Solutions to Sparse Multilevel Matrix Problems

BY TUI H. NOLAN AND MATT P. WAND

University of Technology Sydney

1st November, 2018

Abstract

We define and solve classes of sparse matrix problems that arise in multilevel modeling and data analysis. The classes are indexed by the number of nested units, with two-level problems corresponding to the common situation in which data on level 1 units are grouped within a two-level structure. We provide full solutions for two-level and three-level problems and their derivations provide blueprints for the challenging, albeit rarer in applications, higher-level versions of the problem. Whilst our linear system solutions are a concise recasting of existing results, our matrix inverse sub-block results are novel and facilitate streamlined computation of standard errors in frequentist inference as well as allowing streamlined mean field variational Bayesian inference for models containing higher-level random effects.

Keywords: Best linear unbiased prediction; Linear mixed models; Longitudinal data analysis; Panel data; Small area estimation; Variational inference.

1 Introduction

Higher-level sparse matrices arise in statistical models for multilevel data, such as units grouped according to geographical sub-regions or repeated measures on medical study patients (e.g. Goldstein, 2010). Other areas of statistics and econometrics that use essentially the same types of models are longitudinal data analysis (e.g. Fitzmaurice *et al.*, 2008), panel data analysis (e.g. Baltagi, 2013) and small area estimation (e.g. Rao & Molina, 2015). Linear mixed models (e.g. McCulloch, Searle & Neuhaus, 2008) are the main vehicle for modeling, fitting and inference. While they can be extended to generalized linear mixed models to cater for skewed, categorical and count response data, ordinary linear mixed models for Gaussian responses have the most relevance for the sparse matrix results presented here.

Both frequentist and Bayesian estimates of the fixed and random effects can be expressed succinctly in terms of ridge regression-type expressions involving design matrices (e.g. Henderson, 1975). However, typically the design matrices are sparse and naïve computation of the fixed and random effects estimates for large numbers of groups is inefficient and storage-greedy. For example, a random intercept linear mixed model for data with $1,000$ groups and 100 observations per group involves a random effects design matrix containing 100 million entries of which 99.9% are zeroes. Streamlined computation of the best linear unbiased predictors of the fixed and random effects are well-documented with Section 2.2 of Pinheiro & Bates (2000) being a prime example. The matrix algebraic notion of *QR decomposition* plays a central role in numerically stable least squares-based fitting of linear models (e.g. Gentle, 2007) and also arises in the current context. It is important to note that such computations are performed *after* estimates of the covariance matrix parameters have been obtained via approaches such as minimum norm quadratic unbiased estimation or restricted maximum likelihood. Streamlined computation of covariance matrix estimates is tackled in, for example, Longford (1987). Given the covariance matrix estimates, implementation-ready matrix algebraic results for streamlined standard error calculations are not, to the best of our knowledge, present in the existing literature. These rely on efficient extraction of sub-blocks of the inverses of potentially very large sparse symmetric matrices. Presentation of these results, in the form of four theorems, is our main novel contribution. In the interests of conciseness and digestibility, we do not delve into the linear mixed model ramifications here – which are long-winded due to the various cases that require separate treatment. This article is purely concerned with generic matrix algebraic facts and, whilst motivated by statistical analysis, is totally free of statistical concepts in its main results and derivations. Ramifications for frequentist and variational Bayesian inference are described in Nolan, Menictas & Wand (2018).

Two-level sparse matrices are also known as *block arrowhead* matrices in certain engineering contexts. In Hołubowski *et al.* (2015) it is claimed that such matrices "often appear in areas of applied science and engineering such as head-positioning systems of hard disk drives or kinematic chains of industrial robots". Hołubowski *et al.* (2015) devise an inversion algorithm based on Cholesky decomposition. They are concerned with computation of the entire inverse of a two-level sparse matrix. This inverse matrix is not sparse. In contrast, our results are concerned with determination of inverse matrix sub-blocks matching the non-sparse sections of the original matrix. Hołubowski *et al.* (2015) also provide access, outside of the statistics literature, to other computational contributions for this class of matrix problems.

Motivated by applications described elsewhere (e.g. Nolan, Menictas & Wand, 2018), our main focus in this article is the provision of full, implementable results for both two-level and three-level sparse matrix problems with statistically relevant matrix inverse sub-blocks. Both general situations and least squares form situations, with QR decomposition enhancement, are covered. We believe that it is best to treat each higher-level case separately. If a future application would benefit from the solution to the four-level version of sparse matrix problems treated here then, whilst notationally and algebraically challenging, our two-level and three-level derivations point the way to a solution.

We cover two-level sparse matrix problems in Section 2 and three-level sparse matrix problems in Section 3. Some concluding remarks are made in Section 4.

2 Two-level Sparse Matrix Problems

Two-level sparse matrix problems involve symmetric invertible matrices having generic form:

$$\boldsymbol{A} = \left[\begin{array}{c|c|c|c|c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,2} & \ldots & \boldsymbol{A}_{12,m} \\ \hline \boldsymbol{A}_{12,1}^T & \boldsymbol{A}_{22,1} & \boldsymbol{O} & \ldots & \boldsymbol{O} \\ \hline \boldsymbol{A}_{12,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,2} & \ldots & \boldsymbol{O} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{A}_{12,m}^T & \boldsymbol{O} & \boldsymbol{O} & \ldots & \boldsymbol{A}_{22,m} \end{array}\right]. \tag{1}$$

The dimensions of the sub-blocks of $\boldsymbol{A}$ are:

$$\boldsymbol{A}_{11} \text{ is } p \times p \text{ and, for each } 1 \le i \le m, \boldsymbol{A}_{12,i} \text{ is } p \times q \text{ and } \boldsymbol{A}_{22,i} \text{ is } q \times q.$$

