GMRFLib 2.0 (?)

Henrique Ap. Laureano
henrique.laureano@kaust.edu.sa ∧ http://mynameislaure.github.io/

/KAUST/CEMSE/STAT

Summer Semester
2018

Contents

Markov Random Fields

"Statistical modeling of a finite collection of spatial random variables is often done through a Markov random field (MRF). A MRF is specified through the set of conditional distributions of one component given all the others. This enables one to focus on a single random variable at a time and leads to simple computational procedures for simulating MRFs, in particular for Bayesian inference via Markov chain Monte Carlo (MCMC)." In the Gaussian (*aka* Normal) case, we have the so called Gaussian MRFs (GMRFs).

Gaussian Markov Random Fields

"A GMRF is simply a Gaussian distributed random vector $\boldsymbol{x}$, which obeys some conditional independence properties. That is, for some $i \neq j$, then

$$x_i \perp x_j \mid \boldsymbol{x}_{-\{i,j\}}, \tag{1}$$

meaning that conditioned on $\boldsymbol{x}_{-\{i,j\}}$, x_i and x_j are independent. This conditional independence is represented using an (undirected) labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., n\}$ is the set of vertices, an $\mathcal{E} = \{\{i, j\} : i, j \in \mathcal{V}\}$ is the set of edges in the graph. For all $i, j \in \mathcal{V}$, the edge $\{i, j\}$ is not included in $\mathcal{E}$ if (1) holds, and included otherwise. Figure 1 displays such a graph, where $n = 4$ and $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}$. From this graph we deduce that $x_2 \perp x_4 \mid \boldsymbol{x}_{\{1,3\}}$ and $x_1 \perp x_3 \mid \boldsymbol{x}_{\{2,4\}}$. A central goal is now to specify a GMRF $\boldsymbol{x}$ with conditional independence properties in agreement with some given graph $\mathcal{G}$."

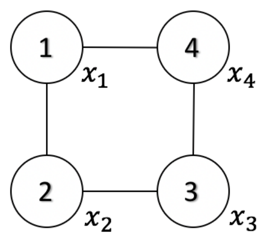

Figure 1: A conditional independence graph.

Theorem 1.
Let $\boldsymbol{x}$ be Gaussian distributed with a symmetric and positive definite (SPD) precision matrix $\boldsymbol{Q}$, then for $i \neq j$

$$x_i \perp x_j \mid \boldsymbol{x}_{-\{i,j\}} \quad \iff \quad Q_{i,j} = 0.$$

"So any SPD precision matrix $\boldsymbol{Q}$ with $Q_{2,4} = Q_{4,2} = Q_{1,3} = Q_{3,1} = 0$ has conditional independence properties as displayed in Figure 1." A precision matrix $\boldsymbol{Q}$ that may correspond to this is presented in (2). "We then say that $\boldsymbol{x}$ is a GMRF with respect to $\mathcal{G}$. A formal definition follows."

$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} Q_{1,1} & Q_{1,2} & 0 & Q_{1,4} \\ Q_{2,1} & Q_{2,2} & Q_{2,3} & 0 \\ 0 & Q_{3,2} & Q_{3,3} & Q_{3,4} \\ Q_{4,1} & 0 & Q_{4,3} & Q_{4,4} \end{pmatrix} \end{array}. \tag{2}$$

Definition 1 (GMRF).
A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ is called a GMRF wrt the labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and SPD precision matrix $\boldsymbol{Q}$, iff its density has the form

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}\,|\boldsymbol{Q}|^{1/2} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\Big) \tag{3}$$

and

$$Q_{i,j} \neq 0 \quad \iff \quad \{i,j\} \in \mathcal{E} \quad \forall \quad i \neq j.$$

"The case where $\boldsymbol{Q}$ is singular still provides a GMRF with an explicit form for its joint density, but the joint density is improper. Such specifications cannot be used as data models, but can be used as priors as long as they yield proper posteriors. Here is a simple example of a (proper) GMRF."

