Markov chain Monte Carlo updating schemes for hidden Gaussian Markov random field models

Ingelin Steinsland

ingelins@math.ntnu.no

Norwegian University of Science and Technology

Background

“The work of this thesis is driven by the need of one-block updating schemes in Markov chain Monte Carlo simulations of latent spatial models, and the opportunity of fast sampling and evaluations of Gaussian Markov random fields. “

Background

“The work of this thesis is driven by the need of one-block updating schemes in Markov chain Monte Carlo simulations of latent spatial models, and the opportunity of fast sampling and evaluations of Gaussian Markov random fields. “

- Latent spatial Gaussian Markov random field (GMRF) models
  - Disease mapping in Germany
  - Functional magnetic resonance imaging (fMRI) of the brain
- One-block MCMC updating schemes

Gaussian Markov random field

A GMRF $x = (x_1, x_2, \ldots, x_n)$ is:

- Multivariate Gaussian distributed
- with a Markov property;
  - neighbourhood structure
  - x_i and x_j conditionally dependent only if they are neighbours

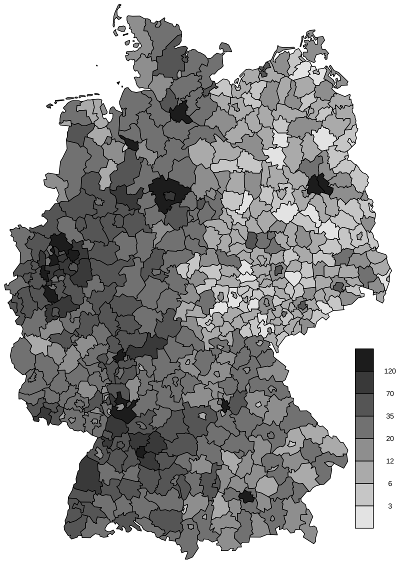
Gaussian Markov random field

A GMRF $x = (x_1, x_2, \ldots, x_n)$ is:

- Multivariate Gaussian distributed
- with a Markov property;
  - neighbourhood structure
  - x_i and x_j conditionally dependent only if they are neighbours
- Gives sparse precision matrix and computational benefits.

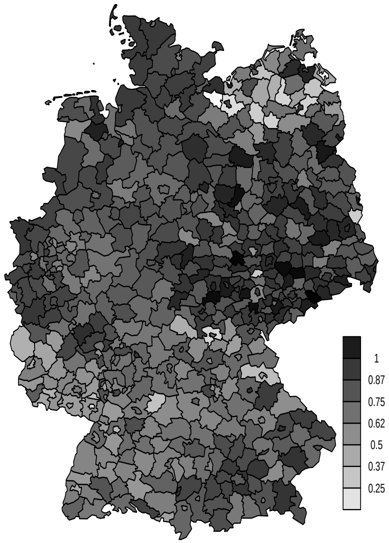
Disease mapping in Germany

- **Data:** All cases of oral cavity cancer mortality in each of Germany's 544 regions, y_i, $i = 1, \ldots, 544$



Disease mapping in Germany

- **Data:** All cases of oral cavity cancer mortality in each of Germany's 544 regions, y_i, $i = 1, \ldots, 544$
- **Of interest:** The log-relative risk, x_i.



Disease mapping in Germany

- **Data:** All cases of oral cavity cancer mortality in each of Germany's 544 regions, y_i, $i = 1, \ldots, 544$
- **Of interest:** The log-relative risk, x_i.
- **Model:**
  - Mutually independent likelihood:

$$y_i | x_i \sim Po(\exp(x_i)) \quad \forall i$$

  - GMRF prior:

$$\pi(x|\kappa) \propto \kappa^{(n-1)/2} \exp(-\frac{1}{2}\kappa \sum_{i \sim j} (x_i - x_j)^2)$$

Disease mapping in Germany

- **Data:** All cases of oral cavity cancer mortality in each of Germany's 544 regions, y_i, $i = 1, \dots, 544$
- **Of interest:** The log-relative risk, x_i.
- **Model:**
  - Mutually independent likelihood:

$$y_i | x_i \sim Po(\exp(x_i)) \quad \forall i$$

  - GMRF prior:

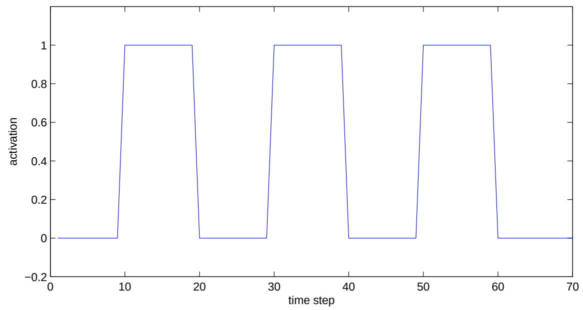
$$\pi(x|\kappa) \propto \kappa^{(n-1)/2} \exp(-\frac{1}{2}\kappa \sum_{i \sim j} (x_i - x_j)^2)$$

- **Posterior:** $\pi(x, \kappa | y) \propto \pi(y|x)\pi(x|\kappa)\pi(\kappa)$

fMRI

functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.

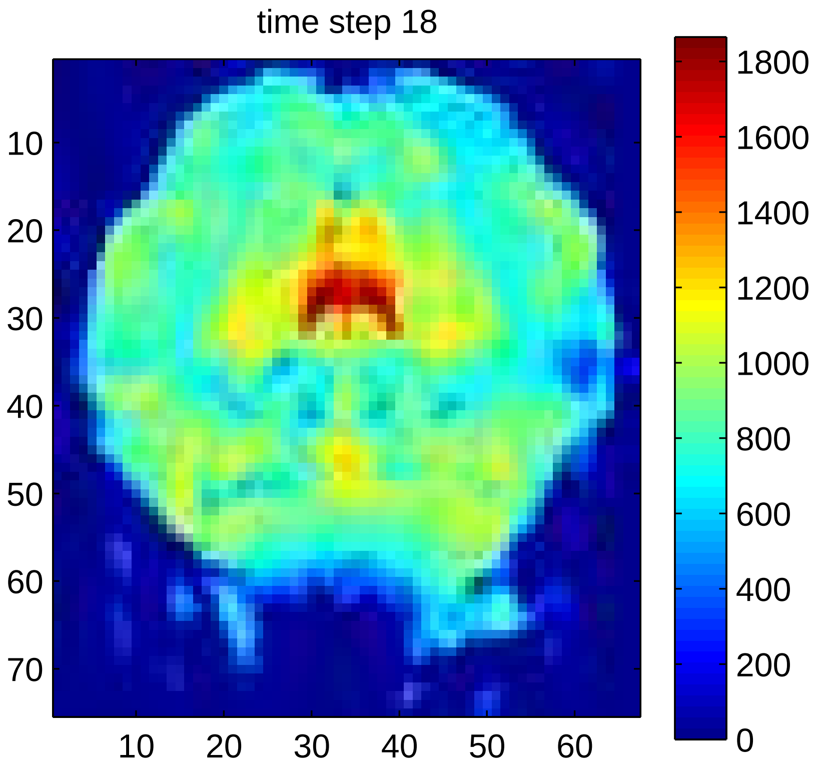
- Stimulus: 8 Hz flickering checkerboard.
- 4 periods (a 30 sec.) rest, 3 periods stimulus.

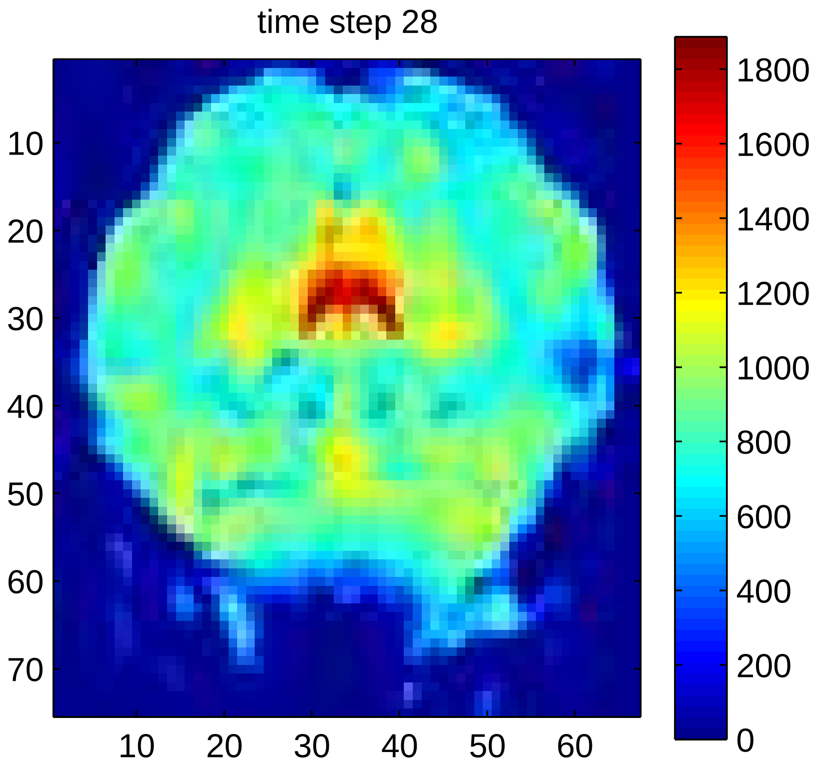


- Cross section of the brain observed every 3rd sec.
- Observe BOLD effects.

fMRI

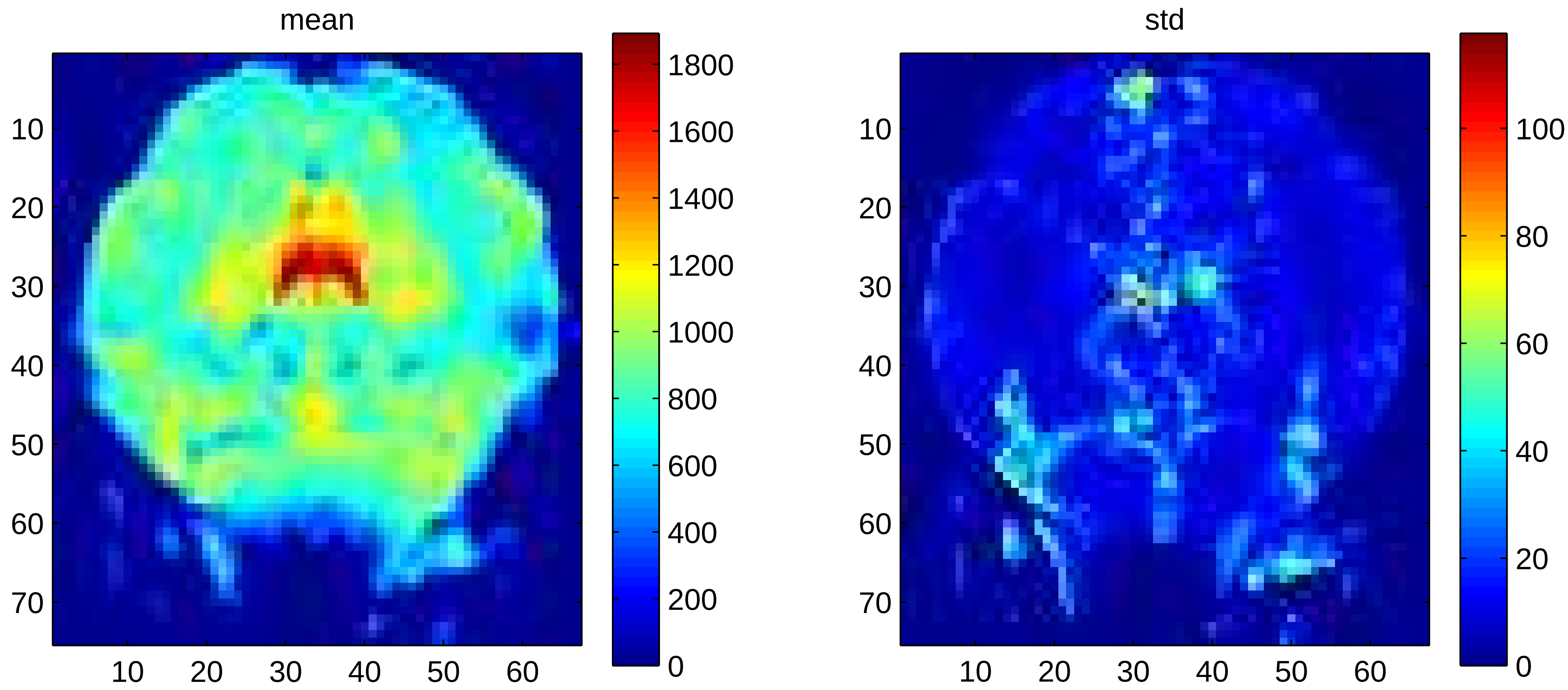
functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.





fMRI

functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.



Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$
- z_t: Transformed stimulus at time step t,
  $t = 1, \ldots 70$
- b_{it}: Activation effect of pixel i at time step t,
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$
- z_t: Transformed stimulus at time step t,
  $t = 1, \ldots 70$
- b_{it}: Activation effect of pixel i at time step t,
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$
- ϵ_{it}: Measurement error of pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $\epsilon \sim N(0, \tau_{Data} I) \rightarrow \; y_{it}|a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it}|a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- z; use estimate from similar studies

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it}|a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF for a: $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j} (a_i - a_j)^2)$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it}|a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF for a: $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j}(a_i - a_j)^2)$
- Time-space GMRF for b:

$$\begin{aligned}\pi(b) \quad &\propto \quad \exp(-\frac{1}{2}\tau_B \sum_{t=1}^{T} \sum_{i \overset{s}{\sim} j} (b_{it} - b_{jt})^2) \\ & \qquad \exp(-\frac{1}{2}\tau_T \sum_{i=1}^{N} \sum_{t \overset{t}{\sim} r} (b_{it} - b_{ir})^2)\end{aligned}$$

Model specification

- $y_{it}|a,b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF for a: $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j}(a_i - a_j)^2)$
- Time-space GMRF for b:

$$\pi(b) \quad \propto \quad \exp(-\frac{1}{2}\tau_B \sum_{t=1}^{T} \sum_{i \overset{s}{\sim} j} (b_{it} - b_{jt})^2)$$

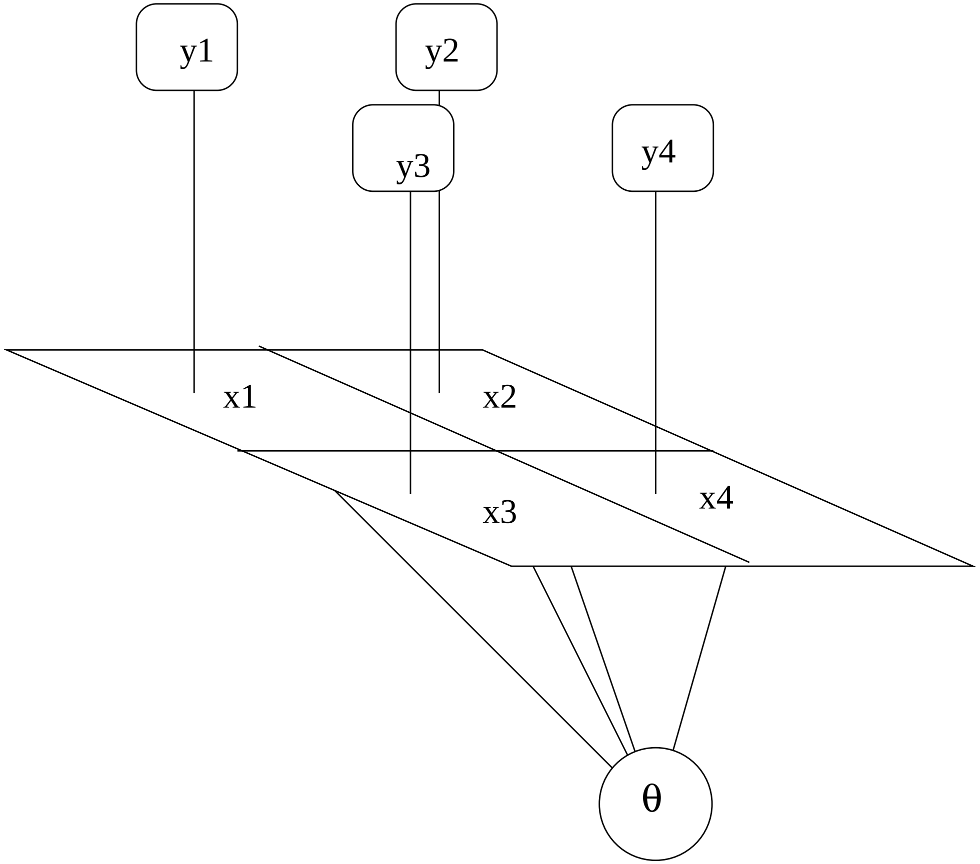
$$\exp(-\frac{1}{2}\tau_T \sum_{i=1}^{N} \sum_{t \overset{t}{\sim} r} (b_{it} - b_{ir})^2)$$

- Posterior, $x = (a, b)$ and $\theta = (\tau_A, \tau_B, \tau_T, \tau_{Data})$:

$$\pi(x, \theta|y) \propto \pi(y|x)\pi(x|\theta)\pi(\theta)$$

Latent GMRF models used

- Mutually independent likelihoods
- $\pi(x|\theta) \sim GMRF$



Metropolis-Hastings with one-block updating scheme

One-block updating scheme: x and θ are updated simultaneously.

