Efficient Fibre Sampling for Statistical Linear Inverse Problems

Martin Hazelton[1]

Department of Mathematics & Statistics
University of Otago

10 December 2024

[1] Email: martin.hazelton@otago.ac.nz

Statistical Inverse Problems

In the age of automated data collection

- Interest is in a process that is observed only indirectly
- Observations provide incomplete information about target variable
  - Aggregated data
  - Summarised data
  - Corrupted data
- Leads to statistical inverse problems
- Problems of this sort are ubiquitous in science and engineering
- Often arise from automated data collection

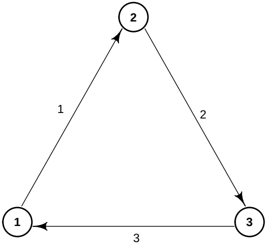
Linear Inverse Problems for Count Data

- For count data, statistical linear inverse problems characterised by
$$\boldsymbol{y} = A\boldsymbol{x} \quad (1)$$
  - $\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0}$ is count vector of interest;
  - $\boldsymbol{y} \in \mathbb{Z}^n_{\geq 0}$ is vector of observed counts.
  - **Configuration matrix** A is $n \times r$ and non-negative integer elements (often binary).
- Typically $r > n$ so linear system (1) will be (heavily) underdetermined.
- Aim is to perform inference for $\boldsymbol{x}$ and/or parameter vector θ describing underlying distribution $f(\boldsymbol{x}|\theta)$.
  - Often prior information or auxiliary data used to regularize problem.

Network Tomography

- $\boldsymbol{x}$ vector path traffic volumes; $\theta = \mathrm{E}[\boldsymbol{x}]$.
- $\boldsymbol{y}$ traffic counts collected at various network locations.
- Inference for $\boldsymbol{x}$ and/or θ is a standard engineering practice:
  - Applications to road traffic and electronic communication systems.

Example



- Assume travel possible between any of $r = 6$ node pairs by direct paths.
- Traffic counts $\boldsymbol{y} = (y_1, y_2, y_3)^{\mathsf{T}}$ observed on $n = 3$ links.
- Collect path volumes in vector $\boldsymbol{x}$.

$$\boldsymbol{y} = A\boldsymbol{x} \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

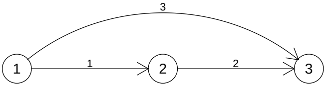
University of Otago
ŌTĀKOU WHAKAIHU WAKA

Network Tomography with Double Counting

Non-binary configuration matrix

- Vehicle counts often collected through inductive loop detectors.
- Moving metal in magnetic field generates current...
- ... but e.g. stop-start motion can result in double counting.

Example



- Nodes 1 and 2 are origins, 2 and 3 are destinations.
- Vehicles on link 3 may be double counted.
- Traffic counts $\boldsymbol{y} = (y_1, y_2, y_3)^{\mathsf{T}}$.

$$\boldsymbol{y} = A\boldsymbol{x} \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Resampling Contingency Tables

- $\boldsymbol{x}$ cell entries in table.
- $\boldsymbol{y}$ marginal totals (or similar).
- Resampling entries $\boldsymbol{x}$ conditional on $\boldsymbol{y}$ can be used to perform exact inference, creating confidentialized cross-tabulations of official statistics, etc.

Example (2 × 3 table)

	y_3	y_4	y_5
y_1	x_1	x_2	x_3
y_2	x_4	x_5	x_6

$$\Rightarrow \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}}_{\boldsymbol{x}}$$

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Resampling Contingency Tables

- $\boldsymbol{x}$ cell entries in table.
- $\boldsymbol{y}$ marginal totals (or similar).
- Resampling entries $\boldsymbol{x}$ conditional on $\boldsymbol{y}$ can be used to perform exact inference, creating confidentialized cross-tabulations of official statistics, etc.

Example (2 × 3 table)

	y_3	y_4	y_5
y_1	x_1	x_2	x_3
y_2	x_4	x_5	x_6

$$\Rightarrow \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}}_{\boldsymbol{x}}$$

Delete redundant row.

Other Applications

- Capture-recapture studies in ecology.
- Multi-list matching problems in public health.
- Biosecurity surveillance.
- Inference for haplotype frequency.

