Gaussian Markov Random Field Models
An introduction

Sara Martino
Department of Mathematical Sciences
NTNU, Norway

November 2008
Outline I

Introduction
Why?

Definition
What is a GMRF?
The precision matrix
Definition of a GMRF
Example: Auto-regressive process
Why are GMRFs important?
Main features of GMRFs

Properties of GMRFs
Interpretation of elements of $Q$
Markov properties
Conditional density
Specification through full conditionals

Computing with GMRF
Gaussian Markov random fields (GMRFs) are simply multivariate Gaussian random variables....

What’s there to learn? Isn’t all just a Gaussian? That is a good question!
Gaussian Markov random fields (GMRFs) are simply multivariate Gaussian random variables.

What’s there to learn? Isn’t all just a Gaussian?
That is a good question!
Gaussians (and most often in the form of a GMRF) are extensively used in statistical models.

However, its excellent computational properties are not often explored.

Lack of knowledge of general purpose and near optimal numerical algorithms.
Why?

- Gaussians (and most often in the form of a GMRF) are extensively used in statistical models.
- However, its excellent computational properties are not often explored.
- Lack of knowledge of general purpose and near optimal numerical algorithms.
Why?

- Gaussians (and most often in the form of a GMRF) are extensively used in statistical models.
- However, its excellent computational properties are not often explored.
- Lack of knowledge of general purpose and near optimal numerical algorithms
Why?

▶ Fast(er) MCMC based inference

Recent developments (the really nice stuff...)
▶ Near instant (compare to MCMC) approximate Bayesian inference is often possible
▶ Deterministic
▶ Relative error
▶ In practice: no "error"
Why?

- Fast(er) MCMC based inference

Recent developments (the *really* nice stuff...)

- Near instant (compare to MCMC) approximate Bayesian inference is often possible
  - Deterministic
  - Relative error
  - In practice: no “error”
Why?

- Fast(er) MCMC based inference

Recent developments (the *really* nice stuff...)
- Near instant (compare to MCMC) approximate Bayesian inference is often possible
- Deterministic
  - Relative error
  - In practice: no “error”
Why?

- Fast(er) MCMC based inference

Recent developments (the *really* nice stuff...)

- Near instant (compare to MCMC) approximate Bayesian inference is often possible
- Deterministic
- Relative error
- In practice: no “error”
Why?

- Fast(er) MCMC based inference

Recent developments (the *really* nice stuff...)
- Near instant (compare to MCMC) approximate Bayesian inference is often possible
- Deterministic
- Relative error
- In practice: no “error”
1. Definition and basic properties
What is a Gaussian Markov random field (GMRF)?

A GMRF is a simple construct

- A normal distributed random vector

\[ \mathbf{x} = (x_1, \ldots, x_n)^T \]

- Additional Markov properties:

\[ x_i \perp x_j \mid \mathbf{x}_{-ij} \]

\( x_i \) and \( x_j \) are conditional independent (CI).
What is a Gaussian Markov random field (GMRF)?

A GMRF is a simple construct

- A normal distributed random vector
  
  \[ \mathbf{x} = (x_1, \ldots, x_n)^T \]

- Additional Markov properties:
  
  \[ x_i \perp x_j \mid \mathbf{x}_{-ij} \]

  \( x_i \) and \( x_j \) are conditional independent (CI).
If $x_i \perp x_j \mid x_{-ij}$ for a set of $\{i, j\}$, then we need to constrain the parametrisation of the GMRF.

- Covariance matrix: difficult
- Precision matrix: easy
If \( x_i \perp x_j \mid \mathbf{x}_{-ij} \) for a set of \( \{i, j\} \), then we need to constrain the parametrisation of the GMRF.