Example 1.
"Let $\{x_t\}$ be a stationary autoregressive process of order one, *i.e.*, $x_t \mid x_{t-1} = \phi x_{t-1} + \epsilon_t$, for $t = 2,\dots,n$, where $|\phi| < 1$ and ϵ_t are independent normally distributed zero mean innovations with unit variance. Further assume that x_1 is normal with mean zero and variance $1/(1\phi^2)$, which is simply the stationary distribution of this process. Then $\boldsymbol{x}$ is a GMRF *wrt* to $\mathcal{G}$ where $\mathcal{E} = \{\{1,2\},\{2,3\},\dots,\{n-1,n\}\}$. The precision matrix has nonzero elements $Q_{i,j} = -\phi$ for $|i-j| = 1, Q_{1,1} = Q_{n,n} = 1$ and $Q_{i,i} = 1+\phi^2$ for $i = 2,\dots,n-1$." Considering, *e.g.*, $n = 5$, the precision matrix $\boldsymbol{Q}$ is given by

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
1 & -\phi & 0 & 0 & 0 \\
-\phi & 1+\phi^2 & -\phi & 0 & 0 \\
0 & -\phi & 1+\phi^2 & -\phi & 0 \\
0 & 0 & -\phi & 1+\phi^2 & -\phi \\
0 & 0 & 0 & -\phi & 1
\end{pmatrix}
\end{array}.$$

"This example nicely illustrates why GMRFs are so useful." "Only $n+2(n1) = 3n2$ of the n^2 terms in $\boldsymbol{Q}$ are nonzero. The sparse precision matrix makes fast $\mathcal{O}(n)$ algorithms for the simulation of autoregressive processes possible."

Basic Properties

Conditional Properties

"Although a GMRF can be seen as a general multivariate Gaussian random variable, some properties simplify and some characteristics are easier to compute. For example, conditional distributions are easier to compute due to the sparse precision matrix. To see this, we split $\mathcal{V}$ into the nonempty sets A and $B = -A$. Partition $\boldsymbol{x}, \boldsymbol{\mu}$ and $\boldsymbol{Q}$ accordingly, *i.e.*,

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \quad \text{and} \quad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}.$$

We also need the notion of a subgraph $\mathcal{G}^A$, which is the graph restricted to A: the graph we obtain after removing all nodes not belonging to A and all edges where at least one node does not belong to A. Then the following theorem holds."

Theorem 2.
Let $\boldsymbol{x}$ be a GMRF wrt $\mathcal{G}$ with mean $\boldsymbol{\mu}$ and SPD precision matrix $\boldsymbol{Q}$. Let $A \subset \mathcal{V}$ and $B = \mathcal{V} \setminus A$ where $A, B \neq \emptyset$. The conditional distribution of $\boldsymbol{x}_A \mid \boldsymbol{x}_B$ is then a GMRF wrt the subgraph $\mathcal{G}^A$ with mean $\boldsymbol{\mu}_{A|B}$ and SPD precision matrix $\boldsymbol{Q}_{A|B}$, where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) \quad \textit{and} \quad \boldsymbol{Q}_{A|B} = \boldsymbol{Q}_{AA}.$$

"The expression for the conditional mean $\boldsymbol{\mu}_{A|B}$ involves the inverse $\boldsymbol{Q}_{AA}^{-1}$, but only in a way such that we can write $\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{b}$, where $\boldsymbol{b}$ is the solution of a sparse linear system $\boldsymbol{Q}_{AA}\boldsymbol{b} = \boldsymbol{Q}_{AB}(\boldsymbol{x}_B - \boldsymbol{\mu}_B)$. Note that the term $\boldsymbol{Q}_{AB}$ is nonzero only for those vertices in A that have an edge to a vertex in B, so usually only a few terms will enter in this matrix-vector product. In the special case $A = \{i\}$, the expressions simplify to

$$\mu_{i|-i} = \mu_i - \sum_{j:j\sim i} \frac{Q_{i,j}}{Q_{i,i}}(x_j - \mu_j) \quad \text{and} \quad Q_{i|-i} = Q_{i,i}. \tag{4}$$

Here we used the notation $j : j \sim i$ to indicate a sum over all vertices j that are neighbors to vertex i, *i.e.*, $\{i, j\} \in \mathcal{E}$. So $Q_{i,i}$ is the conditional precision of x_i and the conditional expectation of x_i is a weighted mean of neighboring x_js with weights $-Q_{i,j}/Q_{i,i}$."

