FROM : Gaussian Markov random fields, Rue & Held, CRC, 2005

5. $A_{ii} + A_{jj} - 2|A_{ij}| > 0$, for $i \neq j$.
6. $\max A_{ii} > \max_{i \neq j} |A_{ij}|$.
7. $\boldsymbol{A}^{-1}$ is SPD.
8. All principal submatrices of $\boldsymbol{A}$ are SPD.

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are SPD, then so is $\boldsymbol{A} + \boldsymbol{B}$, but the converse is not true in general. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are SPD and $\boldsymbol{AB} = \boldsymbol{BA}$, then $\boldsymbol{AB}$ is SPD.

The following conditions are all sufficient and necessary for a symmetric matrix $\boldsymbol{A}$ to be SPD:

1. All the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\boldsymbol{A}$ are strictly positive.
2. There exists a matrix $\boldsymbol{C}$ such that $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{C}^T$. If $\boldsymbol{C}$ is lower triangular it is called the *Cholesky triangle* of $\boldsymbol{A}$.
3. All leading principal submatrices have strictly positive determinants.

A sufficient but not necessary condition for a (symmetric) matrix to be SPD is the *diagonal dominance* criterion:

$$A_{ii} - \sum_{j:j \neq i} |A_{ij}| > 0, \ \forall i.$$

An $n \times n$ matrix $\boldsymbol{A}$ is called *positive semidefinite* iff

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}.$$

If $\boldsymbol{A}$ is also symmetric, then it is called a *symmetric positive semidefinite* (SPSD) matrix. A SPSD matrix $\boldsymbol{A}$ is sometimes denoted by '$\boldsymbol{A} \geq 0$'.

2.1.7 The normal distribution

We now recall the multivariate normal distribution and give some of its basic properties. This makes the difference to a GMRF more clear. Other distributions are defined in Appendix A.

The density of a normal random variable $\boldsymbol{x} = (x_1, \ldots, x_n)^T$, $n < \infty$, with mean $\boldsymbol{\mu}$ ($n \times 1$ vector) and SPD covariance matrix $\boldsymbol{\Sigma}$ ($n \times n$ matrix), is

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^n. \tag{2.3}$$

Here, $\mu_i = \mathrm{E}(x_i)$, $\Sigma_{ij} = \mathrm{Cov}(x_i, x_j)$, $\Sigma_{ii} = \mathrm{Var}(x_i) > 0$ and $\mathrm{Corr}(x_i, x_j) = \Sigma_{ij}/(\Sigma_{ii}\Sigma_{jj})^{1/2}$. We write this as $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A standard normal distribution is obtained if $n = 1$, $\mu = 0$ and $\Sigma_{11} = 1$.

We now divide $\boldsymbol{x}$ into two parts, $\boldsymbol{x} = (\boldsymbol{x}_A^T, \boldsymbol{x}_B^T)^T$, and split $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly:

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

Here are some basic properties of the normal distribution.

1. $\boldsymbol{x}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$
2. $\boldsymbol{\Sigma}_{AB} = \boldsymbol{0}$ iff $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ are independent.
3. The conditional distribution $\pi(\boldsymbol{x}_A | \boldsymbol{x}_B)$ is $\mathcal{N}(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$, where
$$\begin{aligned} \boldsymbol{\mu}_{A|B} &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) \quad \text{and} \\ \boldsymbol{\Sigma}_{A|B} &= \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}. \end{aligned}$$
4. If $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{x}' \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ are independent, then $\boldsymbol{x} + \boldsymbol{x}' \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\mu}', \boldsymbol{\Sigma} + \boldsymbol{\Sigma}')$.

2.2 Definition and basic properties of GMRFs

2.2.1 Definition

Let $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ have a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Define the labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, n\}$ and $\mathcal{E}$ be such that there is no edge between node i and j iff $x_i \perp x_j | \boldsymbol{x}_{-ij}$, where $\boldsymbol{x}_{-ij}$ is short for $\boldsymbol{x}_{-\{i,j\}}$. Then we say that $\boldsymbol{x}$ is a GMRF wrt $\mathcal{G}$.