- Covariance matrix: difficult
- Precision matrix: easy
Conditional independence and the precision matrix

The density of a zero mean Gaussian

$$\pi(x) \propto |Q|^{1/2} \exp \left( -\frac{1}{2} x^T Q x \right)$$

Constraining the parametrisation to obey CI properties

Theorem

$$x_i \perp x_j \mid x_{-ij} \iff Q_{ij} = 0$$

This means in practice that the precision matrix is *sparse*. 
Conditional independence and the precision matrix

The density of a zero mean Gaussian

\[ \pi(x) \propto |Q|^{1/2} \exp \left( -\frac{1}{2} x^T Q x \right) \]

Constraining the parametrisation to obey CI properties

Theorem

\[ x_i \perp x_j \mid x_{-ij} \iff Q_{ij} = 0 \]

This means in practice that the precision matrix is \textit{sparse}. 
Use a (undirected) graph \( G = (\mathcal{V}, \mathcal{E}) \) to represent the CI properties,

- **\( \mathcal{V} \) Vertices**: 1, 2, \ldots, \( n \).
- **\( \mathcal{E} \) Edges** \( \{i, j\} \)
  - No edge between \( i \) and \( j \) if \( x_i \perp x_j | x_{-ij} \).
  - An edge between \( i \) and \( j \) if \( x_i \not\perp x_j | x_{-ij} \).
Use a (undirected) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the CI properties,

- Vertices: $1, 2, \ldots, n$.
- Edges $\{i, j\}$
  - No edge between $i$ and $j$ if $x_i \perp x_j | x_{-ij}$.
  - An edge between $i$ and $j$ if $x_i \not\perp x_j | x_{-ij}$. 
Use a (undirected) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the CI properties,

- **Vertices**: $1, 2, \ldots, n$.
- **Edges** $\{i, j\}$
  - No edge between $i$ and $j$ if $x_i \perp x_j \mid x_{-ij}$.
  - An edge between $i$ and $j$ if $x_i \not\perp x_j \mid x_{-ij}$. 

Gaussian Markov Random Field Models

- Definition
- The precision matrix
Definition of a GMRF

Definition (GMRF)

A random vector \( \mathbf{x} = (x_1, \ldots, x_n)^T \) is called a GMRF wrt the graph \( \mathcal{G} = (\mathcal{V} = \{1, \ldots, n\}, \mathcal{E}) \) with mean \( \mu \) and precision matrix \( Q > 0 \), iff its density has the form

\[
\pi(x) = (2\pi)^{-n/2} |Q|^{1/2} \exp \left( -\frac{1}{2} (x - \mu)^T Q (x - \mu) \right)
\]

and

\[Q_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \quad \text{for all} \quad i \neq j.\]
Simple example of a GMRF

Auto-regressive process of order 1

\[ x_t \mid x_{t-1}, \ldots, x_1 \sim \mathcal{N}(\phi x_{t-1}, 1), \quad t = 2, \ldots, n \]

and \[ x_1 \sim \mathcal{N}(0, (1 - \phi^2)^{-1}). \]

Tridiagonal precision matrix

\[
\begin{pmatrix}
1 & -\phi & & & \\
-\phi & 1 + \phi^2 & -\phi & & \\
& \ddots & \ddots & \ddots & \\
-\phi & 1 + \phi^2 & -\phi & & \\
& & -\phi & 1 & \\
\end{pmatrix}
\]
Simple example of a GMRF

Auto-regressive process of order 1

\[ x_t \mid x_{t-1}, \ldots, x_1 \sim \mathcal{N}(\phi x_{t-1}, 1), \quad t = 2, \ldots, n \]

and \( x_1 \sim \mathcal{N}(0, (1 - \phi^2)^{-1}). \)

Tridiagonal precision matrix

\[
Q = \begin{pmatrix}
1 & -\phi \\
-\phi & 1 + \phi^2 & -\phi \\
& \ddots & \ddots & \ddots \\
& -\phi & 1 + \phi^2 & -\phi \\
& -\phi & 1 & 1
\end{pmatrix}
\]
Usage of GMRFs (I)

**Structural time-series analysis**
- Autoregressive models.
- Gaussian state-space models.
- Computational algorithms based on the Kalman filter and its variants.

