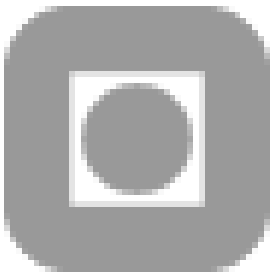
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Marginal Variances for Gaussian Markov Random Fields

by

Håvard Rue

PREPRINT
STATISTICS NO. 1/2005



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2005/S1-2005.pdf
Håvard Rue has homepage: http://www.math.ntnu.no/∼hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Marginal Variances for Gaussian Markov Random Fields

Håvard Rue
Department of Mathematical Sciences
NTNU, Norway

March 14, 2005

Abstract

Gaussian Markov random fields (GMRFs) are specified conditionally by its precision matrix meaning that its inverse, the covariance matrix, is not explicitly known. Computing the often dense covariance matrix directly using matrix inversion is often unfeasible due to time and memory requirement. In this note, we discuss a simple and fast algorithm to compute the marginal variances for a GMRF. We also provide extensions to deal with linear soft and hard constraints, essentially without extra costs.

KEYWORDS: Cholesky triangle, Conditional auto-regressions, Gaussian Markov random fields, Non-homogeneous autoregressive processes, Sparse matrices.

ADDRESS FOR CORRESPONDENCE: H. Rue, Department of Mathematical Sciences, The Norwegian University for Science and Technology, N-7491 Trondheim, Norway.
E-MAIL: Havard.Rue@math.ntnu.no
WWW-ADDRESS: http://www.math.ntnu.no/∼hrue

1 Introduction

A Gaussian Markov random field (GMRF) $\boldsymbol{x} = \{x_i : i \in \mathcal{V}\}$ is a $n = |\mathcal{V}|$-dimensional Gaussian random vector with additional conditional independence or Markov properties. Assume for simplicity that $\mathcal{V} = \{1, \ldots, n\}$. The conditional independence properties can be represented using an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $\mathcal{V}$ and edges $\mathcal{E}$, so that x_i and x_j are conditional independent if and only if $\{i, j\} \notin \mathcal{E}$. We then say that $\boldsymbol{x}$ is a GMRF with respect to $\mathcal{G}$. The edges in $\mathcal{E}$ is in one-to-one correspondence with the non-zero elements of the precision matrix of $\boldsymbol{x}$, $\boldsymbol{Q}$, in the sense that $\{i, j\} \in \mathcal{E}$ if and only if $Q_{ij} \neq 0$ for $i \neq j$. When $\{i, j\} \in \mathcal{E}$ we say that i and j are neighbours, which we denote by $i \sim j$.

GMRFs are also known as conditional auto-regressions (CARs) following seminal work of Besag (1974, 1975). GMRFs (and its intrinsic versions) have a broad usage in statistics, with important applications in structural time-series analysis, analysis of longitudinal and survival data, graphical models, semiparametric regression and splines, image analysis and spatial statistics. For references and examples, see Rue and Held (2005, Ch. 1).

A zero mean GMRFs are often specified implicitly through the full conditionals

$$x_i \mid \boldsymbol{x}_{-i} \sim \mathcal{N}(\sum_j \beta_{ij} x_j, 1/\kappa_i), \quad \kappa_i > 0, \; i = 1, \ldots, n. \tag{1}$$

The parameters $\{\beta_{ij}\}$ and $\{\kappa_i\}$ must satisfy consistency requirements for a joint density to exists, which is that $\boldsymbol{Q} = (Q_{ij})$ where $Q_{ii} = \kappa_i$ and $Q_{ij} = -\kappa_i \beta_{ij}$, is symmetric and positive definite. Then (1) defines a zero mean GMRF with precision matrix $\boldsymbol{Q}$. The number of neighbours to i is typical small and do not depend on n, which implies that the precision matrix $\boldsymbol{Q}$ is sparse with only $\mathcal{O}(n)$ non-zero terms. The specification of the precision matrix through (1) means that the covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$ is only implicitly know. Although we formally can invert $\boldsymbol{Q}$, the dimension n is typically large ($10^3 - 10^5$) so inverting $\boldsymbol{Q}$ directly will be costly and inconvenient.