A two-level sparse matrix linear system is a matrix equation of the form

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}, \tag{2}$$

where we partition the vectors $\boldsymbol{a}$ and $\boldsymbol{x}$ as follows:

$$\boldsymbol{a} \equiv \left[\begin{array}{c} \boldsymbol{a}_1 \\ \hline \boldsymbol{a}_{2,1} \\ \hline \boldsymbol{a}_{2,2} \\ \hline \vdots \\ \hline \boldsymbol{a}_{2,m} \end{array}\right] \quad \text{and} \quad \boldsymbol{x} \equiv \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \vdots \\ \hline \boldsymbol{x}_{2,m} \end{array}\right].$$

The dimensions of the sub-vectors of $\boldsymbol{a}$ and $\boldsymbol{x}$ are:

$$\text{both } \boldsymbol{a}_1 \text{ and } \boldsymbol{x}_1 \text{ are } p \times 1 \text{ and, for } 1 \leq i \leq m, \text{ both } \boldsymbol{a}_{2,i} \text{ and } \boldsymbol{x}_{2,i} \text{ are } q \times 1.$$

Suppose that the entries of $\boldsymbol{A}^{-1}$ are labelled according to

$$\boldsymbol{A}^{-1} = \left[\begin{array}{c|c|c|c|c} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,2} & \ldots & \boldsymbol{A}^{12,m} \\ \hline \boldsymbol{A}^{12,1T} & \boldsymbol{A}^{22,1} & \times & \ldots & \times \\ \hline \boldsymbol{A}^{12,2T} & \times & \boldsymbol{A}^{22,2} & \ldots & \times \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{A}^{12,mT} & \times & \times & \ldots & \boldsymbol{A}^{22,m} \end{array}\right],$$

where $\times$ generically denotes sub-blocks of $\boldsymbol{A}^{-1}$ which are in the same positions as the $\boldsymbol{O}$ blocks in $\boldsymbol{A}$. In applications to multilevel data analysis these sub-blocks are usually not of interest since they correspond to between-group covariances. On the other hand, the sub-blocks of $\boldsymbol{A}^{-1}$ which are in the same position as the non-zero sub-blocks of $\boldsymbol{A}$ are required for obtaining standard errors of within-group fits. In the case of mean field variational inference, these sub-blocks are sufficient for both coordinate ascent and message passing optimal parameter computation with minimal product restrictions. Details of how these sub-blocks of $\boldsymbol{A}^{-1}$ are used in linear mixed model inference are given in Nolan, Menictas & Wand (2018).

The two-level sparse matrix problem involves solving (2), which inherently relies on determining the non-$\times$ sub-blocks of $\boldsymbol{A}^{-1}$. Theorem 1 provides a streamlined solution to this problem such that the number of operations is linear in m. An analogous expression for $|\boldsymbol{A}|$, the determinant of $\boldsymbol{A}$, is also provided.

Theorem 1. *For the two-level sparse matrix problem, the solution to the matrix inverse sub-block problem is:*

$$\begin{aligned} \boldsymbol{A}^{11} &= \left(\boldsymbol{A}_{11} - \sum_{i=1}^{m} \boldsymbol{A}_{12,i}\,\boldsymbol{A}_{22,i}^{-1}\,\boldsymbol{A}_{12,i}^{T}\right)^{-1} \\ \text{and } \boldsymbol{A}^{12,i} &= -(\boldsymbol{A}_{22,i}^{-1}\,\boldsymbol{A}_{12,i}^{T}\,\boldsymbol{A}^{11})^{T},\ \ \boldsymbol{A}^{22,i} = \boldsymbol{A}_{22,i}^{-1}(\boldsymbol{I} - \boldsymbol{A}_{12,i}^{T}\,\boldsymbol{A}^{12,i}), \quad 1 \leq i \leq m. \end{aligned}$$

The determinant of $\boldsymbol{A}$ is

$$|\boldsymbol{A}| = \left|\left(\boldsymbol{A}^{11}\right)^{-1}\right| \prod_{i=1}^{m} |\boldsymbol{A}_{22,i}|.$$

The linear system solution is:

$$\boldsymbol{x}_1 = \boldsymbol{A}^{11}\left(\boldsymbol{a}_1 - \sum_{i=1}^{m} \boldsymbol{A}_{12,i}\,\boldsymbol{A}_{22,i}^{-1}\,\boldsymbol{a}_{2,i}\right) \quad \text{and} \quad \boldsymbol{x}_{2,i} = \boldsymbol{A}_{22,i}^{-1}(\boldsymbol{a}_{2,i} - \boldsymbol{A}_{12,i}^{T}\,\boldsymbol{x}_1), \quad 1 \leq i \leq m.$$

2.1 Least Squares Form and QR-decomposition Enhancement

In statistical applications involving linear mixed models, it is common for $\boldsymbol{A}$ to admit *least squares form* that lends itself to a QR decomposition-based solution. QR decompositions of rectangular matrices are a numerically preferred method for solving least squares problems. A QR-decomposition of a rectangular $n \times p$ ($n \geq p$) matrix $\boldsymbol{X}$ involves representing $\boldsymbol{X}$ as

$$\boldsymbol{X} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R} \\ \boldsymbol{O}\end{bmatrix},$$

where $\boldsymbol{Q}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{R}$ is a $p \times p$ upper-triangular matrix. Such a decomposition affords computational stability for least squares problems.