**Analysis of longitudinal and survival data**
- temporal GMRF priors
- state-space approaches
- spatial GMRF priors

used to analyse longitudinal and survival data.
Usage of GMRFs (I)

Structural time-series analysis

- Autoregressive models.
- Gaussian state-space models.
- Computational algorithms based on the Kalman filter and its variants.

Analysis of longitudinal and survival data

- Temporal GMRF priors
- State-space approaches
- Spatial GMRF priors

Used to analyse longitudinal and survival data.
Usage of GMRFs (II)

Graphical models

- A key model
- Estimate $Q$ and its (associated) graph from data.
- Often used in a larger context.

Semiparametric regression and splines

- Model a smooth curve in time or a surface in space.
- Intrinsic GMRF models and random walk models
- Discretely observed integrated Wiener processes are GMRFs
- GMRFs models for coefficients in B-splines.
Usage of GMRFs (II)

**Graphical models**
- A key model
- Estimate $Q$ and its (associated) graph from data.
- Often used in a larger context.

**Semiparametric regression and splines**
- Model a smooth curve in time or a surface in space.
- Intrinsic GMRF models and random walk models
- Discretely observed integrated Wiener processes are GMRFs
- GMRFs models for coefficients in B-splines.
Usage of GMRFs (III)

**Image analysis**
- The first main area for spatial models
- Image restoration using the Wiener filter.
- Texture modelling and texture discrimination.
- Segmentation
- Deformable templates
- Object identification
- 3D reconstruction
- Restoring ultrasound images
Usage of GMRFs (IV)

Spatial statistics

- Latent GMRF model analysis of spatial binary data
- Geostatistics using GMRFs
- Analysis of data in social sciences and spatial econometrics
- Spatial and space-time epidemiology
- Environmental statistics
- Inverse problems
Main features

- Analytical tractable
- Modelling using conditional independence
- Merging GMRFs using conditioning (hierarchical models)
- Unified framework for
  - understanding
  - representation
  - computation using numerical methods for sparse matrices
- Fits nicely into the MCMC world
- Can construct faster and more reliable block-MCMC algorithms.
- Approximate Bayesian inference
Constraining the parametrisation to obey CI properties

Theorem

\[ x_i \perp x_j \mid x_{-ij} \iff Q_{ij} = 0 \]
Interpretation of elements of $Q$

Let $x$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\mu$ and precision matrix $Q > 0$, then

$$E(x_i \mid x_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j : j \sim i} Q_{ij} (x_j - \mu_j),$$

$$\text{Prec}(x_i \mid x_{-i}) = Q_{ii} \quad \text{and}$$

$$\text{Corr}(x_i, x_j \mid x_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}}, \quad i \neq j.$$
Markov properties

Let $\mathbf{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.

*The pairwise Markov property:*

$$x_i \perp x_j \mid \mathbf{x}_{-ij} \text{ if } \{i, j\} \not\in \mathcal{E} \text{ and } i \neq j.$$
Markov properties

Let $\mathbf{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.

The local Markov property:

$$x_i \perp \mathbf{x}_{-\{i, \text{ne}(i)\}} \mid \mathbf{x}_{\text{ne}(i)} \quad \text{for every } i \in \mathcal{V}. $$
Markov properties

Let \( x \) be a GMRF wrt \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \). Then the following are equivalent.

*The global Markov property:* 

\[
x_A \perp x_B \mid x_C
\]

for all disjoint sets \( A, B \) and \( C \) where \( C \) separates \( A \) and \( B \), and \( A \) and \( B \) are non-empty.
(Induced) subgraph

Let $A \subset \mathcal{V}$

$\mathcal{G}^A$ denote the graph restricted to $A$.

