Dynamic fibre samplers for linear inverse problems

Martin Hazelton[1]

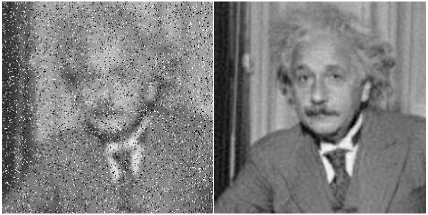
Department of Mathematics & Statistics
University of Otago

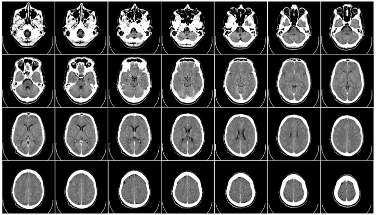
19 January 2022

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

Statistical Inverse Problems

- Interest is in a process that is observed only indirectly.
- Problems of this sort are ubiquitous in science and technology.
- Image deblurring and computed tomography are classic examples.





Linear Inverse Problems for Count Data

- For count data, statistical linear inverse problems characterised by

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

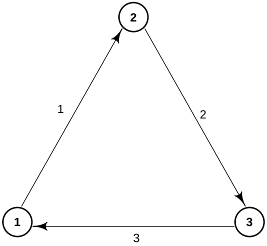
  - $\boldsymbol{x} \in \mathbb{Z}^r_{\geq 0}$ is count vector of interest;
  - $\boldsymbol{y} \in \mathbb{Z}^n_{\geq 0}$ is vector of observed counts.
  - **Configuration matrix** A is $n \times r$ and has binary (or sometimes non-negative integer) elements.
- Typically $r > n$ so linear system (1) will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

UNIVERSITY of OTAGO Te Whare Wānanga o Otāgo NEW ZEALAND

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 $\boldsymbol{x}$.

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

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

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

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

- Data collected over a sequence of observational periods.
- $\boldsymbol{y}$ is vector of recorded counts classified by pattern of sightings.
  - E.g. y_{101} count of animals observed in periods 1 and 3 but not 2.
- True pattern of sightings $\boldsymbol{x}$ differs from $\boldsymbol{y}$ due to misidentifications.

Biosecurity Surveillance

- Inspection schemes for mail items stratified based on their expected risk.
- Each item classified by unknown true compliance status, inclusion/exclusion and compliance assessment at each stage.
- This cross-classification generates a contingency table with cell counts $\boldsymbol{x}$, but we can observe only certain sums $\boldsymbol{y}$ of these entries.

The Conditional Distribution of $\boldsymbol{x}$

- Inference for $\boldsymbol{x}$ based on conditional distribution $f(\boldsymbol{x}|\,\boldsymbol{y})$.
  - Dependence of f on parameter θ 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 $f(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\mathcal{F}_{\boldsymbol{y}}} f(\boldsymbol{x})$.
- Here $\mathcal{F}_{\boldsymbol{y}} = \{\boldsymbol{x} \colon \boldsymbol{y} = A\boldsymbol{x}\} \cap \mathbb{Z}^r_{\geq 0}$.
- This is solution set is called the ***y*-fibre**.

The Geometry of a $\boldsymbol{y}$-fibre

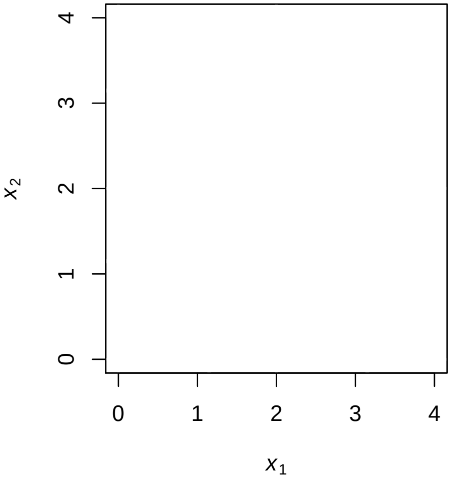
2×3 contingency table example

	2	4	2
3	x_1	x_2	x_3
5	x_4	x_5	x_6

- $\boldsymbol{y} = (3, 5, 2, 4)^{\mathsf{T}}$. (Recall y_5 constraint redundant.)
- Set of feasible counts $\mathcal{F}_{\boldsymbol{y}} = \{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\} \cap \mathbb{Z}^r_{\geq 0}$ can be fully specified by values of x_1, x_2.
- Constraints on these entries:
  - $0 \leq x_1 \leq 2$
  - $0 \leq x_2 \leq 4$
  - $x_1 + x_2 \leq 3$
  - $1 \leq x_1 + x_2$

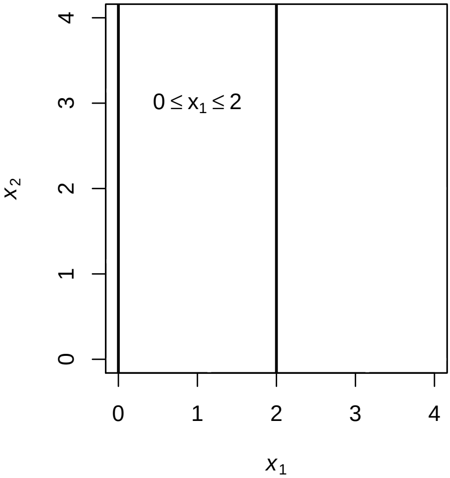
The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example