University of Otago
ŌTĀKOU WHAKAIHU WAKA

The Conditional Distribution of $\boldsymbol{x}$

- Inference for $\boldsymbol{x}$ based on conditional distribution $f(\boldsymbol{x}|\,\boldsymbol{y})$.
  - Dependence of f on parameter $\boldsymbol{\theta}$ suppressed for notational convenience.
- Courtesy of fundamental equation $\boldsymbol{y} = A\boldsymbol{x}$,

$$f(\boldsymbol{x}|\,\boldsymbol{y}) = \frac{f(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{x})I_{\{\boldsymbol{y}=A\boldsymbol{x}\}}}{f(\boldsymbol{y})}$$

- Normalizing constant is $f(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\mathcal{F}_{A,\boldsymbol{y}}} f(\boldsymbol{x})$.
- Here $\mathcal{F}_{A,\boldsymbol{y}} = \{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\} \cap \mathbb{Z}^r_{\geq 0}$.
- This is solution set is called the ***y*-fibre**.

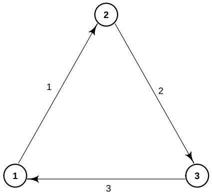
$\mathbb{Z}$-Polytopes

- Continuous version of $\boldsymbol{y}$-fibre is $\{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$.
- This is intersection of linear manifold $\{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\}$ with non-negative orthant $\{\boldsymbol{x} \geq \boldsymbol{0}\}$.
- Hence $\{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$ is a convex polytope.
- Follows that fibre $\mathcal{F}_{A,\boldsymbol{y}} = \{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\} \cap \mathbb{Z}^r_{\geq 0}$ is a **$\mathbb{Z}$-polytope**.
- Assuming A of full rank, then $\mathcal{F}_{A,\boldsymbol{y}}$ is an $r - n$ dimensional object embedded in r-dimensional space.
- Have flexibility in representation.

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Different Projections of a Polytope

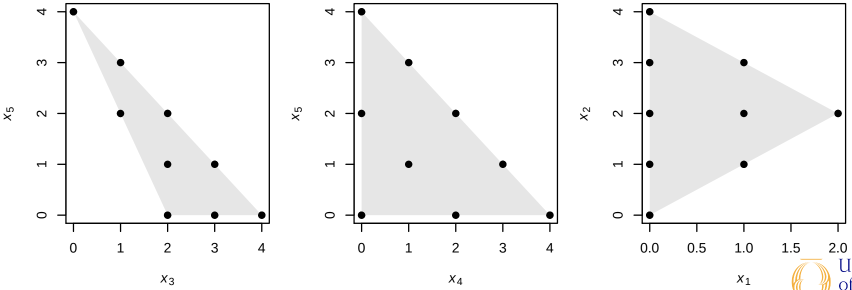
Circuit network example: $r = 5$ and $r - n = 2$



Like earlier example, but last route deleted.

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

Traffic counts $\boldsymbol{y} = (4, 4, 4)^{\mathsf{T}}$ observed.



Inference

- Likelihood is $L(\theta) = f(\boldsymbol{y}|\theta) = \sum_{\boldsymbol{x} \in \mathcal{F}_{A,\boldsymbol{y}}} f(\boldsymbol{x}|\theta)$
- Hence direct resampling of $\boldsymbol{x}$ and likelihood-based inference for θ both require knowledge of $\mathcal{F}_{A,\boldsymbol{y}}$...
- ... but fibres usually far too large to enumerate.

MCMC Based Inference

- **Problem 1:** Resampling $\boldsymbol{x}$ for fixed θ.
  - Applications: contingency table resampling, stochastic EM algorithm
- **Problem 2:** Posterior inference for θ.
  - Sampling $f(\theta|\boldsymbol{x})$ typically straightforward by Gibbs, Metropolis-Hastings algorithms.
  - Iterate sampling from $f(\boldsymbol{x}|\boldsymbol{y},\theta)$ with sampling from $f(\theta|\boldsymbol{x})$.
  - Sampling $f(\boldsymbol{x}|\boldsymbol{y},\theta)$ is challenging step.