Example 2.
"We continue with Example 1. From (4) we obtain the conditional mean and precision of $x_i \mid \boldsymbol{x}_{-i}$,"

$$\begin{aligned}
\mu_{i|-i} &= 0 - \left[\frac{-\phi}{1+\phi^2}(x_{i-1} - 0) + \frac{-\phi}{1+\phi^2}(x_{i+1} - 0)\right] \\
&= 0 - \frac{-\phi}{1+\phi^2}(x_{i-1} + x_{i+1}) \\
&= \frac{\phi}{1+\phi^2}(x_{i-1} + x_{i+1}), \quad \text{and} \quad Q_{i|-i} = 1 + \phi^2, \quad 1 < i < n.
\end{aligned}$$

Markov Properties

"The graph $\mathcal{G}$ of a GMRF is defined through looking at which x_i and x_j are conditionally independent, the so-called *pairwise* Markov property. However, more general Markov properties can be derived from $\mathcal{G}$.

A *path* from vertex i_1 to vertex i_m is a sequence of distinct nodes in $\mathcal{V}, i_1, i_2, \ldots, i_m$, for which $(i_j, i_{j+1}) \in \mathcal{E}$ for $j = 1, \ldots, m-1$. A subset $C \subset \mathcal{V}$ *separates* two nodes $i \notin C$ and $j \notin C$, if every path from i to j contains at least one node from C. Two disjoint sets $A \subset \mathcal{V} \setminus C$ and $B \subset \mathcal{V} \setminus C$ are separated by C , if all $i \in A$ and $j \in B$ are separated by C. In other words, we cannot walk on the graph starting somewhere in A ending somewhere in B without passing through C. The global Markov property, is that

$$\boldsymbol{x}_A \perp \boldsymbol{x}_B \mid \boldsymbol{x}_C$$

for all mutually disjoint sets A, B and C where C separates A and B, and A and B are nonempty."

Theorem 3.
Let $\boldsymbol{x}$ be a GMRF wrt $\mathcal{G}$, then $\boldsymbol{x}$ obeys the global Markov property.

Conditional Specification

"It is common to specify a GMRF implicitly through the so-called full conditionals $\{\pi(x_i \mid \boldsymbol{x}_{-i})\}$." "However, the full conditionals cannot be specified completely arbitrarily, as we must ensure that they correspond to a proper joint density.

A conditional specification defines the full conditional $\{\pi(x_i \mid \boldsymbol{x}_{-i})\}$ as normal with moments

$$\mathbb{E}(x_i \mid \boldsymbol{x}_{-i}) = \mu_i + \sum_{j \neq i} \beta_{i,j}(x_j - \mu_j) \quad \text{and} \quad \text{Precision}(x_i \mid \boldsymbol{x}_{-i}) = \kappa_i > 0. \tag{5}$$

The rationale for such an approach, is that it is easier to specify the full conditionals than the joint distribution. Comparing (5) with (3), we can choose $\boldsymbol{\mu}$ as the mean, $Q_{i,i} = \kappa_i$, $\beta_{i,j} = -Q_{i,j}/Q_{i,i}$ to obtain the same full conditionals. However, since $\boldsymbol{Q}$ is symmetric, we must require that

$$\kappa_i \beta_{i,j} = \kappa_j \beta_{j,i} \tag{6}$$

for all $i \neq j$." "In addition to the symmetry constraint (6), there is a joint requirement that $\boldsymbol{Q}$ is SPD. Unfortunately, this is a *joint* property, which is hard to validate locally. One convenient approach that avoids this problem is to choose a diagonally dominant parametrization that ensures $\boldsymbol{Q}$ to be SPD: $Q_{i,i} > \sum_j |Q_{i,j}|$ for all i. This implies that"

$$\sum_j |\beta_{i,j}| < 1, \quad \forall i.$$

Fixing $\boldsymbol{\mu} = \boldsymbol{0}$, using the full conditionals in Equation (5), and considering the symmetry constraint (6), the density of $\boldsymbol{x}$ can then be expressed as

$$\begin{aligned}
\log \pi(\boldsymbol{x}) &= \text{const} + \frac{1}{2} \log |\boldsymbol{Q}| - \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \quad \text{(starting from (3))} \\
&= \text{const} + \frac{1}{2} \sum_{i \neq j} Q_{i,j} x_i x_j - \frac{1}{2} \sum_{i=1}^{n} Q_{i,i} x_i^2 \\
&= \text{const} - \frac{1}{2} \sum_{i \neq j} \kappa_i \beta_{i,j} x_i x_j - \frac{1}{2} \sum_{i=1}^{n} \kappa_i x_i^2;
\end{aligned}$$

"hence, $\boldsymbol{x}$ is zero mean GMRF provided $\boldsymbol{Q}$ is SPD."

