any pair of points on a fibre. However, if we do find a c0-Eulerian submatrix in $\mathsf{sgn}(U)$, there is no certainty that it can be turned into a cr0-Eulerian submatrix through application of sign changes to some of the columns of $\mathsf{sgn}(U)$. In addition, if $\mathsf{sgn}(U) \neq U$ then it is possible that augmenting paths without the up-and-down property may exist even if $\mathsf{sgn}(U)$ contains a cr0-Eulerian submatrix. As a consequence, only a partial reverse implication to Theorem 4.4 is available.

Theorem 4.6. *Let $A = [A_1 \mid A_2]$ be a configuration matrix partitioned such that A_1 is invertible and U (from 8) has entries in $\{0, \pm 1\}$. Suppose that there is some set of sign changes to the columns of U such that the resulting matrix contains a cr0-Eulerian submatrix with no column of zeroes. Then there exists at least one fibre $\mathcal{F}_{A,\boldsymbol{y}}$ for which the columns of U do not form an augmenting Markov sub-basis.*

Application of Theorem 4.6 in practice requires searching for a cr0-Eulerian submatrix in a comprehensive set of matrices generated by column sign changes to U. Checking for the presence of just a c0-Eulerian submatrix in the original U is sufficient if A is totally unimodular.

Corollary 4.7. *Let $A = [A_1 \mid A_2]$ be a configuration matrix partitioned such that A_1 is invertible, and compute U from (8). Suppose also that A is totally unimodular and U contains a c0-Eulerian submatrix with no column of zeroes. Then there exists at least one fibre $\mathcal{F}_{A,\boldsymbol{y}}$ for which the columns of U do not form an augmenting Markov sub-basis.*

Example 10. Recall that the PLB $\mathcal{M}_U$ in Example 8 contains the c0-Eulerian submatrix E_3. Changing the sign of column 4 of $\mathsf{sgn}(U) = U$ generates the cr0-Eulerian submatrix

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

when Theorem 4.6 applies to prove that $\mathcal{M}_U$ is not an augmenting Markov sub-basis for some fibre. We saw an example of such a fibre in Example 5. Alternatively, note that the underlying configuration matrix A (from Example 2) is totally unimodular, so that we may cite Corollary 4.7 to draw the same conclusion.

Corollary 4.7 is of some practical significance for network tomography. In particular, Hazelton and Bilton (2017) showed that the link-path incidence matrix A is totally unimodular for linear networks (like sections of highway) and star networks (where we observe entry/exit counts for every origin and destination of travel). This result in tandem with Theorem 4.4 provide a checkable if and only if condition for a PLB to guarantee that augmenting paths are available to the sampler.

5. Practical choice of basis

In practice we will be faced with a particular fibre $\mathcal{F}_{A,\boldsymbol{y}}$ from which we wish to sample. We would like to choose a set of basic moves $\mathcal{M}$ so that the sampler $\mathcal{H}(\mathcal{M})$ mixes rapidly. Let us suppose that there are multiple PLBs for which the c0-Eulerian condition of Theorem 4.4 holds. All of these will be c-minimal augmenting Markov sub-bases for the fibre. It follows that $\mathcal{H}(\mathcal{M}_U)$ will mix rapidly for uniform sampling of $\mathcal{F}_{A,\boldsymbol{y}}$. However, $f(\boldsymbol{x} \mid \boldsymbol{y})$ will rarely be uniform in the kinds of applications of linear inverse problems that we have described. Far more commonly, $x_1, \ldots, x_n$ will have independent Poisson distributions, when this posterior distribution will be the product of Poisson distributions truncated to the fibre. We denote this distribution by $p_{\mathcal{F}}(\boldsymbol{x})$ for fibre $\mathcal{F}$. It is of considerable interest to develop methods for deciding which of the available bases will lead to uniformly rapid mixing for this model.

Recall from Definition 3.2 that a sampler mixes uniformly rapidly for a class of distributions if its second largest eigenvalue modulus is bounded away from one uniformly over that class. Let $\boldsymbol{\theta} = \mathsf{E}[\boldsymbol{x}]$ for the Poisson model. If $\boldsymbol{\theta}$ is bounded away from zero in some class of models then all the non-zero transition probabilities for the sampler will be uniformly bounded away from zero. Both the maximum probability loading ρ and path length ℓ will be bounded above, and so uniformly rapid mixing is guaranteed.

As we saw in Example 6, problems arise when some of the Poisson means can be very small. In such circumstances some of the non-zero transition probabilities will be arbitrarily close to zero. This is not a cause for concern if the corresponding transitions are only required to reach elements of the fibre with correspondingly small posterior probability, nor is it a concern if there is an alternative transitional path with much higher probability. However, mixing will be adversely affected should some move that is an essential step on multiple transitional pathways have low probability. In the context of PLBs, this will occur if steps in at least one direction along the ray $\mathcal{R}_{\mathcal{F},\boldsymbol{u}}(\boldsymbol{x})$ have low probability for some $\boldsymbol{u} \in \mathcal{M}_U$. In more detail, suppose that θ_i is small for $1 \le i \le n$, and that $u_i < 0$ (say). Then x_i will likely be zero, when it follows from equations (3) and (4) that a move with positive step size b is not available.

The implications for mixing of $\mathcal{H}(\mathcal{M}_U)$ when sampling from a truncated Poisson model are formalized in the following theorem.