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Random Walk $\mathbb{Z}$-Polytope Samplers

Hit and Run Algorithm

- Want to sample $f(\boldsymbol{x}|\boldsymbol{y})$ (parameter dependence suppressed)
- Recall that support of $f(\boldsymbol{x}|\boldsymbol{y})$ is $\mathbb{Z}$-polytope $\mathcal{F}_{A,\boldsymbol{y}}$.
- Will adopt random walk Metropolis-Hastings sampler.

input

Current state $\boldsymbol{x}$

generate candidate $\boldsymbol{x}^\dagger$

Draw $\boldsymbol{z}$ from set $\mathcal{S} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\}$ of possible moves

Draw step size $b \in \mathbb{Z}$

Define candidate $\boldsymbol{x}^\dagger = \boldsymbol{x} + b\boldsymbol{z} \sim q(\cdot|\boldsymbol{x})$

return $\boldsymbol{x}^\dagger$

accept/reject

Compute $\alpha = \mathbf{1}_{\mathcal{F}_{A,\boldsymbol{y}}}(\boldsymbol{x}^\dagger) \min\left\{1, \frac{f(\boldsymbol{x}^\dagger|\boldsymbol{\theta})q(\boldsymbol{x}|\boldsymbol{x}^\dagger)}{f(\boldsymbol{x}|\boldsymbol{\theta})q(\boldsymbol{x}^\dagger|\boldsymbol{x})}\right\}$

Update $\boldsymbol{x} \leftarrow \boldsymbol{x}^\dagger$ with probability α

return $\boldsymbol{x}$

University of Otago
ŌTAKOU WHAKAIHU WAKA

All the Right Moves

Focus for now on move directions; set move length $b = 1$.

Random walk sampler draws moves from set $\mathcal{S} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\}$.

If a move $\boldsymbol{z}$ is to have any chance of acceptance, require:

1. $A\boldsymbol{x}^\dagger = A(\boldsymbol{x} + \boldsymbol{z}) = \boldsymbol{y}$
   $\Rightarrow A\boldsymbol{z} = 0.$
   - That is, $\boldsymbol{z} \in \ker_{\mathbb{Z}}(A) = \ker(A) \cap \mathbb{Z}^r$.
2. $\boldsymbol{x} + \boldsymbol{z} \geq \mathbf{0}$.
   - Inequality interpreted elementwise (here and henceforth)

Lattice Bases

A set $\mathcal{S} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M\}$ is **lattice basis** if every $\boldsymbol{z} \in \ker_{\mathbb{Z}}(A)$ can be written as a unique integer combination of the basis vectors.

Example

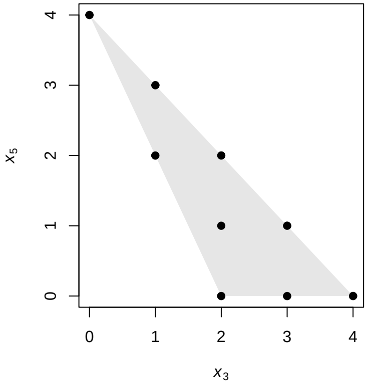
The columns of

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

form a lattice basis for $\mathcal{F}_{A,\boldsymbol{y}}$ in the double counting network tomography problem whenever $\boldsymbol{y} > 0$.

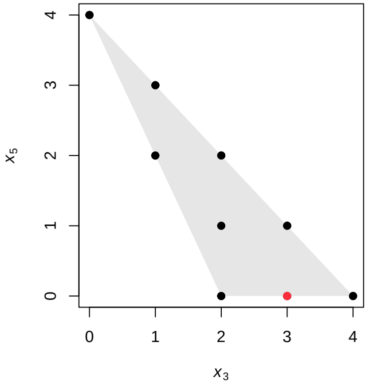
University of Otago
ŌTAKOU WHAKAIHU WAKA

Application to Circuit Network Example



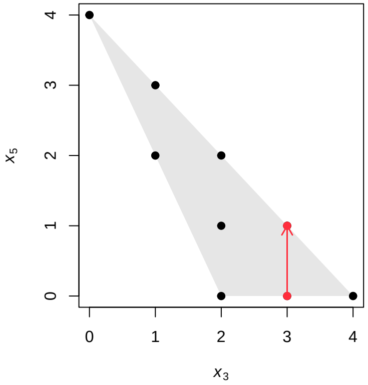
- Lattice basis comprises moves in coordinate directions.