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \; q(x|x^j, \theta^{new})$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \; q(x|x^j, \theta^{new})$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \ q(x|x^j, \theta^{new})$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$
- Return $(x^1, x^2, \dots, x^{niter})$ and $(\theta^1, \theta^2, \dots, \theta^{niter})$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \boxed{q(x|x^j, \theta^{new})}$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$
- Return $(x^1, x^2, \ldots, x^{niter})$ and $(\theta^1, \theta^2, \ldots, \theta^{niter})$

Challenge: To make a good and cheap proposal for x.

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
- Part III: *Overlapping bock proposals for latent Gaussian Markov random fields*
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*

  With Håvard Rue and Sveinung Erland, paper
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
- Part III: *Overlapping bock proposals for latent Gaussian Markov random fields*
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
  Technical report
- Part III: *Overlapping bock proposals for latent Gaussian Markov random fields*
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
- Part III: *Overlapping block proposals for latent Gaussian Markov random fields*
  With Håvard Rue, technical report
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
- Part III: *Overlapping block proposals for latent Gaussian Markov random fields*
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

  Note

Contents

- Introduction
- Part I: *Approximating hidden Gaussian Markov random fields*
- Part II: *Parallel sampling of GMRFs and geostatistical GMRF models*
- Part III: *Overlapping block proposals for latent Gaussian Markov random fields*
- Part IV: *Approximation to hidden Gaussian Markov random fields and overlapping block proposals in parallel*

Constructions of approximations to $\pi(x|y, \theta^{new})$ and/or sampling from these distributions.

Part I: Approximating hidden Gaussian Markov random fields

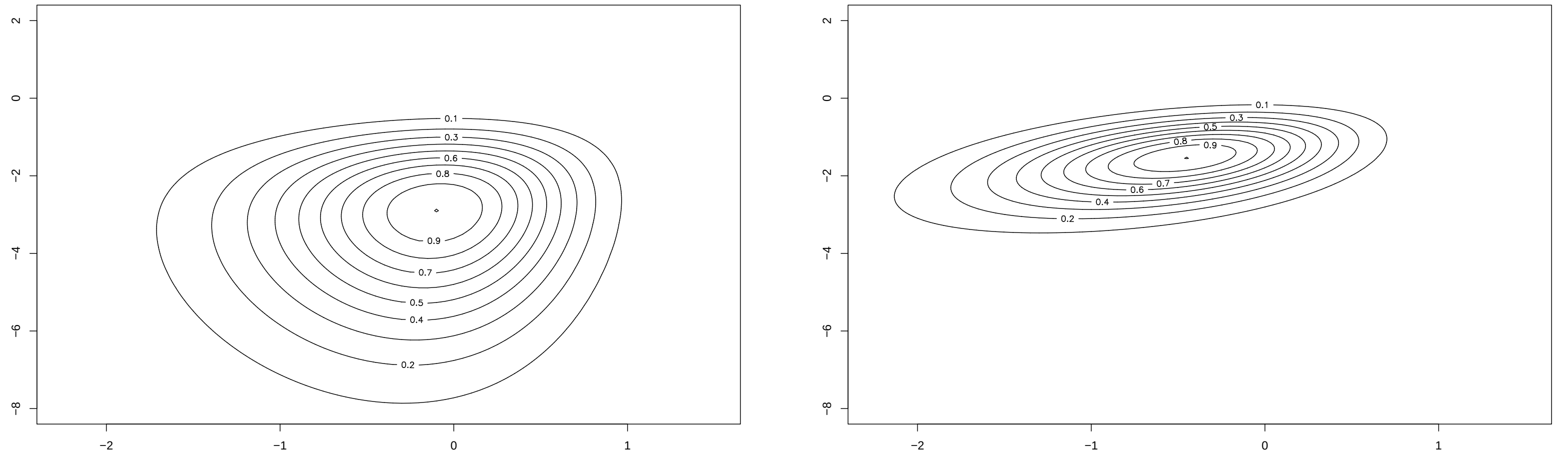
Idea:

- Find an approximation $\widetilde{\pi}(x) \approx \pi(x|y,\theta)$ we can sample from and evaluate.
- Use $q(x|x^{old},\theta^{new}) = \widetilde{\pi}(x)$.

Part I: Approximating hidden Gaussian Markov random fields

$$\pi(x \mid \theta, y) \propto \pi(x \mid \theta) \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i)$$

$$\pi(x|\theta, y) \propto \exp\left(-\frac{1}{2}x^T Q(\theta) x - \sum_i g_i(x_i, y_i)\right)$$



A1: Gaussian approximation at the mode

- If $\pi(y|x,\theta)$ is log-concave then $\pi(x|y,\theta)$ is unimodal.
- Can use Gaussian approximation at the mode, x^m.

A1: Gaussian approximation at the mode

- If $\pi(y|x,\theta)$ is log-concave then $\pi(x|y,\theta)$ is unimodal.
- Can use Gaussian approximation at the mode, x^m.
- Replace $g_i(x_i, y_i)$ by the Taylor expansion at the mode, $a_i + b_i x_i + c_i x_i^2$.

A1: Gaussian approximation at the mode

- If $\pi(y|x,\theta)$ is log-concave then $\pi(x|y,\theta)$ is unimodal.
- Can use Gaussian approximation at the mode, x^m.
- Replace $g_i(x_i, y_i)$ by the Taylor expansion at the mode, $a_i + b_i x_i + c_i x_i^2$.
- $\widetilde{\pi}^{A1}(x)$ is $GMRF(x^m, Q(\theta) + c)$

Improved approximations

$$\pi(x \mid \theta, y) \;=\; \prod_{t=1}^{n} \pi(x_t \mid x_{(t+1):n}, \theta, y)$$

where $x_{t+1:n} = (x_{t+1}, x_{t+2}, \ldots, x_n)$.

Improved approximations

$$\begin{aligned}\pi(x \mid \theta, y) &= \prod_{t=1}^{n} \pi(x_t \mid x_{(t+1):n}, \theta, y) \\ &\propto \prod_{t=1}^{n} \widetilde{\pi}_{A1}(x_t \mid x_{(t+1):n}, \theta, y) \exp\left(-h_t(x_t, y_t)\right)\end{aligned}$$

where

$$h_t(x_t, y_t) = g_t(x_t, y_t) - (a_t + b_t x_t + c_t x_t^2/2).$$

Marginal conditional approximations

$$
\begin{aligned}
\pi(x_t \mid x_{(t+1):n}, \theta, y) &\propto \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \\
&\times \exp\left(-h_t(x_t, y_t)\right) \\
&\times \int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \\
&\quad \exp(-\sum_{j=1}^{t-1} h_j(x_j, y_j))dx_{1:(t-1)}
\end{aligned}
$$

Marginal conditional approximations

$$
\begin{aligned}
\pi(x_t \mid x_{(t+1):n}, \theta, y) \quad &\propto \quad \color{blue}{\pi_G(x_t \mid x_{(t+1):n}, \theta, y)} \\
&\times \quad \exp\left(-h_t(x_t, y_t)\right) \\
&\times \quad \int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \\
&\qquad \exp(-\sum_{j=1}^{t-1} h_j(x_j, y_j)) dx_{1:(t-1)},
\end{aligned}
$$

- A1: Gaussian approximation at the mode.

Marginal conditional approximations

$$\begin{aligned}
\pi(x_t \mid x_{(t+1):n}, \theta, y) &\propto \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \\
&\times \exp\left(-h_t(x_t, y_t)\right) \\
&\times \int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \\
&\qquad \exp(-\sum_{j=1}^{t-1} h_j(x_j, y_j)) dx_{1:(t-1)}
\end{aligned}$$

- A1: Gaussian approximation at the mode.
- A2: Also correct for $h_t(x_i, y_i)$ when sampling x_i.

Marginal conditional approximations

$$
\begin{aligned}
\pi(x_t \mid x_{(t+1):n}, \theta, y) \quad &\propto \quad \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \\
&\times \quad \exp\left(-h_t(x_t, y_t)\right) \\
&\times \quad \int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \\
&\qquad \exp(-\sum_{j=1}^{t-1} h_j(x_j, y_j)) dx_{1:(t-1)}
\end{aligned}
$$

- A1: Gaussian approximation at the mode.
- A2: Also correct for $h_t(x_i, y_i)$ when sampling x_i.
- A3: Also correct for $\int \ldots$.

A2: Correcting for current likelihood-term

- Want to sample from and evaluate

$$\begin{aligned}\pi_{A2}(x_t \mid x_{(t+1):n}, \theta, y) \;\;\propto\;\; & \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \\ & \exp\Big(- h_t(x_t, y_t)\Big)\end{aligned}$$

A2: Correcting for current likelihood-term

- Want to sample from and evaluate

$$\pi_{A2}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \exp\Big(- h_t(x_t, y_t)\Big)$$

- For evaluation reasons we use a log spline approximation between knots instead:

$$\widetilde{\pi}_{A2}(x_t|x_{t+1:n}, \theta, y) \approx \pi_{A2}(x_t|x_{t+1:n}, \theta, y)$$

A2: Correcting for current likelihood-term

- Want to sample from and evaluate

$$\pi_{A2}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \pi_G(x_t \mid x_{(t+1):n}, \theta, y) \exp\Big(- h_t(x_t, y_t)\Big)$$

- For evaluation reasons we use a log spline approximation between knots instead:

$$\widetilde{\pi}_{A2}(x_t|x_{t+1:n}, \theta, y) \approx \pi_{A2}(x_t|x_{t+1:n}, \theta, y)$$

- Tuning parameter: K, number of knots.

A3: Correcting for non-sampled non-Gaussian likelihood terms

- Want to sample from

$$\pi_{A3}(x_t \mid x_{(t+1):n}, \theta, y) \;\; \propto \;\; \widetilde{\pi}_{A2}(x_t | x_{t+1:n}, \theta, y) I(x_t)$$

where $I(x_t)$ is

$$\int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \exp\left(- \sum_{j=1}^{t-1} h_j(x_j, y_j) \right) dx_{1:(t-1)}$$

A3: Correcting for non-sampled non-Gaussian likelihood terms

- Want to sample from

$$\pi_{A3}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \widetilde{\pi}_{A2}(x_t | x_{t+1:n}, \theta, y) I(x_t)$$

where $I(x_t)$ is

$$\int \pi_G(x_{1:(t-1)} \mid x_{t:n}, \theta, y) \exp\left(-\sum_{j=1}^{t-1} h_j(x_j, y_j)\right) dx_{1:(t-1)}$$

$$I(x_t) \;=\; \mathrm{E}\Big[\exp\Big(-\sum_{j=1}^{t-1} h_j(x_j, y_j)\Big)\Big]$$

A3: Correcting for non-sampled non-Gaussian likelihood terms

- Want to sample from

$$\pi_{A3}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \widetilde{\pi}_{A2}(x_t|x_{t+1:n}, \theta, y)I(x_t)$$

- Approximate $I(x_t)$ by sampling:

$$\widehat{I}(x_t) \;=\; \frac{1}{M}\sum_{i=1}^{M}\exp\Big(-\sum_{j\in\mathcal{J}(t)} h_j(x_j^i, y_j)\Big)$$

A3: Correcting for non-sampled non-Gaussian likelihood terms

- Want to sample from

$$\pi_{A3}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \widetilde{\pi}_{A2}(x_t|x_{t+1:n}, \theta, y) I(x_t)$$

- Approximate $I(x_t)$ by sampling:

$$\widehat{I}(x_t) \;=\; \frac{1}{M}\sum_{i=1}^{M}\exp\left(-\sum_{j\in\mathcal{J}(t)} h_j(x_j^i, y_j)\right)$$

- Use only some of the nearest unsampled neighbours.

A3: Correcting for non-sampled non-Gaussian likelihood terms

- Want to sample from

$$\pi_{A3}(x_t \mid x_{(t+1):n}, \theta, y) \;\propto\; \widetilde{\pi}_{A2}(x_t | x_{t+1:n}, \theta, y) I(x_t)$$

- Approximate $I(x_t)$ by sampling:

$$\widehat{I}(x_t) \;=\; \frac{1}{M} \sum_{i=1}^{M} \exp\Big(- \sum_{j \in \mathcal{J}(t)} h_j(x_j^i, y_j) \Big)$$

- Use only some of the nearest unsampled neighbours.
- Tuning parameters: Number of samples, M and neighbours $|\mathcal{J}(t)|$.

Results: Disease mapping in Germany

Average accept-rate	$\kappa = 0.1$	$\kappa = 1$	$\kappa = 10$
$A1$	0.01	0.11	0.47
$A2$	0.94	0.80	0.78
$A3a$	0.96	0.87	0.86
$A3b$	0.99	0.96	0.90

- For $A2$ and $A3$: $K = 20$ knots.
- $A3a$: $M = 1$
- $A3b$: $M = 100$

Approximating $\pi(\kappa|y)$

- Use

$$\pi(\kappa \mid y) = \frac{\pi(x, \kappa \mid y)}{\pi(x \mid \kappa, y)}$$

for any x.