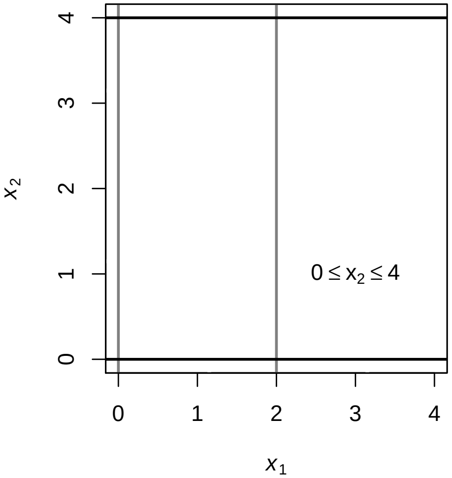
The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example



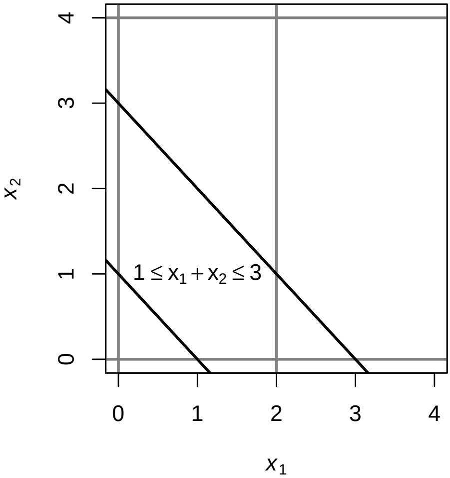
The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example



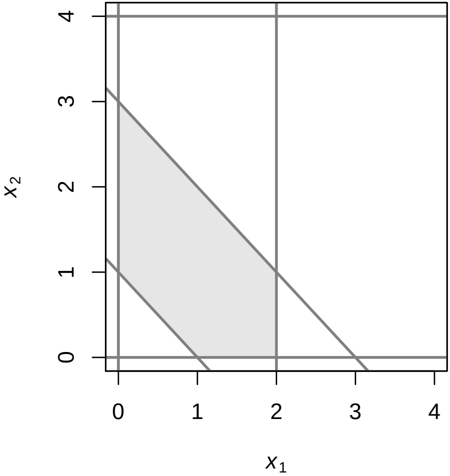
The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example



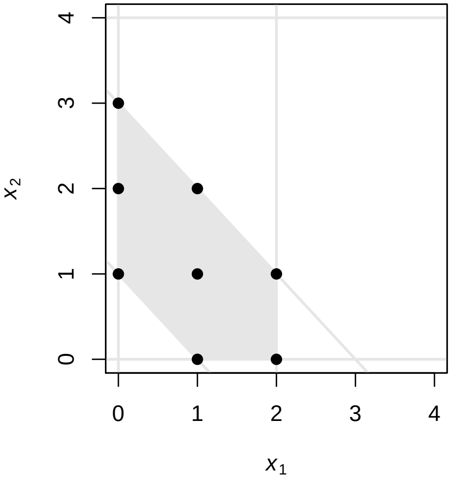
The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example



The Geometry of a *y*-fibre

Constructing the fibre for the 2 × 3 contingency table example

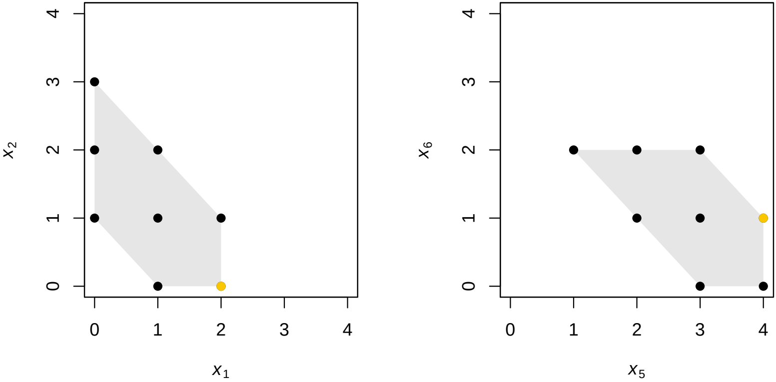


$\mathbb{Z}$-Polytopes

- 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 $\mathcal{F}_{\boldsymbol{y}} = \{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\} \cap \mathbb{Z}^r_{\geq 0}$ is a **$\mathbb{Z}$-polytope.**
- Assuming A of full rank, then $\mathcal{F}_{\boldsymbol{y}}$ is an $r - n$ dimensional object embedded in r-dimensional space.
- Have flexibility in representation.