Application to Circuit Network Example



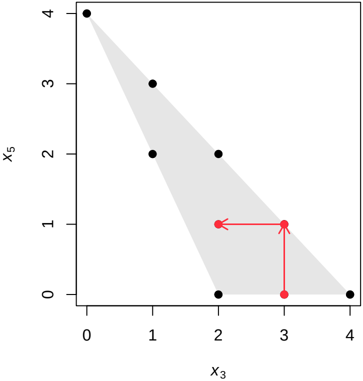
- Lattice basis comprises moves in coordinate directions.

Application to Circuit Network Example



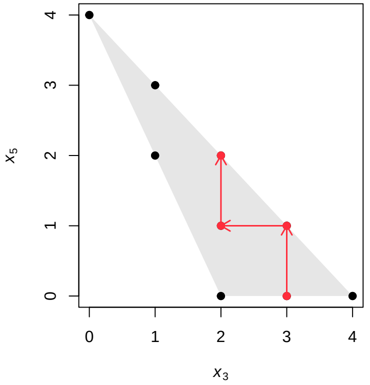
- Lattice basis comprises moves in coordinate directions.

Application to Circuit Network Example



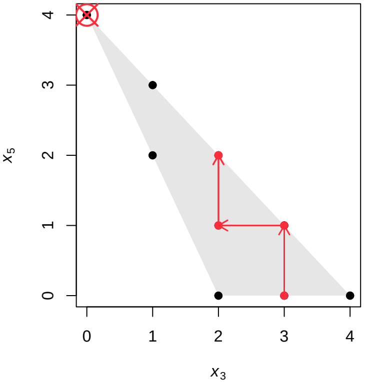
- Lattice basis comprises moves in coordinate directions.

Application to Circuit Network Example



- Lattice basis comprises moves in coordinate directions.

Application to Circuit Network Example



- Lattice basis comprises moves in coordinate directions.
- Random walk cannot visit all points.

Connectedness

- Irreducibility of random walk required for convergence to target posterior.
- This requires that all elements of $\mathcal{F}_{A,\boldsymbol{y}}$ are accessible.
- In other words, the MCMC sampler must be **connected**.
- Connectedness can be very difficult to check in practice.
- As we saw, lattice bases generally do not guarantee connectedness.

Markov Bases and Sub-Bases

- A set of moves $\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\}$ is a **Markov sub-basis** if it guarantees existence of walk between any pair of points on $\mathcal{F}_{A,\boldsymbol{y}}$.
- A set of moves $\mathcal{M}_A$ is a **Markov basis** if it guarantees existence of walk between any pair of points on any fibre (i.e. for all $\boldsymbol{y} \geq \boldsymbol{0}$).
- Computing Markov bases can be very difficult.
- Most successful approach uses **algebraic statistics** (Diaconis and Sturmfels, 1998).
  - But computationally prohibitive in even moderately large applications.

Diaconis, P., & Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, **26(1)**, 363–397.

Mixing ~~Properties~~ Problems with Markov Bases

- Samplers using full Markov bases often mix **very** poorly.
- Full Markov bases can be huge.
  - For given $\boldsymbol{y}$, Markov basis typically contains many useless moves.
  - May wait a long time to select essential move.
- Stanley, C., & Windisch, T. (2018) prove arbitrarily slow convergence to uniform target $f(\boldsymbol{x} \mid \boldsymbol{y})$ in large problems.

Stanley, C., & Windisch, T. (2018). Heat-bath random walks with Markov bases. *Advances in Applied Mathematics*, **92**, 122-143.

Mixing with Markov Sub-Bases

- Markov sub-basis typically much smaller than full Markov basis.
- If moves in a given Markov sub-basis are poorly aligned to polytope geometry, then mixing still slow.
- But Markov sub-bases are not unique...
- ... and mixing can be rapid for well chosen basis.

