# STREAMLINED SOLUTIONS TO MULTILEVEL SPARSE MATRIX PROBLEMS

## TUI H. NOLAN[1] and MATT P. WAND[1]

## Abstract

We define and solve classes of sparse matrix problems that arise in multilevel modelling 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. While 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.

## 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 [4]. Other areas of statistics and econometrics that use essentially the same types of models are longitudinal data analysis [2], panel data analysis [1] and small-area estimation [12]. Linear mixed models [9] are the main vehicles for modelling, 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

---

[1]University of Technology Sydney, P.O. Box 123, Broadway, New South Wales 2007, Australia;
e-mail: tui.h.nolan@student.uts.edu.au, matt.wand@uts.edu.au.

matrices (for example, [6]). 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 1000 groups and 100 observations per group involves a random effects design matrix containing 100 million entries of which 99.9% are zeros. Streamlined computation of the best linear unbiased predictors of the fixed and random effects are well documented, with Pinheiro and Bates [11, Section 2.2] 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 [3], 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 by, for example, Langford [8]. 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 paper is purely concerned with generic matrix algebraic facts and, while motivated by statistical analysis, is totally free of statistical concepts in its main results and derivations. Ramifications for statistical inference are described by Nolan et al. [10].

There is a related, but essentially nonoverlapping, literature concerning inversion of so-called "arrowhead" matrices, which are invertible matrices that have all entries equal to zero except for those on the main diagonal and in one row and one column. Houbowski et al. [7] explain 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" and then develop a fast method for inversion of a generalization of arrowhead matrices known as "block arrowhead" matrices. Other recent contributions of this type are by Nejafi et al. [13] and Stanimirović et al. [14]. It is important to note that while block arrowhead matrices coincide with two-level sparse matrices, the class of problems treated here and their motivating statistical applications are different from the central goal of the arrowhead matrix inversion literature. In multilevel sparse matrix problems, the full matrix inverse is *not* of interest but, instead, inverse matrix sub-blocks matching the nonsparse sections of the original matrix. Arrowhead matrix inversion methods provide much more than is required and, therefore, are overly slow for large sparse two-level matrix problems. In addition, multilevel sparse matrices beyond the two-level case do not have block arrowhead forms. In summary, multilevel sparse matrix problems are, in essence, distinct from arrowhead matrix inversion problems.

Motivated by applications described elsewhere (for example, [10]), our main focus in this paper is on 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, while 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. Appendix A contains proofs of all theorems.

## 2. Two-level sparse matrix problems

We begin by defining a two-level sparse matrix problem.

DEFINITION 2.1. Let $A$ be a symmetric and invertible matrix of the form

$$A = \begin{bmatrix} A_{11} & A_{12,1} & A_{12,2} & \dots & A_{12,m} \\ A_{12,1}^T & A_{22,1} & O & \dots & O \\ A_{12,2}^T & O & A_{22,2} & \dots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{12,m}^T & O & O & \dots & A_{22,m} \end{bmatrix}, \tag{2.1}$$

where the dimensions of the sub-blocks of $A$ are as follows:

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

The two-level sparse matrix problem is defined as follows.

(I) Solve the linear system

$$Ax = a.$$

(II) Obtain the sub-blocks of $A^{-1}$ corresponding to the positions of the sub-blocks $A_{11}$, $A_{12,i}$ and $A_{22,i}$, $1 \le i \le m$, in $A$.

If $p$ and $q$ are small relative to $m$ then matrices defined by $A$ are sparse, since, as $m \to \infty$, the leading term of the fraction of nonzero entries of $A$ is $\{1 + 2(p/q)\}m^{-1}$. In motivating statistical applications, $m$ corresponds to sample size, which often is large, while $p$ and $q$ correspond to dimensions of model parameter spaces, which are moderate in size. In the area of longitudinal data analysis, typical values are $p = q = 2$. The dimension variable $m$ matches the number of subjects in the longitudinal study,

which may be in the hundreds or even thousands. If $m = 1000$, then for $p = q = 2$, only about 0.3% of the entries of $A$ are nonzero. Throughout this paper, we assume that $p$ and $q$ are small relative to $m$. Similar remarks apply to the three-level case treated in Section 3.