Different Projections of a Polytope

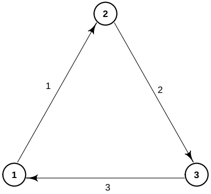
2×3 contingency table example: $r = 6$ and $r - n = 2$



Gold points correspond to table $\boldsymbol{x} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 1 \end{bmatrix}$

Different Projections of a Polytope

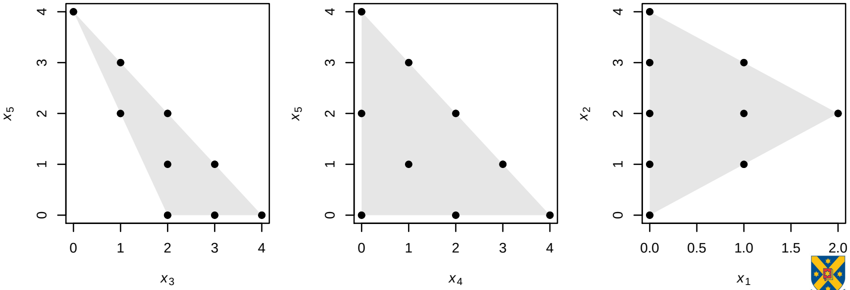
Circuit network example: $r = 5$ and $r - n = 2$



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

Example: how many tables on the same fibre?

	Hair				
Eyes	Black	Brunette	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

Inference

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

Example: how many tables on the same fibre?

	Hair				
Eyes	Black	Brunette	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

Answer: 1,225,914,276,276,768,514

MCMC Based Inference

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

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

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 $\mathcal{F}_{\boldsymbol{y}}$.
- Will adopt random walk Metropolis-Hastings sampler.

input

Current state $\boldsymbol{x}$

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

Draw $\boldsymbol{z}$ from set $\mathcal{S} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_M\}$ 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 $\alpha = \mathbf{1}_{\mathcal{F}_{\boldsymbol{y}}}(\boldsymbol{x}^{\dagger}) \min\left\{1, \frac{f(\boldsymbol{x}^{\dagger}|\boldsymbol{\theta})q(\boldsymbol{x}|\boldsymbol{x}^{\dagger})}{f(\boldsymbol{x}|\boldsymbol{\theta})q(\boldsymbol{x}^{\dagger}|\boldsymbol{x})}\right\}$

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

return $\boldsymbol{x}$

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

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. $\boldsymbol{x} + \boldsymbol{z} \geq \boldsymbol{0}$.
   - Inequality interpreted elementwise (here and henceforth)

Constructing a Lattice Basis

- A **lattice basis** is a basis for $\ker_{\mathbb{Z}}(A)$.
- Partition $A = [A_1|A_2]$ with $n \times n$ matrix A_1 invertible.
  - Partition $\boldsymbol{x} = [\boldsymbol{x}_1|\boldsymbol{x}_2]$ likewise.
- Define matrix

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

- Then

$$AU = [A_1|A_2] \begin{bmatrix} -A_1^{-1}A_2 \\ I_{r-n} \end{bmatrix} = -A_2 + A_2 = 0$$

- Hence columns $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{r-n} \in \ker_{\mathbb{Z}}(A)$ and so form lattice basis.
- Moves $\pm \boldsymbol{u}_i$ correspond to steps in coordinate directions in polytope projection onto column space of A_2.

Application to 2×3 Contingency Table Example

Lattice basis contains $r - n = 2$ vectors, $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$.

Basis vector	$\boldsymbol{u}_1 = (1, -1, 0, -1, 1, 0)^{\mathsf{T}}$	$\boldsymbol{u}_2 = (1, 0, -1, -1, 0, 1)^{\mathsf{T}}$
Effect on table	$\begin{bmatrix} + & - & \cdot \\ - & + & \cdot \end{bmatrix}$	$\begin{bmatrix} + & \cdot & - \\ - & \cdot & + \end{bmatrix}$

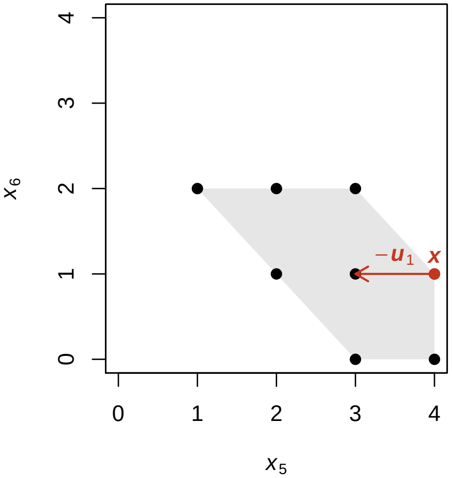
Illustration

$\boldsymbol{x} = (2, 0, 1, 0, 4, 1)^{\mathsf{T}}$, then $\boldsymbol{x} - \boldsymbol{u}_1 = (1, 1, 1, 1, 3, 1) \in \mathcal{F}_{\boldsymbol{y}}$.

$$\boldsymbol{x} - \boldsymbol{u}_1 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 1 \end{bmatrix} - \begin{bmatrix} + & - & \cdot \\ - & + & \cdot \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 1 \end{bmatrix}$$

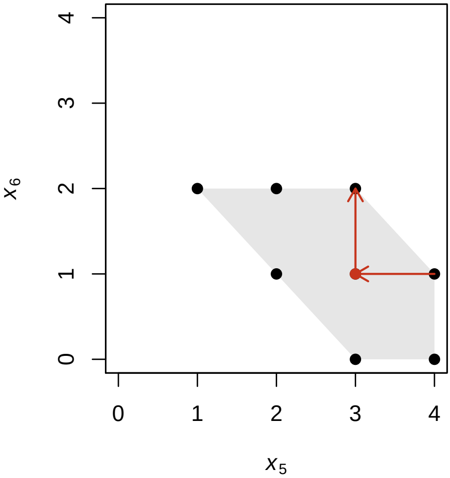
Application to 2 $\times$ 3 Contingency Table Example