Example: Contingency Table Resampling

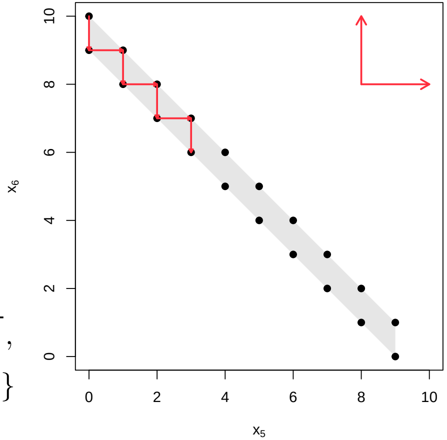
Bad Geometry

$\boldsymbol{y} = (10, 10, 1, 9)^T$

	1	9	10
10	x_1	x_2	x_3
10	x_4	x_5	x_6

A Markov sub-basis is

$$\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{(1, -1, 0, -1, 1, 0)^T, (1, 0, -1, -1, 0, 1)^T\}$$



Example: Contingency Table Resampling

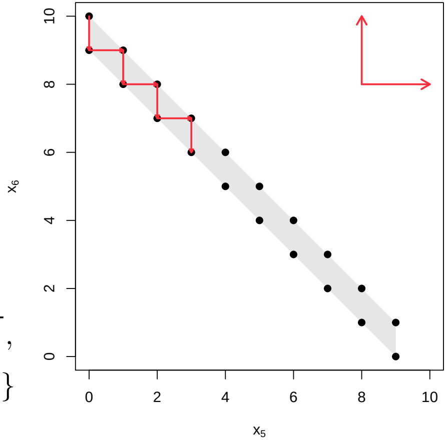
Bad Geometry

$\boldsymbol{y} = (10, 10, 1, 9)^T$

	1	9	10
10	x_1	x_2	x_3
10	x_4	x_5	x_6

A Markov sub-basis is

$$\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{(1, -1, 0, -1, 1, 0)^T, (1, 0, -1, -1, 0, 1)^T\}$$



Mixing arbitrarily slow for $\boldsymbol{y} = (M, M, 1, M-1)^T$ for large M.

University of Otago
Ōtākou Whakaihu Waka

Example: Contingency Table Resampling

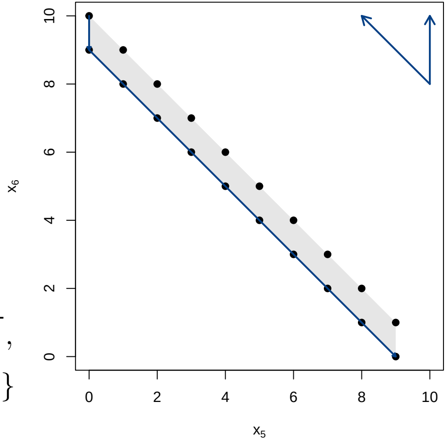
Good Geometry

$\boldsymbol{y} = (10, 10, 1, 9)^T$

	1	9	10
10	x_1	x_2	x_3
10	x_4	x_5	x_6

A Markov sub-basis is

$$\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{(0, 1, -1, 0, -1, 1)^T, (1, 0, -1, -1, 0, 1)^T\}$$



Augmenting Markov Sub-Bases

- $\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}}$ is an **augmenting Markov sub-basis** if any two points on $\mathcal{F}_{A,\boldsymbol{y}}$ are connected by walk utilizing each move at most once.
- If $|\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}}| = \dim(\ker_{\mathbb{Z}}(A))$, then sampler mixes rapidly for uniform target $f(\boldsymbol{x} \mid \boldsymbol{y})$. (Stanley, C., & Windisch, T., 2018)
  - Technically, consider sequence of sampling problems with counts $m\boldsymbol{y}$, $m = 1, 2, \ldots$

Stanley, C., & Windisch, T. (2018). Heat-bath random walks with Markov bases. *Advances in Applied Mathematics*, **92**, 122-143.