Theorem 5.1. *Consider a sampler $\mathcal{H}(\mathcal{M}_U)$ and a finite set $\mathcal{F}$ connected by a PLB $\mathcal{M}_U$.*

(i) *Given $\eta > 0$, let $\Theta = \{\boldsymbol{\theta} \colon \boldsymbol{\theta} \ge \eta\}$. Then $\mathcal{H}(\mathcal{M}_U)$ mixes uniformly rapidly for $\mathcal{D}_\Theta = \{p_{\mathcal{F}}(\boldsymbol{x} \mid \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}$.*

(ii) *Given $\eta > 0$, fix $\boldsymbol{\theta}^0 \ge \eta\mathbf{1}$ and $1 \le j \le n$. Define $\Theta = \{\boldsymbol{\theta}^m \colon m = 1,2,3\ldots\}$ such that $\theta_j^m = \theta_j^0/m$ and $\theta_i^m = \theta_i^0$ otherwise. Then $\mathcal{H}(\mathcal{M}_U)$ may not mix uniformly rapidly for $\mathcal{D}_\Theta = \{p_{\mathcal{F}}(\boldsymbol{x} \mid \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}$.*

(iii) *Given $\eta > 0$, fix $\boldsymbol{\theta}^0 \ge \eta\mathbf{1}$. Define $\Theta = \{\boldsymbol{\theta}^m \colon m = 1,2,3\ldots\}$ such that $\theta_i^m = \theta_i^0/m$ for $i = 1,\ldots,n$ and $\theta_i^m = \theta_i^0$ for $i = n+1,\ldots,r$. Assume that there exists $\boldsymbol{x} \in \mathcal{F}$ such that $x_1 = \cdots = x_n = 0$. Then $\mathcal{H}(\mathcal{M}_U)$ does not mix uniformly rapidly for $\mathcal{D}_\Theta = \{p_{\mathcal{F}}(\boldsymbol{x} \mid \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}$.*

Example 11. Continuing on with the network tomography problem introduced in Example 2, consider employing a Gibbs hit-and-run sampler $\mathcal{H}(\mathcal{M}_U)$ implemented using the augmenting Markov sub-basis from Example 9. Suppose the observed counts are $\boldsymbol{y} = (4,4,2,2)^T$, producing a fibre $\mathcal{F}_{A,\boldsymbol{y}}$ of size 55. We examine uniform sampling from $\mathcal{F}_{A,\boldsymbol{y}}$, and sampling when the underlying model for $\boldsymbol{x}$ is independent Poisson with $\mathsf{E}[\boldsymbol{x}] = \boldsymbol{\theta} = (1,1,1,1,\theta_5,1,1,1)^\mathsf{T}$. The second and third columns of U in (9) both have a -1 entry in position 5, and so the size of θ_5 controls the ease with which we may sample in directions $-\boldsymbol{u}_2$ and $-\boldsymbol{u}_3$.

We compute the second largest eigenvalue modulus λ from the transition matrices of the sampler for a range of values of θ_5 between 0 and 1. The corresponding values of the spectral gap $1 - \lambda$ are plotted

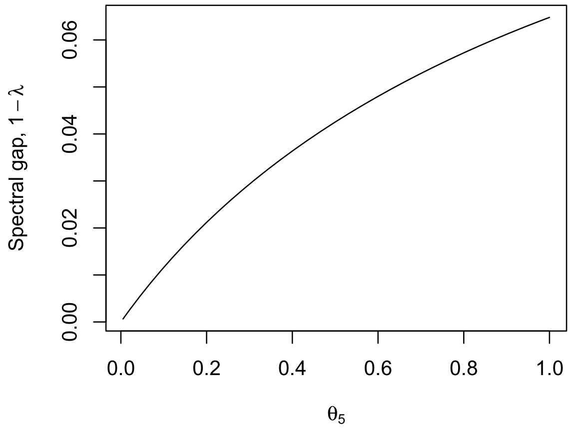

Figure 3. Plot of the spectral gap as a function of θ_5 when sampling from a product of independent Poisson distributions with means $\boldsymbol{\theta} = (1, 1, 1, 1, \theta_5, 1, 1, 1)$ truncated to the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ with $\boldsymbol{y} = (4, 4, 2, 2)^T$. For comparison, the spectral gap for uniform sampling is $1 - \lambda = 0.120$.

in Figure 3. Despite the fact that we are using an augmenting Markov basis, mixing times become very large when θ_5 is small.

Consider these results in the context of the bound presented in (5). Since we are using an augmenting Markov basis, the maximum canonical path length ℓ is no greater than $r - n = 4$. The differences in mixing times are driven instead by the maximum edge loading, ρ. Using augmenting paths, for uniform sampling $\rho = 4.093$ with the move $\boldsymbol{u}_1$ being the most overloaded transition. It is the same move that is most overloaded for Poisson sampling with $\theta_5 = 1$, when $\rho = 202.3$. For $\theta_5 = 0.1$ and $\theta_5 = 0.01$ we obtain respectively $\rho = 14536$ and $\rho = 1423382$. In both cases the most overloaded transition is $-\boldsymbol{u}_2$.

For sampling for Poisson models, Theorem 5.1 indicates that when faced with a choice of partition lattice bases satisfying the c0-Eulerian condition of Theorem 4.4, we should if possible select one for which the columns of the configuration matrix forming A_1 correspond to components of $\boldsymbol{x}$ that do not have small means. Indeed, intuitively we may want to choose a partition for which we select the components with largest means to form A_1, subject to the requirement the columns of the resulting matrix U form an augmenting Markov basis.