Approximating $\pi(\kappa|y)$

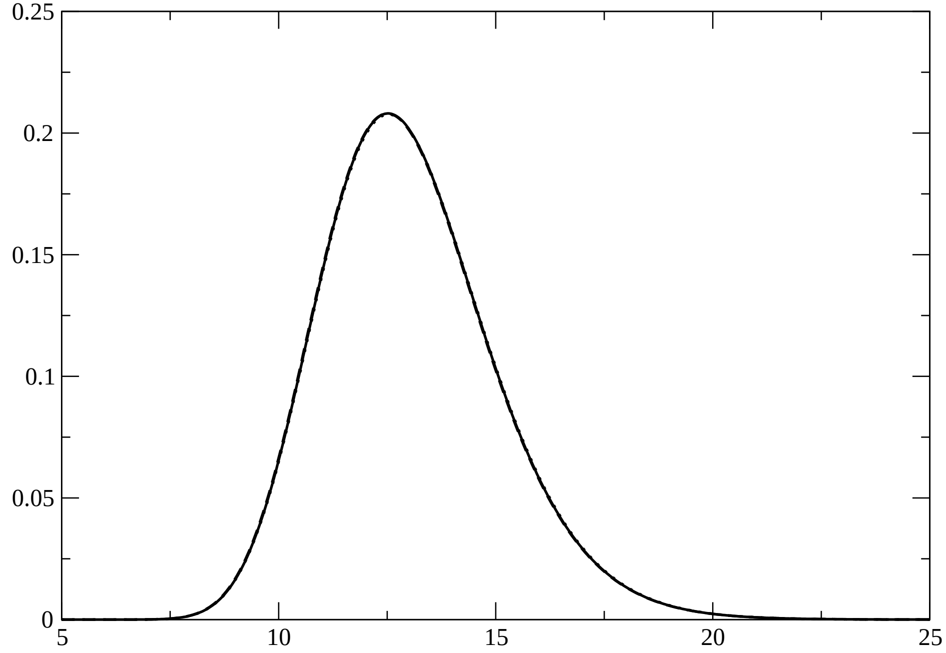
- Use
$$\pi(\kappa \mid y) = \frac{\pi(x, \kappa \mid y)}{\pi(x \mid \kappa, y)}$$
for any x.
- Use $\widetilde{\pi}(x|\kappa, y, x^{\mathrm{m}}(\kappa))$
$$\widetilde{\pi}(\kappa \mid y) \propto \left. \frac{\pi(\kappa)\pi(x|\kappa)\pi(y|x)}{\widetilde{\pi}(x|\kappa, y, x^{\mathrm{m}}(\kappa))} \right|_{x=x^{\mathrm{m}}(\kappa)}$$

Approximating $\pi(\kappa|y)$

- Use
$$\pi(\kappa \mid y) = \frac{\pi(x, \kappa \mid y)}{\pi(x \mid \kappa, y)}$$
for any x.
- Use $\widetilde{\pi}(x|\kappa, y, x^{\mathrm{m}}(\kappa))$
$$\widetilde{\pi}(\kappa \mid y) \propto \left. \frac{\pi(\kappa)\pi(x|\kappa)\pi(y|x)}{\widetilde{\pi}(x|\kappa, y, x^{\mathrm{m}}(\kappa))} \right|_{x=x^{\mathrm{m}}(\kappa)}$$
- Can use $q(\kappa|\kappa^{old}) = \widetilde{\pi}(\kappa \mid y)$
$\Rightarrow$ Metropolised independence sampler

Results: Approximating $\pi(\kappa|y)$

- Estimated $\pi(\kappa|y)$ using $A1$, $A2$ and $A3a$:



- Acceptance rates Metropolised independence samplers: $A1$: 0.43, $A2$: 0.82 and $A3a$: 0.86.

Part III: Overlapping block proposals for latent GMRFs

- Background: Too expensive to sample from an n-dim. distribution.
- Here: Construct a proposal for x which we can sample from and evaluate working only with smaller blocks.

Proposal from blocks

- Can use one full scan of a Gibbs sampler as $q(x|x^{old}, \theta^{new})$.
- Block Gibbs sampler used to get better convergence.

Traditional blocking

- x: 100×100 GMRF, $E(x) = 0$ but $x^0 = 3$
- 5×5 neighbourhood
- A GMRF approximation to correlation function:

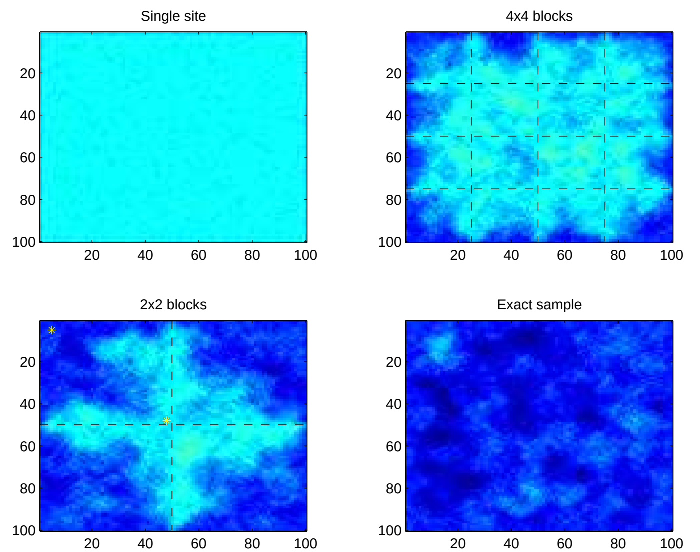
$$\rho(x_i, x_j) = \exp(\frac{-3d(x_i, x_j)}{r})$$

Traditional blocking

- x: 100×100 GMRF, $E(x) = 0$ but $x^0 = 3$
- 5×5 neighbourhood
- A GMRF approximation to correlation function:

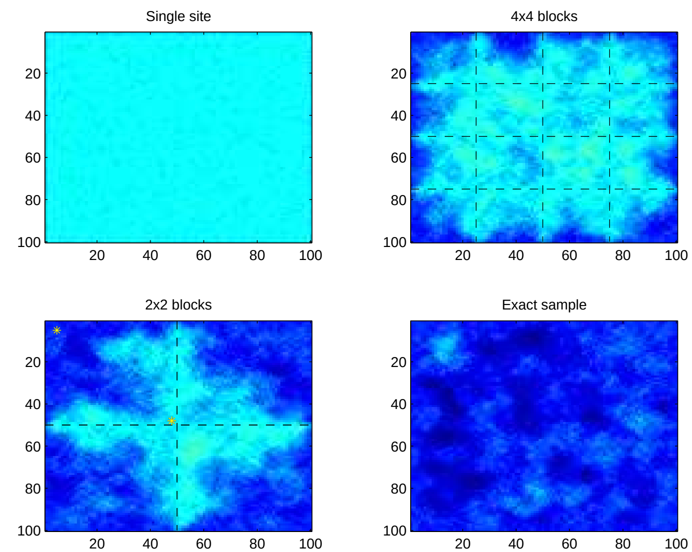
$$\rho(x_i, x_j) = \exp(\frac{-3d(x_i, x_j)}{r})$$

First iteration ($r = 40$)

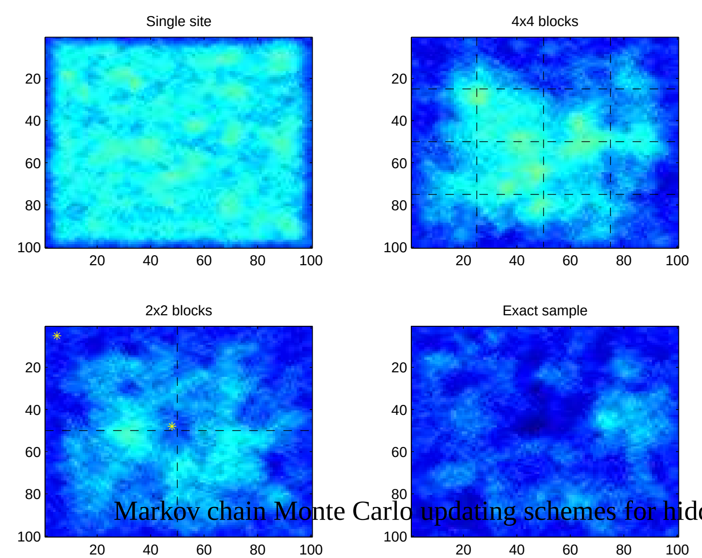


Traditional blocking

First iteration ($r = 40$)

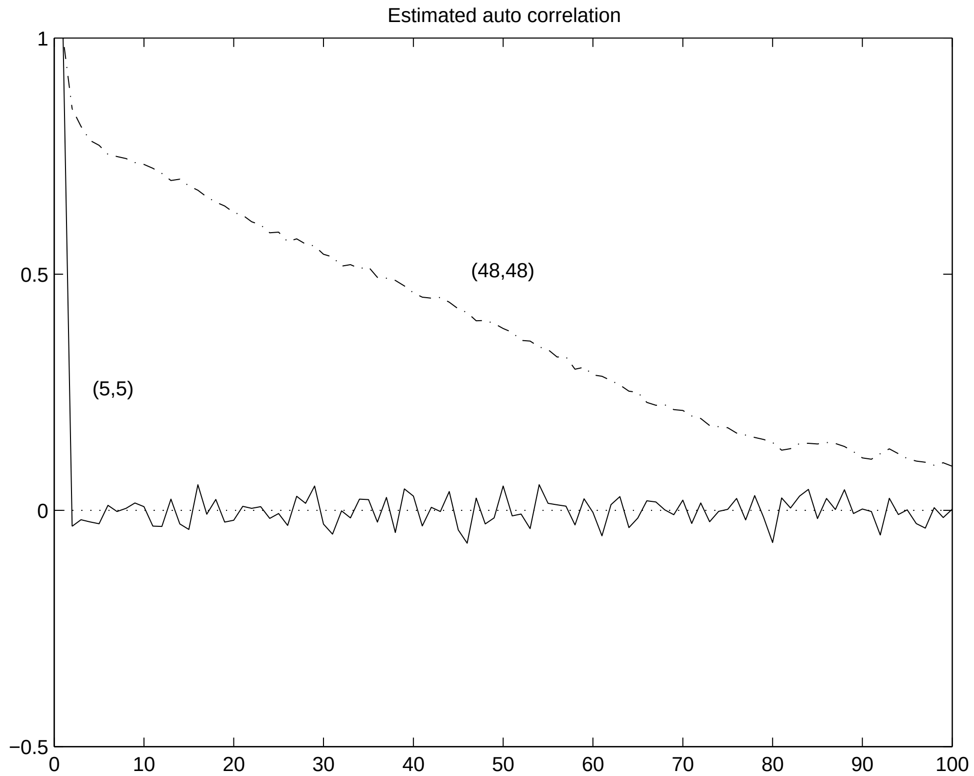


200th iteration ($r = 40$)



Traditional blocking

Estimated autocorrelation:

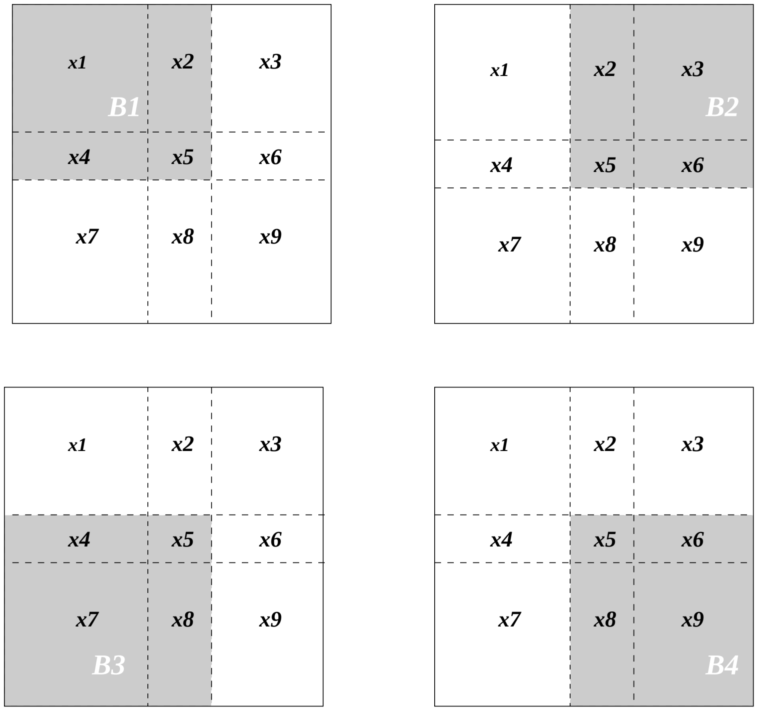


Overlapping block Gibbs sampler

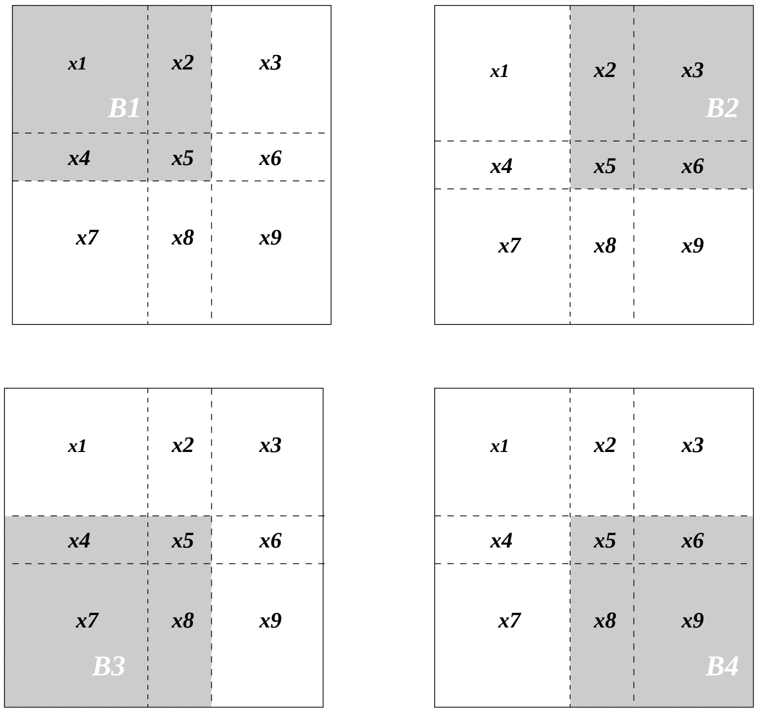
Idea: Let the blocks overlap.

Overlapping block Gibbs sampler

Idea: Let the blocks overlap.

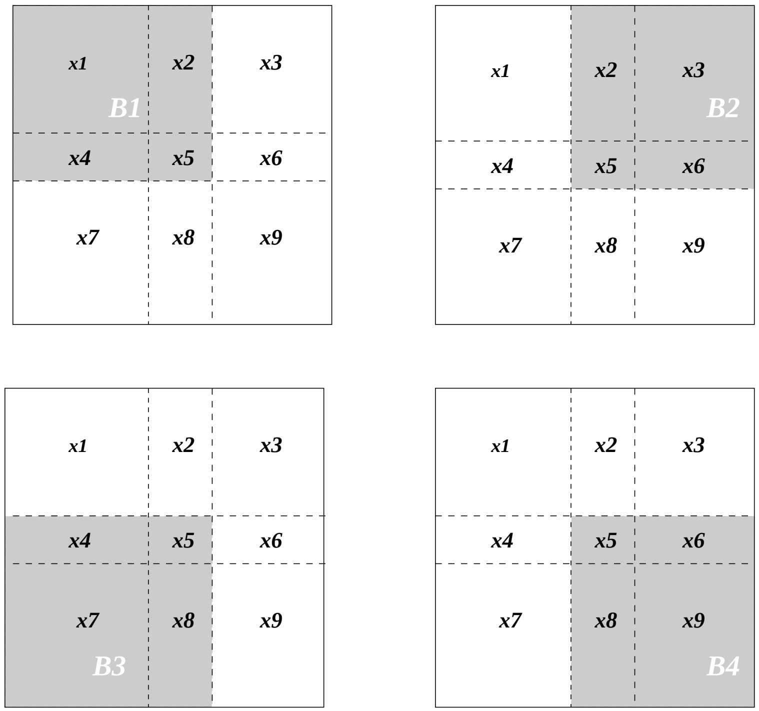


Overlapping block Gibbs sampler



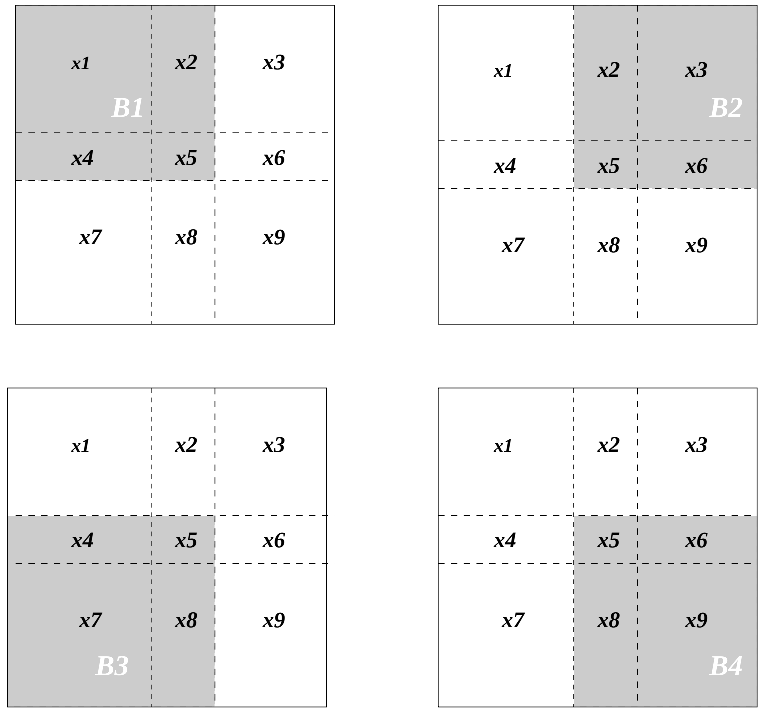
- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$