Walking on sunshine



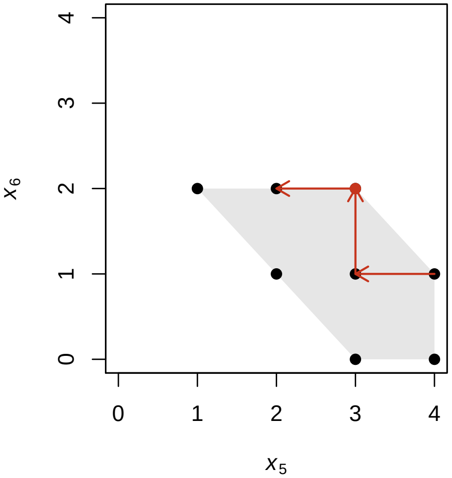
Application to 2 × 3 Contingency Table Example

Walking on sunshine



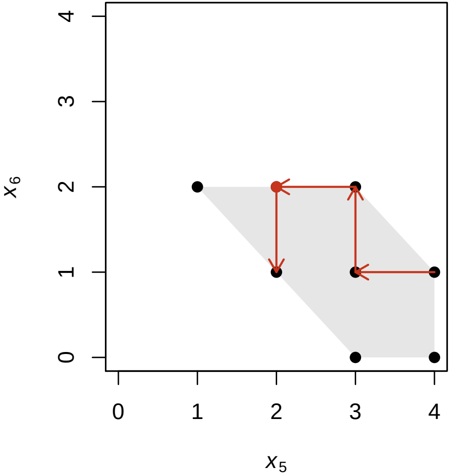
Application to 2 × 3 Contingency Table Example

Walking on sunshine



Application to 2 × 3 Contingency Table Example

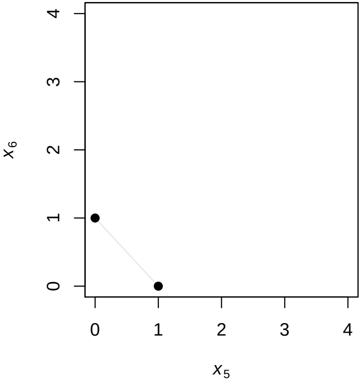
Walking on sunshine



A Sparse Contingency Table Example

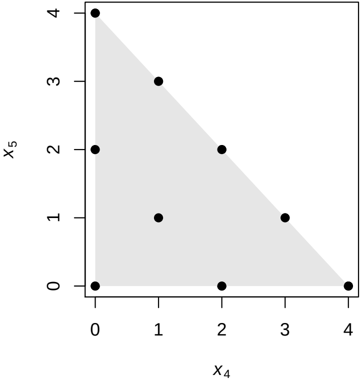
Road to nowhere

	0	1	1
1	x_1	x_2	x_3
1	x_4	x_5	x_6



- $\boldsymbol{u}_1 = (1, -1, 0, -1, 1, 0)^\mathsf{T}$
- $\boldsymbol{u}_2 = (1, 0, -1, -1, 0, 1)^\mathsf{T}$
- E.g. $\boldsymbol{x} = (0, 1, 0, 0, 0, 1)$
- All of $\boldsymbol{x} \pm \boldsymbol{u}_i$ will have negative entry

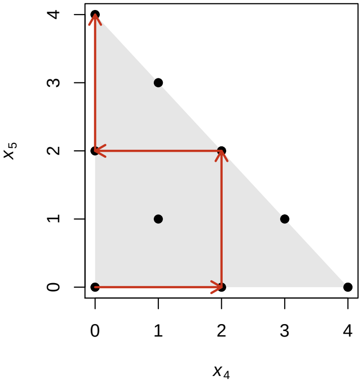
Application to Circuit Network Example



- Lattice basis comprises moves in coordinate directions.
- Impossible to change parity of entries of $\boldsymbol{x}$.
- Random walk cannot visit all points.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Application to Circuit Network Example



- Lattice basis comprises moves in coordinate directions.
- Impossible to change parity of entries of $\boldsymbol{x}$.
- Random walk cannot visit all points.

Connectedness

- Irreducibility of random walk required for convergence to target posterior.
- This requires that all elements of $\mathcal{F}_{\boldsymbol{y}}$ 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

$\mathcal{B} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\}$ is a **Markov (sub-)basis** if for all $\boldsymbol{x}^a, \boldsymbol{x}^b \in \mathcal{F}_{\boldsymbol{y}}$

$$\boldsymbol{x}^b = \boldsymbol{x}^a + \sum_{i=1}^{L} \epsilon_i \boldsymbol{z}_i \quad \text{and} \quad \boldsymbol{x}^a + \sum_{i=1}^{K} \epsilon_i \boldsymbol{z}_i \in \mathcal{F}_{\boldsymbol{y}} \text{ for } K = 1, 2, \ldots, L$$

where $\epsilon_1, \ldots, \epsilon_L \in \{-1, 1\}$.