For sampling in mark-recapture models, these comments suggest that we partition A so that the columns forming A_1 correspond to histories without misidentification, since we expect these to be the most probable members of the fibre. What is more, this will lead to a partition in which A_1 is the identity matrix, as in Example 1. The corresponding PLB will then be a c-minimal augmenting Markov sub-basis by Corollary 4.5, and will mix uniformly rapidly when employing natural Poisson models.

Such a strategy will not necessarily work well for network tomography problems. Consider for example a network in which every node is both an origin and destination of travel, and we have counts on every link of the network. In such a circumstance we can again partition the configuration matrix such that A_1 is the identity. However, the corresponding entries of $\boldsymbol{x}$ will be flows between neighbouring nodes. In many applications in road transport engineering these trips will be unusually short, and so will have small mean traffic flows. An alternative partition of A may lead to much better mixing for the sampler.

6. Highway data example

To illustrate we consider volume network tomography for a section of New Zealand's State Highway 16 running through the city of Auckland in New Zealand. An abstracted form of the network is displayed in Figure 4. Drivers may join the highway at any of the first 4 nodes (representing locations of on-ramps); and may leave the network at any node from 3 to 8 (representing locations of off-ramps). Traffic counts on the 7 network links are available over a 15 minute period: $\boldsymbol{y} = (2991, 3352, 3977, 4576, 3849, 2458, 978)^{\mathsf{T}}$. The traffic counts for the 21 feasible origin-destination pairs are collected in the vector $\boldsymbol{x}$ (ordered lexicographically). We adopt an independent Poisson model for the components of $\boldsymbol{x}$ with mean $\mathsf{E}[\boldsymbol{x}] = \boldsymbol{\theta}$. In this example we will sample from $f(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is based on an existing estimate. Its entries vary in size from 33.0 to 825.0, with the exception of $\theta_7 = 0.5$. This last value corresponds to origin-destination pair $(2, 3)$, which in reality is an implausibly short journey for which to join the highway.

The link path incidence matrix is displayed below:

$$A = \begin{bmatrix} 1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0 \\ 0&1&1&1&1&1&0&1&1&1&1&1&1&1&1&1&1&0&0&0&0 \\ 0&0&1&1&1&1&0&0&1&1&1&1&0&1&1&1&1&1&1&1&1 \\ 0&0&0&1&1&1&0&0&0&1&1&1&0&0&1&1&1&0&1&1&1 \\ 0&0&0&0&1&1&0&0&0&0&1&1&0&0&0&1&1&0&0&1&1 \\ 0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&1&0&0&0&1 \end{bmatrix}. \tag{10}$$

Columns 1, 7, 13, 18, 19, 20, and 21 correspond to short paths through the network. If the matrix is partitioned so that those columns form A_1, then it is straightforward to check any other column of the matrix can be expressed as a linear combination of the columns of A_1 with non-negative coefficients. Corollary 4.5 therefore applies, so that the corresponding PLB is a c-minimal augmenting Markov sub-basis for this problem. This basis is also the Markov basis found by 4ti2. Nevertheless, A has 18 further partitions generating PLBs that can also be shown to be c-minimal augmenting Markov sub-bases courtesy of Theorem 4.4.

Recall that θ_7 is small. Following Theorem 5.1 and ensuing discussion, we expect a hit-and-run sampler to mix better if we employ a PLB for which the underlying A_1 does not contain the 7th column of A. This is the case for two of the 18 PLBs just described. In particular, we can form a PLB using a partition in which A_1 comprises columns 1, 8, 13, 19, 20, and 21.

We ran hit-and-run samplers using each of these bases, which we label respectively PLB1 and PLB2. These methods were implemented in R (R Core Team, 2022); the computing code is supplied as supplementary material (Hazelton et al., 2024). In each case we allowed 10^6 iterations. Effective sample sizes for each component of $\boldsymbol{x}$ are displayed using a barplot (with logarithmic y-scale) in Figure 5. For the first seven and last eight components of $\boldsymbol{x}$ there is relatively little difference between the performance using the two bases. However, for components 8–12 of $\boldsymbol{x}$ the effective sample sizes using PLB2

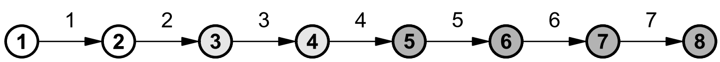

Figure 4. Abstraction of a section of highway network in Auckland, New Zealand. White nodes are origins of traffic flow (on-ramps); light grey nodes are both origins and destinations of traffic (both on- and off- ramps); and dark grey nodes are solely destinations of travel (off-ramps). Link lengths are not to scale: in reality, the separation between nodes 2 and 3 is much smaller than other inter-node distances.