Overlapping block Gibbs sampler



- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}, x_5^{B2}, x_6^{B2}) \sim \pi(x_{B_2} | x_1^{i+1}, x_4^{B1}, x_7^i, x_8^i, x_9^i)$

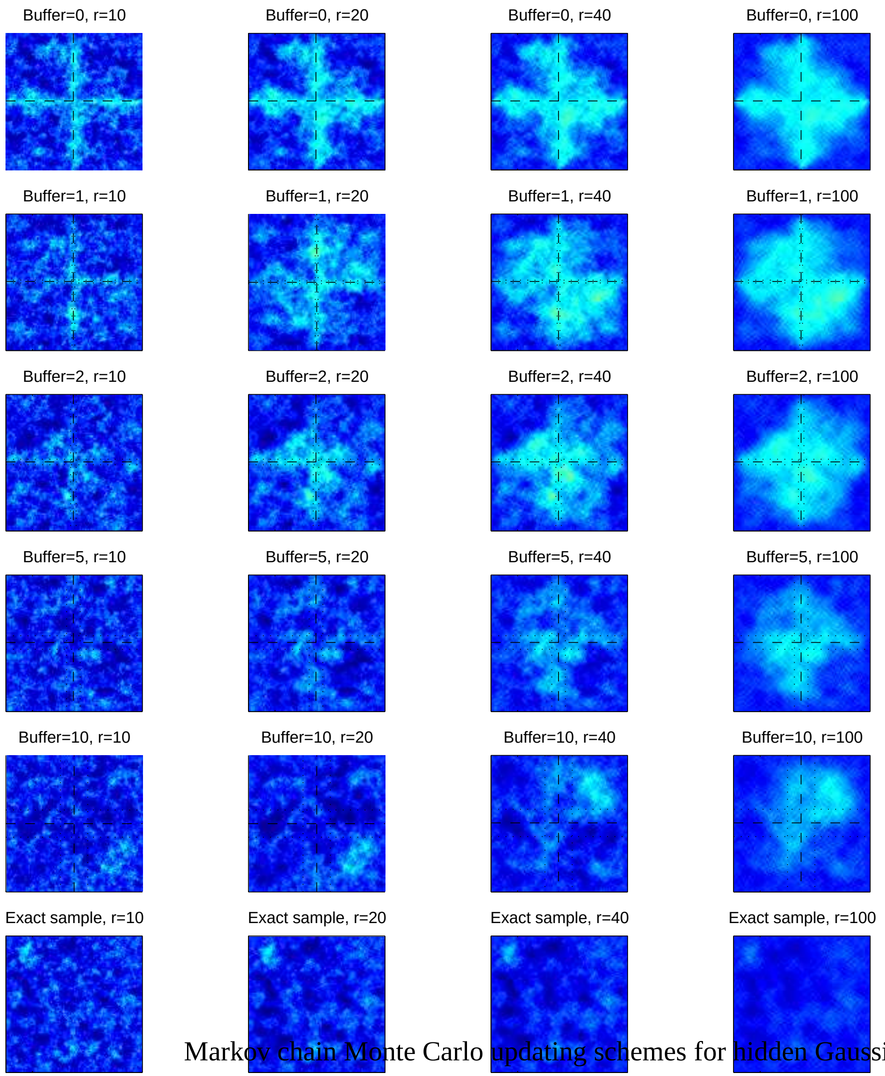
Overlapping block Gibbs sampler



- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}, x_5^{B2}, x_6^{B2}) \sim \pi(x_{B_2} | x_1^{i+1}, x_4^{B1}, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_4^{i+1}, x_5^{B3}, x_7^{i+1}, x_8^{B3}) \sim \pi(x_{B_3} | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_6^{B2}, x_9^i)$
  - Sample $(x_5^{i+1}, x_6^{i+1}, x_8^{i+1}, x_9^{i+1}) \sim \pi(x_{B_4} | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_4^{i+1}, x_7^{i+1})$
- Return x^{i+1}

Does it work?

As previous example with $r = \{10, 20, 40, 100\}$ and buffer $= \{0, 1, 2, 5, 10\}$

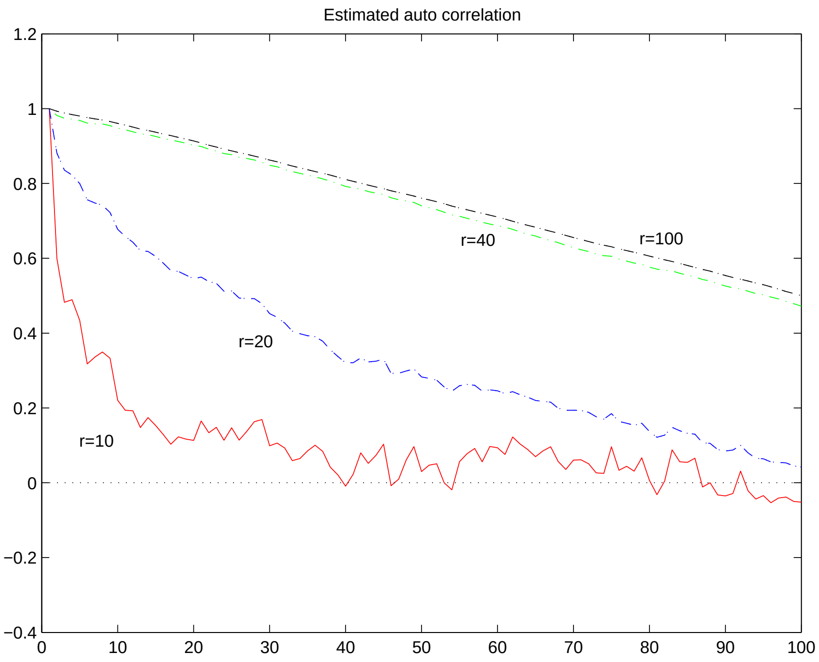


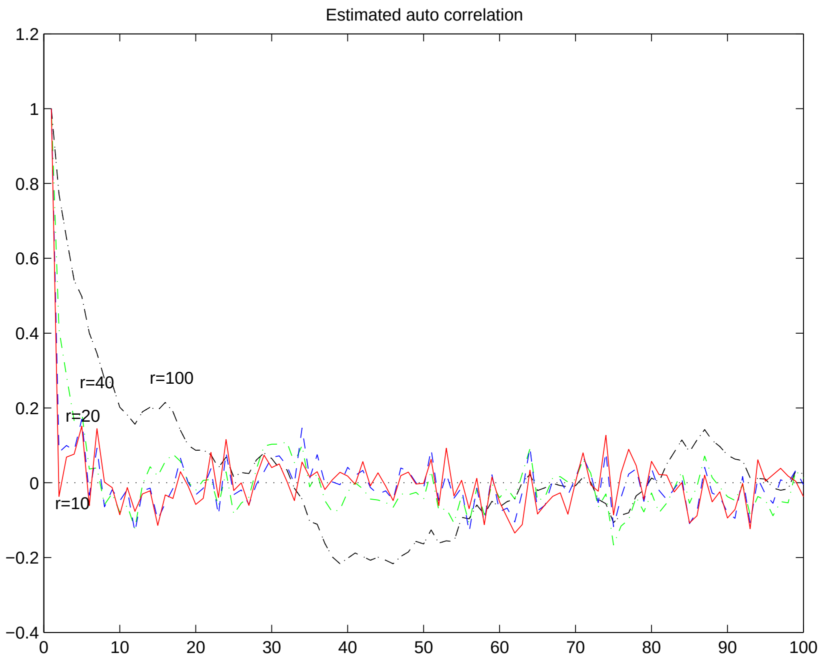
Does it work?

Estimated auto-correlation function at pixel $(48, 48)$.

Left: Block Gibbs without buffers

Right: With buffer five



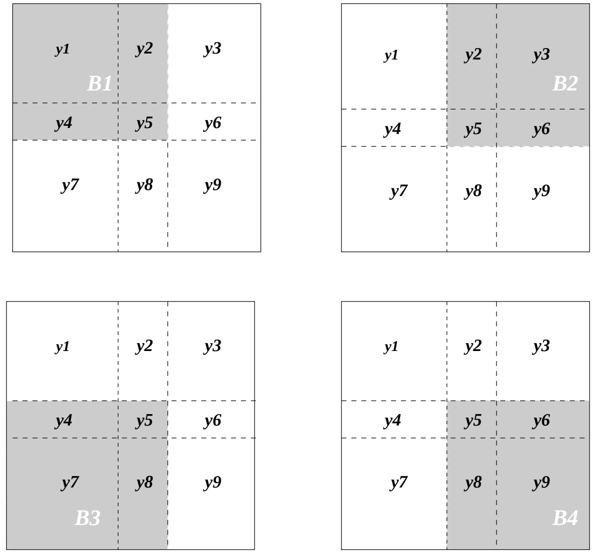


Transition probability

Hard to calculate the transition probability:

Transition probability

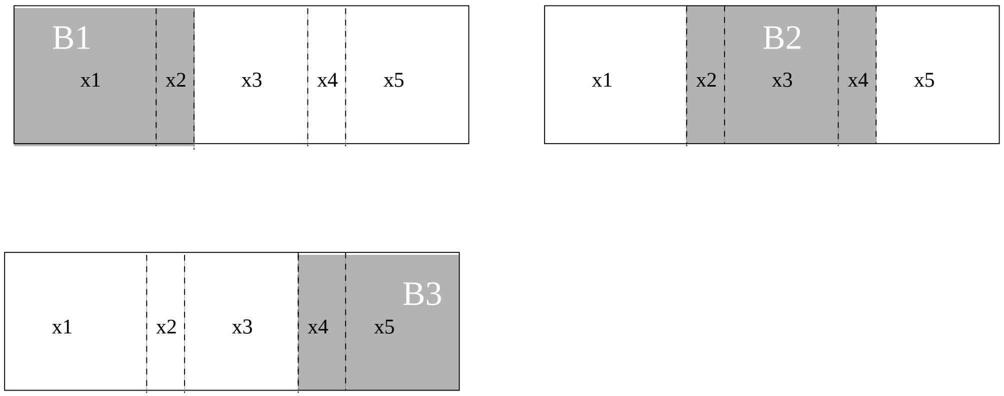
Hard to calculate the transition probability:



$$
\begin{aligned}
q(x|x') \quad = \quad & \int [\pi_1(x_1', x_2^{B1}, x_4^{B1}, x_5^{B1} | x_3, x_6, x_7, x_8, x_9) \\
& \pi_2(x_2', x_3', x_5^{B2}, x_6^{B2} | x_1', x_4^{B1}, x_7, x_8, x_9) \\
& \pi_3(x_4', x_5^{B3}, x_7', x_8^{B3} | x_1', x_2', x_3', x_6^{B2}, x_9) \\
& \pi_4(x_5', x_6', x_8', x_9' | x_1', x_2', x_3', x_4', x_7')] dx_2^{B1} dx_4^{B1} dx_5^{B1} dx_5^{B2} dx_5^{B3} dx_6^{B2} dx_8^{B3}
\end{aligned}
$$

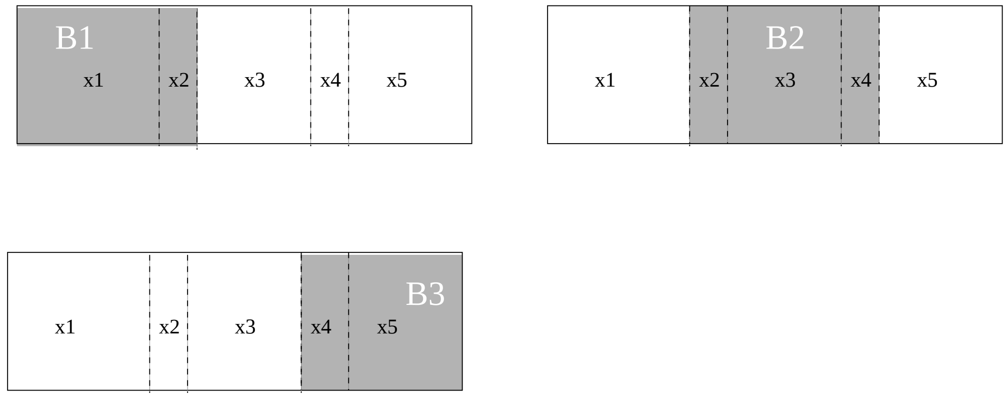
Transition probability

Time series blocking:



Transition probability

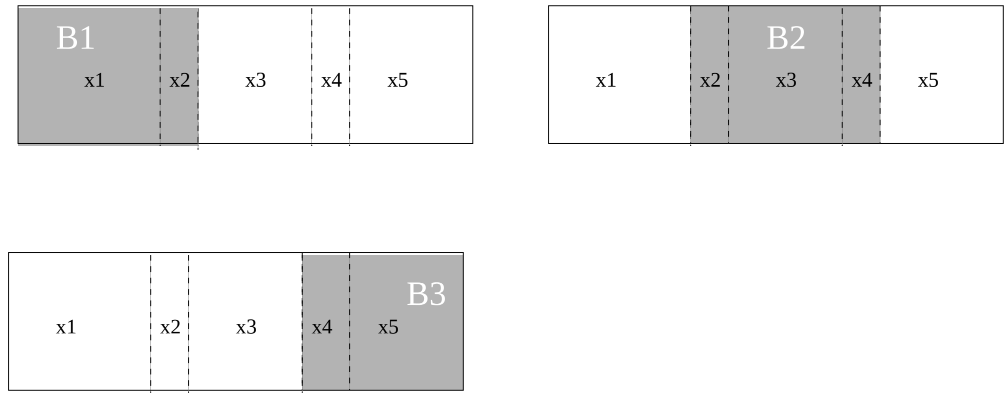
Time series blocking:



$$
\begin{aligned}
q(x|x^{'}) \quad = \quad & \int [\pi_1(x_1^{'}, x_2^{B1} | x_3, x_4, x_5) \\
& \pi_2(x_2^{'}, x_3^{'}, x_4^{B2} | x_1^{'}, x_5) \\
& \pi_3(x_4^{'}, x_5^{'} | x_1^{'}, x_2^{'}, x_3^{'})] dx_2^{B1} dx_4^{B2}
\end{aligned}
$$