In this note we discuss a simple and fast algorithm to compute the diagonal of $\boldsymbol{\Sigma}$, the marginal variances, applicable for GMRFs with large dimension. The motivation for this work is a not-well-known matrix identity which appeared in a IEEE conference proceedings (Takahashi et al., 1973), see also Erisman and Tinney (1975). In our setting, the identity is as follows. Let $\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$ be the Cholesky-decomposition of $\boldsymbol{Q}$ where $\boldsymbol{D}$ is a diagonal matrix and where $\boldsymbol{V}$ is a lower triangular matrix with ones at the diagonal, then

$$\boldsymbol{\Sigma} = \boldsymbol{D}^{-1}\boldsymbol{V}^{-1} + (\boldsymbol{I} - \boldsymbol{V}^T)\boldsymbol{\Sigma}. \tag{2}$$

The upper triangle defines recursions for Σ_{ij} which provided the basis for fast computations of the marginal variances of x_1 to x_n. However, the identity (2) give little insight in how Σ_{ij} depend on elements of $\boldsymbol{Q}$ and the graph $\mathcal{G}$. We will therefore in Section 2 derive the recursions defined in (2) "statistically", starting from a simulation algorithm for GMRFs and using the relation between $\boldsymbol{Q}$ and its Cholesky triangle given by the global Markov property. We use the same technique to prove Theorem 1 which locate a set of indices for which the recursions are to be solved to obtain the marginal variances. (A similar result was also given in Takahashi et al. (1973)). Our Theorem 2 shows under what conditions this set is tight. We also generalise the recursions to compute marginal variances for GMRFs defined with additional soft and hard linear constraints, for example under a sum-to-zero constraint. Practical issues appearing when implementing the algorithm using the Cholesky triangle of $\boldsymbol{Q}$ computed using sparse matrix libraries, are also discussed.

The recursions for Σ_{ij} are applicable to a GMRF with respect any graph $\mathcal{G}$ and generalise the well known (fixed-interval) Kalman recursions for smoothing applicable for dynamic models. The computational effort to solve the recursions depends on both the neighbourhood structure in $\mathcal{G}$ and

the size n. For typical spatial applications, the costs is $\mathcal{O}(n\log(n)^2)$ when the Cholesky triangle of $\boldsymbol{Q}$ is available.

The outline of this note is as follows. We derive the recursions in Section 2 and the corrections needed to account for soft and hard linear constraints in Section 3. Practical issues are discussed in Section 4 which discuss how to compute the recursions efficiently using numerical algorithms for sparse matrices. In Section 5 we discuss a geostatistical application using the recursions to assess the error in spatial predictions. We end with a discussion in Section 6.

2 The Recursions

Let $\boldsymbol{L}$ be the Cholesky triangle of $\boldsymbol{Q}$ such that $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^T$ and $\boldsymbol{L}$ is lower triangular. The Cholesky triangle is the starting point both for producing (unconditional and conditional) samples from a zero mean GMRF and to evaluate the log-density for any configuration. Refer to Rue and Held (2005, Ch. 2) for algorithms and further details. In short, (unconditional) samples are found as the solution of $\boldsymbol{L}^T\boldsymbol{x} = \boldsymbol{z}$ where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The log-density is computed using that $\log|\boldsymbol{Q}| = 2\sum_i \log L_{ii}$.

Using that the solution of $\boldsymbol{L}^T\boldsymbol{x} = \boldsymbol{z}$ is a sample from a zero mean GMRF with precision matrix $\boldsymbol{Q}$, we obtain that

$$x_i \mid x_{i+1}, \ldots, x_n \;\sim\; \mathcal{N}(-\frac{1}{L_{ii}} \sum_{k=i+1}^{n} L_{ki}x_k, 1/L_{ii}^2), \quad i = n, \ldots, 1. \tag{3}$$

Eq. (3) provide a sequential representation of the GMRF backward in "time" i, as

$$\pi(\boldsymbol{x}) = \prod_{i=n}^{1} \pi(x_i \mid x_{i+1}, \ldots, x_n).$$

Let $\boldsymbol{L}_{i:n}$ be the lower-right $(n-i-1)\times(n-i-1)$ submatrix of $\boldsymbol{L}$. It follows directly from (3) that $\boldsymbol{L}_{i:n}\boldsymbol{L}_{i:n}^T$ is the precision matrix of $\boldsymbol{x}_{i:n} = (x_i, \ldots, x_n)^T$. The non-zero pattern in $\boldsymbol{L}$ is important for the recursions, see Rue and Held (2005, Ch. 2) for further details about the relation between $\boldsymbol{Q}$ and $\boldsymbol{L}$. Zeros in the i'th column of $\boldsymbol{L}$, L_{ki} relates directly to the conditional independence properties of $\pi(\boldsymbol{x}_{i:n})$. We have for $i < k$

$$-\frac{1}{2}\boldsymbol{x}_{i:n}^T\boldsymbol{L}_{i:n}\boldsymbol{L}_{i:n}^T\boldsymbol{x}_{i:n} = -x_i x_k L_{ii} L_{ki} + \text{remaining terms}$$

hence $L_{ki} = 0$ is equivalent to x_i and x_k are conditional independent given $x_{i+1}, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n$. This is similar to the fact that $Q_{ij} = 0$ is equivalent to x_i and x_j are conditional independent given the remaining elements of $\boldsymbol{x}$. To ease the notation, define the set

$$F(i,k) = \{i+1, \ldots, k-1, k+1, \ldots, n\}, \quad 1 \le i \le k \le n$$

which is the future of i except k. Then for $i < k$

$$x_i \perp x_k \mid \boldsymbol{x}_{F(i,k)} \Longleftrightarrow L_{ki} = 0. \tag{4}$$