- remove all nodes not belonging to $A$, and
- all edges where at least one node does not belong to $A$

Example

$A = \{1, 2\}$, then

$\mathcal{V}^A = \{1, 2\}$ and $\mathcal{E}^A = \{\{1, 2\}\}$
Let $A \subseteq \mathcal{V}$
$\mathcal{G}^A$ denote the graph restricted to $A$.
- remove all nodes not belonging to $A$, and
- all edges where at least one node does not belong to $A$

**Example**
$A = \{1, 2\}$, then

$$\mathcal{V}^A = \{1, 2\} \quad \text{and} \quad \mathcal{E}^A = \{\{1, 2\}\}$$
Conditional density I

Let $\mathcal{V} = A \cup B$ where $A \cap B = \emptyset$, and

$$x = \begin{pmatrix} x_A \\ x_B \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \quad Q = \begin{pmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{pmatrix}.$$

**Result:** $x_A|x_B$ is then a GMRF wrt the subgraph $G^A$ with parameters $\mu_{A|B}$ and $Q_{A|B} > 0$, where

$$\mu_{A|B} = \mu_A - Q_{AA}^{-1}Q_{AB}(x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}.$$
Conditional density I

Let $\mathcal{V} = A \cup B$ where $A \cap B = \emptyset$, and

$$
\mathbf{x} = \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_B \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \quad Q = \begin{pmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{pmatrix}.
$$

**Result:** $\mathbf{x}_A | \mathbf{x}_B$ is then a GMRF wrt the subgraph $\mathcal{G}^A$ with parameters $\mu_{A|B}$ and $Q_{A|B} > 0$, where

$$
\mu_{A|B} = \mu_A - Q_{AA}^{-1}Q_{AB}(\mathbf{x}_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}.
$$
Conditional density II

\[ \mu_{A|B} = \mu_A - Q_{AA}^{-1}Q_{AB}(x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}. \]

- Have explicit knowledge \( Q_{A|B} \) as the principal matrix \( Q_{AA} \).
- The subgraph \( G^A \) does not change the structure, only removes nodes and edges in \( A \).
- The conditional mean only depends on nodes in \( A \cup \text{ne}(A) \).
- If \( Q_{AA} \) is sparse, then \( \mu_{A|B} \) is the solution of a sparse linear system

\[ Q_{AA}(\mu_{A|B} - \mu_A) = -Q_{AB}(x_B - \mu_B) \]
Conditional density II

\[ \mu_{A|B} = \mu_A - Q_{AA}^{-1}Q_{AB}(x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}. \]

- Have explicit knowledge \( Q_{A|B} \) as the principal matrix \( Q_{AA} \).
- The subgraph \( G^A \) does not change the structure, only removes nodes and edges in \( A \).
- The conditional mean only depends on nodes in \( A \cup \text{ne}(A) \).
- If \( Q_{AA} \) is sparse, then \( \mu_{A|B} \) is the solution of a sparse linear system
  \[ Q_{AA}(\mu_{A|B} - \mu_A) = -Q_{AB}(x_B - \mu_B) \]
Conditional density II

\[ \mu_{A|B} = \mu_A - Q^{-1}_{AA}Q_{AB}(x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}. \]

- Have explicit knowledge \( Q_{A|B} \) as the principal matrix \( Q_{AA} \).
- The subgraph \( G^A \) does not change the structure, only removes nodes and edges in \( A \).
- The conditional mean only depends on nodes in \( A \cup \text{ne}(A) \).
- If \( Q_{AA} \) is sparse, then \( \mu_{A|B} \) is the solution of a sparse linear system

\[ Q_{AA}(\mu_{A|B} - \mu_A) = -Q_{AB}(x_B - \mu_B) \]
Conditional density II

\[ \mu_{A|B} = \mu_A - Q_{AA}^{-1} Q_{AB} (x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}. \]

- Have explicit knowledge \( Q_{A|B} \) as the principal matrix \( Q_{AA} \).
- The subgraph \( G^A \) does not change the structure, only removes nodes and edges in \( A \).
- The conditional mean only depends on nodes in \( A \cup \text{ne}(A) \).
- If \( Q_{AA} \) is sparse, then \( \mu_{A|B} \) is the solution of a sparse linear system
  \[ Q_{AA}(\mu_{A|B} - \mu_A) = -Q_{AB}(x_B - \mu_B) \]
Conditional density III
Conditional density III
Conditional density III