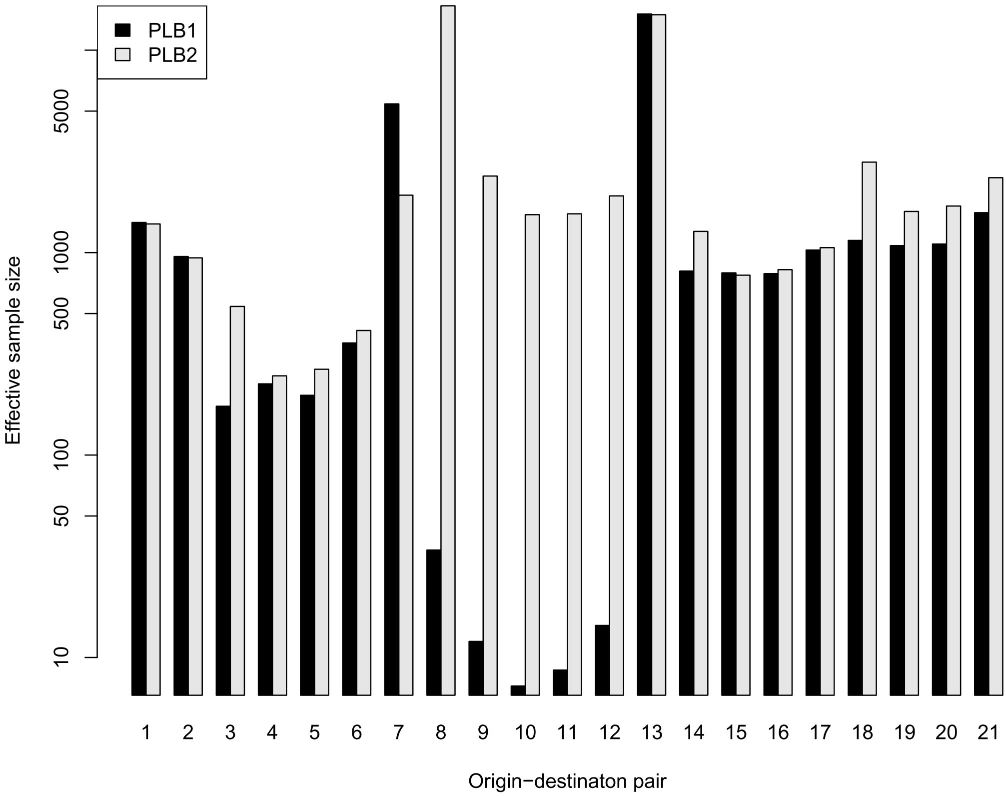

Figure 5. Results of volume tomography for the Auckland highway network. The barplot displays the effective sample size for each component of $\boldsymbol{x}$ (origin-destination traffic volumes) for hit-and-run samplers using two PLBs.

is orders of magnitude better than those obtained using PLB1. Updates to these components are slowed by a transition bottleneck relating to basis vectors with a ± 1 entry in position 7.

7. Conclusions

Markov chain Monte Carlo methods are only practical if the sampler employed mixes sufficiently rapidly. Identification of poor mixing behaviour is usually relatively straightforward using standard diagnostic tools (e.g. Brooks and Roberts, 1998). However, judging the number of iterations required before commencement, or modifying an existing sampler to improve the convergence rate, are more challenging issues. For fibre sampling in linear inverse problems for count data there is little existing theory in the literature to provide guidance on these matters.

Full Markov bases for fibre sampling can be very large. For example, the minimal Markov basis found by `4ti2` for network tomography for the Abilene internet network (Fang, Vardi and Zhang, 2007) contains 10,705 vectors, even though the system contains only 12 nodes and the configuration matrix is of dimension 42×144. As a second example, the minimal Markov basis for resampling a $4 \times 4 \times 4$ contingency table conditional on 2-marginals (for which the configuration matrix is of dimension 37×64) contains 148,968 vectors (Hemmecke and Malkin, 2005). In practical MCMC runs it is chastening to recognize that each sampling direction in such large bases may be used only a few times, or perhaps not at all. There are evidently challenges in developing theory that can provide useful practical guidance in such circumstances.

Nevertheless, there are a variety of important types of linear inverse problem for which small Markov sub-bases, matching the dimension of the fibre, are available. These include models for accounting for identification errors in mark-recapture data, and network tomography for highway and transit systems. If the sampling directions available in such a small basis are sufficient to ensure the existence of augmenting sample paths between all pairs of points on a fibre, then rapid mixing is guaranteed under uniform sampling. In this article we have identified and studied a class of lattice bases with this property when the matrix U computed from (8) is integral and contains no c0-Eulerian submatrices. This condition can be checked fairly quickly for moderately-sized problems in practice.

We have also examined mixing for Poisson models. Our Theorem 5.1 provides clear practical guidance into choice of basis. In particular, when partitioning A so as to form our basis we want to ensure that the columns forming A_1 correspond to coordinates of $\boldsymbol{x}$ that are likely to take relatively large values. This ties in naturally with an analysis of the limits on step lengths provided by equations (3) and (4). Specifically step sizes are constrained by the magnitude of those aforementioned coordinates. Geometrically, this can be thought of in terms of rotating the polytope encapsulating the fibre so that its longer edges are aligned to the coordinate directions corresponding to the columns of A_2.

Appendix: Proofs of theorems

A.1. Proof of Theorem 4.4

Lemma A.1. *A matrix $W \in \{0, \pm 1\}^{n\times r}$ that contains a cr0-Eulerian submatrix cannot have its columns reordered so that the* 1*s all precede the* -1*s in every row.*

Proof. The proof is by contradiction. Let W be a matrix containing a cr0-Eulerian submatrix, V. Suppose that we can reorder the columns of W such that the 1s all precede the -1s in every row.

Under this reordering, let $\boldsymbol{v}_{\cdot j}$ be the first column of V that contains a non-zero entry. The entries of $\boldsymbol{v}_{\cdot j}$ sum to zero, so it must contain at least one 1 and one -1.

Let i be an index such that $\boldsymbol{v}_{i,j} = -1$. The entries of the ith row also sum to zero, so the ith row must also contain a 1.