Before we define a GMRF formally, let us investigate the connection between the graph $\mathcal{G}$ and the parameters of the normal distribution. Since the mean $\boldsymbol{\mu}$ does not have any influence on the pairwise conditional independence properties of $\boldsymbol{x}$, we can deduce that this information must be 'hidden' solely in the covariance matrix $\boldsymbol{\Sigma}$. It turns out that the inverse covariance matrix, the *precision matrix* $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ plays the key role.

Theorem 2.2 *Let $\boldsymbol{x}$ be normal distributed with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$. Then for $i \neq j$,*

$$x_i \perp x_j \mid \boldsymbol{x}_{-ij} \quad \Longleftrightarrow \quad Q_{ij} = 0.$$

This is a nice and useful result. It simply says that the nonzero pattern of $\boldsymbol{Q}$ determines $\mathcal{G}$, so we can read off from $\boldsymbol{Q}$ whether x_i and x_j are conditionally independent. We will return to this in a moment. On the other hand, for a given graph $\mathcal{G}$, we know the nonzero terms in $\boldsymbol{Q}$. This can be used to provide a parameterization of $\boldsymbol{Q}$, being aware that we also require $\boldsymbol{Q} > 0$.

Before providing the proof of Theorem 2.2, we state the formal definition of a GMRF.

Definition 2.1 (GMRF) *A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ is called a GMRF wrt a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and*

precision matrix $\boldsymbol{Q} > 0$, *iff its density has the form*

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{Q}|^{1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\Big) \tag{2.4}$$

and

$$Q_{ij} \neq 0 \iff \{i,j\} \in \mathcal{E} \quad \textit{for all} \quad i \neq j.$$

If $\boldsymbol{Q}$ is a completely dense matrix then $\mathcal{G}$ is fully connected. This implies that any normal distribution with SPD covariance matrix is also a GMRF and vice versa. We will focus on the case when $\boldsymbol{Q}$ is sparse, as it is here that the nice properties of GMRFs are really useful.

Proof. [Theorem 2.2] We partition $\boldsymbol{x}$ as $(x_i, x_j, \boldsymbol{x}_{-ij})$ and then use the multivariate version of the factorization criterion (Theorem 2.1) on $\pi(x_i, x_j, \boldsymbol{x}_{-ij})$. Fix $i \neq j$ and assume $\boldsymbol{\mu} = \boldsymbol{0}$ without loss of generality. From (2.4) we get

$$\begin{aligned}\pi(x_i, x_j, \boldsymbol{x}_{-ij}) &\propto \exp\Big(-\frac{1}{2}\sum_{k,l} x_k Q_{kl} x_l\Big) \\ &\propto \exp\Big(-\frac{1}{2}\underbrace{x_i x_j (Q_{ij}+Q_{ji})}_{\text{term 1}} - \frac{1}{2}\underbrace{\sum_{\{k,l\}\neq\{i,j\}} x_k Q_{kl} x_l}_{\text{term 2}}\Big).\end{aligned}$$

Term 2 does not involve $x_i x_j$ while term 1 involves $x_i x_j$ iff $Q_{ij} \neq 0$. Comparing with (2.2) in Theorem 2.1, we see that

$$\pi(x_i, x_j, \boldsymbol{x}_{-ij}) = f(x_i, \boldsymbol{x}_{-ij}) g(x_j, \boldsymbol{x}_{-ij})$$

for some functions f and g, iff $Q_{ij} = 0$. The claim then follows. □

We have argued that the natural way to describe a GMRF is by its precision matrix $\boldsymbol{Q}$. The elements of $\boldsymbol{Q}$ have nice conditional interpretations.

Theorem 2.3 *Let* $\boldsymbol{x}$ *be a GMRF wrt* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *with mean* $\boldsymbol{\mu}$ *and precision matrix* $\boldsymbol{Q} > 0$, *then*

$$E(x_i \mid \boldsymbol{x}_{-i}) = \mu_i - \frac{1}{Q_{ii}}\sum_{j:j\sim i} Q_{ij}(x_j - \mu_j), \tag{2.5}$$

$$Prec(x_i \mid \boldsymbol{x}_{-i}) = Q_{ii} \quad \textit{and} \tag{2.6}$$

$$Corr(x_i, x_j \mid \boldsymbol{x}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \quad i \neq j. \tag{2.7}$$