Gaussian Markov Random Field Models
- Properties of GMRFs
- Conditional density
Specification through full conditionals

An alternative to specifying a GMRF by its mean and precision matrix, is to specify it implicitly through the full conditionals

\[ \{ \pi(x_i|\mathbf{x}_{-i}), \ i = 1, \ldots, n \} \]

This approach was pioneered by Besag (1974, 1975)
Also known as conditional autoregressions or CAR-models.
Specification through full conditionals

An alternative to specifying a GMRF by its mean and precision matrix, is to specify it implicitly through the full conditionals

\[ \{ \pi(x_i|x_{-i}), \ i = 1, \ldots, n \} \]

This approach was pioneered by Besag (1974,1975)
Also known as conditional autoregressions or CAR-models.
Example: AR-process

Specify full conditionals

\[ E(x_1 | x_{-1}) = 0.3x_2 \]
\[ E(x_3 | x_{-3}) = 0.2x_2 + 0.4x_4 \]
\[ Prec(x_1 | x_{-1}) = 1 \]
\[ Prec(x_3 | x_{-3}) = 2 \]

and so on.
Example: Spatial model

Specify full conditionals for each $i$ (assuming zero mean)

$$E(x_i \mid x_{-i}) = \sum_j w_{ij} x_j \quad \text{Prec}(x_i \mid x_{-i}) = \kappa_i$$
Potential problems

- Even though the “full conditionals” are well defined, are we sure that the *joint density* exists and is unique?
- What are the compatibility conditions required for a joint GMRF to exist with the prescribed Markov properties.
Potential problems

- Even though the “full conditionals” are well defined, are we sure that the joint density exists and is unique?
- What are the compatibility conditions required for a joint GMRF to exists with the prescribed Markov properties.
In general

Specify the full conditionals as normals with

\[
E(x_i \mid x_{-i}) = \mu_i - \sum_{j:j \sim i} \beta_{ij} (x_j - \mu_j) \quad \text{and} \quad (1)
\]

\[
\text{Prec}(x_i \mid x_{-i}) = \kappa_i > 0 \quad (2)
\]

for \( i = 1, \ldots, n \), for some \( \{\beta_{ij}, i \neq j\} \), and vectors \( \mu \) and \( \kappa \).

Clearly, \( \sim \) is defined implicitly by the nonzero terms of \( \{\beta_{ij}\} \).
In general

Specify the full conditionals as normals with

$$E(x_i | x_{-i}) = \mu_i - \sum_{j: j \sim i} \beta_{ij} (x_j - \mu_j) \quad \text{and}$$

$$\text{Prec}(x_i | x_{-i}) = \kappa_i > 0$$

for $i = 1, \ldots, n$, for some $\{\beta_{ij}, i \neq j\}$, and vectors $\mu$ and $\kappa$. Clearly, $\sim$ is defined implicitly by the nonzero terms of $\{\beta_{ij}\}$. 

E(x_i | x_{-i}) = \mu_i - \sum_{j:j \sim i} \beta_{ij} (x_j - \mu_j) \quad \text{and} \quad \text{Prec}(x_i | x_{-i}) = \kappa_i > 0

Must be a joint density \( \pi(\mathbf{x}) \) with these as the full conditionals. Comparing term by term with:

\[
\mu_{A|B} = \mu_A - Q^{-1}_{AA} Q_{AB} (x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}.
\]

we get that

\[
Q_{ii} = \kappa_i, \quad \text{and} \quad Q_{ij} = \kappa_i \beta_{ij}
\]

Q is symmetric, i.e.,

\[
\kappa_i \beta_{ij} = \kappa_j \beta_{ji},
\]

We have a candidate for a joint density provided \( Q > 0 \).
\[ E(x_i \mid x_{-i}) = \mu_i - \sum_{j:j \sim i} \beta_{ij}(x_j - \mu_j) \quad \text{and} \quad \text{Prec}(x_i \mid x_{-i}) = \kappa_i > 0 \]