- $A\boldsymbol{z}_i = \boldsymbol{0}$ for $\boldsymbol{z}_i \in \mathcal{B}$.
- For all intermediate points on walk, $\boldsymbol{x}^a + \sum_{i=1}^{K} \epsilon_i \boldsymbol{z}_i \geq \boldsymbol{0}$.
- MCMC sampler is connected if proposed moves drawn from $\mathcal{B}$.
- A full Markov basis will ensure connectivity for any $\boldsymbol{y}$-fibre.
- A Markov sub-basis is specific to a given $\boldsymbol{y}$-fibre.

Markov Bases and Algebraic Statistics

- Computing Markov bases is very difficult in all but toy problems.
- Most successful approach to date uses **algebraic statistics**...
- ...following seminal work of Diaconis and Sturmfels (1998).
- Idea is to represent $\boldsymbol{x} \geq \mathbf{0}$ by monomial:

$$T(\boldsymbol{x}) := \boldsymbol{t}^{\boldsymbol{x}} = t_1^{x_1} t_2^{x_2} \cdots t_r^{x_r}$$

- A move $\boldsymbol{z}$ represented by monomial difference $\boldsymbol{t}^{\boldsymbol{z}^+} - \boldsymbol{t}^{\boldsymbol{z}^-}$ where $\boldsymbol{z}^+$ and $\boldsymbol{z}^-$ contain respectively positive and negative parts of $\boldsymbol{z}$.
- Markov basis for sampling defined by Gröbner basis for toric ideal of monomial differences.
- Implemented using `4ti2` software.

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

Problems with Markov bases

1. Finding full Markov basis usually computationally infeasible in even moderately large problems.
2. Samplers using full Markov bases can mix **very** poorly.
    - For given $\boldsymbol{y}$, Markov basis typically contains many useless moves.
    - Full Markov bases take no account of polytope geometry.

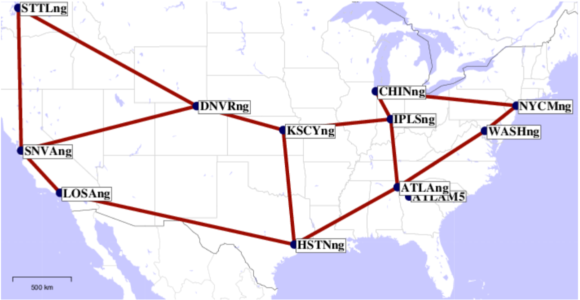
Examples of unwieldy Markov bases

Contingency Tables

- A full Markov basis for an $I \times J$ contingency table has $\frac{1}{4}IJ(I-1)(J-1)$ elements.
- Hence for 20×20 table, Markov basis has more than $35,000$ elements.
- In almost all cases, an adequate Markov sub-basis can be found with 361 elements.

Examples of unwieldy Markov bases

Network Tomography



- 12 nodes, $r = 132$ paths, $n = 42$ links.
- Using `4ti2`, took more than 9 hours to find a Markov basis containing 10,705 vectors.

Dynamic Markov Bases

- Idea is to avoid computing full Markov basis *ab initio*.
- At each step, find a suitable set of 'local moves'.
- So long as union of all such sets forms a Markov basis (in a sensible way), the resulting random walk should be connected.
- Seminal work in this area by Dobra (2012) specific to contingency tables, and ignored geometry of polytopes.
- Our idea is to find a geometrically aware dynamic Markov basis using collections of lattice bases.

Dobra, A. (2012). Dynamic Markov bases. *Journal of Computational and Graphical Statistics*, **21(2)**, 496–517.

Lattice Bases as Local Moves

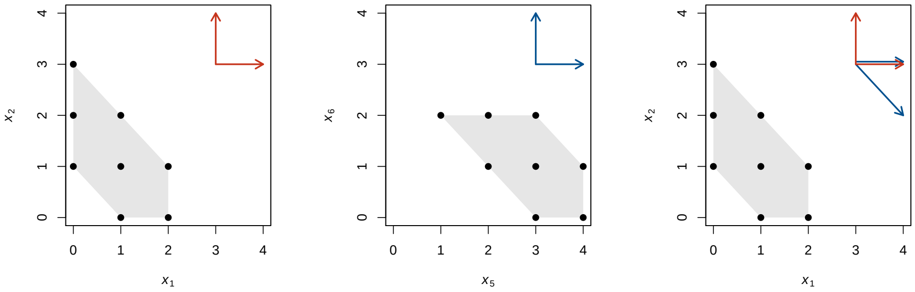
- Idea is to use lattice bases to provide sets of local moves.
- Recall lattice bases not unique.
- Let π denote a partition of $\{1, \ldots, r\}$ into two subsets, K_1 and K_2, of size n and $r - n$ respectively.
- Let A_i^π denote submatrix of A formed by columns indexed by K_i for $i = 1, 2$.
- Let $\Pi = \{\pi \colon |A_1^\pi| \neq 0\}$.
- For $\pi \in \Pi$, lattice basis $\mathcal{B}_L^\pi$ defined by columns of

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

- Corresponds to coordinate moves with respect to columns of A_2^π.

Different Lattice Bases for 2 × 3 contingency table

	2	4	2
3	x_1	x_2	x_3
5	x_4	x_5	x_6



Critical Theory

All you need is ~~love~~ lattice bases

Recall:

- $\mathcal{B}_L^\pi$ is lattice basis corresponding to partition $\pi \in \Pi$
- Π set of partitions for which A_1^π is invertible.