The diagonal elements of $\boldsymbol{Q}$ are the conditional precisions of x_i given $\boldsymbol{x}_{-i}$, while the off-diagonal elements, with a proper scaling, provide information about the conditional correlation between x_i and x_j, given $\boldsymbol{x}_{-ij}$. These results should be compared to the interpretation of the

elements of the covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})$; As $\mathrm{Var}(x_i) = \Sigma_{ii}$ and $\mathrm{Corr}(x_i, x_j) = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$, the covariance matrix gives information about the marginal variance of x_i and the marginal correlation between x_i and x_j. The marginal interpretation given by $\boldsymbol{\Sigma}$ is intuitive and directly informative, as it reduces the interpretation from an n-dimensional distribution to a one- or two-dimensional distribution. The interpretation provided by $\boldsymbol{Q}$ is hard (or nearly impossible) to interpret marginally, as we have to integrate out $\boldsymbol{x}_{-i}$ or $\boldsymbol{x}_{-ij}$ from the joint distribution parameterized in terms of $\boldsymbol{Q}$. In matrix terms this is immediate; by definition $\boldsymbol{Q}^{-1} = \boldsymbol{\Sigma}$, and Σ_{ii} depends generally on all elements in $\boldsymbol{Q}$, and visa versa.

Proof. [Theorem 2.3] First recall that a univariate normal random variable x_i with mean γ and precision κ has density proportional to

$$\exp\big(-\frac{1}{2}\kappa x_i^2 + \kappa x_i \gamma\big). \tag{2.8}$$

Assume for the moment that $\boldsymbol{\mu} = \boldsymbol{0}$ and apply (2.1) to (2.4):

$$\begin{aligned}\pi(x_i \mid \boldsymbol{x}_{-i}) &\propto \exp\big(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}\big) \\ &\propto \exp\big(-\frac{1}{2}x_i^2 Q_{ii} - x_i\sum_{j:j\sim i} Q_{ij}x_j\big).\end{aligned} \tag{2.9}$$

Comparing (2.8) and (2.9) we see that $\pi(x_i|\boldsymbol{x}_{-i})$ is normal. Comparing the coefficients for the quadratic term, we obtain (2.6). Comparing the coefficients for the linear term, we obtain

$$\mathrm{E}(x_i \mid \boldsymbol{x}_{-i}) = -\frac{1}{Q_{ii}}\sum_{j:j\sim i} Q_{ij}x_j.$$

If $\boldsymbol{x}$ has mean $\boldsymbol{\mu}$, then $\boldsymbol{x}-\boldsymbol{\mu}$ has mean zero, hence replacing x_i and x_j by $x_i - \mu_i$ and $x_j - \mu_j$, respectively, gives (2.5). To show (2.7), we proceed similarly and consider

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij}) \propto \exp\big(-\frac{1}{2}(x_i, x_j)\begin{pmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj}\end{pmatrix}\begin{pmatrix} x_i \\ x_j\end{pmatrix} + \text{linear terms}\big). \tag{2.10}$$

We compare this density with the density of the bivariate normal random variable $(x_i, x_j)^T$ with covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})$, which has density proportional to

$$\exp\big(-\frac{1}{2}(x_i, x_j)\begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj}\end{pmatrix}^{-1}\begin{pmatrix} x_i \\ x_j\end{pmatrix} + \text{linear terms}\big). \tag{2.11}$$

Comparing (2.10) with (2.11), we obtain

$$\begin{pmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{pmatrix},$$

which implies that $\Sigma_{ii} = Q_{jj}/\Delta$, $\Sigma_{jj} = Q_{ii}/\Delta$, and $\Sigma_{ij} = -Q_{ij}/\Delta$ where $\Delta = Q_{ii}Q_{jj} - Q_{ij}^2$. Using these expressions and the definition of conditional correlation we obtain

$$\begin{aligned} \text{Corr}(x_i, x_j \mid \boldsymbol{x}_{-ij}) &= -\frac{Q_{ij}/\Delta}{\sqrt{(Q_{jj}/\Delta)(Q_{ii}/\Delta)}} \\ &= -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}. \end{aligned}$$