Transition probability

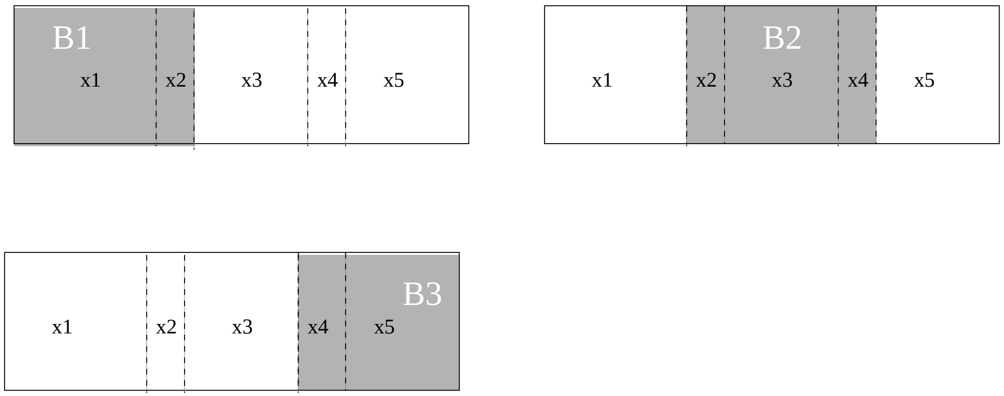
Time series blocking:



$$
\begin{aligned}
q(x|x') &= \int [\pi_1(x_1', x_2^{B1} | x_3, x_4, x_5) \\
&\quad \pi_2(x_2', x_3', x_4^{B2} | x_1', x_5) \\
&\quad \pi_3(x_4', x_5' | x_1', x_2', x_3')] dx_2^{B1} dx_4^{B2} \\
&= \pi(x_1' | x_3, x_4, x_5) \cdot \pi(x_2', x_3' | x_1', x_5) \cdot \pi(x_4', x_5' | x_1', x_2', x_3')
\end{aligned}
$$

Transition probability

Time series blocking:



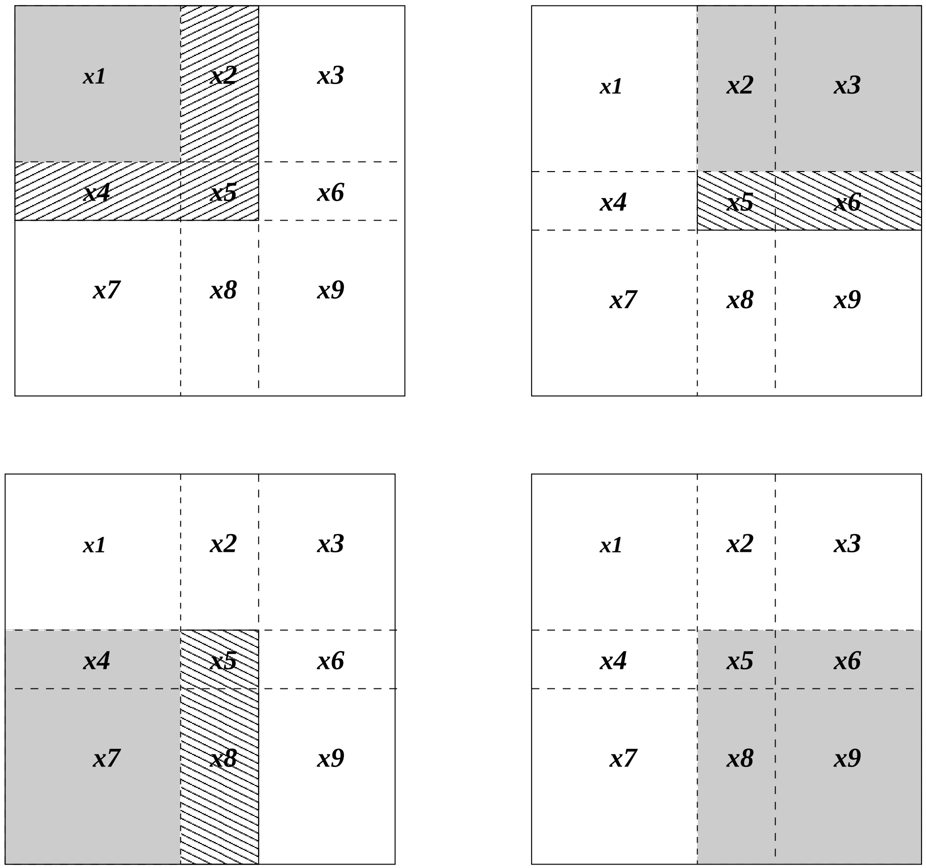
$$q(x^{'}|x) \quad = \quad \pi(x_1^{'}|x_3, x_4, x_5) \cdot \pi(x_2^{'}, x_3^{'}|x_1^{'}, x_5) \cdot \pi(x_4^{'}, x_5^{'}|x_1^{'}, x_2^{'}, x_3^{'})$$

Can use that:

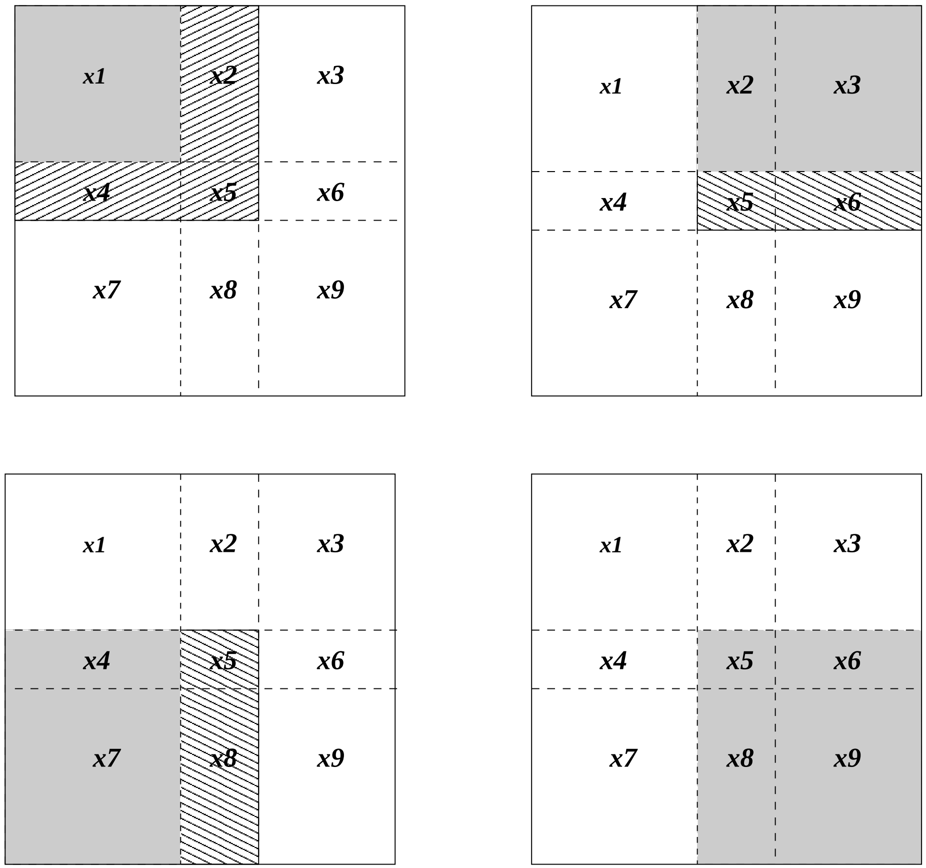
$$\pi(x_1^{'}|x_3, x_4, x_5) = \frac{\pi(x_1^{'}, x_2^{B1}|x_3, x_4, x_5)}{\pi(x_2^{B1}|x_1^{'}, x_3, x_4, x_5)}$$

for any x_2^{B1}

Partial conditional block sampler

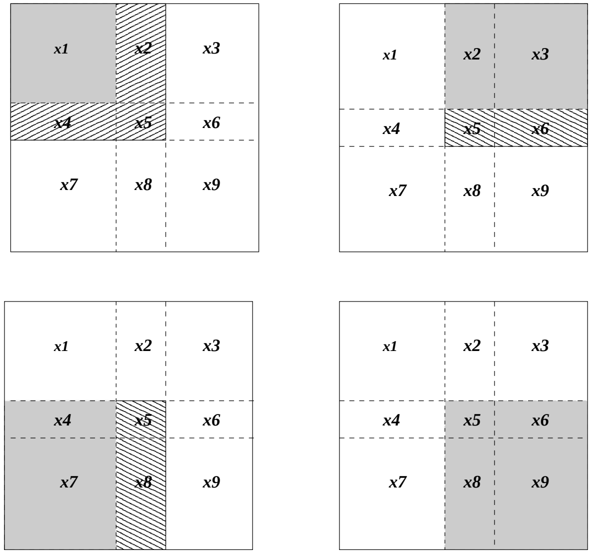


Partial conditional block sampler



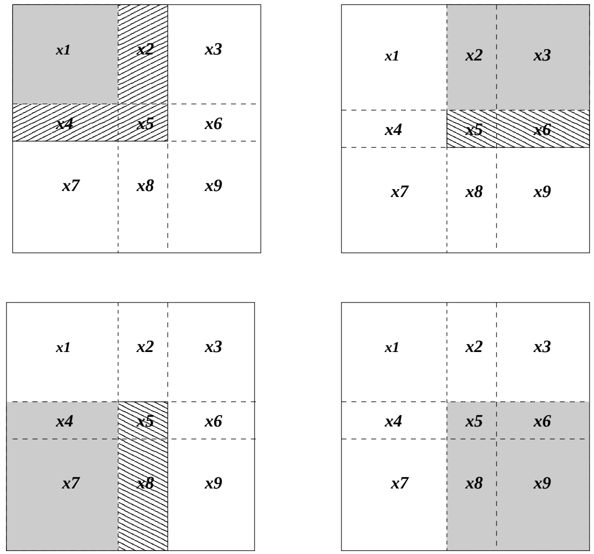
- Given x^0
- for $i = 0 : (niter - 1)$
  - Sample $(x_1^{i+1}) \sim \pi(x_1|x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$

Partial conditional block sampler



- Given x^0
- for $i = 0 : (niter - 1)$
  - Sample $(x_1^{i+1}) \sim \pi(x_1 | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}) \sim \pi(x_2, x_3 | x_1^{i+1}, x_4^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_4^{i+1}, x_7^{i+1}) \sim \pi(x_4, x_7 | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_6^i, x_9^i)$
  - Sample $(x_5^{i+1}, x_6^{i+1}, x_8^{i+1}, x_9^{i+1}) \sim \pi(x_5, x_6, x_8, x_9 | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_4^{i+1}, x_7^{i+1})$
- Return $((x_1^1, x_2^1, \ldots, x_9^1), (x_1^2, x_2^2, \ldots, x_9^2), \ldots, (x_1^{niter}, x_2^{niter}, \ldots, x_9^{niter}))$

Partial conditional block sampler



Transition probability:

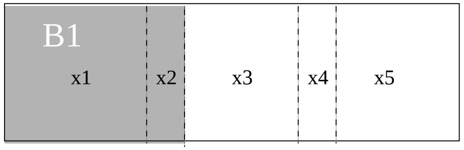
$$
\begin{aligned}
q(x|x^{'}) \;&=\; \pi(x_1'|x_3, x_6, x_7, x_8, x_9) \\
&\quad\; \pi(x_2', x_3'|x_1', x_4, x_7, x_8, x_9) \\
&\quad\; \pi(x_4', x_7'|x_1', x_2', x_3', x_6, x_9) \\
&\quad\; \pi(x_5', x_6', x_8', x_9'|x_1', x_2', x_3', x_4', x_7')
\end{aligned}
$$

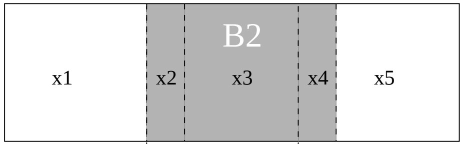
Opposite reversal

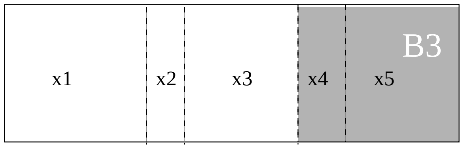
- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.

Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.
- Sample first a direction $i = \{0, 1\}$
  - if $i == 0$ use $q_0 : B_1 \rightarrow B_2 \rightarrow B_3$
  - if $i == 1$ use $q_1 : B_3 \rightarrow B_2 \rightarrow B_1$







Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.
- Sample first a direction $i = \{0, 1\}$
  - if $i == 0$ use $q_0 : B_1 \rightarrow B_2 \rightarrow B_3$
  - if $i == 1$ use $q_1 : B_3 \rightarrow B_2 \rightarrow B_1$
- use acceptance probability

$$\alpha_{i,1-i}(y|x) = \min\left\{1, \frac{\pi(x')q_{1-i}(x|x')}{\pi(x)q_i(x'|x)}\right\}$$

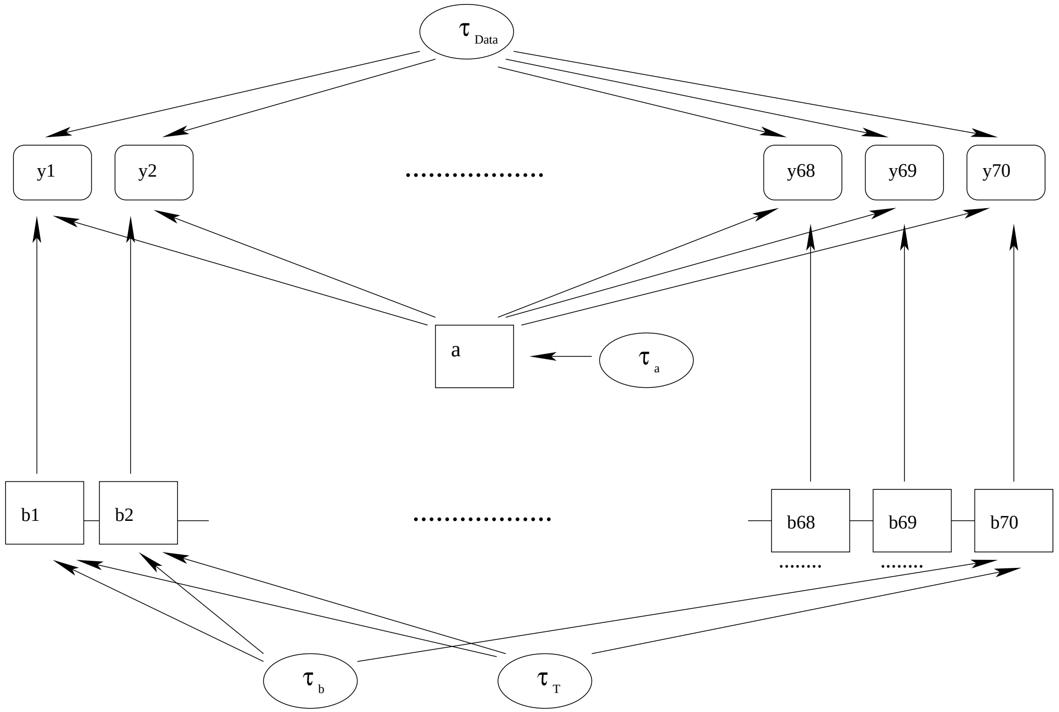
Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.
- Sample first a direction $i = \{0, 1\}$
  - if $i == 0$ use $q_0 : B_1 \rightarrow B_2 \rightarrow B_3$
  - if $i == 1$ use $q_1 : B_3 \rightarrow B_2 \rightarrow B_1$
- use acceptance probability