Suppose that $\boldsymbol{x}$ is chosen to minimize the least squares criterion

$$||\boldsymbol{b} - \boldsymbol{B}\boldsymbol{x}||^2 \equiv (\boldsymbol{b} - \boldsymbol{B}\boldsymbol{x})^T(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{x}) \tag{3}$$

for matrices

$$\boldsymbol{B} \equiv \left[\begin{array}{c|c|c|c|c} \boldsymbol{B}_1 & \dot{\boldsymbol{B}}_1 & \boldsymbol{O} & \ldots & \boldsymbol{O} \\ \hline \boldsymbol{B}_2 & \boldsymbol{O} & \dot{\boldsymbol{B}}_2 & \ldots & \boldsymbol{O} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{B}_m & \boldsymbol{O} & \boldsymbol{O} & \ldots & \dot{\boldsymbol{B}}_m \end{array}\right] \quad \text{and} \quad \boldsymbol{b} \equiv \left[\begin{array}{c} \boldsymbol{b}_1 \\ \hline \boldsymbol{b}_2 \\ \hline \vdots \\ \hline \boldsymbol{b}_m \end{array}\right] \tag{4}$$

with sub-matrices and sub-vectors having dimensions:

$$\boldsymbol{B}_i \text{ is } n_i \times p, \quad \dot{\boldsymbol{B}}_i \text{ is } n_i \times q \quad \text{and} \quad \boldsymbol{b}_i \text{ is } n_i \times 1 \quad \text{for } 1 \le i \le m,$$

where $n_i > \max(p, q)$. Then it is easily verified that the $\boldsymbol{x}$ that minimizes (3) is the solution to the two-level sparse linear system (2) with

$$\boldsymbol{A} = \boldsymbol{B}^T\boldsymbol{B} \quad \text{and} \quad \boldsymbol{a} = \boldsymbol{B}^T\boldsymbol{b}. \tag{5}$$

The non-zero sub-blocks of $\boldsymbol{A}$ and the sub-vectors of $\boldsymbol{a}$ are

$$\boldsymbol{A}_{11} = \sum_{i=1}^{m} \boldsymbol{B}_i^T\boldsymbol{B}_i, \quad \boldsymbol{a}_1 = \sum_{i=1}^{m} \boldsymbol{B}_i^T\boldsymbol{b}_i$$

and

$$\boldsymbol{A}_{12,i} = \boldsymbol{B}_i^T\dot{\boldsymbol{B}}_i, \quad \boldsymbol{A}_{22,i} = \dot{\boldsymbol{B}}_i^T\dot{\boldsymbol{B}}_i, \quad \boldsymbol{a}_{2,i} = \dot{\boldsymbol{B}}_i^T\boldsymbol{b}_i, \quad 1 \le i \le m.$$

Theorem 2 extends Theorem 1 by employing a QR decomposition approach for the purpose of numerical stability. Here, and later, we use the following notation for matrices $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_d$ each having the same number of columns:

$$\underset{1 \le i \le d}{\text{stack}}(\boldsymbol{M}_i) \equiv \begin{bmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_d \end{bmatrix}.$$

Theorem 2. *Suppose that $\boldsymbol{A}$ and $\boldsymbol{a}$ admit the forms defined by* (4) *and* (5)*. Then the two-level sparse matrix problem may be solved using the following QR decomposition-based approach:*

1. *For $i = 1, \ldots, m$:*
   (a) *Decompose $\dot{\boldsymbol{B}}_i = \boldsymbol{Q}_i \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix}$ such that $\boldsymbol{Q}_i^{-1} = \boldsymbol{Q}_i^T$ and $\boldsymbol{R}_i$ is upper-triangular.*
   (b) *Then obtain*

$$\begin{aligned} &\boldsymbol{c}_{0i} \equiv \boldsymbol{Q}_i^T\boldsymbol{b}_i, \quad \boldsymbol{c}_{1i} \equiv \textit{first } q \textit{ rows of } \boldsymbol{c}_{0i}, \quad \boldsymbol{c}_{2i} \equiv \textit{remaining rows of } \boldsymbol{c}_{0i}, \\ &\boldsymbol{C}_{0i} \equiv \boldsymbol{Q}_i^T\boldsymbol{B}_i, \quad \boldsymbol{C}_{1i} \equiv \textit{first } q \textit{ rows of } \boldsymbol{C}_{0i} \quad \textit{and} \quad \boldsymbol{C}_{2i} \equiv \textit{remaining rows of } \boldsymbol{C}_{0i}. \end{aligned}$$

2. *Decompose $\underset{1 \le i \le m}{\text{stack}}(\boldsymbol{C}_{2i}) = \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix}$ such that $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ and $\boldsymbol{R}$ is upper-triangular and let*

$$\boldsymbol{c} \equiv \textit{first } p \textit{ rows of } \boldsymbol{Q}^T \left\{ \underset{1 \le i \le m}{\text{stack}}(\boldsymbol{c}_{2i}) \right\}.$$

3. *The solutions are*

$$\boldsymbol{x}_1 = \boldsymbol{R}^{-1}\boldsymbol{c}, \quad \boldsymbol{A}^{11} = \boldsymbol{R}^{-1}\boldsymbol{R}^{-T}$$

and, for $1 \le i \le m$*,*

$$\boldsymbol{x}_{2,i} = \boldsymbol{R}_i^{-1}(\boldsymbol{c}_{1i} - \boldsymbol{C}_{1i}\boldsymbol{x}_1), \quad \boldsymbol{A}^{12,i} = -\boldsymbol{A}^{11}(\boldsymbol{R}_i^{-1}\boldsymbol{C}_{1i})^T, \quad \boldsymbol{A}^{22,i} = \boldsymbol{R}_i^{-1}(\boldsymbol{R}_i^{-T} - \boldsymbol{C}_{1i}\boldsymbol{A}^{12,i}).$$

4. *The determinant of* $\boldsymbol{A}$ *is*

$$|\boldsymbol{A}| = \left\{ \big(\textit{product of the diagonal entries of } \boldsymbol{R}\big) \prod_{i=1}^{m} \big(\textit{product of the diagonal entries of } \boldsymbol{R}_i\big) \right\}^2.$$

Remarks:

1. In Theorem 2, Step 1 involves determination of m upper triangular matrices $\boldsymbol{R}_i$, $1 \le i \le m$, via QR-decomposition which is a standard procedure within most computing environments. Each of the matrix inversions in Step 3 involve $\boldsymbol{R}_i^{-1}$, which can be achieved rapidly via back-solving.