□

2.2.2 Markov properties of GMRFs

We have defined the graph $\mathcal{G}$ from checking if $x_i \perp x_j | \boldsymbol{x}_{-ij}$ or not. Theorem 2.2 says this is the same as checking if the corresponding off-diagonal entry of the precision matrix, Q_{ij}, is zero or not. Hence $\mathcal{G}$ is constructed from the nonzero pattern of $\boldsymbol{Q}$. An interesting and useful property of a GMRF is that more information regarding conditional independence can be extracted from $\mathcal{G}$. We consider now the *local Markov property* and the *global Markov property*, additional to the *pairwise Markov property* used to define $\mathcal{G}$. It turns out that all these properties are equivalent for a GMRF.

Theorem 2.4 *Let $\boldsymbol{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.*
The pairwise Markov property:

$$x_i \perp x_j \mid \boldsymbol{x}_{-ij} \quad \text{if } \{i,j\} \notin \mathcal{E} \text{ and } i \neq j.$$

The local Markov property:

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

The global Markov property:

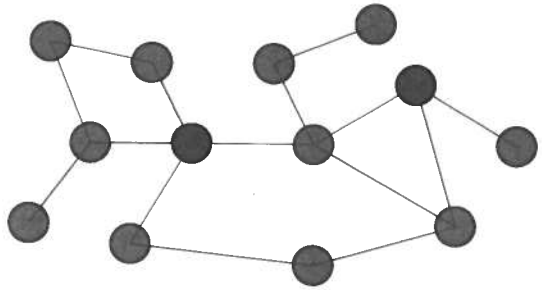
$$\boldsymbol{x}_A \perp \boldsymbol{x}_B \mid \boldsymbol{x}_C \tag{2.12}$$

for all disjoint sets A, B and C where C separates A and B, and A and B are non-empty.

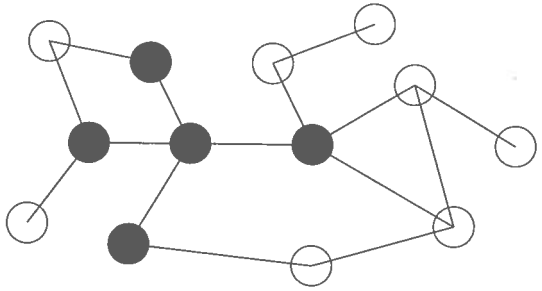
Figure 2.3 illustrates Theorem 2.4. The proof is a consequence of a more general result, stating the equivalence of the various Markov properties under some conditions satisfied for GMRFs. A simpler proof can be constructed in the Gaussian case (Speed and Kiiveri, 1986), but we omit it here.

The global Markov property immediately implies the local and pairwise Markov property, but the converse is a bit surprising. Note that the union of A, B, and C does not need to be $\mathcal{V}$, so properties of the marginal distribution can also be derived from $\mathcal{G}$.

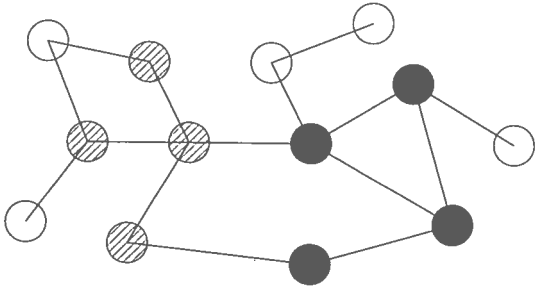
If C in (2.12) is empty, then $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ are independent.



(*a*) Pairwise



(*b*) Local



(*c*) Global

Figure 2.3 *Illustration of the various Markov properties. (a) The pairwise Markov property; the two black nodes are conditionally independent given the gray nodes. (b) The local Markov property; the black and white nodes are conditionally independent given the gray nodes. (c) The global Markov property; the black and striped nodes are conditionally independent given the gray nodes.*

2.2.3 Conditional properties of GMRFs

We will now discuss an important result of GMRFs; the conditional distribution for a subset $\boldsymbol{x}_A$ of $\boldsymbol{x}$ given the rest $\boldsymbol{x}_{-A}$. In this context the canonical parameterization will be useful, a parameterization that is easily updated under successive conditioning. Although all computations can be expressed with matrices, we will also consider a more graph-oriented view in Appendix B, which allows for efficient computation of the conditional densities.