It is however not easy to verify that $x_i \perp x_k \mid \boldsymbol{x}_{F(i,k)}$ without computing $\boldsymbol{L}$ and check if $L_{ki} = 0$ or not. However, the global Markov property provides a sufficient condition for L_{ki} to be zero; If i and $k > i$ are separated by $F(i,k)$ in $\mathcal{G}$, then $x_i \perp x_k \mid \boldsymbol{x}_{F(i,k)}$ and $L_{ki} = 0$. This sufficient criteria depends only on the graph $\mathcal{G}$. If we use this to conclude that $L_{ki} = 0$, then this is true for all $\boldsymbol{Q} > 0$ with fixed graph $\mathcal{G}$. In particular, if $k \sim i$ then L_{ki} is non-zero in general. This imply that the Cholesky triangle is in general more dense than the lower triangle of $\boldsymbol{Q}$.

To obtain the recursions for $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$, we note that (3) implies that

$$\Sigma_{ij} = \delta_{ij}/L_{ii}^2 - \frac{1}{L_{ii}} \sum_{k \in \mathcal{I}(i)}^{n} L_{ki}\Sigma_{kj}, \quad j \geq i,\ i = n, \ldots, 1, \tag{5}$$

where $\mathcal{I}(i)$ as those k where L_{ki} is non-zero,

$$\mathcal{I}(i) = \{k > i\ :\ L_{ki} \neq 0\} \tag{6}$$

and δ_{ij} is one if $i = j$ and zero otherwise. We can compute all covariances directly using (5) but the order of the indices are important; The outer loop is i from n to 1 and the inner loop is j from n to i. The first and last computed covariance is then Σ_{nn} and Σ_{11}, respectively.

It is possible to derive from (1) a similar set of equations to (5) which relates covariances to elements of $\boldsymbol{Q}$ instead of elements of $\boldsymbol{L}$, see Besag (1981). However, these equations does not define recursions.

Example 1 *Let $n = 3$, $\mathcal{I}(1) = \{2, 3\}$, $\mathcal{I}(2) = \{3\}$, then (5) gives*

$$\begin{aligned}
\Sigma_{33} &= \frac{1}{L_{33}^2} \\
\Sigma_{23} &= -\frac{1}{L_{22}} \left(L_{32}\Sigma_{33}\right) \\
\Sigma_{22} &= \frac{1}{L_{22}^2} - \frac{1}{L_{22}} \left(L_{32}\Sigma_{32}\right) \\
\Sigma_{13} &= -\frac{1}{L_{11}} \left(L_{21}\Sigma_{23} + L_{31}\Sigma_{33}\right) \\
\Sigma_{12} &= -\frac{1}{L_{11}} \left(L_{21}\Sigma_{22} + L_{31}\Sigma_{32}\right) \\
\Sigma_{11} &= \frac{1}{L_{11}^2} - \frac{1}{L_{11}} \left(L_{21}\Sigma_{21} + L_{31}\Sigma_{31}\right)
\end{aligned}$$

where we also need to use that $\boldsymbol{\Sigma}$ is symmetric.

Our aim is to compute the marginal variances $\Sigma_{11}, \ldots, \Sigma_{nn}$. In order to do so, we need to compute Σ_{ij} (or Σ_{ji}) for all ij in some set $\mathcal{S}$ as evident from (5). Let the elements in $\mathcal{S}$ be unordered, meaning that if $ij \in \mathcal{S}$ then also $ji \in \mathcal{S}$. If the recursions can be solved by only computing Σ_{ij} for all $ij \in \mathcal{S}$ we say that the recursions are solvable using $\mathcal{S}$, or simply $\mathcal{S}$ is solvable. It is evident from (5) that $\mathcal{S}$ must satisfy

$$ij \in \mathcal{S} \text{ and } k \in \mathcal{I}(i) \implies kj \in \mathcal{S} \tag{7}$$

Further, we need that $ii \in \mathcal{S}$ for $i = 1, \ldots, n$. Of course $\mathcal{S} = \mathcal{V} \times \mathcal{V}$ is a such set, but we want $|\mathcal{S}|$ to be minimal to avoid unnecessary computations. A such minimal set depends however on the numerical values in $\boldsymbol{L}$, or $\boldsymbol{Q}$ implicitly. Denote by $\mathcal{S}(\boldsymbol{Q})$ a minimal set. The following result identify a solvable set $\mathcal{S}^*$ containing the union of $\mathcal{S}(\boldsymbol{Q})$ for all $\boldsymbol{Q} > 0$ with a fixed graph $\mathcal{G}$.