Our solution to the two-level sparse matrix problem benefits from the following notation for the sub-matrices of $A^{-1}$, $a$ and $x$:

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12,1} & A^{12,2} & \dots & A^{12,m} \\ A^{12,1T} & A^{22,1} & \times & \dots & \times \\ A^{12,2T} & \times & A^{22,2} & \dots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A^{12,mT} & \times & \times & \dots & A^{22,m} \end{bmatrix}, \quad a \equiv \begin{bmatrix} a_1 \\ a_{2,1} \\ a_{2,2} \\ \vdots \\ a_{2,m} \end{bmatrix} \quad \text{and} \quad x \equiv \begin{bmatrix} x_1 \\ x_{2,1} \\ x_{2,2} \\ \vdots \\ x_{2,m} \end{bmatrix}, \quad (2.2)$$

where $\times$ generically denotes sub-blocks of $A^{-1}$ which are in the same positions as the $O$ blocks in $A$. The dimensions of the sub-vectors of $a$ and $x$ are as follows:

both $a_1$ and $x_1$ are $p \times 1$ and, for $1 \le i \le m$, both $a_{2,i}$ and $x_{2,i}$ are $q \times 1$.

Armed with the notation in (2.2), part (II) of Definition 2.1 can be expressed as follows.

Obtain the matrix $A^{11}$ and, for each $1 \le i \le m$, obtain the matrices $A^{12,i}$ and $A^{22,i}$.

In applications involving multilevel data analysis, the sub-blocks corresponding to the $\times$ symbols are usually not of interest, since they correspond to between-group covariances. On the other hand, the sub-blocks of $A^{-1}$ which are in the same position as the nonzero sub-blocks of $A$ 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 $A^{-1}$ are used in linear mixed model inference are given by Nolan et al. [10].

Theorem 2.2 provides a streamlined solution to this problem, such that the number of operations is linear in $m$. An analogous expression for $|A|$, the determinant of $A$, is also provided.

THEOREM 2.2. *Consider the two-level sparse matrix problem given by Definition 2.1 and suppose that all $A_{22,i}$, $1 \le i \le m$, are invertible. The solution to part (II) of Definition 2.1 is*

$$A^{11} = \left( A_{11} - \sum_{i=1}^{m} A_{12,i} A_{22,i}^{-1} A_{12,i}^{T} \right)^{-1},$$

$$A^{12,i} = -(A_{22,i}^{-1} A_{12,i}^{T} A^{11})^{T}, \quad and \quad A^{22,i} = A_{22,i}^{-1}(I - A_{12,i}^{T} A^{12,i}), \quad 1 \le i \le m.$$

*The determinant of $A$ is*

$$|A| = |(A^{11})^{-1}| \prod_{i=1}^{m} |A_{22,i}|.$$

*The solution to part* (I) *of Definition* 2.1 *is*

$$x_1 = A^{11}\left(a_1 - \sum_{i=1}^{m} A_{12,i} A_{22,i}^{-1} a_{2,i}\right) \quad and \quad x_{2,i} = A_{22,i}^{-1}(a_{2,i} - A_{12,i}^{T} x_1), \quad 1 \leq i \leq m.$$

A proof of Theorem 2.2 is given in Appendix A.1.

**2.1. Least squares form and QR-decomposition enhancement** In statistical applications involving linear mixed models, it is common for $A$ to admit a "least squares form" that lends itself to a solution based on QR decomposition. 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 $X$ involves representing $X$ as

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

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

Suppose that $x$ is chosen to minimize the least squares criterion

$$\|b - Bx\|^2 \equiv (b - Bx)^T (b - Bx) \tag{2.3}$$

for matrices

$$B \equiv \left[ \begin{array}{c|c|c|c|c} B_1 & \dot{B}_1 & O & \dots & O \\ \hline B_2 & O & \dot{B}_2 & \dots & O \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline B_m & O & O & \dots & \dot{B}_m \end{array} \right] \quad and \quad b \equiv \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \tag{2.4}$$

with sub-matrices and sub-vectors having the following dimensions:

$$B_i \text{ is } n_i \times p, \quad \dot{B}_i \text{ is } n_i \times q \quad and \quad b_i \text{ is } n_i \times 1, \quad for \ 1 \leq i \leq m,$$

such that $B$ is full rank. Then it is easily verified that the $x$ that minimizes (2.3) is the solution to

$$Ax = a, \quad \text{where } A = B^T B \text{ and } a = B^T b. \tag{2.5}$$

Moreover, $A$ is a two-level sparse matrix of the form given by (2.1). The nonzero sub-blocks of $A$ and the sub-vectors of $a$ are

$$A_{11} = \sum_{i=1}^{m} B_i^T B_i, \quad a_1 = \sum_{i=1}^{m} B_i^T b_i,$$

and

$$A_{12,i} = B_i^T \dot{B}_i, \quad A_{22,i} = \dot{B}_i^T \dot{B}_i, \quad a_{2,i} = \dot{B}_i^T b_i, \quad 1 \le i \le m.$$

The form of the matrix $B$ in (2.4) arises in statistical models containing both fixed effects and random effects with two-level structure (for example, [4]). Full details on this motivational connection are given by Nolan et al. [10].