Conditional distribution

We split the indices into the nonempty sets A and denote by B the set $-A$, so that

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix}. \tag{2.13}$$

Partition the mean and precision accordingly,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \text{and} \quad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}. \tag{2.14}$$

Our next result, is a generalization of Theorem 2.3.

Theorem 2.5 *Let $\boldsymbol{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$. Let $A \subset \mathcal{V}$ and $B = \mathcal{V} \setminus A$ where $A, B \neq \emptyset$. The conditional distribution of $\boldsymbol{x}_A | \boldsymbol{x}_B$ is then a GMRF wrt the subgraph $\mathcal{G}^A$ with mean $\boldsymbol{\mu}_{A|B}$ and precision matrix $\boldsymbol{Q}_{A|B} > 0$, where*

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1} \boldsymbol{Q}_{AB} (\boldsymbol{x}_B - \boldsymbol{\mu}_B) \tag{2.15}$$

and

$$\boldsymbol{Q}_{A|B} = \boldsymbol{Q}_{AA}.$$

This is a powerful result for two reasons. First, we have explicit knowledge of $\boldsymbol{Q}_{A|B}$ through the principal matrix $\boldsymbol{Q}_{AA}$, so no computation is needed to obtain the conditional precision matrix. Constructing the subgraph $\mathcal{G}^A$ does not change the structure; it just removes all nodes not in A and the corresponding edges. This is important for the computational issues that will be discussed in Section 2.3. Secondly, since Q_{ij} is zero unless $j \in \text{ne}(i)$, the conditional mean only depends on values of $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ in $A \cup \text{ne}(A)$. This is a great advantage if A is a small subset of $\mathcal{V}$ and in striking contrast to the corresponding general result for the normal distribution, see Section 2.1.7.

Example 2.2 *To illustrate Theorem 2.5, we compute the mean and precision of x_i given $\boldsymbol{x}_{-i}$, which are found using $A = \{i\}$ as (2.5) and (2.6). This result is frequently used for single-site Gibbs sampling in GMRF models, to which we return in Section 4.1.*

Proof. [Theorem 2.5] The proof is similar to Theorem 2.3, but uses matrices. Assume $\boldsymbol{\mu} = \boldsymbol{0}$ and compute the conditional density,

$$\begin{aligned} \pi(\boldsymbol{x}_A \mid \boldsymbol{x}_B) &\propto \exp\Big(-\frac{1}{2} (\boldsymbol{x}_A, \boldsymbol{x}_B) \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix} \Big) \\ &\propto \exp\Big(-\frac{1}{2} \boldsymbol{x}_A^T \boldsymbol{Q}_{AA} \boldsymbol{x}_A - (\boldsymbol{Q}_{AB} \boldsymbol{x}_B)^T \boldsymbol{x}_A \Big). \end{aligned}$$

Comparing this with the density of a normal with precision $\boldsymbol{P}$ and mean $\boldsymbol{\gamma}$,

$$\pi(\boldsymbol{z}) \propto \exp\Big(-\frac{1}{2} \boldsymbol{z}^T \boldsymbol{P} \boldsymbol{z} + (\boldsymbol{P}\boldsymbol{\gamma})^T \boldsymbol{z} \Big),$$

we see that $\boldsymbol{Q}_{AA}$ is the conditional precision matrix and the conditional mean is given by the solution of

$$\boldsymbol{Q}_{AA} \boldsymbol{\mu}_{A|B} = -\boldsymbol{Q}_{AB} \boldsymbol{x}_B.$$

Note that $\boldsymbol{Q}_{AA} > 0$ since $\boldsymbol{Q} > 0$. If $\boldsymbol{x}$ has mean $\boldsymbol{\mu}$ then $\boldsymbol{x} - \boldsymbol{\mu}$ has mean zero, hence (2.15) follows. The subgraph $\mathcal{G}^A$ follows from the nonzero elements of $\boldsymbol{Q}_{AA}$. □

To compute the conditional mean $\boldsymbol{\mu}_{A|B}$, we need to *solve* the linear system

$$\boldsymbol{Q}_{AA} (\boldsymbol{\mu}_{A|B} - \boldsymbol{\mu}_A) = -\boldsymbol{Q}_{AB} (\boldsymbol{x}_B - \boldsymbol{\mu}_B)$$

but *not* necessarily invert $\boldsymbol{Q}_{AA}$. We postpone the discussion of this numerical issue until Section 2.3.