Theorem 1 *The union of $\mathcal{S}(\boldsymbol{Q})$ for all $\boldsymbol{Q} > 0$ with fixed graph $\mathcal{G}$, is a subset of*

$$\mathcal{S}^* = \{ij \in \mathcal{V} \times \mathcal{V}\ :\ j \geq i,\ i \text{ and } j \text{ are not separated by } F(i, j)\}$$

and the recursions in (5) are solvable using $\mathcal{S}^*$.

Proof. *We first note that* $ii \in \mathcal{S}^*$*, for* $i = 1, \ldots, n$*, since* i *and* i *are not separated by* $F(i,i)$*. We will now verify that the recursions are solvable using* $\mathcal{S}^*$*. The global Markov property ensure that if* $ij \notin \mathcal{S}^*$ *then* $L_{ji} = 0$ *for all* $\boldsymbol{Q} > 0$ *with fixed graph* $\mathcal{G}$*. We use this to replace* $\mathcal{I}(i)$ *with* $\mathcal{I}^*(i) = \{k > i : ik \in \mathcal{S}^*\}$ *in (7), which is legal since* $\mathcal{I}(i) \subseteq \mathcal{I}^*(i)$ *and the difference only identify terms* L_{ki} *which are zero. It is now sufficient to show that*

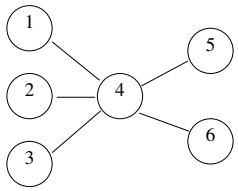
$$ij \in \mathcal{S}^* \text{ and } ik \in \mathcal{S}^* \implies kj \in \mathcal{S}^* \tag{8}$$

which implies (7). Eq. (8) is trivially true for $i \leq k = j$*. Fix now* $i < k < j$*. Then* $ij \in \mathcal{S}^*$ *says that there exists a path* $i, i_1, \ldots, i_n, j$*, where* $i_1, \ldots, i_n$ *are all smaller than* i*, and* $ik \in \mathcal{S}^*$ *says that exists a path* $i, i'_1, \ldots, i'_{n'}, k$*, where* $i'_1, \ldots, i'_{n'}$ *are all smaller than* i*. Then there is a path from* k *to* i *and from* i *to* j *where all nodes are less or equal to* i*, but then also less than* k *since* $i < k$*. Hence,* k *and* j *are not separated by* $F(k,j)$ *so* $kj \in \mathcal{S}^*$*. Finally, since* $\mathcal{S}^*$ *contains* $11, \ldots, nn$ *and only depend on* $\mathcal{G}$*, it must contain the union of all* $\mathcal{S}(\boldsymbol{Q})$ *since each* $\mathcal{S}(\boldsymbol{Q})$ *is minimal.* ■

An alternative interpretation of $\mathcal{S}^*$, is that it identify from the graph $\mathcal{G}$ only, all possible non-zero elements in $\boldsymbol{L}$. Some of these might turn out to be zero depending on the conditional independence properties of the marginal density for $\boldsymbol{x}_{i:n}$ for $i = n, \ldots, 1$, see (4). In particular, if $j \sim i$ and $j > i$ then $ij \in \mathcal{S}^*$. This provides the lower bound for the size of $\mathcal{S}^*$,

$$|\mathcal{S}^*| \geq n + |\mathcal{E}|. \tag{9}$$

Example 2 *Let* $\boldsymbol{x} = (x_1, \ldots, x_6)^T$ *be a GMRF with respect to the graph*



then the set of the possible non-zero terms in $\boldsymbol{L}$*, is*

$$\mathcal{S}^* = \{11, 22, 33, 41, 42, 43, 44, 54, 55, 64, 65, 66\}. \tag{10}$$

The only element in $\mathcal{S}^*$ *where the corresponding element in* $\boldsymbol{Q}$ *is zero, is* 65*, this because* 5 *and* 6 *are not separated by* $F(5,6) = \emptyset$ *in* $\mathcal{G}$ *(due to* 4*), so* $|\mathcal{S}^*| = n + |\mathcal{E}| + 1$*.*

The recursions are particular simple if $\mathcal{G}$ is decomposable allowing us to be more specific about the set $\mathcal{S}^*$. Let p be a permutation of the vertices, meaning that vertex i is labelled p_i after the permutation.

Theorem 2 *If the graph* $\mathcal{G}$ *is decomposable then there exists a permutation* p *of the vertices, such that*

$$|\mathcal{S}^*| = n + |\mathcal{E}| \tag{11}$$

and $\mathcal{S}^*$ *is the union of* $\mathcal{S}(\boldsymbol{Q})$ *for all* $\boldsymbol{Q} > 0$ *with fixed graph* $\mathcal{G}$*.*

Note that (11) is the lower bound of $|\mathcal{S}^*|$, see (9), hence $\mathcal{S}^*$ reduces to $\{ij \in \mathcal{V} \times \mathcal{V} \ : \ i \sim j \text{ or } i = j\}$ after the vertices in $\mathcal{G}$ are permuted by p.