2. Calculations such as $\boldsymbol{Q}_i^T\boldsymbol{b}_i$ do not require storage of $\boldsymbol{Q}_i$ and ordinary matrix multiplication. Standard matrix algebraic programming languages are such that information concerning $\boldsymbol{Q}_i$ is stored in a compact form from which matrices such as $\boldsymbol{Q}_i^T\boldsymbol{b}_i$ can be efficiently obtained.

3. Pinheiro & Bates (2000; Section 2.2) make use of this QR decomposition-based approach for fitting two-level linear mixed models. However, their descriptions are restricted to the $\boldsymbol{x}_1$ and $\boldsymbol{x}_{2,i}$ formulae, not those for the sub-blocks of $\boldsymbol{A}$.

3 Three-level Sparse Matrix Problems

An example of a three-level sparse matrix is:

$$\boldsymbol{A} = \left[\begin{array}{c|c|c|c|c|c|c|c}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,11} & \boldsymbol{A}_{12,12} & \boldsymbol{A}_{12,2} & \boldsymbol{A}_{12,21} & \boldsymbol{A}_{12,22} & \boldsymbol{A}_{12,23} \\ \hline
\boldsymbol{A}_{12,1}^T & \boldsymbol{A}_{22,1} & \boldsymbol{A}_{12,1,1} & \boldsymbol{A}_{12,1,2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{A}_{12,11}^T & \boldsymbol{A}_{12,1,1}^T & \boldsymbol{A}_{22,11} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{A}_{12,12}^T & \boldsymbol{A}_{12,1,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,12} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{A}_{12,2}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,2} & \boldsymbol{A}_{12,2,1} & \boldsymbol{A}_{12,2,2} & \boldsymbol{A}_{12,2,3} \\ \hline
\boldsymbol{A}_{12,21}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,1}^T & \boldsymbol{A}_{22,21} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{A}_{12,22}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,22} & \boldsymbol{O} \\ \hline
\boldsymbol{A}_{12,23}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,3}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,23}
\end{array}\right].$$

In general, a three-level sparse matrix consists of the following components:

- A $p \times p$ matrix $\boldsymbol{A}_{11}$, which is designated the $(1,1)$-block position.
- A set of partitioned matrices $\big\{\big[\ \boldsymbol{A}_{12,i} \ \big|\ \boldsymbol{A}_{12,ij} \ \big|\ \ldots \ \big|\ \boldsymbol{A}_{12,in_i}\ \big]\big\}_{1 \le i \le m}$, which is designated the $(1,2)$-block position. For each $1 \le i \le m$, $\boldsymbol{A}_{12,i}$ is $p \times q_1$, and for each $1 \le j \le n_i$, $\boldsymbol{A}_{12,ij}$ is $p \times q_2$.
- A $(2,1)$-block, which is simply the transpose of the $(1,2)$-block.

- A block diagonal structure along the $(2,2)$-block position, where each sub-block is a two-level sparse matrix, as defined in (1). For each $1 \le i \le m$, $\boldsymbol{A}_{22,i}$ is $q_1 \times q_1$, and for each $1 \le j \le n_i$, $\boldsymbol{A}_{12,\,i,\,j}$ is $q_1 \times q_2$ and $\boldsymbol{A}_{22,ij}$ is $q_2 \times q_2$.

In the example above, $m = 2$, $n_1 = 2$ and $n_2 = 3$ for the matrix $\boldsymbol{A}$. To enhance digestibility we will use these values for m, n_1 and n_2 throughout our discussion regarding the three-level sparse matrix problem. However, this can be easily generalised to any three-level sparse matrix, where m and $\{n_i\}_{1\le i\le m}$ are arbitrary. A three-level sparse matrix linear system involving $\boldsymbol{A}$ is also defined by (2). That is,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a} \tag{6}$$

where we partition the vectors $\boldsymbol{a}$ and $\boldsymbol{x}$ as follows:

$$\boldsymbol{a} \equiv \left[\begin{array}{c} \boldsymbol{a}_1 \\ \hline \boldsymbol{a}_{2,1} \\ \hline \boldsymbol{a}_{2,11} \\ \hline \boldsymbol{a}_{2,12} \\ \hline \boldsymbol{a}_{2,2} \\ \hline \boldsymbol{a}_{2,21} \\ \hline \boldsymbol{a}_{2,22} \\ \hline \boldsymbol{a}_{2,23} \end{array}\right] \quad \text{and} \quad \boldsymbol{x} \equiv \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,11} \\ \hline \boldsymbol{x}_{2,12} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \boldsymbol{x}_{2,21} \\ \hline \boldsymbol{x}_{2,22} \\ \hline \boldsymbol{x}_{2,23} \end{array}\right].$$

The dimensions of the partitioned vectors are:

- $\boldsymbol{a}_1$ and $\boldsymbol{x}_1$ are $p \times 1$ vectors;
- for each $1 \le i \le m$, $\boldsymbol{a}_{2,i}$ and $\boldsymbol{x}_{2,i}$ are $q_1 \times 1$ vectors;
- for each $1 \le i \le m$ and $1 \le j \le n_i$, $\boldsymbol{a}_{2,ij}$ and $\boldsymbol{x}_{2,ij}$ are $q_2 \times 1$ vectors.