The canonical parameterization

The canonical parameterization for a GMRF will be useful for successive conditioning.

Definition 2.2 (Canonical parameterization) *A GMRF $\boldsymbol{x}$ wrt $\mathcal{G}$ with canonical parameters $\boldsymbol{b}$ and $\boldsymbol{Q} > 0$ has density*

$$\pi(\boldsymbol{x}) \propto \exp\Big(-\frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} \Big),$$

i.e., the precision matrix is $\boldsymbol{Q}$ and the mean is $\boldsymbol{\mu} = \boldsymbol{Q}^{-1} \boldsymbol{b}$. We write the canonical parameterization as

$$\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \boldsymbol{Q}).$$

The relation to the normal distribution, is that $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}) = \mathcal{N}_C(\boldsymbol{Q}\boldsymbol{\mu}, \boldsymbol{Q})$.

Partition the indices into two nonempty sets A and B, and partition $\boldsymbol{x}$, $\boldsymbol{b}$ and $\boldsymbol{Q}$ accordingly as in (2.13) and (2.14). Two lemmas follow easily.

Lemma 2.1 *Let $\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \boldsymbol{Q})$, then*

$$\boldsymbol{x}_A \mid \boldsymbol{x}_B \sim \mathcal{N}_C(\boldsymbol{b}_A - \boldsymbol{Q}_{AB} \boldsymbol{x}_B, \boldsymbol{Q}_{AA}).$$

Lemma 2.2 *Let* $\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \boldsymbol{Q})$ *and* $\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{P}^{-1})$, *then*

$$\boldsymbol{x} \mid \boldsymbol{y} \sim \mathcal{N}_C(\boldsymbol{b} + \boldsymbol{P}\boldsymbol{y}, \boldsymbol{Q} + \boldsymbol{P}). \tag{2.16}$$

These results are useful for computing conditional densities with several sources of conditioning, for example, conditioning on observed data and a subset of variables. We can successively update the canonical parameterization, without explicitly computing the mean, until we actually need it. Computing the mean requires the solution of $\boldsymbol{Q}\boldsymbol{\mu} = \boldsymbol{b}$, but only matrix-vector products are required to update the canonical parameterization.

2.2.4 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 $\{\pi(x_i|\boldsymbol{x}_{-i})\}$. This approach was pioneered by Besag (1974, 1975) and the models are also known by the name *conditional autoregressions*, abbreviated as *CAR* models. We will now discuss this possibility and the specific conditions we must impose on the full conditionals to correspond to a valid GMRF.

Suppose we specify the full conditionals as normals with

$$\mathrm{E}(x_i \mid \boldsymbol{x}_{-i}) = \mu_i - \sum_{j:j\sim i} \beta_{ij}(x_j - \mu_j) \quad \text{and} \tag{2.17}$$

$$\mathrm{Prec}(x_i \mid \boldsymbol{x}_{-i}) = \kappa_i > 0 \tag{2.18}$$

for $i = 1, \ldots, n$, for some $\{\beta_{ij}, i \neq j\}$, and vectors $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$. Clearly, $\sim$ is defined implicitly by the nonzero terms of $\{\beta_{ij}\}$. These full conditionals must be consistent so that there exists a joint density $\pi(\boldsymbol{x})$ that will give rise to these full conditional distributions. Since $\sim$ is symmetric, this immediate gives the requirement that if $\beta_{ij} \neq 0$ then $\beta_{ji} \neq 0$. Comparing term by term with (2.5) and (2.6), we see that if we choose the entries of the precision matrix $\boldsymbol{Q}$ as

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

and also require that $\boldsymbol{Q}$ is symmetric, i.e.,

$$\kappa_i \beta_{ij} = \kappa_j \beta_{ji},$$

then we have a candidate for a joint density giving the specified full conditionals provided $\boldsymbol{Q} > 0$. The next result says that this candidate is unique.