Example 3.
"The image in Figure 2(a) is a 256×256 gamma camera image of a phantom designed

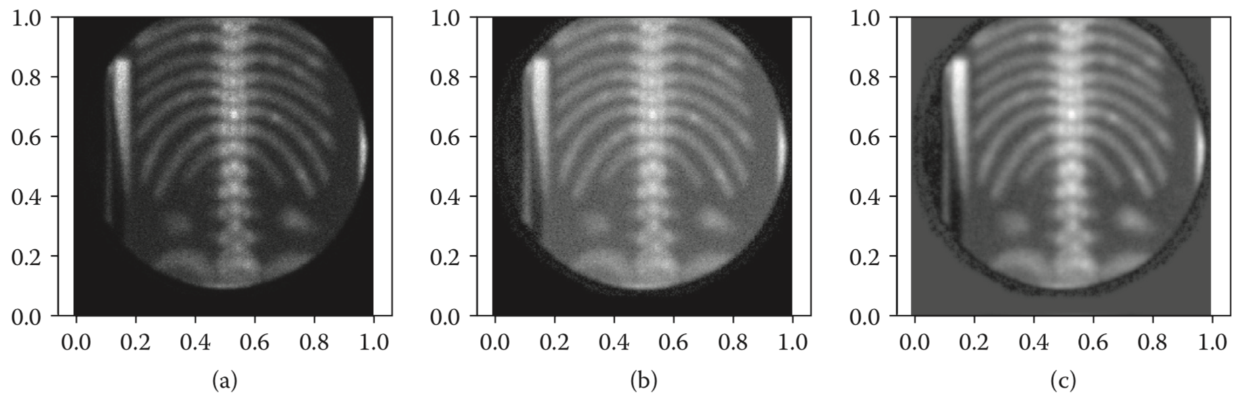

Figure 2: Panel (a) shows the raw x-ray image, (b) shows the square-root transformed image, and (c) shows the restored image (the posterior mean).

to reflect structure expected from cancerous bones. Each pixel in the circular part of the image, $\mathcal{I}$, represent the gamma radiation count, where a black pixel represents (essentially) zero counts and a white pixel the maximum count. The image is quite noisy and the task in this example is to (try to) remove the noise. The noise process is quite accurately described by a Poisson distribution, so that for each pixel i, the recorded count y_i relates to the true signal η_i, as $y_i \sim \text{Poisson}(\eta_i)$. For simplicity, we will use the approximation that"

$$\sqrt{y_i} \mid \eta_i \sim \mathcal{N}\left(\sqrt{\eta_i}, \frac{1}{4}\right), \quad i \in \mathcal{I}$$

"and the square-root transformed image is displayed in Figure 2(b). Taking a Bayesian approach, we need to specify a prior distribution for the (square-root-transformed) image $\boldsymbol{x} = (x_1, \dots, x_n)^\top$, where $x_i = \sqrt{\eta_i}$. (We need η_i to be (somewhat) larger than zero for this approximation to be adequate.) Although this is a daunting problem in general, for such noise-removal tasks it is usually sufficient to specify the prior to be informative for how the true image behaves locally. Since the image itself is locally smooth, we might specify the prior through the full conditionals (5). Using the full conditionals we only need to answer questions like: *What if we do not know the true signal in pixel i, but all others; what is then our belief in x_i?* One choice, is to set $\beta_{i,j}$ to zero unless j is one of the four nearest neighbors of i; $N_4(i)$, say. As we have no particular preference for direction, we might take for each i,

$$\beta_{i,j} = \frac{\delta}{4}, \quad j \in N_4(i)$$

where δ is considered as fixed. Further, we take κ_i to be common (and unknown) for all i, and restrict δ to $|\delta| < 1$ so that the (prior) precision matrix is diagonally dominant. (We ignore here some corrections at the boundary where a boundary pixels may have less than four neighbors.) We take further $\boldsymbol{\mu} = \mathbf{0}$ and a (conjugate) $\Gamma(a, b)$ prior for κ (with density $\propto \kappa^{a-1} \exp(-b\kappa)$), and then the posterior for $(\boldsymbol{x}, \kappa)$ reads