Proof. *Rose et al. (1976) shows if (11) holds then the graph is decomposable and for all decomposable graphs there exists a permutation p of the vertices such that (11) holds. Assume for simplicity that $p_i = i$ for all i. From the lower bound (9) we know that $\mathcal{S}^*$ equals $\{ij \in \mathcal{V} \times \mathcal{V} \,:\, i \sim j \text{ or } i = j\}$. We will now construct one particular $\boldsymbol{Q}$, denoted by $\widehat{\boldsymbol{Q}}$ such that $\mathcal{S}(\widehat{\boldsymbol{Q}}) = \mathcal{S}^*$ and from this we can deduce that $\mathcal{S}^*$ is the union of $\mathcal{S}(\boldsymbol{Q})$ for all $\boldsymbol{Q} > 0$ with fixed graph $\mathcal{G}$. Let $\widehat{L}_{ji} = 1$, $j \geq i$, for all $ij \in \mathcal{S}^*$ and define $\widehat{\boldsymbol{Q}} = \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^T$ so that*

$$\widehat{Q}_{ij} = |\{k \in \mathcal{V} \,:\, k \sim i \text{ or } k = i, k \sim j \text{ or } k = j, \text{ and } k \leq \min(i,j)\}|. \tag{12}$$

If $i \sim j$, then $k = \min(i,j)$ ensure that the set in (12) is non-empty, hence $\widehat{Q}_{ij} \neq 0$. For $i \not\sim j$ we know that $ij \notin \mathcal{S}^*$, and then $\widehat{Q}_{ij} = 0$ which can be seen as follows; Assume $\widehat{Q}_{ij} > 0$ and $i < j$, then there must exists a k such that the set in (12) is non-empty. For a such k there is a path from i to k and from k to j where $k < i$ hence $ij \in \mathcal{S}^*$ which is a contradiction. Since $\widehat{Q}_{ij} \neq 0$ if and only if $ij \in \mathcal{S}^*$ then $\mathcal{I}(i) = \mathcal{I}^*(i)$. From (5) it is evident that to compute Σ_{ii} we need Σ_{ij} for all $j \in \mathcal{I}^*(i)$. To compute $\Sigma_{11}, \ldots, \Sigma_{nn}$, we need Σ_{ij} for all $ij \in \cup_{i=1}^n \mathcal{I}^*(i)$. Since $\mathcal{S}^* = \bigcup_{i=1}^n \{\mathcal{I}^*(i) \cup ii\}$ we must compute Σ_{ij} for all $ij \in \mathcal{S}^*$ hence $\mathcal{S}(\widehat{\boldsymbol{Q}}) = \mathcal{S}^*$. ■

The maximum cardinality search-algorithm of Tarjan and Yannakakis (1984) can be used to find the permutation need to obtain (11).

Example 3 *A homogeneous autoregressive model of order p satisfy*

$$x_i \mid x_1, \ldots, x_{i-1} \;\sim\; \mathcal{N}(\sum_{j=1}^{p} \phi_j x_{i-j}, 1), \quad i = 1, \ldots, n,$$

for some parameters $\{\phi_j\}$ where for simplicity we assume that $x_{-1}, \ldots, x_{-p+1}$ are fixed. Let $\{y_i\}$ be independent Gaussian observations of x_i such that $y_i \sim \mathcal{N}(x_i, 1)$. Then $\boldsymbol{x}$ conditioned on the observations is Gaussian where the precision matrix $\boldsymbol{Q}$ is a band-matrix with band-width p and $\boldsymbol{L}$ is lower triangular with the same bandwidth. When $\{\phi_j\}$ are such that $Q_{ij} \neq 0$ for all $|i-j| \leq p$, then the graph is decomposable and the permutation needed in Theorem 2 is the identity. In this case the recursions correspond to the (fixed-interval) smoothing recursions derived from the Kalman filter for (Gaussian) linear state-space models.

Although the situation is particular simple for decomposable graphs, most GMRFs are defined with respect to graphs that are not decomposable. This is the case for GMRFs used in spatial or spatio-temporal applications, but also for GMRFs used in temporal models outside the state-space framework. Additional to be able to identify the set $\mathcal{S}^*$ efficiently, we also need to compute the Cholesky triangle $\boldsymbol{L}$. It is important to have efficiently algorithms for these tasks as the dimension of GMRFs is typically large. Fortunately, algorithms that compute $\boldsymbol{L}$ efficiently also minimise (approximately) the size of $\mathcal{S}^*$ and then also the cost of solving the recursions. We return to this and other practical issues in Section 5, after discussing how to compute marginal variances for GMRFs with additional linear constraints.