But $\boldsymbol{v}_{\cdot j}$ 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 W, it also has a row with a -1 that precedes a 1. This contradicts our hypothesis, and so W cannot be ordered so the the 1s precede the -1s in every row. □

We now prove the converse.

Lemma A.2. *A matrix $W \in \{0, \pm 1\}^{n\times r}$ that cannot have its columns reordered so that the* 1*s all precede the* -1*s in every row must contain a cr0-Eulerian submatrix.*

Proof. Let $<$ over $\{0, \pm 1\}^r$ define a binary relation such that $\boldsymbol{u} < \boldsymbol{v}$ if $u_k = 1$ and $v_k = -1$ for some $k \in \{1, \dots, n\}$. Let $\prec$ be the transitive closure of $<$. It is therefore a binary relation over $\{0, \pm 1\}^r$ where $\boldsymbol{u} \prec \boldsymbol{v}$ if $\boldsymbol{u} < \boldsymbol{v}$ or if $\boldsymbol{u} < \cdots < \boldsymbol{v}$.

If $\prec$ is a strict partial order over $\mathcal{W}$, the set of columns of W, then any reordering of $\mathcal{W}$ 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 $\mathcal{W}$, 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 $\mathcal{W}$. 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 $\boldsymbol{w}_{\cdot j_1} \prec \boldsymbol{w}_{\cdot j_1}$, then either $\boldsymbol{w}_{\cdot j_1} < \boldsymbol{w}_{\cdot j_1}$, or $\boldsymbol{w}_{\cdot j_1} < \cdots < \boldsymbol{w}_{\cdot j_m} < \boldsymbol{w}_{\cdot j_1}$ for some $m \in \mathbb{Z}^+$ and some indices $j_1, \ldots, j_m$. It cannot be $\boldsymbol{w}_{\cdot j_1} < \boldsymbol{w}_{\cdot j_1}$, since that would imply that there is an index k such that $\boldsymbol{w}_{k,j_1} = 1$ and $\boldsymbol{w}_{k,j_1} = -1$, which is a contradiction. Therefore the second condition must be true, and we choose $\{\boldsymbol{w}_{\cdot j_1}, \ldots, \boldsymbol{w}_{\cdot j_m}\}$ such that m is the minimum over all such sequences. 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 $\boldsymbol{w}_{\cdot j_1} < \cdots < \boldsymbol{w}_{\cdot j_p} < \boldsymbol{w}_{\cdot j_{q+1}} < \cdots < \boldsymbol{w}_{\cdot j_m} < \boldsymbol{w}_{\cdot j_1}$ and m would not be minimal.

From the definition of $<$, for every $k \in \{1, \ldots, m\}$ there is i_k such that $\boldsymbol{w}_{i_k, j_k} = 1$ and $\boldsymbol{w}_{i_k, j_{k+1}} = -1$. 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 $\boldsymbol{w}_{i_p, j_{p-1}} = 1$ and $\boldsymbol{w}_{i_p, j_p} = -1$, and $\boldsymbol{w}_{i_q, j_{q-1}} = 1$, and $\boldsymbol{w}_{i_q, j_q} = -1$. Then we would have $\boldsymbol{w}_{\cdot j_1} < \cdots < \boldsymbol{w}_{\cdot j_{p-1}} < \boldsymbol{w}_{\cdot j_q} < \cdots < \boldsymbol{w}_{\cdot j_m} < \boldsymbol{w}_{\cdot j_1}$ and m would not be minimal.

We construct a submatrix V of W by taking the i_kth rows and j_kth columns of W for $k \in \{1, \ldots, m\}$. We claim that the entries of this matrix not already defined by $\boldsymbol{v}_{i_k, j_k} = 1$ and $\boldsymbol{v}_{i_k, j_{k+1}} = -1$ are all 0. To the contrary, if one of these entries $\boldsymbol{v}_{i_q, j_q} = 1$ where $p \neq q$ and $p + 1 \neq q$, then $\boldsymbol{w}_{\cdot j_q} < \boldsymbol{w}_{\cdot j_{p+1}}$, which if $q < p$ produces $\boldsymbol{w}_{\cdot j_1} < \cdots < \boldsymbol{w}_{\cdot j_q} < \boldsymbol{w}_{\cdot j_{p+1}} < \cdots < \boldsymbol{w}_{\cdot i_m} < \boldsymbol{w}_{\cdot i_1}$; or if $p < q$ produces $\boldsymbol{w}_{\cdot j_q} < \cdots < \boldsymbol{w}_{\cdot j_{p+1}} < \boldsymbol{w}_{\cdot j_q}$ and m was not minimal. On the other hand, if one of these entries $\boldsymbol{v}_{i_q, j_q} = -1$ where $p \neq q$ and $p + 1 \neq q$, then $\boldsymbol{w}_{\cdot j_p} < \boldsymbol{w}_{\cdot j_q}$, which if $p < q$ produces $\boldsymbol{w}_{\cdot j_1} < \cdots < \boldsymbol{w}_{\cdot j_p} < \boldsymbol{w}_{\cdot j_q} < \cdots < \boldsymbol{w}_{\cdot j_m} < \boldsymbol{w}_{\cdot j_1}$, and m was not minimal; or if $q < p$ produces $\boldsymbol{w}_{\cdot j_q} < \cdots < \boldsymbol{w}_{\cdot j_p} < \boldsymbol{w}_{\cdot j_q}$, 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 W, so W contains a cr0-Eulerian submatrix. $\square$

Lemmata A.1 and A.2 are combined into the following theorem.