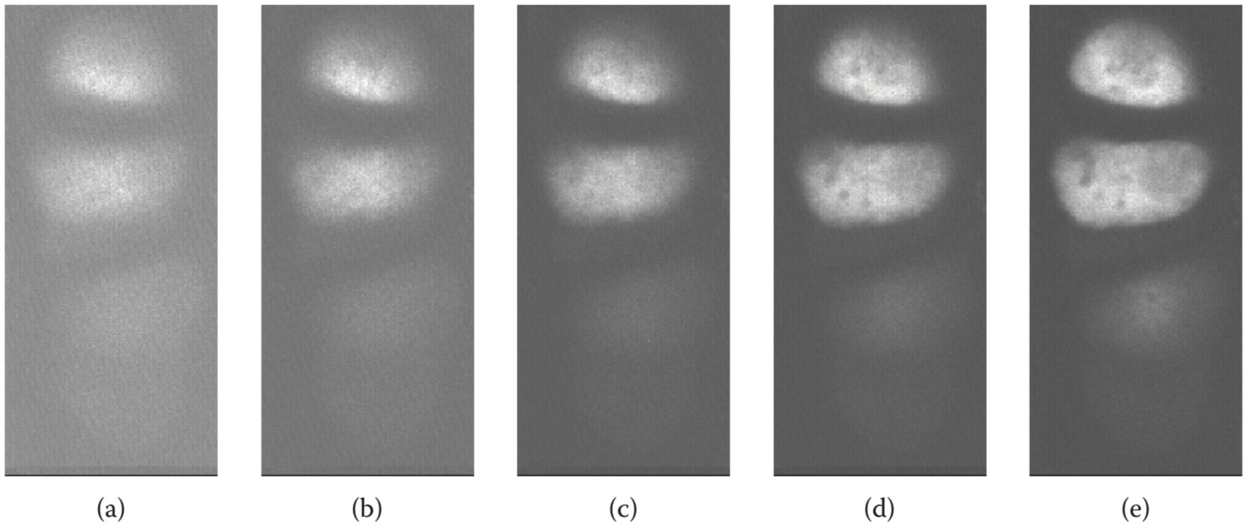

Figure 3: Panels (a) to (e) show five consecutive frames of a three-dimensional confocal microscopy image.

$$\pi(\boldsymbol{x}, \kappa \mid \boldsymbol{y}) \propto \pi(\boldsymbol{x}, \kappa)\pi(\kappa) \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i)$$
$$\propto \kappa^{a-1} \exp(-b\kappa) \, |\boldsymbol{Q}_{prior}(\kappa)|^{1/2} \exp\Big(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_{post}(\kappa)\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x}\Big).$$

Here, $b_i = 4\sqrt{y_i}$ for $i \in \mathcal{I}$ and zero otherwise, $\boldsymbol{Q}_{post}(\kappa) = \boldsymbol{Q}_{prior}(\kappa) + \boldsymbol{D}$ where $\boldsymbol{D}$ is a diagonal matrix where $D_{i,i} = 4$ if $i \in \mathcal{I}$ and zero otherwise, and

$$\boldsymbol{Q}_{prior}(\kappa)_{i,j} = \kappa \begin{cases} 1, & i = j \\ \delta/4, & j \in N_4(i) \\ 0, & \text{otherwise.} \end{cases}$$

Conditioned on κ and the observations, then $\boldsymbol{x}$ is a GMRF with precision matrix $\boldsymbol{Q}_{post}$ and where the mean $\boldsymbol{\mu}_{post}$ is given by the solution of"

$$\boldsymbol{Q}_{post}\boldsymbol{\mu}_{post} = \boldsymbol{b}.$$

Multivariate GMRFs

"To fix ideas, we will consider a generalization of Example 3 where the observations are now sequences of images. The sequence can either be a movie where each frame in the sequence is indexed by time, or the height where recorded a three-dimensional object as a set of two-dimensional images. Other examples include a temporal version of spatial models of disease counts in each administrative region of a country. Figure 3 shows five consecutive frames of three-dimensional cells taken by confocal microscopy. The first frame has a lot of noise, but the signal gets stronger farther up in the image stack. We consider the same problem as for Example 3; we want to estimate the true signal in the presence of the noise. The five frames represent the same three-dimensional object, but at different height.