Suppose that the entries $\boldsymbol{A}^{-1}$ are labelled according to

$$\boldsymbol{A}^{-1} = \left[\begin{array}{c|c|c|c|c|c|c|c} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,11} & \boldsymbol{A}^{12,12} & \boldsymbol{A}^{12,2} & \boldsymbol{A}^{12,21} & \boldsymbol{A}^{12,22} & \boldsymbol{A}^{12,23} \\ \hline \boldsymbol{A}^{12,1T} & \boldsymbol{A}^{22,1} & \boldsymbol{A}^{12,1,1} & \boldsymbol{A}^{12,1,2} & \times & \times & \times & \times \\ \hline \boldsymbol{A}^{12,11T} & \boldsymbol{A}^{12,1,1T} & \boldsymbol{A}^{22,11} & \times & \times & \times & \times & \times \\ \hline \boldsymbol{A}^{12,12T} & \boldsymbol{A}^{12,1,2T} & \times & \boldsymbol{A}^{22,12} & \times & \times & \times & \times \\ \hline \boldsymbol{A}^{12,2T} & \times & \times & \times & \boldsymbol{A}^{22,2} & \boldsymbol{A}^{12,2,1} & \boldsymbol{A}^{12,2,2} & \boldsymbol{A}^{12,2,3} \\ \hline \boldsymbol{A}^{12,21T} & \times & \times & \times & \boldsymbol{A}^{12,2,1T} & \boldsymbol{A}^{22,21} & \times & \times \\ \hline \boldsymbol{A}^{12,22T} & \times & \times & \times & \boldsymbol{A}^{12,2,2T} & \times & \boldsymbol{A}^{22,22} & \times \\ \hline \boldsymbol{A}^{12,23T} & \times & \times & \times & \boldsymbol{A}^{12,2,3T} & \times & \times & \boldsymbol{A}^{22,23} \end{array}\right].$$

As in Section 2, $\times$ denotes the blocks of $\boldsymbol{A}^{-1}$ that are not of interest. Solving the three-level sparse matrix problem in (6) requires computation of the non-$\times$ components of $\boldsymbol{A}^{-1}$. Theorem 3 presents the solution for any matrix that has the same sparsity structure as $\boldsymbol{A}$ when m and $\{n_i\}_{1\le i\le m}$ are arbitrary. An analogous expression for $|\boldsymbol{A}|$ is also provided.

Theorem 3. *For $1 \le i \le m$, define*

$$\boldsymbol{h}_{2,i} \equiv \boldsymbol{a}_{2,i} - \sum_{j=1}^{n_i} \boldsymbol{A}_{12,i,j}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{a}_{2,ij}, \quad \boldsymbol{H}_{12,i} \equiv \boldsymbol{A}_{12,i} - \sum_{j=1}^{n_i} \boldsymbol{A}_{12,ij}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{A}_{12,i,j}^T$$

$$\text{and} \quad \boldsymbol{H}_{22,i} \equiv \boldsymbol{A}_{22,i} - \sum_{j=1}^{n_i} \boldsymbol{A}_{12,i,j}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{A}_{12,i,j}^T.$$

Then the solution to the matrix inverse sub-block problem is:

$$\begin{aligned}
\boldsymbol{A}^{11} &= \left(\boldsymbol{A}_{11} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{A}_{12,ij}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{A}_{12,ij}^T - \sum_{i=1}^{m} \boldsymbol{H}_{12,i}\, \boldsymbol{H}_{22,i}^{-1}\, \boldsymbol{H}_{12,i}^T \right)^{-1}, \\
\boldsymbol{A}^{12,i} &= -(\boldsymbol{H}_{22,i}^{-1}\, \boldsymbol{H}_{12,i}^T\, \boldsymbol{A}^{11})^T, \quad \boldsymbol{A}^{22,i} = \boldsymbol{H}_{22,i}^{-1}(\boldsymbol{I} - \boldsymbol{H}_{12,i}^T \boldsymbol{A}^{12,i}), \quad 1 \le i \le m, \\
\boldsymbol{A}^{12,ij} &= -\big\{\boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{A}_{12,ij}^T\, \boldsymbol{A}^{11} + \boldsymbol{A}_{12,i,j}^T\, \boldsymbol{A}^{12,i\,T}\big)\big\}^T, \\
\boldsymbol{A}^{12,i,j} &= -\big\{\boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{A}_{12,ij}^T\, \boldsymbol{A}^{12,i} + \boldsymbol{A}_{12,i,j}^T\, \boldsymbol{A}^{22,i}\big)\big\}^T, \\
\boldsymbol{A}^{22,ij} &= \boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{I} - \boldsymbol{A}_{12,ij}^T\, \boldsymbol{A}^{12,ij} - \boldsymbol{A}_{12,i,j}^T\, \boldsymbol{A}^{12,i,j}\big), \quad 1 \le i \le m,\, 1 \le j \le n_i.
\end{aligned}$$

The determinant of $\boldsymbol{A}$ is

$$|\boldsymbol{A}| = \left|(\boldsymbol{A}^{11})^{-1}\right| \prod_{i=1}^{m} \left(\left| \boldsymbol{A}_{22,i} - \sum_{j=1}^{n_i} \boldsymbol{A}_{12,i,j}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{A}_{12,i,j}^T \right| \prod_{j=1}^{n_i} |\boldsymbol{A}_{22,ij}| \right).$$