Theorem A.3. *A matrix with $\{0, \pm 1\}$ entries cannot have its columns reordered so that the 1s all precede the -1s in every row if and only if it contains a cr0-Eulerian submatrix.*

Having established this result, we now establish which column PLB-defining matrices it is applicable to.

Theorem A.4. *Let $W \in \{0, \pm 1\}^{n \times r}$ be a matrix that contains no c0-Eulerian submatrix. For all $\sigma = (\sigma_1, \ldots, \sigma_r) \in \{-1, 1\}^r$, define W^σ to be the matrix obtained by multiplying the ith column of W by σ_i for $i = 1, \ldots, r$. Then W^σ contains no cr0-Eulerian submatrix.*

Proof. Suppose that the theorem is false, and we have a matrix W^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 W^σ was constructed by multiplying columns of W by ± 1, the original W can be found by performing the same multiplications on W^σ, and W 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 W_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 W 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. □

We are now in a position to prove Theorem 4.4. To that end, let $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$. Since the columns of U form a lattice basis for $\ker_{\mathbb{Z}}(A)$ we can write

$$\boldsymbol{x}' - \boldsymbol{x} = \sum_{i=1}^{r-n} \epsilon_i a_i \boldsymbol{u}_i,$$

where $\epsilon_i \in \{\pm 1\}$ gives the sign and $a_i \in \mathbb{Z}_{\geq \mathbf{0}}$ gives the number of copies of $\boldsymbol{u}_i$ 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

$$\boldsymbol{x}' - \boldsymbol{x} = \sum_{i=1}^{r-n} a_i \boldsymbol{u}_i^\epsilon \tag{11}$$

without a sign. From Theorem A.4, this matrix $\mathsf{sgn}(U^\epsilon)$ contains no cr0-Eulerian submatrix. By Theorem A.3, we can reorder the columns of $\mathsf{sgn}(U^\epsilon)$ such that no -1 precedes a 1 in any row. This same reordering ensures that no negative entry in U^ϵ precedes a positive entry in any row.

This ordering of the columns of U gives the order in which the moves should be applied when conducting the walk from $\boldsymbol{x}$ to $\boldsymbol{x}'$. The final $r-n$ rows of U^ϵ each contain just a single non-zero entry, and so the corresponding coordinates of $\boldsymbol{x} + \sum_{i=1}^{s} a_i \boldsymbol{u}_i^\epsilon$ can never be negative for any $1 \leq s \leq r-n$. Moreover, a decrease in any of the first n coordinates comes after all increases in that coordinate, and so these coordinates cannot become negative since $\boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$. Since each basis vector appears only once in (11), that equation describes an augmenting path from $\boldsymbol{x}$ to $\boldsymbol{x}'$.

This holds for all pairs $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{A,\boldsymbol{y}}$ for all $\boldsymbol{y}$, therefore the columns of U form a Markov basis for A, proving (i). Since we can find an augmenting path to connect any pair of points, the columns of U form an augmenting Markov sub-basis for any fibre. Finally, this Markov sub-basis is evidently c-minimal for any fibre of dimension $r-n$, proving (ii).

A.2. Proof of the remaining results

Proof of Corollary 4.5. The columns of A_1 span the column space of A. As a consequence, A_1 is invertible. Now, $A_1 C = A_2$ where $C = A_1^{-1} A_2$ is a non-negative integer matrix. Moreover, since A_1 and A_2 are binary matrices, C much also be a binary matrix. It follows that the columns of U form a PLB, and also that top n rows of $\mathsf{sgn}(U)$ are 0 or -1. There can be no Eulerian submatrices involving any of the bottom $r-n$ rows of U (because this is the identity matrix). Moreover, there can be no c0-Eulerian submatrices of the top n rows on $\mathsf{sgn}(U)$, since there can be no cancellations to give zero column sums. The result therefore follows from Theorem 4.4. □

Proof of Theorem 4.6. Let $\mathcal{I}$ and $\mathcal{J}$ respectively index the rows and columns of U forming the cr0-Eulerian submatrix. Define $\boldsymbol{w} = \sum_{j \in \mathcal{J}} \boldsymbol{u}_i$. Also, define $\boldsymbol{w}^+$ and $\boldsymbol{w}^-$ such that $w_i^+ = \max\{w_i, 0\}$ and $w_i^- = -\min\{w_i, 0\}$ for all $i = 1, \ldots, r$. Then both $\boldsymbol{w}^+$ and $\boldsymbol{w}^-$ are on the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ where $\boldsymbol{y} = A\boldsymbol{w}^+$, since $\boldsymbol{w} \in \ker_{\mathbb{Z}}(A)$. Consider the walk from $\boldsymbol{w}^-$ to $\boldsymbol{w}^+$. Because the row sums of the submatrix are zero, $\boldsymbol{w}^-$ is zero in all components indexed by $\mathcal{I}$. However, the zero column sum property means that each vector in the set $\{\boldsymbol{u}_j : j \in \mathcal{J}\}$ has a negative entry in at least one of the rows indexed by $\mathcal{I}$. As a consequence none of those vectors can form the first step on the walk. Any walk on the fibre $\mathcal{F}_{A,\boldsymbol{y}}$ must then involve a diversion, and any column of $\boldsymbol{u}_j$ with $j \notin \mathcal{J}$ must appear with both positive and negative signs so as to cancel out by the end of the walk. Such a walk is not augmenting, completing the proof. □

Proof of Corollary 4.7. If $A = [A_1 \mid A_2]$ is totally unimodular then $A_1^{-1}A_2$ is evidently also totally unimodular, and therefore U has the same property.