We can use this information when we specify the full conditionals. It is then both easier and more natural to specify a multivariate version of the full conditionals (5), which we now will describe. Let $\boldsymbol{x}_i$ represent all the $p = 5$ observations at pixel i

$$\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}, x_{i,5})^\top.$$

Here, $x_{i,2}$ is the pixel at location i in frame 2 and so on. The conditional specification (5) extends naturally to

$$\mathbb{E}(\boldsymbol{x}_i \mid \boldsymbol{x}_{-i}) = \boldsymbol{\mu}_i - \sum_{j:j\sim i} \boldsymbol{\beta}_{i,j}(\boldsymbol{x}_j - \boldsymbol{\mu}_j) \quad \text{and} \quad \text{Precision}(\boldsymbol{x}_i \mid \boldsymbol{x}_{-i}) = \boldsymbol{\kappa}_i > 0, \tag{7}$$

for some $p \times p$ matrices $\{\boldsymbol{\beta}_{i,j}\}$ and $\{\boldsymbol{\kappa}_i\}$. In this formulation, we can now specify that our knowledge of $x_{i,3}$ might benefit of knowing $x_{i,2}$ and $x_{i,4}$. These pixels are in the same $\boldsymbol{x}_i$ vector, although they represent the ith pixel at the previous and next frame. Additionally, we can have dependency from neighbors within the same frame, such as $\{x_{j,3}, j \in N_4(i)\}$. In short, we can specify how $\boldsymbol{x}_i$ depends on $\{\boldsymbol{x}_j\}, j \neq i$, and thinking about neighbors that are (small p-) vectors.

The conditional specification in this example motivates the introduction of a multivariate GMRF, which we denote as MGMRF_p. Its definition is a direct extension of (1). Let $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top)^\top$ be Gaussian distributed, where each $\boldsymbol{x}_1$ is a p-vector. Similarly, let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_n^\top)^\top$ denote the mean and $\widetilde{\boldsymbol{Q}} = (\widetilde{\boldsymbol{Q}}_{i,j})$ the precision matrix with $p \times p$ elements $\widetilde{\boldsymbol{Q}}_{i,j}$."

Definition 2 (MGMRF_p).
A random vector $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top)^\top$ where $dim(\boldsymbol{x}_i) = p$, is called a $MGMRF_p$ wrt $\mathcal{G} = (\mathcal{V} = \{1, \ldots, n\}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and SPD precision matrix $\widetilde{\boldsymbol{Q}}$, iff its density has the form

$$\begin{aligned}
\pi(\boldsymbol{x}) &= (2\pi)^{-np/2} \left|\widetilde{\boldsymbol{Q}}\right|^{1/2} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \widetilde{\boldsymbol{Q}}(\boldsymbol{x}-\boldsymbol{\mu})\Big) \\
&= (2\pi)^{-np/2} \left|\widetilde{\boldsymbol{Q}}\right|^{1/2} \exp\Big(-\frac{1}{2}\sum_{ij}(\boldsymbol{x}_i-\boldsymbol{\mu}_i)^\top \widetilde{\boldsymbol{Q}}_{i,j}(\boldsymbol{x}_j-\boldsymbol{\mu}_j)\Big)
\end{aligned}$$

and

$$\widetilde{\boldsymbol{Q}}_{i,j} \neq \boldsymbol{0} \quad \Longleftrightarrow \quad \{i,j\} \in \mathcal{E} \quad \forall \quad i \neq j.$$

"It is important to note that a size n MGMRF_p is just another GMRF of dimension np; so all our previous results and forthcoming sparse matrix algorithms for GMRFs also apply for a MGMRF_p. However, some results have easier interpretation using the block formulation, such as

$$\boldsymbol{x}_i \perp \boldsymbol{x}_j \mid \boldsymbol{x}_{-\{i,j\}} \quad \Longleftrightarrow \quad \widetilde{\boldsymbol{Q}}_{i,j} = \mathbf{0}$$

and

$$\mathbb{E}(\boldsymbol{x}_i \mid \boldsymbol{x}_{-i}) = \boldsymbol{\mu}_i - \widetilde{\boldsymbol{Q}}_{i,j}^{-1} \sum_{j:j\sim i} \widetilde{\boldsymbol{Q}}_{i,j}(\boldsymbol{x}_j - \boldsymbol{\mu}_j) \quad \text{and} \quad \text{Precision}(\boldsymbol{x}_i \mid \boldsymbol{x}_{-i}) = \widetilde{\boldsymbol{Q}}_{i,i}. \tag{8}$$