Example: Contingency Table Resampling

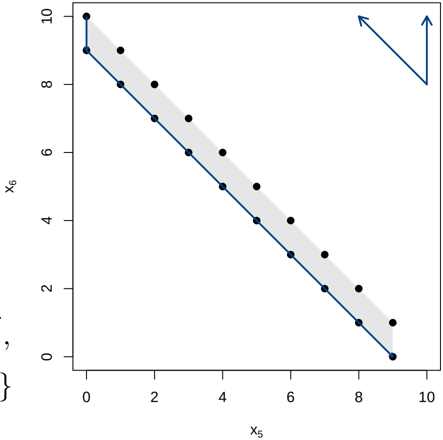
Déjà vu

$\boldsymbol{y} = (10, 10, 1, 9)^{\mathsf{T}}$

	1	9	10
10	x_1	x_2	x_3
10	x_4	x_5	x_6

A Markov sub-basis is

$$\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}} = \{(0, 1, -1, 0, -1, 1)^{\mathsf{T}}, (1, 0, -1, -1, 0, 1)^{\mathsf{T}}\}$$



$\mathcal{M}_{\mathcal{F}_{A,\boldsymbol{y}}}$ is an augmenting Markov basis

Finding Small Augmenting Markov Sub-Bases

Not easy!

- In principle the Graver basis for A generates an augmenting Markov sub-basis for any $\boldsymbol{y} \geq \mathbf{0}$.
  - Graver basis arises in integer programming problems
  - Larger than full Markov basis
- For rapid mixing, we need a Markov sub-basis that is:
  - augmenting;
  - small.
- Lattice bases are small...
- ... and are sometimes (augmenting?) Markov bases (but when?)

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Lattice Bases as Augmenting Markov Sub-Bases

Theory

- H et al. (2024), H & Karimi (2024) found checkable conditions for lattice basis U to be an augmenting Markov basis.
- Uses **dominating matrices**
  - M is a c-mixed matrix if every column contains both positive and negative entries.
  - A matrix M is dominating if it does not contain a square c-mixed matrix.
- **Theorem**: Columns of U form a Markov basis if and only if U is dominating.
- **Corollary**: every lattice Markov basis is augmenting, and hence gives rise to rapid mixing for uniform fibre sampling.

Hazelton, M.L., McVeagh, M.R., Tuffley, C. and van Brunt, B. (2024). *Bernoulli*, **30(4)**, 2676–2699.

Hazelton, M.L., & Karimi, M. (2024). *Statistics and Probability Letters*, **209**, 110106.

Lattice Bases as Augmenting Markov Sub-Bases
Example

For double counting network tomography problem, consider lattice bases:

$$U_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 0 & 2 \\ 0 & -1 \end{bmatrix} \qquad U_2 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 1 & 0 \\ 0 & 2 \\ 0 & -1 \end{bmatrix}$$

- Columns of U_1 do not form a Markov basis.
  - U_1 contains a square mixed matrix.
  - No path from $\boldsymbol{x} = (0,0,0,2,0)^\mathsf{T}$ to $\boldsymbol{x}' = (0,0,0,0,1)^\mathsf{T}$ on fibre $\mathcal{F}_{A,\boldsymbol{y}}$ with $\boldsymbol{y} = (0,0,2)^\mathsf{T}$.
- Columns of U_2 form an augmenting Markov basis.

University of Otago
ŌTĀKOU WHAKAIHU WAKA

Scalable Fibre Samplers

Research Directions

- Approximate samplers using lattice bases.
  - Theory and methods.
- Mixing properties for Poisson models?

Thanks to ...

Collaborators

Masoud Karimi (Islamic Azad University, Bojnourd Branch)
Bruce van Brunt (Massey U.)
Chris Tuffley (Massey U.)
Linus Fromm (Honours, U. Otago)
Mike McVeagh (PhD, Massey U.)
Andrew Robinson (U. Melbourne)

Funding

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN
ROYAL SOCIETY TE APĀRANGI

University of Otago
ŌTAKOU WHAKAIHU WAKA

To Learn More ...

Journal Articles

Hazelton, M.L., McVeagh, M.R., and van Brunt, B. (2021). Geometrically aware dynamic Markov Bases for statistical linear inverse problems. *Biometrika*, **108(3)**, 609-626.

Hazelton, M.L., McVeagh, M.R., Tuffley, C. and van Brunt, B. (2024). Some rapidly mixing hit-and-run samplers for latent counts in linear inverse problems. *Bernoulli*, **30(4)**, 2676–2699.

Hazelton, M.L., and Karimi, M. (2024). When lattice bases are Markov bases. *Statistics and Probability Letters*, **209**, 110106.

R Package `LinInvCount`

`github.com/MartinLHazelton/LinInvCount`