The linear system solution is:

$$\begin{aligned}
\boldsymbol{x}_1 &= \boldsymbol{A}^{11} \left(\boldsymbol{a}_1 - \sum_{i=1}^{m} \boldsymbol{H}_{12,i}\, \boldsymbol{H}_{22,i}^{-1}\, \boldsymbol{h}_{2,i} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{A}_{12,ij}\, \boldsymbol{A}_{22,ij}^{-1}\, \boldsymbol{a}_{2,ij} \right), \\
\boldsymbol{x}_{2,i} &= \boldsymbol{H}_{22,i}^{-1}\left(\boldsymbol{h}_{2,i} - \boldsymbol{H}_{12,i}^T\, \boldsymbol{x}_1\right), \quad 1 \le i \le m, \\
\boldsymbol{x}_{2,ij} &= \boldsymbol{A}_{22,ij}^{-1}\left(\boldsymbol{a}_{2,ij} - \boldsymbol{A}_{12,ij}^T\, \boldsymbol{x}_1 - \boldsymbol{A}_{12,i,j}^T\, \boldsymbol{x}_{2,i}\right), \quad 1 \le i \le m,\, 1 \le j \le n_i.
\end{aligned}$$

3.1 Least Squares Form and QR-decomposition Enhancement

The three-level sparse matrix problem also lends itself to QR-decomposition enhancement. For the special case of $\boldsymbol{A}$, with $m = 2$, $n_1 = 2$ and $n_2 = 3$, the least squares criterion has the form (3) with

$$\boldsymbol{B} \equiv \left[\begin{array}{c|c|c|c|c|c|c|c}
\boldsymbol{B}_{11} & \dot{\boldsymbol{B}}_{11} & \ddot{\boldsymbol{B}}_{11} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{B}_{12} & \dot{\boldsymbol{B}}_{12} & \boldsymbol{O} & \ddot{\boldsymbol{B}}_{12} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{B}_{21} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dot{\boldsymbol{B}}_{21} & \ddot{\boldsymbol{B}}_{21} & \boldsymbol{O} & \boldsymbol{O} \\ \hline
\boldsymbol{B}_{22} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dot{\boldsymbol{B}}_{22} & \boldsymbol{O} & \ddot{\boldsymbol{B}}_{22} & \boldsymbol{O} \\ \hline
\boldsymbol{B}_{23} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dot{\boldsymbol{B}}_{23} & \boldsymbol{O} & \boldsymbol{O} & \ddot{\boldsymbol{B}}_{23}
\end{array}\right] \quad \text{and} \quad \boldsymbol{b} \equiv \left[\begin{array}{c}
\boldsymbol{b}_{11} \\ \hline \boldsymbol{b}_{12} \\ \hline \boldsymbol{b}_{21} \\ \hline \boldsymbol{b}_{22} \\ \hline \boldsymbol{b}_{23}
\end{array}\right]. \qquad (7)$$

For general values of m and $\{n_i\}_{1 \le i \le m}$, the forms of $\boldsymbol{B}$ and $\boldsymbol{b}$ are

$$\boldsymbol{B} \equiv \left[\underset{1 \le i \le m}{\text{stack}} \left\{ \underset{1 \le j \le n_i}{\text{stack}} (\boldsymbol{B}_{ij}) \right\} \middle| \underset{1 \le i \le m}{\text{blockdiag}} \left\{ \left[\underset{1 \le j \le n_i}{\text{stack}} (\dot{\boldsymbol{B}}_{ij}) \,\middle|\, \underset{1 \le j \le n_i}{\text{blockdiag}} (\ddot{\boldsymbol{B}}_{ij}) \right] \right\} \right]$$

and

$$\boldsymbol{b} \equiv \underset{1\le i\le m}{\text{stack}}\left\{\underset{1\le j\le n_i}{\text{stack}}(\boldsymbol{b}_{ij})\right\}.$$

For each $1 \le i \le m$ and $1 \le j \le n_i$, the dimensions of the sub-blocks of $\boldsymbol{B}$ and $\boldsymbol{b}$ are:

$$\boldsymbol{B}_{ij} \text{ is } o_{ij} \times p, \quad \dot{\boldsymbol{B}}_{ij} \text{ is } o_{ij} \times q_1, \quad \ddot{\boldsymbol{B}}_{ij} \text{ is } o_{ij} \times q_2 \quad \text{and} \quad \boldsymbol{b}_{ij} \text{ is } o_{ij} \times 1$$

where $o_{ij} > \max(p, q_1, q_2)$. Then the $\boldsymbol{x}$ that minimizes (3) is the solution to the three-level sparse linear system (6) with

$$\boldsymbol{A} = \boldsymbol{B}^T\boldsymbol{B} \quad \text{and} \quad \boldsymbol{a} = \boldsymbol{B}^T\boldsymbol{b}. \tag{8}$$

For general $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ (where m and $\{n_i\}_{1\le i\le m}$ are arbitrary), the non-zero components of $\boldsymbol{A}$ and the sub-vectors of $\boldsymbol{a}$ are, for $1 \le i \le m$,

$$\boldsymbol{A}_{11} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{B}_{ij}^T\boldsymbol{B}_{ij}, \quad \boldsymbol{a}_1 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{B}_{ij}^T\boldsymbol{b}_{ij}, \quad \boldsymbol{A}_{22,i} = \sum_{j=1}^{n_i}\dot{\boldsymbol{B}}_{ij}^T\dot{\boldsymbol{B}}_{ij},$$

$$\boldsymbol{A}_{12,i} = \sum_{j=1}^{n_i}\boldsymbol{B}_{ij}^T\dot{\boldsymbol{B}}_{ij}, \quad \boldsymbol{a}_{2,i} = \sum_{j=1}^{n_i}\dot{\boldsymbol{B}}_{ij}^T\boldsymbol{b}_{ij}$$

and

$$\boldsymbol{A}_{22,ij} = \ddot{\boldsymbol{B}}_{ij}^T\ddot{\boldsymbol{B}}_{ij}, \quad \boldsymbol{A}_{12,\,i,\,j} = \dot{\boldsymbol{B}}_{ij}^T\ddot{\boldsymbol{B}}_{ij}, \quad \boldsymbol{A}_{12,ij} = \boldsymbol{B}_{ij}^T\ddot{\boldsymbol{B}}_{ij}, \quad \boldsymbol{a}_{2,ij} = \ddot{\boldsymbol{B}}_{ij}^T\boldsymbol{b}_{ij}, \quad 1 \le i \le m,\ 1 \le j \le n_i.$$