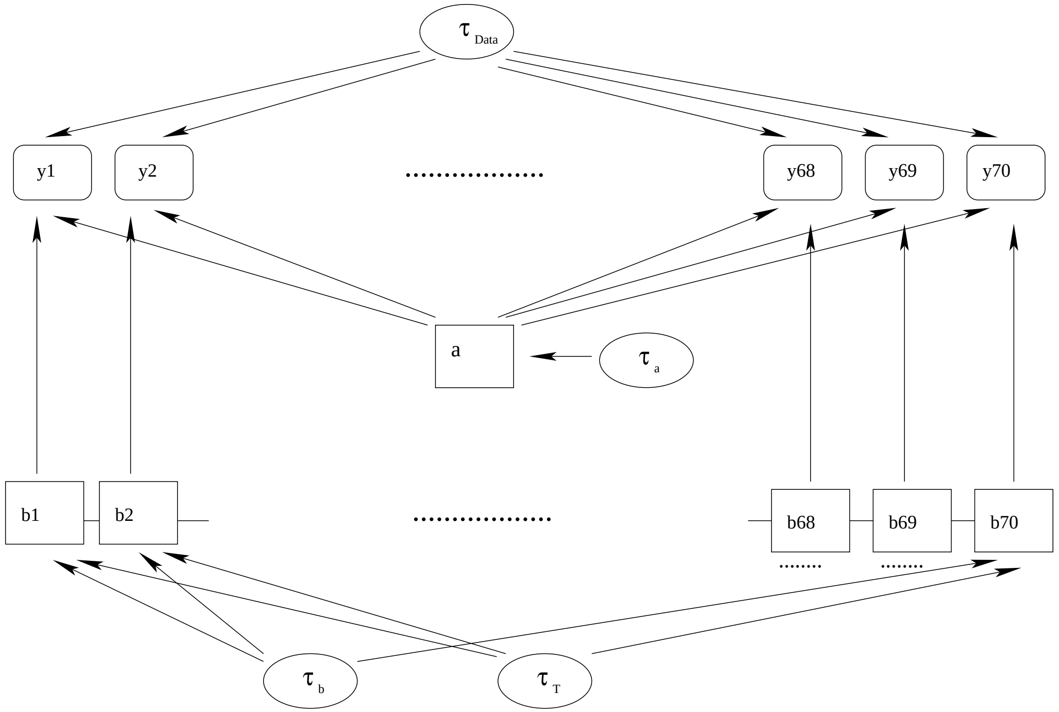
$$\alpha_{i,1-i}(y|x) = \min\left\{1, \frac{\pi(x')q_{1-i}(x|x')}{\pi(x)q_i(x'|x)}\right\}$$

- This gives $\alpha = 1$ for overlapping block Gibbs proposal, but generally not for a partial conditioning sampler

DAG fMRI



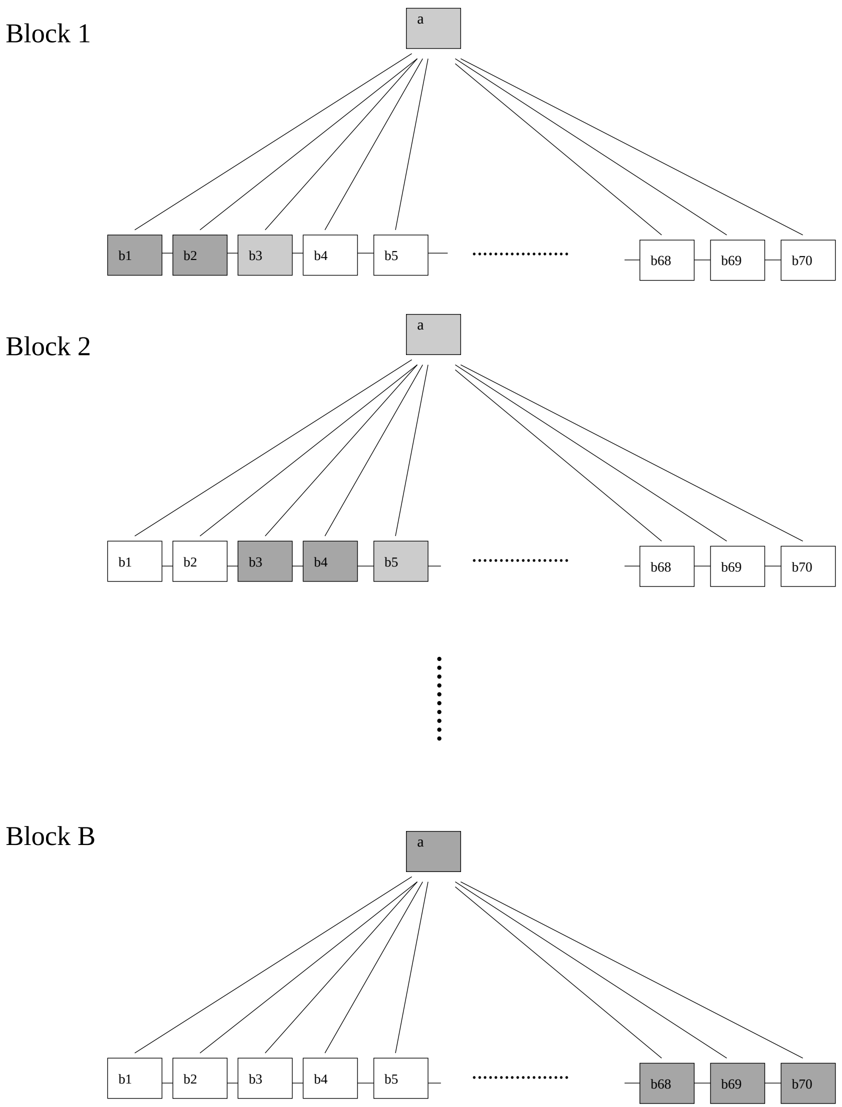
DAG fMRI



- Dimension (a, b) full problem 356 775.
- Dimension (a, b) reduced problem 111 825.

Solution fMRI

Sampling scheme:



Algorithm

- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$ $niter = 20000$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$ Independent random walk, τ_{Data} estimated beforehand

Algorithm

- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping block Gibbs proposal $x^{new} \sim q_i(x|x^{old}, \theta^{new})$
    - Each block: a and five b_t.
    - Overlap: a and two b_t.

Algorithm

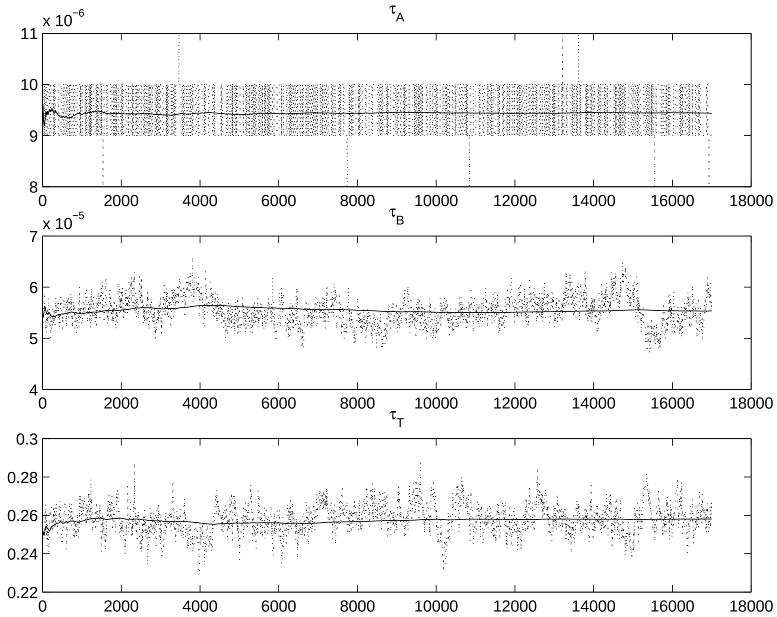
- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping block Gibbs proposal $x^{new} \sim q_i(x|x^{old}, \theta^{new})$
  - Calculate acceptance probability

$$\alpha = \min(1, \frac{\pi(y|x^{new})\pi(x^{new}|\theta^{new})\pi(\theta^{new})q(\theta^j|\theta^{new})q_i(x^j|x^{new}, \theta^j)}{\pi(y|x^j)\pi(x^j|\theta^j)\pi(\theta^j)q(\theta^{new}|\theta^j)q_{1-i}(x^{new}|x^j, \theta^{new})})$$

  - Sample $u \sim \text{Unif}(0, 1)$
  - if$(u < \alpha)$
    - $\theta^{j+1} = \theta^{new}$
    - $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$
    - $x^{j+1} = x^j$
- Return $((\theta^1, x^1), (\theta^2, x^2), \ldots, (\theta^n, x^n))$.

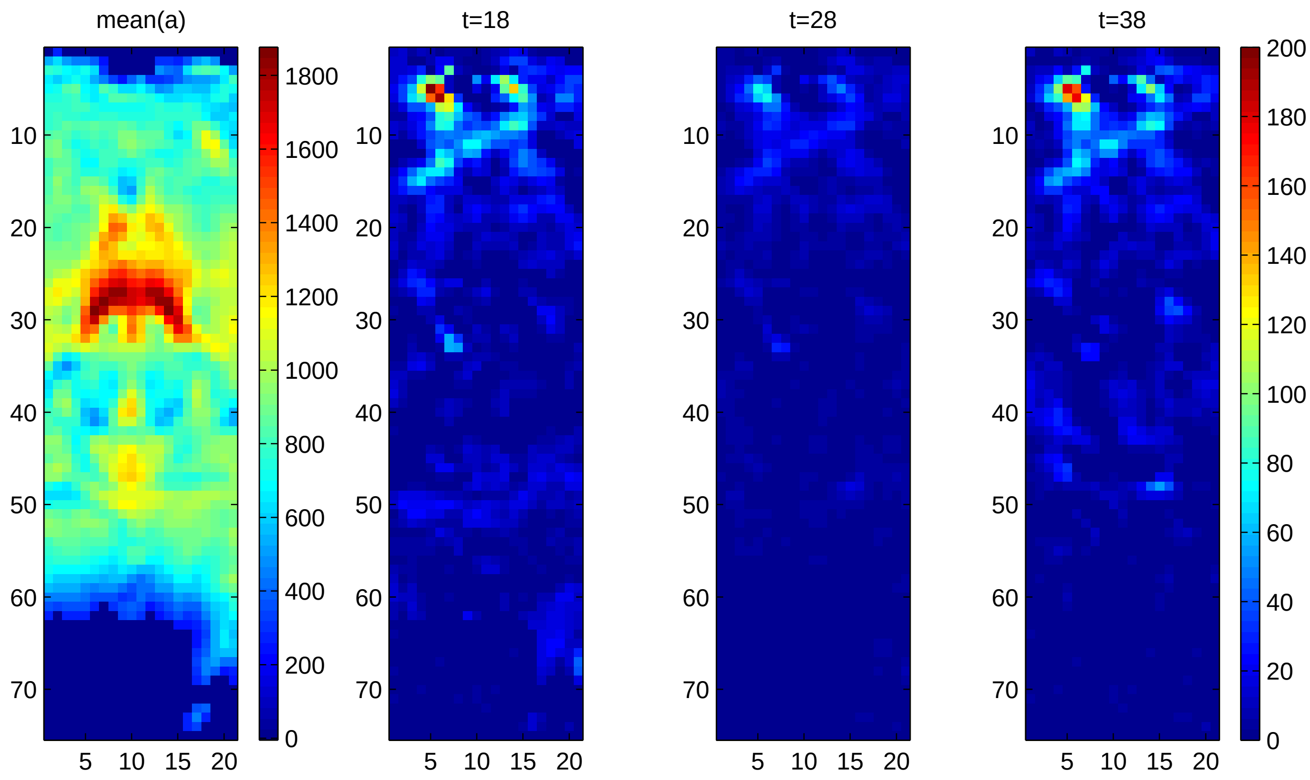
Results fMRI

Trace plots hyper-parameters:



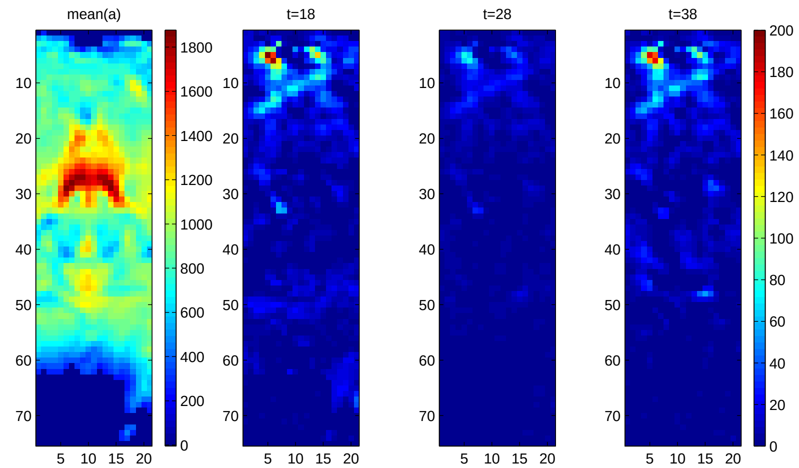
Results fMRI

Estimated mean a and b_{18}, b_{28} and b_{38}:

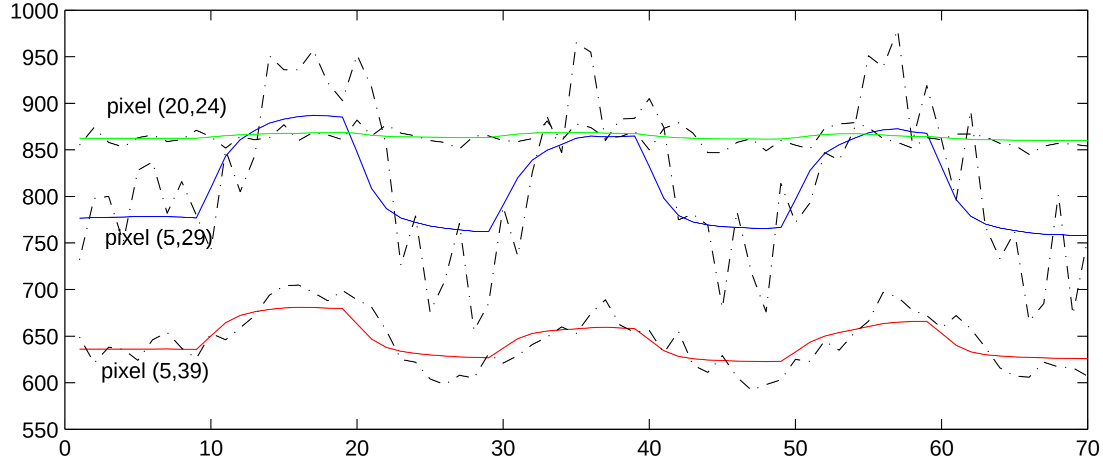


Results fMRI

Estimated mean a and b_{18}, b_{28} and b_{38}:



Estimate for some pixels in time:



Overlapping approximated blocks proposal

- Can sample each block from an approximation to $\pi(x_B|x_{-B})$.
- Enable us to make inference from hidden GMRF models (as in part I).
- Use approximations from Part I.

Overlapping approximated blocks proposal

- Given θ^0 and y^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta^{new}|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping approximated blocks proposal $x^{new} \sim q_i^A(x|x^{old}, \theta^{new})$.

Overlapping approximated blocks proposal

- Given θ^0 and y^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta^{new}|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping approximated blocks proposal $x^{new} \sim q_i^A(x|x^{old}, \theta^{new})$.
  - Calculate

$$\alpha = \min(1, \frac{\pi(y|x^{new})\pi(x^{new}|\theta^{new})\pi(\theta^{new})q(\theta^j|\theta^{new})q_i^A(x^j|x^{new}, \theta^j)}{\pi(y|x^j)\pi(x^j|\theta^j)\pi(\theta^j)q(\theta^{new}|\theta^j)q_{1-i}^A(x^{new}|x^j, \theta^{new})})$$

Overlapping approximated blocks proposal

- Given θ^0 and y^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta^{new}|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping approximated blocks proposal $x^{new} \sim q_i^A(x|x^{old}, \theta^{new})$.
  - Calculate

$$\alpha = \min(1, \frac{\pi(y|x^{new})\pi(x^{new}|\theta^{new})\pi(\theta^{new})q(\theta^j|\theta^{new})q_i^A(x^j|x^{new}, \theta^j)}{\pi(y|x^j)\pi(x^j|\theta^j)\pi(\theta^j)q(\theta^{new}|\theta^j)q_{1-i}^A(x^{new}|x^j, \theta^{new})})$$

  - $u \sim \text{Unif}(0, 1)$
  - if($u < \alpha$)
    - $\theta^{j+1} = \theta^{new}$
    - $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$
    - $x^{j+1} = x^j$
- Return $\theta = (\theta^1, \theta^2, \ldots, \theta^n)$ and $x = (x^1, x^2, \ldots, x^{niter})$.