3 Correcting for hard and soft linear constraints

We will now demonstrate how we can correct the marginal variances computed in (5) to account for additional linear constraints, for example a simple sum-to-zero constraint. Let $\boldsymbol{A}$ be a $k \times n$ matrix of rank k. The goal is now to compute the marginal variances of the GMRF under the linear constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{e}$. If $\boldsymbol{e}$ is fixed we denote the constraint as hard, and if $\boldsymbol{e}$ is a realisation of $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{e}}, \boldsymbol{\Sigma}_{\boldsymbol{e}})$, $\boldsymbol{\Sigma}_{\boldsymbol{e}} > 0$, we denote the constraint as soft.

A constrained GMRF is also a GMRF meaning that recursions (5) are still valid using the Cholesky triangle for the constrained GMRF. Since linear constraints destroy the sparseness of the precision matrix it will not allow fast computation of the marginal variances. However, the covariance matrix under hard linear constraints, $\widetilde{\boldsymbol{\Sigma}}$, relates to the unconstrained covariance matrix $\boldsymbol{\Sigma}$ as

$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \boldsymbol{Q}^{-1}\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\right)^{-1} \boldsymbol{A}\boldsymbol{Q}^{-1}. \tag{13}$$

There is a similar relation with a soft constraint (Rue and Held, 2005, Ch. 2). Assume a hard constraint in the following. It is evident from (13) that

$$\widetilde{\Sigma}_{ii} = \Sigma_{ii} - \left(\boldsymbol{Q}^{-1}\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\right)^{-1} \boldsymbol{A}\boldsymbol{Q}^{-1}\right)_{ii}, \quad i = 1, \ldots, n.$$

Hence, we can compute the diagonal of $\boldsymbol{\Sigma}$ and then correct it to account for the hard constraints. Define the $n \times k$ matrix $\boldsymbol{W}$ as $\boldsymbol{Q}^{-1}\boldsymbol{A}^T$ which is found from solving $\boldsymbol{Q}\boldsymbol{W} = \boldsymbol{A}^T$ for each of the k columns of $\boldsymbol{W}$. As the Cholesky triangle to $\boldsymbol{Q}$ is available, the j'th column of $\boldsymbol{W}$, $\boldsymbol{W}_j$, is found by solving $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{A}_j^T$ and then solving $\boldsymbol{L}^T\boldsymbol{W}_j = \boldsymbol{v}$. We now see that $\widetilde{\Sigma}_{ii} = \Sigma_{ii} - C_{ii}$ where $\boldsymbol{C} = \boldsymbol{W}\left(\boldsymbol{A}\boldsymbol{W}\right)^{-1}\boldsymbol{W}^T$. We only need the diagonal of $\boldsymbol{C}$. Let $\boldsymbol{V} = \boldsymbol{W}(\boldsymbol{A}\boldsymbol{W})^{-1}$, and then $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{W}^T$ and $C_{ii} = \sum_{l=1}^{k} V_{il}W_{il}$. The cost of computing $\boldsymbol{V}$ and $\boldsymbol{W}$ is for large k dominated by factorising the (dense) $k \times k$ matrix $\boldsymbol{A}\boldsymbol{W}$, which is cubic in k. As long as k is not to large it is nearly free to correct for linear soft and hard constraints.

A special case of the hard constraint is to condition on a subset, B say, of the nodes in $\mathcal{G}$. This is however equivalent to compute the marginal variances for $\boldsymbol{x}_A|\boldsymbol{x}_B$ where $\boldsymbol{x} = (\boldsymbol{x}_A, \boldsymbol{x}_B)$. In most cases it is more efficient not to use (13), but utilise that $\boldsymbol{x}_A|\boldsymbol{x}_B$ is a GMRF with precision matrix $\boldsymbol{Q}_{AA}$ and mean $\boldsymbol{\mu}$ given by the solution of $\boldsymbol{Q}_{AA}\boldsymbol{\mu} = -\boldsymbol{Q}_{AB}\boldsymbol{x}_B$. (Note that solving for $\boldsymbol{\mu}$ require only the Cholesky triangle of $\boldsymbol{Q}_{AA}$ which is needed in any case for the recursions.) The marginal variances are then computed using (5), possibly correcting for additional linear constraints using (13).

4 Practical issues

Since the precision matrix $\boldsymbol{Q}$ is a sparse matrix we can take advantage of numerical algorithms for sparse symmetric positive definite matrices. Such algorithms are very efficient and makes it possible to factorise precision matrices of dimension $10^3 - 10^5$ without to much effort. A major benefit is that these algorithms also minimise (approximately) the size of $\mathcal{S}^*$ and then also the cost of solving the recursions. Rue (2001) and Rue and Held (2005) discuss numerical algorithms for sparse matrices from a statistical perspective and how to apply them for GMRFs.