Definition (Unimodular Matrix)

A matrix A is unimodular if every invertible maximal square submatrix of A has determinant ± 1.

Theorem

If A unimodular then $\bigcup_{\pi \in \Pi} \mathcal{B}_L^\pi$ is a Markov basis.

Designing a Dynamic Lattice Basis Sampler

- Look at Markov process $\{(\boldsymbol{x}^t, \pi^t) \colon t = 1, 2, \ldots\}$.
- Let conditional distribution of π^t depend on π^{t-1} but not $\boldsymbol{x}^{t-1}$, to avoid upsetting balance equations.
- Connectedness of walk is assured if all $\pi \in \Pi$ have non-zero probability.

Naive approach: randomly select π from Π at each iteration. But...

1. Need to recalculate lattice bases from scratch unacceptably slow.
2. Does not take account of polytope geometry to facilitate mixing.

New Lattice Bases Via Single Column Updating

- Update π be potentially exchanging swapping a pair of columns i and j between K_1 and K_2.
- Then current lattice basis U can be updated to

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

where $C = A_1^{-1} A_2$, and updated version is

$$\tilde{C} = C - \frac{1}{c_{ij}} (\boldsymbol{c}_j - \boldsymbol{e}_i)(\boldsymbol{c}_i + \boldsymbol{e}_j)^\mathsf{T}$$

courtesy of the Sherman-Morrison formula.

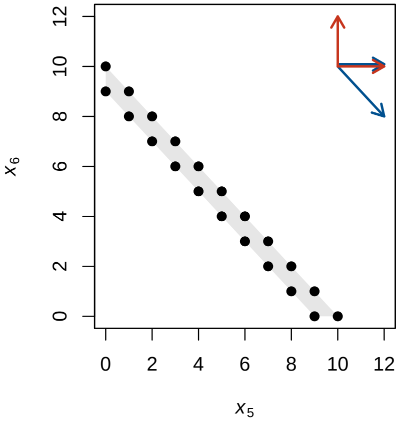
- Note that the interchange of columns is feasible if and only if $c_{ij} \neq 0$ (required to ensure A_1 remains invertible).

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Geometrically Aware Lattice Bases

	1	10	10
11	x_1	x_2	x_3
10	x_4	x_5	x_6

- Colour identifies moves in two different lattice bases.
- Choice of basis affects rate of mixing of sampler.

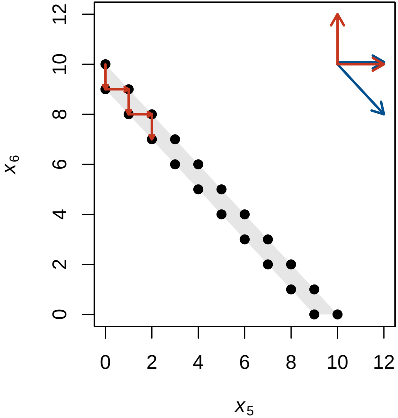


UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Geometrically Aware Lattice Bases

	1	10	10
11	x_1	x_2	x_3
10	x_4	x_5	x_6

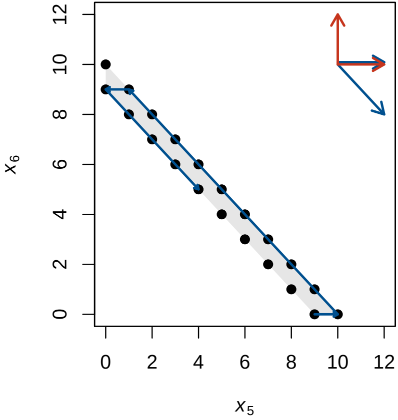
- Colour identifies moves in two different lattice bases.
- Choice of basis affects rate of mixing of sampler.



Geometrically Aware Lattice Bases

	1	10	10
11	x_1	x_2	x_3
10	x_4	x_5	x_6

- Colour identifies moves in two different lattice bases.
- Choice of basis affects rate of mixing of sampler.



Identifying Geometrically Advantageous Lattice Bases

- Consider sampling in direction $\boldsymbol{u} \in \mathcal{B}_L$.
- For feasible $\boldsymbol{x}^\dagger = \boldsymbol{x} + b\boldsymbol{u}$, require $\boldsymbol{x} + b\boldsymbol{u} \geq \mathbf{0}$.
- $b_{\min}(\boldsymbol{x}) = -\left\lfloor \min_{i:u_i>0}\{x_i/|u_i|\} \right\rfloor$, $b_{\max}(\boldsymbol{x}) = \left\lfloor \min_{i:u_i<0}\{x_i/|u_i|\} \right\rfloor$.
- Advantageous polytope geometry corresponds to representation where $b_{\max}(\boldsymbol{x}) - b_{\min}(\boldsymbol{x})$ is relatively large.
- To optimize, choose partition such that entries of $\boldsymbol{x}_1$ are relatively large.
- Corresponding to maximizing slack in linear inequality $A_2\boldsymbol{x}_2 \leq \boldsymbol{y}$.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Sampling Partitions