Must be a joint density \( \pi(x) \) with these as the full conditionals. Comparing term by term with:

\[
\mu_{A|B} = \mu_A - Q_{AA}^{-1}Q_{AB}(x_B - \mu_B) \quad \text{and} \quad Q_{A|B} = Q_{AA}.
\]

we get that

\[
Q_{ii} = \kappa_i, \quad \text{and} \quad Q_{ij} = \kappa_i \beta_{ij}
\]

\( Q \) is symmetric, i.e.,

\[
\kappa_i \beta_{ij} = \kappa_j \beta_{ji},
\]

We have a candidate for a joint density provided \( Q > 0 \).
Theorem

Given the normal full conditionals seen before then $x$ is a GMRF wrt a labelled graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with mean $\mu$ and precision matrix $Q = Q_{ij}$ where

$$Q_{ij} = \begin{cases} 
\kappa_i \beta_{ij}, & i \neq j \\
\kappa_i, & i = j 
\end{cases}$$

provided $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}$ for $i \neq j$ and $Q > 0$
2. Computing with GMRF
Sparse precision matrix

Recall that

\[ E(x_i - \mu_i \mid x_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \sim i} Q_{ij}(x_i - \mu_j), \quad \text{Prec}(x_i \mid x_{-i}) = Q_{ii} \]

In most cases:

- Total number of neighbours is \( \mathcal{O}(n) \).
- Only \( \mathcal{O}(n) \) of the \( n^2 \) terms in \( Q \) will be non-zero.
- Use this to construct exact simulation algorithms for GMRFs, using numerical algorithms for sparse matrices.
Sparse precision matrix

Recall that

\[ E(x_i - \mu_i \mid x_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \sim i} Q_{ij} (x_i - \mu_j), \quad \text{Prec}(x_i \mid x_{-i}) = Q_{ii} \]

In most cases:

- Total number of neighbours is \( O(n) \).
- Only \( O(n) \) of the \( n^2 \) terms in \( Q \) will be non-zero.
- Use this to construct exact simulation algorithms for GMRFs, using numerical algorithms for sparse matrices.
Sparse precision matrix

Recall that

\[
E(x_i - \mu_i \mid x_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \sim i} Q_{ij} (x_i - \mu_j), \quad \text{Prec}(x_i \mid x_{-i}) = Q_{ii}
\]

In most cases:

- Total number of neighbours is \(O(n)\).
- Only \(O(n)\) of the \(n^2\) terms in \(Q\) will be non-zero.
- Use this to construct exact simulation algorithms for GMRFs, using numerical algorithms for sparse matrices.
Example of a typical precision matrix
Example of a typical precision matrix
Example of a typical precision matrix

Each node have on the average 7 neighbours.
Which computations do we want to perform on a GMRF

- Unconditional simulation
  Simulate $\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{Q}^{-1})$

- Conditional simulation:
  - Condition on a subset of variables
    Simulate $\mathbf{x}|\mathbf{x}_A$
  - Condition on linear constraints
    Simulate $\mathbf{x}|\mathbf{Ax} = \mathbf{e}$ with $\mathbf{e}$ known
  - Condition on linear constraints with normal noise
    Simulate $\mathbf{x}|\mathbf{Ax} = \mathbf{e}$ with $\mathbf{e} \sim \mathcal{N}(0, \Sigma_e)$ known

- Evaluation of the log-density in all cases.

- Compute marginal variances without having to invert $\mathbf{Q}$
Which computations do we want to perform on a GMRF

- **Unconditional simulation**
  \[ x \sim \mathcal{N}(\mu, Q^{-1}) \]

- **Conditional simulation:**
  - Condition on a subset of variables
    \[ x|_{x_A} \]
  - Condition on linear constraints
    \[ x|_{Ax = e} \text{ with } e \text{ known} \]
  - Condition on linear constraints with normal noise
    \[ x|_{Ax = e} \text{ with } e \sim \mathcal{N}(0, \Sigma_e) \text{ known} \]

- Evaluation of the log-density in all cases.