Let U^* denote the c0-Eulerian submatrix. This is totally unimodular. Following Ghouila-Houri (1962), each column of U^* can be assigned a multiplier $\epsilon = \pm 1$ such that the sum of the signed columns $\sum_i (\epsilon_i \boldsymbol{u}_i^*) = \boldsymbol{v}$ is a vector with all components in $\{0, \pm 1\}$. Since U^* is Eulerian, all entries of $\boldsymbol{v}$ are even. Therefore $\boldsymbol{v}$ can only be the zero vector. The multipliers $\{\epsilon_i\}$ introduced previously give the required change of column signs to turn U^* from a c0-Eulerian matrix into a cr0-Eulerian matrix. The result then follows from Theorem 4.6. □

Proof of Theorem 5.1. Throughout this proof we focus on the second largest eigenvalue of the transition matrix, λ_2, courtesy of the standard arguments described in Section 3.

(i) Recall from (6),

$$\lambda_2(\boldsymbol{\theta}) \leq 1 - \frac{1}{\rho \ell}.$$

Since $\mathcal{F}$ is finite, the maximum canonical path length ℓ is bounded above. Also, $\inf_{e \in E} Q(e)$ is bounded away from zero for all $\boldsymbol{\theta} \in \Theta$ since $p_{\mathcal{F}}$ is bounded away from zero on $\mathcal{D}_\Theta$. Hence ρ is bounded above, and so $\lambda_2(\boldsymbol{\theta}) \leq 1 - \epsilon$ for some $\epsilon > 0$.

(ii) By Theorem 2 of Sinclair (1992) the second largest eigenvalue is bounded below by $1 - 2\Phi \leq \lambda_2(\boldsymbol{\theta})$, where Φ is the bottleneck ratio (or conductance of the graphical representation) of the sampler. Consider Example 2 where the configuration matrix is reordered as

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

The corresponding PLB is given by the columns of

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

Consider $\boldsymbol{x} = (0,0,0,0,0,2,0,0)^{\mathsf{T}}$ which lies on the fibre $\mathcal{F}_{\boldsymbol{y}} = \{\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0} : \boldsymbol{y} = A\boldsymbol{x}\}$ for $\boldsymbol{y} = (2,2,2,0)^{\mathsf{T}}$. The sampler is connected over this fibre, since U has no c0-Eulerian sub-matrices and its columns therefore form a Markov basis. Take $\boldsymbol{\theta}^0 = \eta \mathbf{1}$. Then $p_{\mathcal{F}}(\boldsymbol{x} \mid \boldsymbol{\theta}^m)$ is bounded away from 0 and 1 for all m. Define $Z = \inf\{\mathbb{P}(\boldsymbol{y} \mid \boldsymbol{\theta}^m) : m = 1,2,\ldots\}$, and observe that $Z > 0$. Any transition away from $\boldsymbol{x}$ requires an increase in x_1, the probability of which is bounded above by

$\mathbb{P}(x_1 = 1 \mid \boldsymbol{\theta}^m, \boldsymbol{y}) \leq \mathbb{P}(x_1 \mid \boldsymbol{\theta}^m)/Z \leq Z^{-1}\eta/m$. It follows that $\Phi \leq Z^{-1}\eta/m$ whence $\lambda_2(\boldsymbol{\theta}^m) \geq 1 - \epsilon$ for sufficiently large m for any $\epsilon > 0$. This provides an instance (with $j = 1$) where the sampler does not mix uniformly rapidly on $\mathcal{D}_\Theta$.

(iii) Take $\boldsymbol{x} \in \mathcal{F}$ with $x_1 = \cdots = x_n = 0$. Then $p_{\mathcal{F}}(\boldsymbol{x} \mid \boldsymbol{\theta}^m) > 0$ for all m. Define $Z = \inf\{\mathbb{P}(\boldsymbol{y} \mid \boldsymbol{\theta}^m)\colon m = 1, 2, \ldots\}$, and observe that $Z > 0$. However, any transition away from $\boldsymbol{x}$ must involve at least one of the first n coordinates of $\boldsymbol{x}$ increasing, the probability of which is bounded above by Z^{-1}/m. It follows that the bottleneck ratio satisfies $\Phi \leq Z^{-1}/m$, whence $\lambda_{\boldsymbol{\theta}^m} \geq 1 - \epsilon$ for sufficiently large m for any $\epsilon > 0$. The sampler therefore does not mix uniformly rapidly on $\mathcal{D}_\Theta$. □

Acknowledgements

The authors acknowledge financial support from Marsden Grant MAU1734/UOO1734, administered by the Royal Society of New Zealand. The first author also acknowledges financial support from Marsden Grant UOO2024. The authors thank the reviewers for their helpful comments.

Supplementary Material

SM-code.r (DOI: 10.3150/23-BEJ1690SUPP; .zip). R code to implement the numerical examples.

References

4ti2team (2015). 4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces (version 1.6.7). Available at www.4ti2.de.

Airoldi, E.M. and Blocker, A.W. (2013). Estimating latent processes on a network from indirect measurements. *J. Amer. Statist. Assoc.* **108** 149–164. MR3174609 https://doi.org/10.1080/01621459.2012.756328

Airoldi, E.M. and Haas, B. (2011). Polytope samplers for inference in ill-posed inverse problems. In *International Conference on Artificial Intelligence and Statistics* **15** 110–118.