An important ingredient in sparse matrix algorithms, is to permute the vertices to minimise (approximately) the number of non-zero terms in $\boldsymbol{L}$. The idea, is that if L_{ji} is known to be zero, then L_{ji} is not computed. It turns out that the set $\mathcal{S}^*$ is exactly the set vertices for which L_{ji} is computed, see Rue and Held (2005, Sec. 2.4.1). Hence, an efficient permutation to compute $\boldsymbol{L}$ minimise (approximately) $|\mathcal{S}^*|$, hence is also an efficient permutation for solving the recursions. However, this implies that we have little control over which other Σ_{ij}'s that are computed in the recursions, apart from the diagonal and for $i \sim j$.

Permutation schemes based on the idea of nested dissection are particular useful in statistical applications. The idea is to find a small separating subset that divides the graph into two (roughly) equal parts, label the nodes in the separating set after the others, and continuous recursively. For such a permutation, the computational complexity to compute $\boldsymbol{L}$ for a GMRF on a square $m \times m$ lattice with a local neighbourhood, is $\mathcal{O}(n^{3/2})$ for $n = m^2$. This also the optimal complexity in the order sense. The number of possible non-zero terms in $\boldsymbol{L}$ is $\mathcal{O}(n\log(n))$ which corresponds to the

size of $\mathcal{S}^*$. The complexity for solving the recursions can estimated from these numbers; We need to compute $\mathcal{O}(n\log(n))$ covariances, each involving on average $\mathcal{O}(\log(n))$ terms in $\mathcal{I}^*(i)$, which in total gives a cost of $\mathcal{O}(n\log(n)^2)$ operations. For a local GMRF on a $m \times m \times m$ cube with $n = m^3$, then the size of $\mathcal{S}^*$ is $\mathcal{O}(n^{4/3})$ and the cost of solving the recursions is then $\mathcal{O}(n^{5/3})$ which is dominated by the cost of factorising $\boldsymbol{Q}$ which is $\mathcal{O}(n^2)$.

A practical concern arise when numerical libraries return a list with the non-zero elements in $\boldsymbol{L}$, but the set $\mathcal{S}^*$ or $\mathcal{S}(\boldsymbol{Q})$ is needed by the recursions. In fact, any easily obtainable solvable set $\mathcal{S}(\boldsymbol{Q})^+$ where $\mathcal{S}(\boldsymbol{Q}) \subseteq \mathcal{S}(\boldsymbol{Q})^+ \subseteq \mathcal{S}^*$ is acceptable. A simple approach to obtain *a* $\mathcal{S}(\boldsymbol{Q})^+$ is the following. Let $\mathcal{S}_0 = \{j \geq i \ : \ L_{ji} \neq 0\}$. Traverse the set $\mathcal{S}_0$ with i from n to 1 as the outer loop, and j from n to i such that $ij \in \mathcal{S}_0$. For each ij, check for each $k \in \mathcal{I}(i)$ if $kj \in \mathcal{S}_0$. If this is not true, then add kj to $\mathcal{S}_0$. Repeat this procedure until no changes appear in $\mathcal{S}_0$. By construction, $\mathcal{S}_0 \subseteq \mathcal{S}^*$ and $\mathcal{S}_0$ is solvable, hence we may use $\mathcal{S}(\boldsymbol{Q})^+ = \mathcal{S}_0$. Two iterations are often sufficient to obtain $\mathcal{S}(\boldsymbol{Q})^+$, where the last verify only that $\mathcal{S}_0$ is solvable. Alternatively, $\mathcal{S}^*$ can either computed directly or extracted from an intermediate result in the sparse matrix library, if this is easily accessible.

Needless to say, solving the recursions efficiently requires very careful implementation in a appropriate language, but this is the rule not the exception working with sparse matrices.

5 Example

In this example we will use the algorithm to compute the marginal variance to assess the error doing spatial prediction on a (square) spatial domain, also knows as (Bayesian) *kriging*. We assume the mean function is a linear combination of p know functions with parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ with a Gaussian prior. For the (intrinsic) Gaussian field, we assume a variogram corresponding to the de Wijs process. Although the precision matrix for the de Wijs process is dense, there exists an accurate intrinsic GMRF approximation. Mondal and Besag (2004) shows that the variogram of the intrinsic GMRF defined by the full conditionals,

$$\mathrm{E}(x_{ij} \mid \boldsymbol{x}_{-ij}) = \frac{1}{4}\left(x_{i+1,j} + x_{i-1,j} + x_{i,j+1} + x_{i,j-1}\right), \quad \mathrm{Prec}(x_{ij} \mid \boldsymbol{x}_{-ij}) = \kappa > 0$$

converge to that one of the de Wijs process as the lattice resolution increases. The precision κ is for simplicity assumed fixed. Let $\boldsymbol{y}$ be k noisy observations of $\{x_i : i \in \mathcal{Y}\}$ where $y_i = x_i + \epsilon_i$ where $\{\epsilon_i\}$ are independent Gaussians with known variances $\{\sigma_i^2\}$. The posterior for $(\boldsymbol{x}, \boldsymbol{\beta})$ is then

$$\pi(\boldsymbol{x}, \boldsymbol{\beta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{x})\pi(\boldsymbol{\beta}) \prod_{i \in \mathcal{Y}} \pi(y_i \mid x_i).$$

The task is to compute the posterior mean and posterior marginal variance for $\boldsymbol{x}$ and $\boldsymbol{\beta}$. Note that $\boldsymbol{x}$ and $\boldsymbol{\beta}$, jointly, is a GMRF where each of the p (global) nodes corresponding to $\boldsymbol{\beta}$ have edges to each of the n nodes corresponding to $\boldsymbol{x}$.