Theorem 4 provides a QR decomposition enhancement of Theorem 3 for the least squares forms situation.

Theorem 4. *Suppose that* $\boldsymbol{A}$ *and* $\boldsymbol{a}$ *admit the least squares forms defined by* (7) *and* (8)*, where each* m *and* $\{n_i\}_{1\le i\le m}$ *is chosen arbitrarily. Then the three-level sparse matrix problem may be solved using the following QR decomposition-based approach:*

1. *For* $i = 1, \ldots, m$*:*

   (a) *For* $j = 1, \ldots, n_i$*:*

      i. *Decompose* $\ddot{\boldsymbol{B}}_{ij} = \boldsymbol{Q}_{ij}\begin{bmatrix}\boldsymbol{R}_{ij}\\ \boldsymbol{0}\end{bmatrix}$ *such that* $\boldsymbol{Q}_{ij}^{-1} = \boldsymbol{Q}_{ij}^T$ *and* $\boldsymbol{R}_{ij}$ *is upper-triangular.*

      ii. *Then obtain*

$$\begin{aligned}
&\boldsymbol{d}_{0ij} \equiv \boldsymbol{Q}_{ij}^T\boldsymbol{b}_{ij}, \quad \boldsymbol{d}_{1ij} \equiv \textit{first } q_2 \textit{ rows of } \boldsymbol{d}_{0ij}, \quad \boldsymbol{d}_{2ij} \equiv \textit{remaining rows of } \boldsymbol{d}_{0ij},\\
&\boldsymbol{D}_{0ij} \equiv \boldsymbol{Q}_{ij}^T\boldsymbol{B}_{ij}, \quad \boldsymbol{D}_{1ij} \equiv \textit{first } q_2 \textit{ rows of } \boldsymbol{D}_{0ij}, \quad \boldsymbol{D}_{2ij} \equiv \textit{remaining rows of } \boldsymbol{D}_{0ij},\\
&\dot{\boldsymbol{D}}_{0ij} \equiv \boldsymbol{Q}_{ij}^T\dot{\boldsymbol{B}}_{ij}, \quad \dot{\boldsymbol{D}}_{1ij} \equiv \textit{first } q_2 \textit{ rows of } \dot{\boldsymbol{D}}_{0ij}, \quad \dot{\boldsymbol{D}}_{2ij} \equiv \textit{remaining rows of } \dot{\boldsymbol{D}}_{0ij}.
\end{aligned}$$

   (b) i. *Decompose* $\underset{1\le j\le n_i}{\text{stack}}(\dot{\boldsymbol{D}}_{2ij}) = \boldsymbol{Q}_i\begin{bmatrix}\boldsymbol{R}_i\\ \boldsymbol{0}\end{bmatrix}$ *such that* $\boldsymbol{Q}_i^{-1} = \boldsymbol{Q}_i^T$ *and* $\boldsymbol{R}_i$ *is upper-triangular.*

      ii. *Then obtain*

$$\begin{aligned}
&\boldsymbol{c}_{0i} \equiv \boldsymbol{Q}_i^T\left\{\underset{1\le j\le n_i}{\text{stack}}(\boldsymbol{d}_{2ij})\right\}, \quad \boldsymbol{c}_{1i} \equiv \textit{first } q_1 \textit{ rows of } \boldsymbol{c}_{0i}, \quad \boldsymbol{c}_{2i} \equiv \textit{remaining rows of } \boldsymbol{c}_{0i}\\
&\boldsymbol{C}_{0i} \equiv \boldsymbol{Q}_i^T\left\{\underset{1\le j\le n_i}{\text{stack}}(\boldsymbol{D}_{2ij})\right\}, \quad \boldsymbol{C}_{1i} \equiv \textit{first } q_1 \textit{ rows of } \boldsymbol{C}_{0i}, \quad \boldsymbol{C}_{2i} \equiv \textit{remaining rows of } \boldsymbol{C}_{0i}.
\end{aligned}$$

2. *Decompose* $\underset{1\le i\le m}{\text{stack}}(\boldsymbol{C}_{2i}) = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}$ *such that* $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ *and* $\boldsymbol{R}$ *is upper-triangular and let*

$$\boldsymbol{c} \equiv \textit{first } p \textit{ rows of } \boldsymbol{Q}^T\left\{\underset{1\le i\le m}{\text{stack}}(\boldsymbol{c}_{2i})\right\}.$$