- Compute marginal variances without having to invert \( Q \)
Which computations do we want to perform on a GMRF

- Unconditional simulation
  Simulate $\mathbf{x} \sim \mathcal{N}(\mu, Q^{-1})$

- Conditional simulation:
  - Condition on a subset of variables
    Simulate $\mathbf{x}|\mathbf{x}_A$
  - Condition on linear constraints
    Simulate $\mathbf{x}|A\mathbf{x} = \mathbf{e}$ with $\mathbf{e}$ known
  - Condition on linear constraints with normal noise
    Simulate $\mathbf{x}|A\mathbf{x} = \mathbf{e}$ with $\mathbf{e} \sim \mathcal{N}(0, \Sigma_e)$ known

- Evaluation of the log-density in all cases.

- Compute marginal variances without having to invert $Q$
**Cholesky factorisation**

If $A > 0$ be a $n \times n$ positive definite matrix, then there exists a unique Cholesky triangle $L$, such that $L$ is a lower triangular matrix, and

$$A = LL^T$$

Computing $L$ costs $n^3/3$ flops.

This factorisation is the basis for solving systems like

$$Ax = b \quad \text{or} \quad AX = B$$

for $k$ right hand sides, or equivalently, computing

$$x = A^{-1}b \quad \text{or} \quad X = A^{-1}B$$
Cholesky factorisation

If $A > 0$ be a $n \times n$ positive definite matrix, then there exists a unique Cholesky triangle $L$, such that $L$ is a lower triangular matrix, and

$$A = LL^T$$

Computing $L$ costs $n^3/3$ flops.
This factorisation is the basis for solving systems like

$$Ax = b \quad \text{or} \quad AX = B$$

for $k$ right hand sides, or equivalently, computing

$$x = A^{-1}b \quad \text{or} \quad X = A^{-1}B$$
How can we take advantage of the sparse precision matrix

All algorithms to do simulation and to evaluate lo-densities are based on a factorization of the precision matrix

\[ Q = L^T L \]

- It is faster to factorise a sparse \( Q \) compared to a dense \( Q \).
- The speedup depends on the “pattern” in \( Q \), not only the number of non-zero terms.
The result

In most cases, the cost is

- $O(n)$ for temporal GMRFs
- $O(n^{3/2})$ for spatial GMRFs
- $O(n^2)$ for spatio-temporal GMRFs

including evaluation of the log-density.

Condition on $k$ linear constraints, add $O(k^3)$.

These are general algorithms only depending on the graph $G$ not the numerical values in $Q$.

The core is numerical algorithms for sparse matrices.
The result

In most cases, the cost is
- $O(n)$ for temporal GMRFs
- $O(n^{3/2})$ for spatial GMRFs
- $O(n^2)$ for spatio-temporal GMRFs

including evaluation of the log-density.
Condition on $k$ linear constraints, add $O(k^3)$.

These are general algorithms only depending on the graph $G$ not the numerical values in $Q$.
The core is numerical algorithms for sparse matrices.
The result

In most cases, the cost is
- $O(n)$ for temporal GMRFs
- $O(n^{3/2})$ for spatial GMRFs
- $O(n^2)$ for spatio-temporal GMRFs

including evaluation of the log-density. Condition on $k$ linear constraints, add $O(k^3)$.

These are general algorithms only depending on the graph $G$ not the numerical values in $Q$.

The core is numerical algorithms for sparse matrices.
The result

In most cases, the cost is

- $O(n)$ for temporal GMRFs
- $O(n^{3/2})$ for spatial GMRFs
- $O(n^2)$ for spatio-temporal GMRFs

including evaluation of the log-density.

Condition on $k$ linear constraints, add $O(k^3)$.

These are general algorithms only depending on the graph $G$ not the numerical values in $Q$.

The core is numerical algorithms for sparse matrices.