Theorem 2.3 extends Theorem 2.2 by employing a QR-decomposition approach for the purpose of numerical stability. Here, and later, we use the following notation for matrices $M_1, \ldots, M_d$ each having the same number of columns:

$$\operatorname*{stack}_{1 \le i \le d}(M_i) \equiv \begin{bmatrix} M_1 \\ \vdots \\ M_d \end{bmatrix}.$$

THEOREM 2.3. *Suppose that $A$ and $a$ admit the forms defined by (2.4) and (2.5), where $B$ is full rank. 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 $\dot{B}_i = Q_i \begin{bmatrix} R_i \\ 0 \end{bmatrix}$ such that $Q_i^{-1} = Q_i^T$ and $R_i$ is upper-triangular,*
  (b) *then obtain*

$$c_{0i} \equiv Q_i^T b_i, \quad c_{1i} \equiv \text{first } q \text{ rows of } c_{0i}, \quad c_{2i} \equiv \text{remaining rows of } c_{0i},$$
$$C_{0i} \equiv Q_i^T B_i, \quad C_{1i} \equiv \text{first } q \text{ rows of } C_{0i} \quad \text{and} \quad C_{2i} \equiv \text{remaining rows of } C_{0i}.$$

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

$$c \equiv \text{first } p \text{ rows of } Q^T \{ \operatorname*{stack}_{1 \le i \le m}(c_{2i}) \}.$$

(3) *The solutions are*

$$x_1 = R^{-1}c, \quad A^{11} = R^{-1}R^{-T},$$

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

$$x_{2,i} = R_i^{-1}(c_{1i} - C_{1i}x_1), \quad A^{12,i} = -A^{11}(R_i^{-1}C_{1i})^T, \quad A^{22,i} = R_i^{-1}(R_i^{-T} - C_{1i}A^{12,i}).$$

(4) *The determinant of $A$ is*

$$|A| = \Big\{ (\text{product of the diagonal entries of } R) \\ \times \prod_{i=1}^{m} (\text{product of the diagonal entries of } R_i) \Big\}^2.$$

A proof of Theorem 2.3 is given in Appendix A.2.

TABLE 1. Averages (standard deviations) of elapsed computing times in seconds for solving the two-level sparse matrix problem naïvely and with the streamlined approach provided by Theorem 2.3. The fourth column lists the ratios of the median computing times.

| $m$ | Naïve | Streamlined | Naïve/streamlined |
|---|---|---|---|
| 100 | 0.103 (0.0173) | 0.00253 (0.000541) | 40.8 |
| 200 | 0.768 (0.0214) | 0.00429 (0.000574) | 179.0 |
| 400 | 6.050 (0.1110) | 0.00798 (0.000586) | 758.0 |
| 800 | 48.300 (0.7830) | 0.01530 (0.001650) | 3150.0 |
| 1600 | 386.000 (6.0200) | 0.03000 (0.001900) | 12900.0 |

REMARK 2.4.

(1) In Theorem 2.3, Step 1 involves determination of $m$ upper-triangular matrices $R_i$, $1 \le i \le m$, via QR decomposition which is a standard procedure within most computing environments. Each of the matrix inversions in Step 3 involve $R_i^{-1}$, which can be achieved rapidly via back-solving.
(2) Calculations such as $Q_i^T b_i$ do not require storage of $Q_i$ and ordinary matrix multiplication. Standard matrix algebraic programming languages are such that information concerning $Q_i$ is stored in a compact form, from which matrices such as $Q_i^T b_i$ can be efficiently obtained.
(3) Pinheiro and Bates [11, 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 $x_1$ and $x_{2,i}$ formulae, and not those for the sub-blocks of $A^{-1}$.