Part II & IV Computational efficiency and parallelisation

- Computational benefits of GMRFs
- Parallelisation opportunity of methods in Part I and III.
- Use methods from numerical linear analysis.

Exact sampling from multivariate Gaussian distribution

- $x \sim N(0, Q^{-1}) \Rightarrow \pi(x) \propto \exp(-\frac{1}{2}x^T Q x)$
- $Q = LL^T$, L is the Choleskey factor, lower triangular
- $\pi(x) \propto \exp(-\frac{1}{2}x^T LL^T x)$
- $z \sim N(0, I)$ i.i.d. standard Gaussian.
- $\pi(z) \propto \exp(-\frac{1}{2}z^T z)$
- The solution of $L^T x = z$ is a sample from our GMRF.

Exact sampling from multivariate Gaussian distribution

- $x \sim N(0, Q^{-1}) \Rightarrow \pi(x) \propto \exp(-\frac{1}{2}x^T Q x)$
- $Q = LL^T$, L is the Choleskey factor, lower triangular
- $\pi(x) \propto \exp(-\frac{1}{2}x^T LL^T x)$
- $z \sim N(0, I)$ i.i.d. standard Gaussian.
- $\pi(z) \propto \exp(-\frac{1}{2}z^T z)$
- The solution of $L^T x = z$ is a sample from our GMRF.

Computational complexity: $\mathcal{O}(n^3)$.

Why GMRF

- The Markov property makes Q sparse.
- **Precision matrix:** $Q_{ij} = 0 \Rightarrow x_i$ and x_j are conditional independent given $\forall x_k, k \neq i, j$
- **Choleskey factor:** $L^T_{ij} = 0 \; (i < j) \Rightarrow x_i$ and x_j are conditional independent given $\forall x_k \mid k \neq j \wedge k > i$.

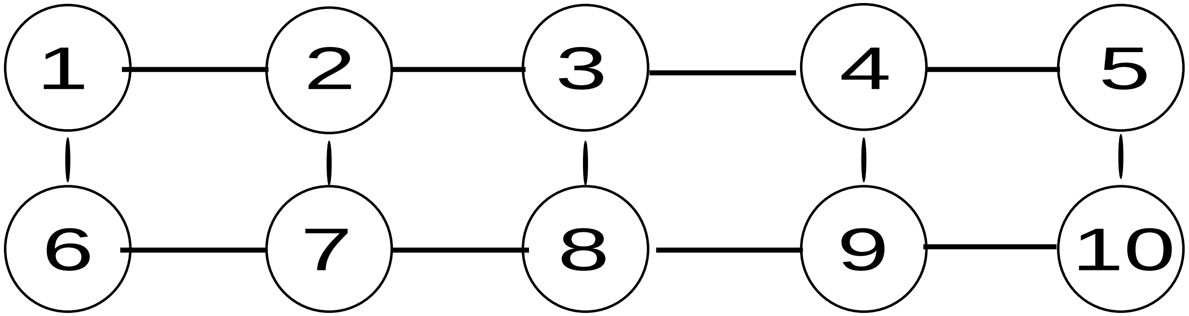
Why GMRF

- The Markov property makes Q sparse.
- **Precision matrix:** $Q_{ij} = 0 \Rightarrow x_i$ and x_j are conditional independent given $\forall x_k, k \neq i, j$
- **Choleskey factor:** $L^T_{ij} = 0 \; (i < j) \Rightarrow x_i$ and x_j are conditional independent given $\forall x_k \mid k \neq j \wedge k > i$.

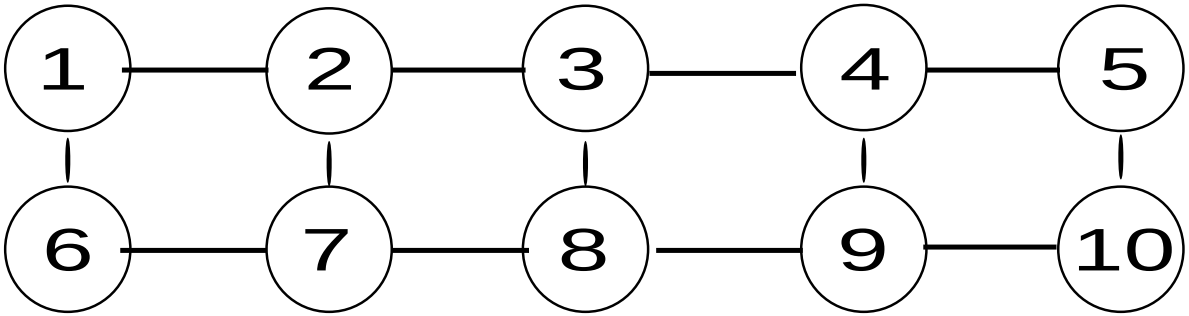
Sparse $Q \Rightarrow$ sparse L?

Graph

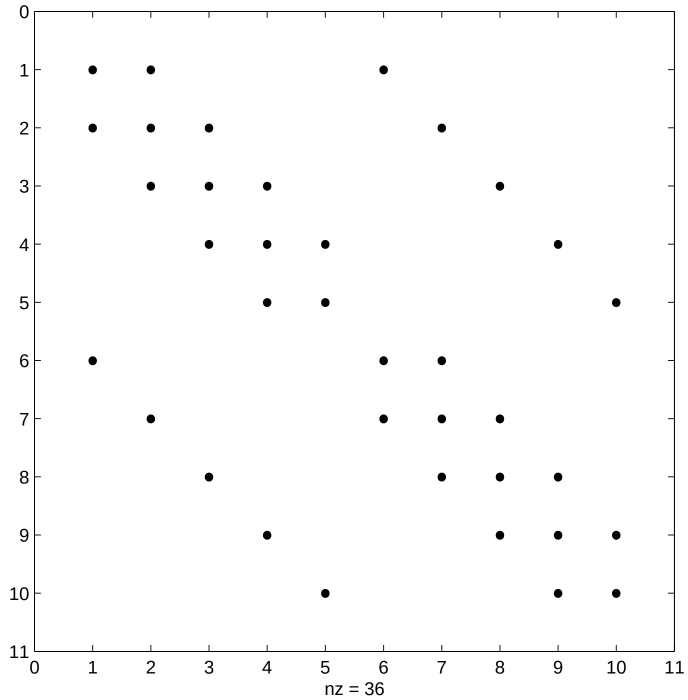
- Each variable is a node.
- Edges between neighbours



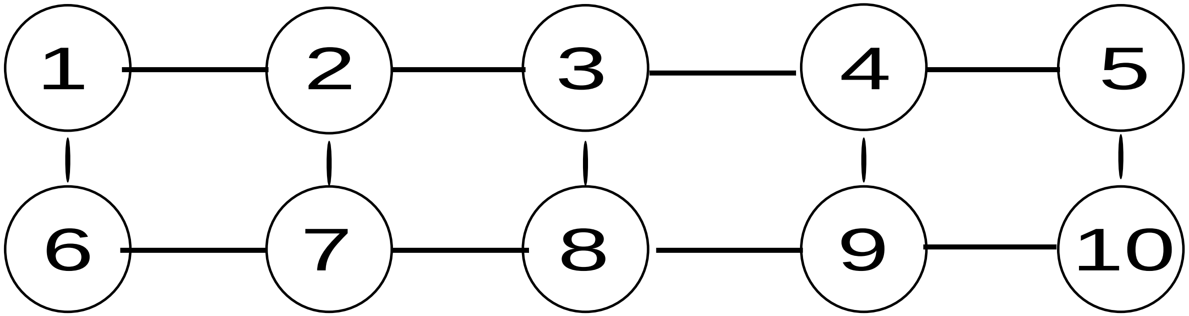
Graph



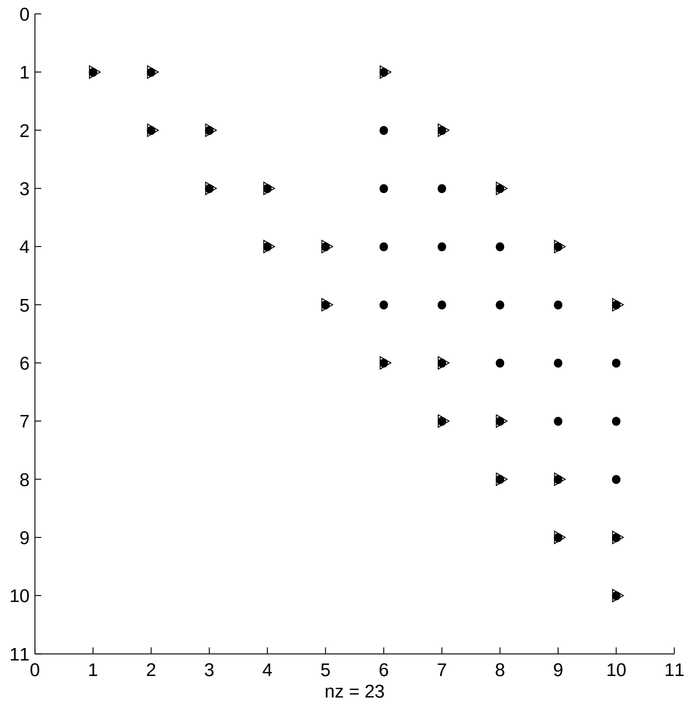
Precision matrix:



Graph

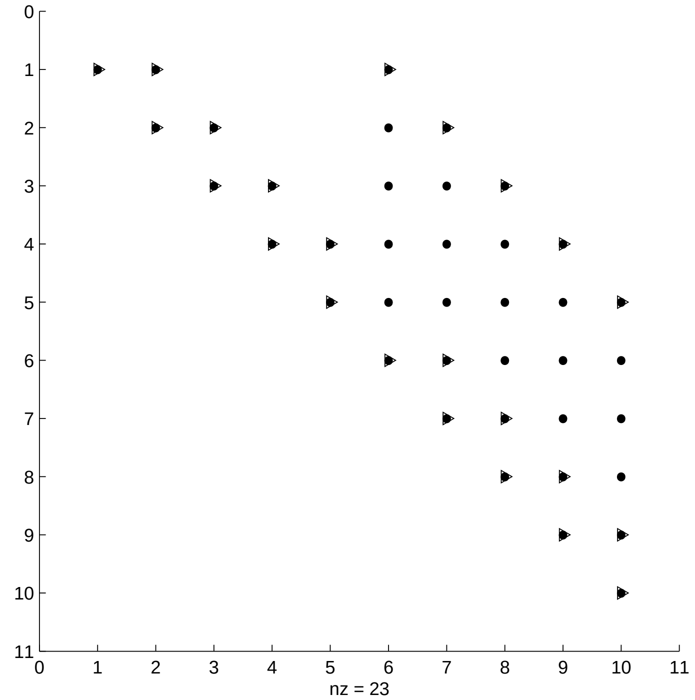


Choleskey factor L^T, $\triangle$ non-zeros in Q.



Graph

Choleskey factor L^T, $\triangle$ non-zeros in Q.



Fill-in: The elements that are zero in Q, but non-zero in L.

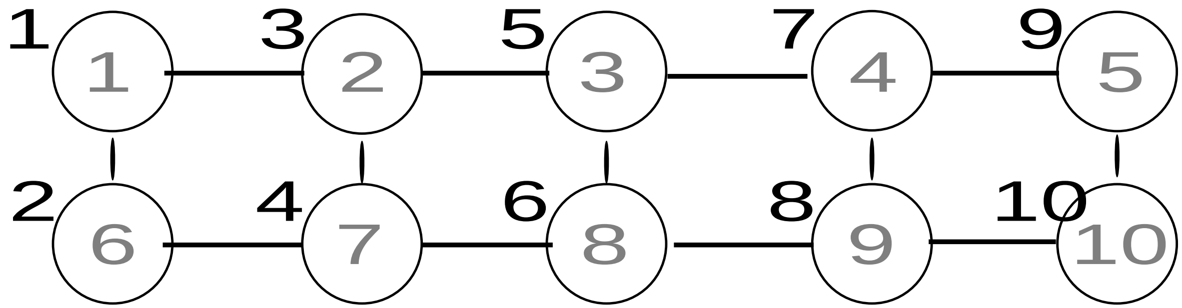
Reduce fill-in $\Rightarrow$ cheaper calculations.

Reordering

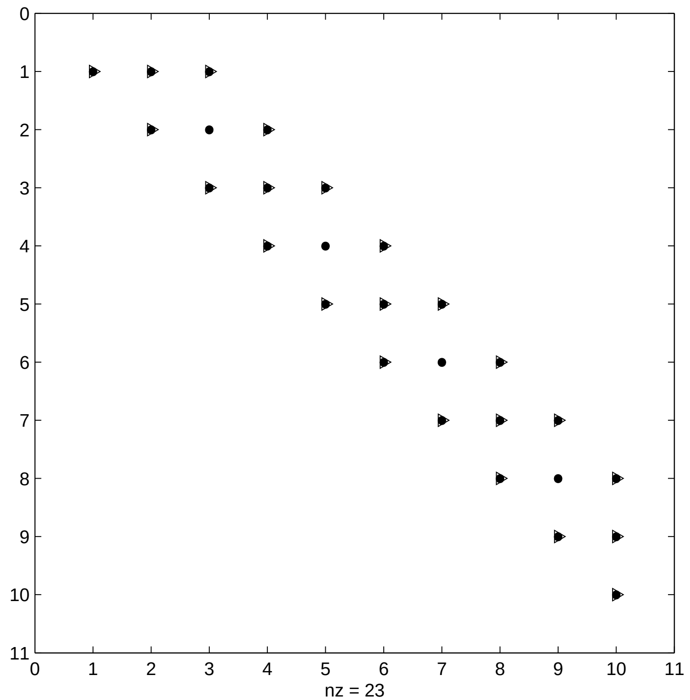
- The ordering of the variables are dummy.
- A reordering can reduce the fill-in.

Reordering

Reordered graph:

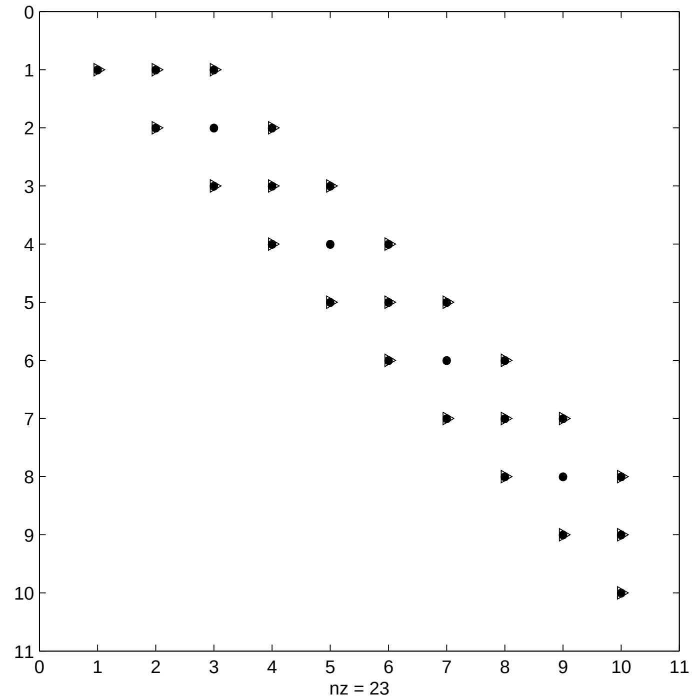


New Choleskey factor L, $\triangle$ non-zeros in Q.