Theorem 2.6 *Given the* n *normal full conditionals with conditional mean and precision as in (2.17) and (2.18), then* $\boldsymbol{x}$ *is a GMRF wrt a labelled graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *with mean* $\boldsymbol{\mu}$ *and precision matrix* $\boldsymbol{Q} = (Q_{ij})$,

where

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

provided $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}, i \neq j$, *and* $\boldsymbol{Q} > 0$.

To prove this result we need *Brook's lemma*.

Lemma 2.3 (Brook's lemma) *Let* $\pi(\boldsymbol{x})$ *be the density for* $\boldsymbol{x} \in \mathbb{R}^n$ *and define* $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \pi(\boldsymbol{x}) > 0\}$. *Let* $\boldsymbol{x}, \boldsymbol{x}' \in \Omega$, *then*

$$\frac{\pi(\boldsymbol{x})}{\pi(\boldsymbol{x}')} = \prod_{i=1}^{n} \frac{\pi(x_i|x_1, \ldots, x_{i-1}, x'_{i+1}, \ldots, x'_n)}{\pi(x'_i|x_1, \ldots, x_{i-1}, x'_{i+1}, \ldots, x'_n)} \tag{2.19}$$

$$= \prod_{i=1}^{n} \frac{\pi(x_i|x'_1, \ldots, x'_{i-1}, x_{i+1}, \ldots, x_n)}{\pi(x'_i|x'_1, \ldots, x'_{i-1}, x_{i+1}, \ldots, x_n)}. \tag{2.20}$$

If we fix $\boldsymbol{x}'$ then (2.19) (and (2.20)) represents $\pi(\boldsymbol{x})$, up to a constant of proportionality, using the set of full conditionals $\{\pi(x_i|\boldsymbol{x}_{-i})\}$. The constant of proportionality is found using that $\pi(\boldsymbol{x})$ integrates to unity.

Proof. [Brook's lemma] Start with the identity

$$\frac{\pi(x_n|x_1, \ldots, x_{n-1})\, \pi(x_1, \ldots, x_{n-1})}{\pi(x'_n|x_1, \ldots, x_{n-1})\, \pi(x_1, \ldots, x_{n-1})} = \frac{\pi(x_1, \ldots, x_{n-1}, x_n)}{\pi(x_1, \ldots, x_{n-1}, x'_n)}$$

from which it follows that

$$\pi(x_1, \ldots, x_n) = \frac{\pi(x_n|x_1, \ldots, x_{n-1})}{\pi(x'_n|x_1, \ldots, x_{n-1})} \pi(x_1, \ldots, x_{n-1}, x'_n).$$

Express the last term on the rhs similarly to obtain

$$\begin{aligned} \pi(x_1, \ldots, x_n) = {} & \frac{\pi(x_n|x_1, \ldots, x_{n-1})}{\pi(x'_n|x_1, \ldots, x_{n-1})} \\ & \times \frac{\pi(x_{n-1}|x_1, \ldots, x_{n-2}, x'_n)}{\pi(x'_{n-1}|x_1, \ldots, x_{n-2}, x'_n)} \\ & \times \pi(x_1, \ldots, x_{n-2}, x'_{n-1}, x'_n). \end{aligned}$$

By repeating this process (2.19) follows. The alternative (2.20) is proved similarly starting with

$$\pi(x_1, \ldots, x_n) = \frac{\pi(x_1|x_2, \ldots, x_n)}{\pi(x'_1|x_2, \ldots, x_n)} \pi(x'_1, x_2, \ldots, x_n)$$

and proceeding forward. □

Proof. [Theorem 2.6] Assume $\boldsymbol{\mu} = \boldsymbol{0}$ and fix $\boldsymbol{x}' = \boldsymbol{0}$. Then (2.19) simplifies to

$$\log \frac{\pi(\boldsymbol{x})}{\pi(\boldsymbol{0})} = -\frac{1}{2} \sum_{i=1}^{n} \kappa_i x_i^2 - \sum_{i=2}^{n} \sum_{j=1}^{i-1} \kappa_i \beta_{ij} x_i x_j. \tag{2.21}$$