- What to assign high probability to partitions π with large $\boldsymbol{x}_1$.
- Problem: sampling distribution of π should not depend on $\boldsymbol{x}$.
- Resolution: use proxy for typical size of entries of $\boldsymbol{x}$.
- Example: use unconditional mean $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{x}|\theta]$.
- Let $\phi \sim \mathrm{N}(\boldsymbol{\mu}, \alpha \mathrm{diag}(\boldsymbol{\mu}))$ be vector of fitnesses for columns of A.
- Select fittest columns for A_1, subject to invertibility...
- ... but chance of selecting any column ordering ensures connectivity requirements for walk.
- Tuning parameter α determines probability of visiting 'sub-optimal' lattice bases.
  - $\alpha = 0$ only uses 'best basis' (irreducibility not assured)
  - $\alpha = \infty$ ignores polytope geometry entirely

Sampling Partitions

Algorithm for single column updates

Input

Current state $\boldsymbol{x}$

Current partition π and corresponding basis vectors U

begin

Draw $\boldsymbol{\phi} \sim \mathrm{N}(\boldsymbol{\mu}, \alpha \mathrm{diag}(\boldsymbol{\mu}))$

Sample $i^\dagger$ from discrete uniform distribution on K_1

Sample $j^\dagger$ from discrete uniform distribution on $\{j \in K_2 : c_{i^\dagger j} \neq 0\}$

if $\phi_{j^\dagger} \geq \phi_{i^\dagger}$ **then**

Update U

Update π by swapping $i^\dagger$ and $j^\dagger$ between K_1 and K_2

return π, U

30 × 15 Contingency Table Application (1/4)

Book crossing data

	au	at	ca	fi	fr	de	it	my	nl	nz	pt	sg	es	uk	us
0062502174	4	0	5	0	0	0	0	0	0	2	0	0	0	0	23
0310205719	0	0	0	0	1	0	0	1	0	0	0	0	0	1	74
0316777730	0	1	7	0	1	3	0	1	0	0	0	1	0	2	97
0375501347	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39
0375704027	1	0	3	0	0	0	0	0	0	1	1	1	0	0	24
0380470845	0	0	2	0	0	2	0	0	0	0	0	0	0	1	26
0385315090	0	0	3	0	0	0	0	1	0	0	0	0	1	0	25
0440207622	0	0	9	0	0	1	0	0	0	0	0	1	1	2	64
0440212723	0	0	2	0	0	0	0	0	0	0	0	0	0	0	32
0441104029	0	0	7	0	0	0	0	0	1	0	1	0	2	0	39
0446357421	1	0	2	0	0	0	0	0	3	0	0	0	1	0	26
044651652X	3	1	32	1	0	2	0	7	0	1	2	0	2	7	286
0446604232	1	0	7	0	0	1	1	0	1	0	0	0	0	0	69
0446606324	3	0	16	0	0	2	0	0	0	0	0	0	1	1	125
0446606812	4	0	39	0	0	1	0	5	0	2	3	1	0	1	285
0451207947	0	0	2	0	0	0	0	1	0	0	0	2	0	0	31
0451524934	3	1	5	2	0	2	0	5	0	1	7	0	3	6	149
0515132187	0	0	28	0	0	1	0	0	0	0	0	0	0	1	140
0552146153	1	1	1	0	1	0	0	0	2	0	0	0	4	17	8
0553258915	2	0	3	0	0	0	0	0	0	0	0	0	0	0	31
0553565915	0	0	5	0	0	0	0	0	0	0	0	0	0	0	54
0553573705	0	0	3	0	0	1	0	1	0	0	0	0	0	0	29
0590453653	3	0	6	0	0	0	0	0	0	2	0	0	0	0	42
0671524097	0	0	3	0	1	0	0	0	0	0	0	0	0	0	26
0671683993	0	0	21	0	0	2	0	0	0	0	0	0	0	0	76
0752844059	0	0	12	0	0	1	0	0	0	2	0	0	0	6	2
0786817879	0	0	8	0	0	1	0	1	2	0	2	0	0	1	31
0812550285	1	0	5	0	0	0	0	0	1	1	1	0	0	0	25
1573227889	0	0	2	1	1	1	0	0	1	0	0	0	3	5	29
8826703132	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0

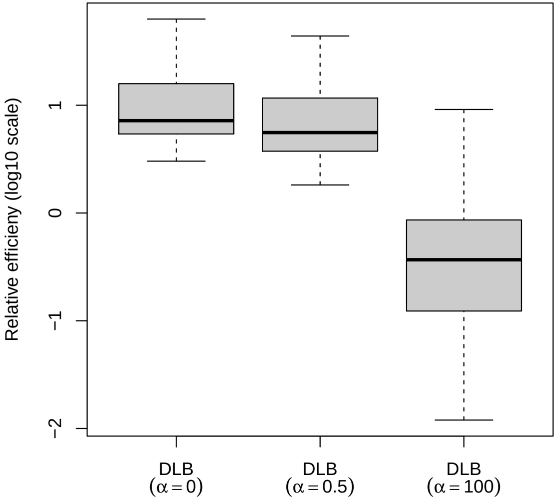
30 × 15 Contingency Table Application (2/4)

Methods for comparison

- Dynamic lattice base sampler with $\alpha = 0$ (not dynamic!)
- Dynamic lattice base sampler with $\alpha = 0.5$
- Dynamic lattice base sampler with $\alpha = 100$ (ignores geometry)
- Full Markov basis (45675 vectors)