From Equation (8), we can obtain the consistency requirements for the conditional specification Equation (7) by choosing

$$\widetilde{\boldsymbol{Q}}_{i,j} = \begin{cases} \boldsymbol{\kappa}_i \boldsymbol{\beta}_{i,j}, & i \neq j \\ \boldsymbol{\kappa}_i, & i = j. \end{cases}$$

Since $\widetilde{\boldsymbol{Q}}_{i,j} = \widetilde{\boldsymbol{Q}}_{j,i}^{\top}$, then we have the requirement that $\boldsymbol{\kappa}_i \boldsymbol{\beta}_{i,j} = \boldsymbol{\beta}_{j,i}^{\top} \boldsymbol{\kappa}_j$ for $i \neq j$, additionally to $\boldsymbol{\kappa}_i > 0 \quad \forall i$. Finally, there is also the "global" requirement that $\widetilde{\boldsymbol{Q}}$ must be SPD, which is equivalent to $(\boldsymbol{I} + (\boldsymbol{\beta}_{i,j}))$ being SPD."

Exact Algorithms for GMRFs

GMRFs have "a nice connection with very efficient numerical algorithms for sparse matrices. This connection allows for exact algorithms for GMRFs. We will now discuss this connection, starting with various exact algorithms to efficiently sample from GMRFs. This includes solving tasks like unconditional and conditional sampling, sampling under linear hard and soft constraints, evaluating the log-density of a (possibly constrained) GMRF at a particular value, and computing marginal variances for (possibly constrained) GMRFs. Although all these tasks are formally "just matrix algebra," we need to ensure that we take advantage of the sparse precision matrix $\boldsymbol{Q}$ in all steps so computations can make use of the efficient numerical algorithms for sparse matrices developed in the computational sciences literature. Further, we can derive all the algorithms for sparse matrices by considering conditional independence properties of GMRFs. The core of all algorithms is the Cholesky factorization $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^{\top}$ of the precision matrix $\boldsymbol{Q}$, where $\boldsymbol{L}$ is a lower-triangular matrix."

Why are Exact Algorithms Important?

"Exact efficient algorithms are generally preferable when they exist, even though they apparently require algorithms that are more involved than simple iterative ones. Computational feasibility is important even for statistical modeling, as a statistical model is not of much use if we cannot do inference efficiently enough to satisfy the end-user.

Sampling from a GMRF can be done exactly using the Cholesky factorization of the precision matrix". "In the spatial case, it turns out that we can (typically) sample a GMRF exactly at the cost of $\mathcal{O}(n^{3/2})$ operations". "The exact algorithm can further produce independent samples at the cost of $\mathcal{O}(n \log n)$ each."

Some Basic Linear Algebra

"Let A be SPD, then there exists a unique (Cholesky) factorization $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, where L is lower triangular and called the Cholesky triangle. This factorization is the starting point for solving $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$ by first solving $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{b}$ and then $\boldsymbol{L}^\top\boldsymbol{y} = \boldsymbol{v}$. The first linear system $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{b}$ is solved directly using forward substitution

$$v_i = \frac{1}{L_{i,i}}\Big(b_i - \sum_{j=1}^{i-1} L_{i,j}v_j\Big), \quad i = 1, \dots, n,$$

whereas $\boldsymbol{L}^\top\boldsymbol{y} = \boldsymbol{v}$ is solved using backward substitution

$$y_i = \frac{1}{L_{i,i}}\Big(v_i - \sum_{j=1+1}^{n} L_{j,i}y_j\Big), \quad i = n, \dots, 1.$$

Computing $\boldsymbol{A}^{-1}\boldsymbol{Y}$, where $\boldsymbol{Y}$ is a $n \times k$ matrix, is done by computing $\boldsymbol{A}^{-1}\boldsymbol{Y}_j$ for each of the k columns $\boldsymbol{Y}_j$ using the algorithm above. Note that $\boldsymbol{A}$ needs to be factorized only once. Note that in the case $\boldsymbol{Y} = \boldsymbol{I}$ (and $k = n$), the inverse of $\boldsymbol{A}$ is computed".

Sampling from a GMRF

```
[1] "2018-08-01 15:23:01 +03"
```