Reordering

New Choleskey factor L, $\triangle$ non-zeros in Q.

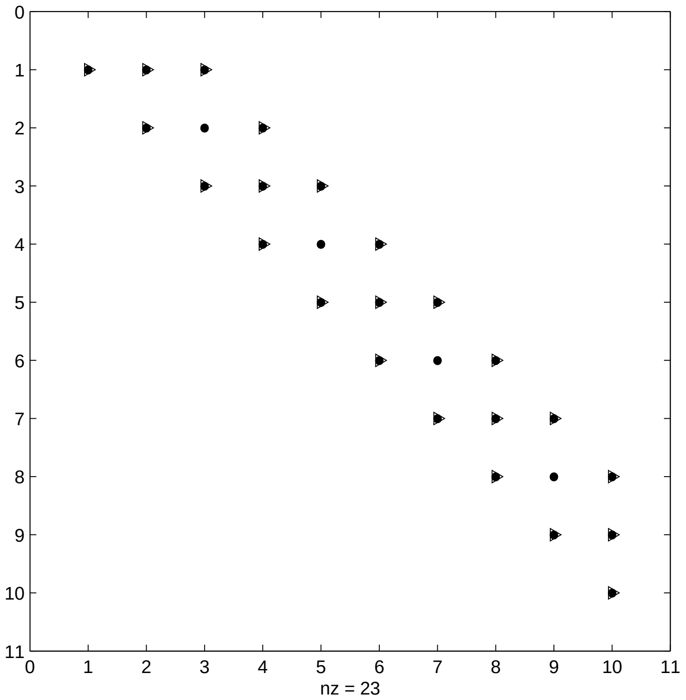


Original ordering: fill-in = 16

Reordered graph: fill-in = 4

Reordering

New Choleskey factor L, $\triangle$ non-zeros in Q.



Original ordering: fill-in = 16
Reorder graph: fill-in = 4

- Computational complexity spatial GMRF: $\mathcal{O}(n^{1.5})$

Fast sampling GMRF

Algorithm:

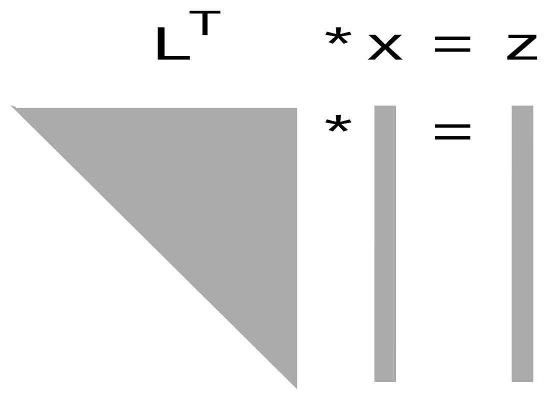
- Reorder (reduce fill-in)
- Calculate L
- Solve $L^T x = z$

Fast sampling GMRF

Algorithm:

- Reorder (reduce fill-in)
- Calculate L
- Solve $L^T x = z$

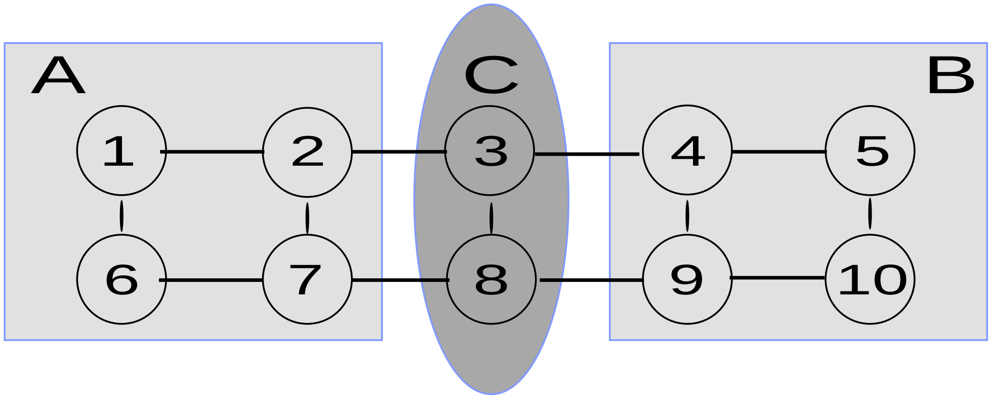
Triangular system:



$$\pi(x) = \pi(x_n)\pi(x_{n-1}|x_n)\ldots\pi(x_1|x_2, x_3, \ldots, x_n)$$

Parallel sampling of GMRF

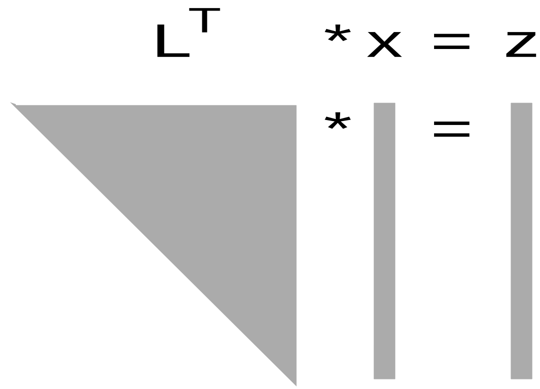
Intuitive idea:



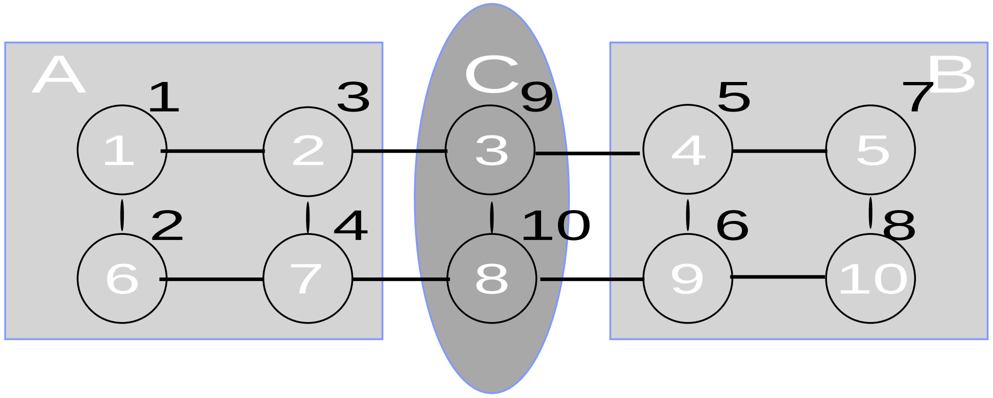
- If we have a sample $x_C \sim \pi_C(x)$.
- $x_A \sim \pi_{A|C}(x)$ and $x_B \sim \pi_{B|C}(x)$ are independent
- $x_A|x_C$ and $x_B|x_C$ can be sampled in parallel.
- $x^* = (x_A, x_C, x_B)$ a sample from our GMRF.

Markov property used to get conditional independent sets

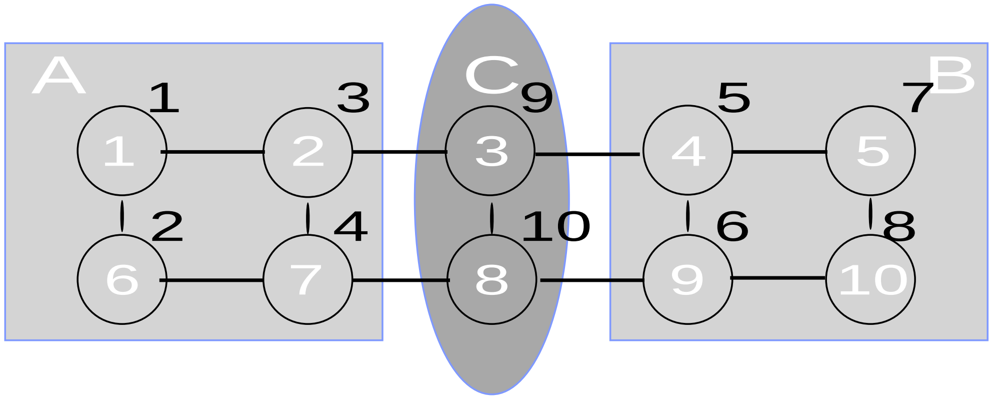
Parallel solving of triangular system



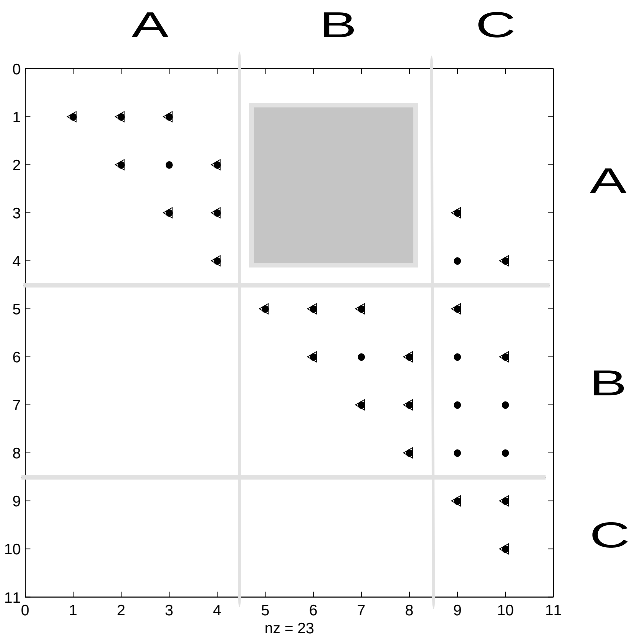
Reordering $x_r = (x_A, x_B, x_C)$:



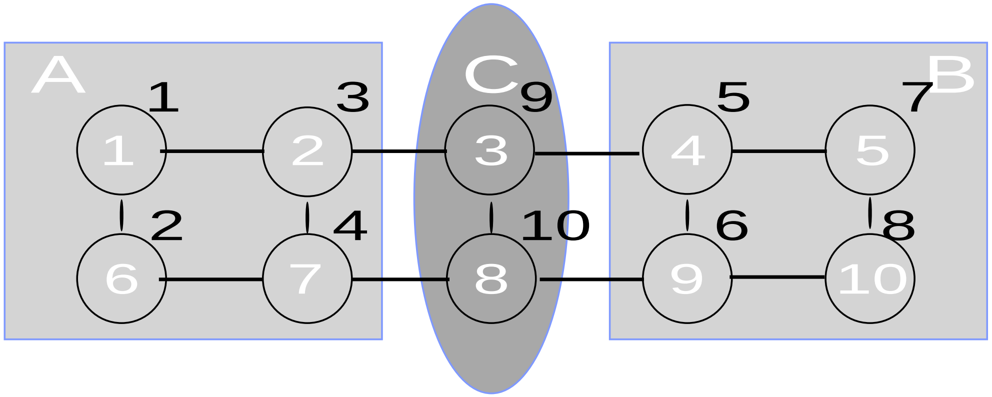
Parallel solving of triangular system

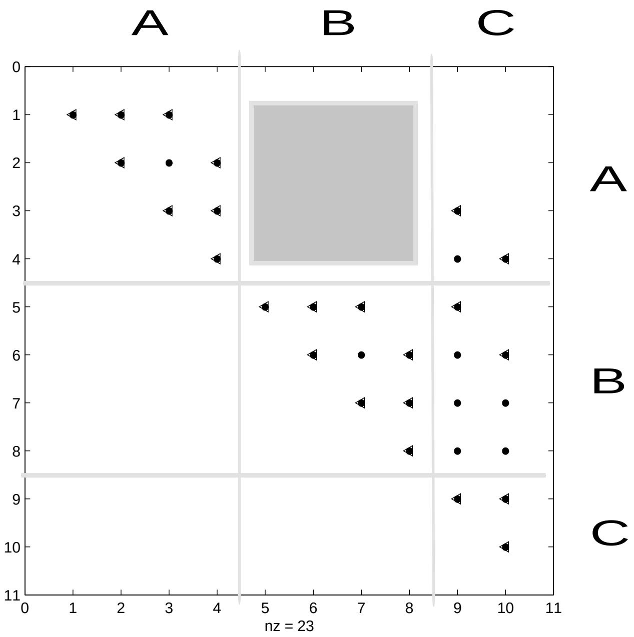


Choleskey factor L^T



Parallel solving of triangular system

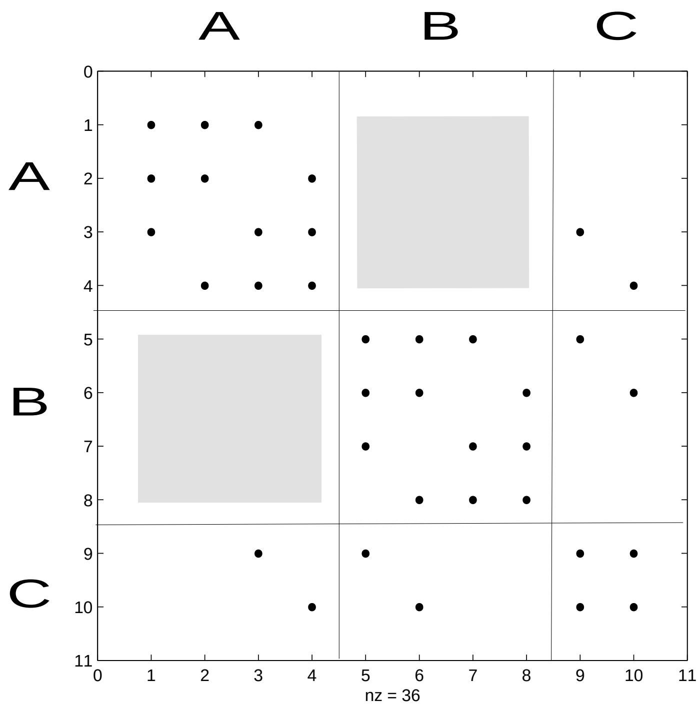




AB zero $\Rightarrow$ A and B can be calculated in parallel.

Parallel Choleskey factorisation

Precision matrix for x_r:



- Choleskey decomposition is done by "column wise right looking elimination" $\Rightarrow$ Choleskey part OK.

Parallel sampling of $A1$, $A2$ and $A3$

- $A1$: The approx. is a GMRF $\Rightarrow$ parallelisation OK.

Parallel sampling of $A1$, $A2$ and $A3$

- $A1$: The approx. is a GMRF $\Rightarrow$ parallelisation OK.
- $A2$: The approx. has the same Markov property as the prior GMRF $\Rightarrow$ can parallelise by finding segmentation sets.

Parallel sampling of $A1$, $A2$ and $A3$

- $A1$: The approx. is a GMRF $\Rightarrow$ parallelisation OK.
- $A2$: The approx. has the same Markov property as the prior GMRF $\Rightarrow$ can parallelise by finding segmentation sets.
- $A3$: The approx. has a Markov property given by the prior and sampling neighbourhood $\mathcal{J}$ $\Rightarrow$ can parallelise by finding segmentation sets.

Parallel overlapping block proposals

- When sampling from GMRF $\Rightarrow$ parallelisation OK.

Parallel overlapping block proposals

- When sampling from GMRF $\Rightarrow$ parallelisation OK.
- Using $A2$ and $A3 \Rightarrow$ the ordering is restricted by the buffers.

Summary

Background:

- Spatial latent GMRF models describe a large class of problems.
- One-block updating schemes important for mixing of M-H.

Summary

Challenge: Proposal for x, $q(x|x^{old}, \theta^{new})$

- Part I: Made approximations to $\pi(x|\theta)$ and used as proposal.
- Part III: Made a proposal from overlapping blocks. Blocks sampled either:
  - exact (Gaussian likelihood)
  - from an approximation.
- Markov property gives large computational benefits and parallelisability.