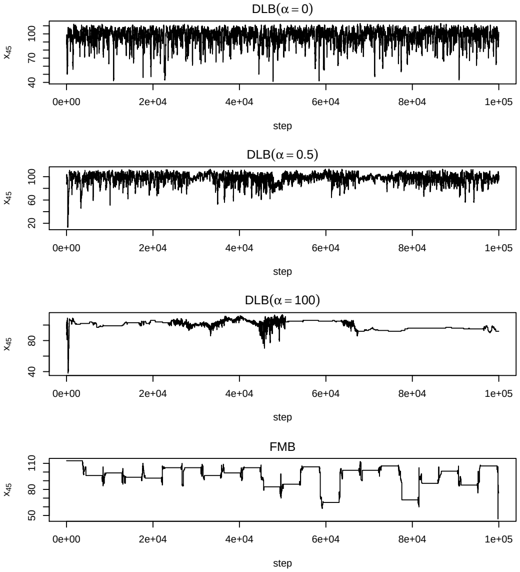
30 × 15 Contingency Table Application (3/4)

Efficiencies relative to full Markov basis



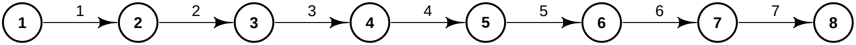
30 × 15 Contingency Table Application (4/4)

Example trace plots



Network Tomography Application (1/3)

Section of A6, Leicester



- Looking at travel in one direction.
- Paths connect each node with any subsequent node.
- $n = 7$ links and $r = 28$ paths.
- $\boldsymbol{y} = (1087, 1008, 1068, 1204, 1158, 1151, 1143)^{\mathsf{T}}$.
- $\boldsymbol{x} \sim \mathrm{Pois}(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda}^{\mathsf{T}} =$ (83.0, 25.0, 19.0, 89.0, 10.0, 9.0, 825.0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 5.0, 1.0, 2.0, 74.0, 0.5, 36.0, 2.0, 105.0, 10.0, 0.1, 69.0, 5.0, 38.0, 15.0).
- Chain initialized by solving integer programming problem.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Network Tomography Application (2/3)

Section of A6, Leicester

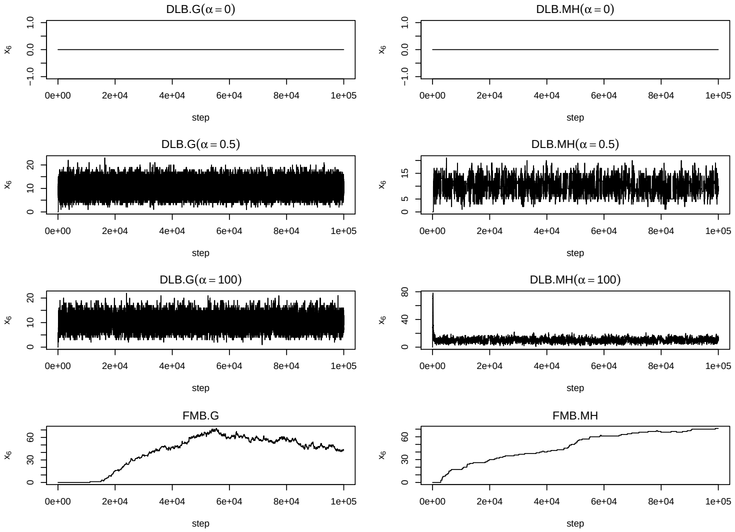
$$A = \begin{pmatrix}
1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&1&1&1&1&0&1&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&1&0&0&1&1&1&1&0&1&1&1&1&1&1&1&1&0&0&0&0&0&0\\
0&0&0&0&1&1&1&0&0&0&1&1&1&0&0&1&1&1&0&1&1&1&1&1&1&0&0&0\\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

- Can partition $A = [I \,|A_2]$ using appropriate column reordering.
- Resultant lattice basis is also a Markov basis, with minimal 21 elements.
- However, columns of A_1 correspond to paths comprising single links (1, 8, 14, 19, 23, 26, 28), many of which do not carry heavy flows.
- This is default Markov basis found by `4ti2`, but not necessarily a good one geometrically.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Network Tomography Application (3/3)

Trace plots



Non-Unimodular Configuration Matrices

- When A is not-unimodular, the union of lattice bases will still often be a Markov basis.
  - In that case our dynamic fibre sampler can be applied directly.
- Sadly, impossible to check whether that result holds in sizeable applications.
- Can fix the theoretical hole by introducing occasional moves based on integer-weighted combinations of lattice basis vectors.
- Sampler performance remains excellent.

Research Questions

- Methods for choosing tuning parameter in practice?
- Theory on mixing properties?

Thanks to ...

Collaborators

Alan Lee (U. Auckland)
Bruce van Brunt (Massey U.)
Chris Tuffley (Massey U.)
Jenny Wilcock (U. Canterbury)
Matt Schofield (U. Otago)
Mike McVeagh (PhD, Massey U.)
Rina Parry (AgResearch)
Timothy Bilton (AgResearch)

Funding

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

ROYAL SOCIETY TE APĀRANGI

To Learn More ...

Journal Article

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. `https://doi.org/10.1093/biomet/asaa083`.

R Package `DynamicLatticeBasis`

`github.com/MartinLHazelton/DynamicLatticeBasis`