Aoki, S., Hara, H. and Takemura, A. (2012). *Markov Bases in Algebraic Statistics*. *Springer Series in Statistics*. New York: Springer. MR2961912 https://doi.org/10.1007/978-1-4614-3719-2

Baumert, S., Ghate, A., Kiatsupaibul, S., Shen, Y., Smith, R.L. and Zabinsky, Z.B. (2009). Discrete hit-and-run for sampling points from arbitrary distributions over subsets of integer hyperrectangles. *Oper. Res.* **57** 727–739. MR2554262 https://doi.org/10.1287/opre.1080.0600

Brooks, S.P. and Roberts, G.O. (1998). Diagnosing convergence of Markov chain Monte Carlo algorithms. *Stat. Comput.* **8** 319–335.

Castro, R., Coates, M., Liang, G., Nowak, R. and Yu, B. (2004). Network tomography: Recent developments. *Statist. Sci.* **19** 499–517. MR2185628 https://doi.org/10.1214/088342304000000422

Chen, Y., Dinwoodie, I.H. and Sullivant, S. (2006). Sequential importance sampling for multiway tables. *Ann. Statist.* **34** 523–545. MR2275252 https://doi.org/10.1214/009053605000000822

De Loera, J.A., Hemmecke, R. and Lee, J. (2015). On augmentation algorithms for linear and integer-linear programming: From Edmonds-Karp to Bland and beyond. *SIAM J. Optim.* **25** 2494–2511. MR3429745 https://doi.org/10.1137/151002915

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156 https://doi.org/10.1214/aos/1030563990

Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli* **9** 1093–1108. MR2046819 https://doi.org/10.3150/bj/1072215202

Fang, J., Vardi, Y. and Zhang, C.-H. (2007). An iterative tomogravity algorithm for the estimation of network traffic. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 12–23. Beachwood, OH: IMS. MR2459176 https://doi.org/10.1214/074921707000000030

Ghouila-Houri, A. (1962). Caractérisation des matrices totalement unimodulaires. *C. R. Acad. Sci. Paris* **254** 1192–1194. MR0132752

Hazelton, M.L. (2015). Network tomography for integer-valued traffic. *Ann. Appl. Stat.* **9** 474–506. MR3341124 https://doi.org/10.1214/15-AOAS805

Hazelton, M.L. and Bilton, T.P. (2017). Polytope samplers for network tomography. *Aust. N. Z. J. Stat.* **59** 495–511. MR3760156 https://doi.org/10.1111/anzs.12216

Hazelton, M.L., McVeagh, M.R. and van Brunt, B. (2021). Geometrically aware dynamic Markov bases for statistical linear inverse problems. *Biometrika* **108** 609–626. MR4298767 https://doi.org/10.1093/biomet/asaa083

Hazelton, M., McVeagh, M., Tuffley, C. and van Brunt, B. (2024). Supplement to "Some rapidly mixing hit-and-run samplers for latent counts in linear inverse problems." https://doi.org/10.3150/23-BEJ1690SUPP

Hemmecke, R. and Malkin, P. (2005). Computing generating sets of lattice ideals. arXiv preprint arXiv:math/0508359.

Kaipio, J. and Somersalo, E. (2006). *Statistical and Computational Inverse Problems* **160**. Berlin: Springer.

Levin, D.A., Peres, Y. and Wilmer, E.L. (2009). *Markov Chains and Mixing Times* **107**. Providence: American Mathematical Soc.

Link, W.A., Yoshizaki, J., Bailey, L.L. and Pollock, K.H. (2010). Uncovering a latent multinomial: Analysis of mark-recapture data with misidentification. *Biometrics* **66** 178–185. MR2756704 https://doi.org/10.1111/j.1541-0420.2009.01244.x

Rapallo, F. (2003). Algebraic Markov bases and MCMC for two-way contingency tables. *Scand. J. Stat.* **30** 385–397. MR1983132 https://doi.org/10.1111/1467-9469.00337

Schofield, M.R. and Bonner, S.J. (2015). Connecting the latent multinomial. *Biometrics* **71** 1070–1080. MR3436732 https://doi.org/10.1111/biom.12333

Sinclair, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1** 351–370. MR1211324 https://doi.org/10.1017/S0963548300000390

Smith, R.L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.* **32** 1296–1308. MR0775260 https://doi.org/10.1287/opre.32.6.1296

Stanley, C. and Windisch, T. (2018). Heat-bath random walks with Markov bases. *Adv. in Appl. Math.* **92** 122–143. MR3699116 https://doi.org/10.1016/j.aam.2017.08.002

Takemura, A. and Aoki, S. (2004). Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. *Ann. Inst. Statist. Math.* **56** 1–17. MR2053726 https://doi.org/10.1007/BF02530522

R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Tebaldi, C. and West, M. (1998). Bayesian inference on network traffic using link count data (with discussion). *J. Amer. Statist. Assoc.* **93** 557–576. MR1631325 https://doi.org/10.2307/2670105

Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *J. Amer. Statist. Assoc.* **91** 365–377. MR1394093 https://doi.org/10.2307/2291416

Windisch, T. (2016). Rapid mixing and Markov bases. *SIAM J. Discrete Math.* **30** 2130–2145. MR3573306 https://doi.org/10.1137/15M1022045

Received November 2022 and revised October 2023