3. *The solutions are, for* $1 \le i \le m$,

$$\boldsymbol{x}_1 = \boldsymbol{R}^{-1}\boldsymbol{c}, \quad \boldsymbol{A}^{11} = \boldsymbol{R}^{-1}\boldsymbol{R}^{-T},$$

$$\boldsymbol{x}_{2,i} = \boldsymbol{R}_i^{-1}(\boldsymbol{c}_{1i} - \boldsymbol{C}_{1i}\,\boldsymbol{x}_1), \quad \boldsymbol{A}^{12,i} = -\boldsymbol{A}^{11}\left(\boldsymbol{R}_i^{-1}\boldsymbol{C}_{1i}\right)^T, \quad \boldsymbol{A}^{22,i} = \boldsymbol{R}_i^{-1}(\boldsymbol{R}_i^{-T} - \boldsymbol{C}_{1i}\,\boldsymbol{A}^{12,i}),$$

and, for $1 \le i \le m$, $1 \le j \le n_i$,

$$\boldsymbol{x}_{2,ij} = \boldsymbol{R}_{ij}^{-1}(\boldsymbol{d}_{1ij} - \boldsymbol{D}_{1ij}\,\boldsymbol{x}_1 - \dot{\boldsymbol{D}}_{1ij}\,\boldsymbol{x}_{2,i}), \quad \boldsymbol{A}^{12,ij} = -\left\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij}\,\boldsymbol{A}^{11} + \dot{\boldsymbol{D}}_{1ij}\,\boldsymbol{A}^{12,iT})\right\}^T,$$

$$\boldsymbol{A}^{12,\,i,\,j} = -\left\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij}\,\boldsymbol{A}^{12,i} + \dot{\boldsymbol{D}}_{1ij}\,\boldsymbol{A}^{22,i})\right\}^T \textit{ and } \boldsymbol{A}^{22,ij} = \boldsymbol{R}_{ij}^{-1}(\boldsymbol{R}_{ij}^{-T} - \boldsymbol{D}_{1ij}\,\boldsymbol{A}^{12,ij} - \dot{\boldsymbol{D}}_{1ij}\,\boldsymbol{A}^{12,\,i,\,j}).$$

4. *The determinant of* $\boldsymbol{A}$ *is*

$$|\boldsymbol{A}| = \left[\big(\textit{product of the diagonal entries of } \boldsymbol{R}\big)\prod_{i=1}^{m}\left\{\big(\textit{product of the diagonal entries of } \boldsymbol{R}_i\big) \times \prod_{j=1}^{n_i}\big(\textit{product of the diagonal entries of } \boldsymbol{R}_{ij}\big)\right\}\right]^2.$$

Remarks:

1. As in Theorem 2, Step 1(a) of Theorem 4 involves determination of $\sum_{i=1}^m n_i$ upper-triangular matrices $\boldsymbol{R}_{ij}$, for $1 \le i \le m$, $1 \le j \le n_i$, via QR-decomposition. In Step 1(b), m upper-triangular matrices $\boldsymbol{R}_i$, for $1 \le i \le m$, are also constructed. A final QR-decomposition is applied in Step 2. Each of the inversions in Step 3 can be solved rapidly via back-solving.

2. The solutions for $\boldsymbol{x}_1$, $\boldsymbol{A}^{11}$ and $\boldsymbol{x}_{2,i}, \boldsymbol{A}^{12,i}, \boldsymbol{A}^{22,i}, 1 \le i \le m$, have the same forms as in Theorem 4 for two-level sparse matrices. The solutions for $\boldsymbol{x}_{2,ij}$, $\boldsymbol{A}^{12,ij}$, $\boldsymbol{A}^{12,\,i,\,j}$ and $\boldsymbol{A}^{22,ij}$ are suggestive of a hierarchical pattern emerging for four-level and higher-level classes of the problem.

4 Conclusion

In this short communication we have conveyed the essence of higher-level sparse matrix problems as viewed through the prism of fitting and inference for multilevel statistical models. Both time-honored best linear unbiased prediction and new-fashioned mean field variational Bayes approaches benefit from our four theorems for the two-level and three-level situations, with details given in Nolan, Menictas & Wand (2018). Future extension to higher-level situations is aided by our results and derivations.

Acknowledgements

We are grateful to Gioia di Credico, Luca Maestrini, Marianne Menictas and Simon Wood for their assistance. This research was partially supported by the Australian Research Council Discovery Project DP140100441.

References

Baltagi, BH (2013), *Econometric Analysis of Panel Data*, Chichester, U.K.: John Wiley & Sons.

Fitzmaurice, G, Davidian, M, Verbeke, G and Molenberghs, G (eds.) (2008), *Longitudinal Data Analysis*, Boca Raton, Florida: Chapman & Hall/CRC.

Gentle, JE (2007), *Matrix Algebra*, New York: Springer.

Goldstein, H (2010), *Multilevel Statistical Models*, 4th edn. Chichester, U.K.: John Wiley & Sons.

Henderson, CR (1975), 'Best linear unbiased estimation and prediction under a selection model', *Biometrics*, **31**, 423–447.

Hołubowski W, Kurzyk D & Trawiński T (2015), 'A fast method for computing the inverse of symmetric block arrowhead matrices', *Applied Mathematics and Information Sciences*, **9**, 319–324.

Lee, CYY & Wand, MP (2016), 'Streamlined mean field variational Bayes for longitudinal and multilevel data analysis', *Biometrical Journal*, **58**, 868–895.

Longford, NT (1987), 'A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects', *Biometrika*, **74**, 817–827.

Nolan, TH, Menictas, M & Wand, MP (2018), 'Streamlined computing for variational inference with higher-level random effects'. In progress.

McCulloch, CE, Searle, SR & Neuhaus JM (2008), *Generalized, Linear, and Mixed Models*, 2nd edn. New York: John Wiley & Sons.

Pinheiro, JC & Bates, DM (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York.

Rao, JNK & Molina, I (2015), *Small Area Estimation*, 2nd edn. Hoboken, New Jersey: John Wiley & Sons.

Wand, MP (2017), 'Fast approximate inference for arbitrarily large semiparametric regression models via message passing', *Journal of the American Statistical Association*, **112**, 137–168.