An advantage with GMRF models is that the size of $|\mathcal{Y}|$ does not influence the computational costs, neither is there any requirements that all σ_i^2 should be equal. There is neither any real cost due to p. To compute the conditional mean we need to permute the vertices and factorise $\boldsymbol{Q}$, while computing the marginal variance makes use of these quantities in the recursions. The obtained CPU-usage are displayed in Table 1 for a 50×50 to 200×200 lattice and $p = 5$. The computations where performed using the free software library GMRFLib (Rue and Follestad, 2002) and a 2.6MHz laptop. The speed is impressive considering the dimension of the matrices, which range from $2\,500 \times 2\,500$ to $40\,000 \times 40\,000$. However, the cost could be reduced further, as to compute the permutation for the 200×200 lattice require about 3.5 seconds. We know that the p nodes corresponding to $\boldsymbol{\beta}$ has to be numbered last and using this information we could have decreased the time to compute

CPU-time	50×50	100×100	200×200
$\mathrm{E}(x_{ij} \mid \boldsymbol{y})$	0.06	0.35	6.0
$\mathrm{Var}(x_{ij} \mid \boldsymbol{y})$	0.04	0.29	3.7

Table 1: Cpu-time in seconds use to compute on a 2.6MHz laptop the conditional mean and marginal variance for a $m \times m$ lattice, for $m = 50, 100$ and 200, with additional 5 global nodes.

the permutation to about $1/10 \times 3.5$ seconds. With this adjustment, the computational cost using 400×400 lattice, i.e. a $160\,000 \times 160\,000$ matrix, is about 10 seconds each for both computing the conditional expectation and the marginal variance.

We are not limited to only use the de Wijs process but other intrinsic GMRFs (Besag and Kooperberg, 1995; Rue and Held, 2005) and non-intrinsic covariance functions can be used as well. Rue and Tjelmeland (2002) demonstrate how to fit Gaussian fields with Exponential, Gaussian, Spherical and Matérn covariance functions to GMRFs. For these choices there is no limiting argument as for the de Wijs process, but each value of the correlation range is fitted separately. However, the size of the neighbourhood must be 5×5 in order to obtain acceptable fits. The CPU-usage does not increase that much for these choices compared to the values in Table 1.

6 Discussion

In this note we have discussed an algorithm to compute the marginal variances for a GMRF specified by a sparse precision matrix. Extensions are given to correct for hard and soft constraints, essentially without extra costs. The derivation of the recursions starts from an equivalent formulation of the GMRF as a non-homogeneous autoregressive process, and then use properties of conditional independence and the interpretation of the elements in the Cholesky triangle of the precision matrix. The algorithm is potentially very useful as GMRFs have found applications in many fields (see Rue and Held (2005, Ch. 1)) and because marginal variances are of direct interest for statisticians. Although we always can estimate marginal variances from independent realisations from the (constrained) GMRF, exact computations are in general preferable. The nice connection between GMRFs and numerical methods for sparse matrices makes such computations feasible in practise for even huge GMRFs, and easily accessible using the library GMRFLib; a free software library for GMRFs (Rue and Follestad, 2002).

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36(2):192–225.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.

Besag, J. (1981). On a system of two-dimentional recurrence equations. *Journal of the Royal Statistical Society, Series B*, 43(3):302–309.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

Erisman, A. M. and Tinney, W. F. (1975). On computing certain elements of the inverse of a sparse matrix. *Communications of the ACM*, 18(3):177–179.

Mondal, D. and Besag, J. (2004). Variogram calculations for first-order intrinsic autoregressions. Technical Report, Department of Statistics, University of Washington, Seattle.

Rose, D., Tarjan, R., and Lueker, G. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5(2):90–123.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B*, 63(2):325–338.

Rue, H. and Follestad, T. (2002). GMRFLib: A C-library for fast and exact simulation of Gaussian Markov random fields. Statistics Report No. 1, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–50.

Takahashi, K., Fagan, J., and Chen, M. S. (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. In *8th PICA Conference proceedings*, pages 63–69. IEEE Power Engineering Society. Papers presented at the 1973 Power Industry Computer Application Conference in Minneapolis, Minnesota.

Tarjan, R. E. and Yannakakis, M. (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing*, 13(3):566–579.