Table 1 summarizes the results of a numerical study, for which 100 $A$ matrices of the form (2.1) were randomly generated and the solution to the streamlined two-level sparse matrix problem using Theorem 2.3 was compared to the naïve solution, where the sparse structure was ignored, for increasingly large versions of the problem. Throughout the study, $p = q = 2$ and $m$ ranged over the set $\{100, 200, 400, 800, 1600\}$. The $n_i$ values were generated uniformly on the set $\{30, \dots, 60\}$. To aid maximal speed, both approaches were implemented in the low-level language `Fortran 77`, with `LINPACK` subroutines used for numerical linear algebra. The naïve approach involved matrix inversion via Gaussian elimination. The study was run on a MacBook Air laptop computer with a 2.2 GHz processor and 8 GB of random access memory. The table lists the average and standard deviation times in seconds across the 100 replications.

From Table 1 we see that the streamlined solutions are delivered in a small fraction of a second, even for very large versions of the problem. By contrast, naïve computation takes several minutes for the $m = 1600$ case, and is more than ten thousand times slower.

### 3. Three-level sparse matrix problems

Three-level sparse matrix problems are such that the $A$ matrix is symmetric and invertible, but the two-level structure is repeated down the main diagonal. The notation for the general case becomes difficult to digest, so we start with a concrete example of such an $A$:

$$A = \begin{bmatrix} A_{11} & A_{12,1} & A_{12,11} & A_{12,12} & A_{12,2} & A_{12,21} & A_{12,22} & A_{12,23} \\ A_{12,1}^T & A_{22,1} & A_{12,1,1} & A_{12,1,2} & O & O & O & O \\ A_{12,11}^T & A_{12,1,1}^T & A_{22,11} & O & O & O & O & O \\ A_{12,12}^T & A_{12,1,2}^T & O & A_{22,12} & O & O & O & O \\ A_{12,2}^T & O & O & O & A_{22,2} & A_{12,2,1} & A_{12,2,2} & A_{12,2,3} \\ A_{12,21}^T & O & O & O & A_{12,2,1}^T & A_{22,21} & O & O \\ A_{12,22}^T & O & O & O & A_{12,2,2}^T & O & A_{22,22} & O \\ A_{12,23}^T & O & O & O & A_{12,2,3}^T & O & O & A_{22,23} \end{bmatrix}. \quad (3.1)$$

The general three-level sparse matrix problem is defined as follows.

DEFINITION 3.1. Let $A$ be a symmetric and invertible matrix with partitioning

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix},$$

where $A_{11}$ is $p \times p$ and

$$A_{12} = \{[A_{12,i} \mid A_{12,ij} \mid \ldots \mid A_{12,in_i}]\}_{1 \le i \le m},$$

such that for each $1 \le i \le m$, $A_{12,i}$ is $p \times q_1$, and for each $1 \le j \le n_i$, $A_{12,ij}$ is $p \times q_2$. The lower right block is

$$A_{22} = \underset{1 \le i \le m}{\text{blockdiag}} \left( \begin{bmatrix} A_{22,i} & A_{12,i,1} & A_{12,i,2} & \ldots & A_{12,i,n_i} \\ A_{12,i,1}^T & A_{22,i1} & O & \ldots & O \\ A_{12,i,2}^T & O & A_{22,i2} & \ldots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{12,i,n_i}^T & O & O & \ldots & A_{22,in_i} \end{bmatrix} \right)$$

where, for each $1 \le i \le m$, $A_{22,i}$ is $q_1 \times q_1$, and for each $1 \le j \le n_i$, $A_{12,i,j}$ is $q_1 \times q_2$ and $A_{22,ij}$ is $q_2 \times q_2$. The three-level sparse matrix problem is defined as follows.

(I) Solve the linear system

$$Ax = a.$$

(II) Obtain the sub-blocks of $A^{-1}$ corresponding to the positions of the sub-blocks $A_{11}$, $A_{12,i}$, $A_{22,i}$, $1 \leq i \leq m$, and $A_{12,ij}$, $A_{12,i,j}$, $A_{22,ij}$, $1 \leq i \leq m$, $1 \leq j \leq n_i$, in $A$.

For the example three-level sparse matrix given by (3.1), we have $m = 2$, $n_1 = 2$ and $n_2 = 3$. 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 generalized to any three-level sparse matrix, where $m$ and $\{n_i\}_{1 \leq i \leq m}$ are arbitrary.

The solution to the thee-level sparse matrix problem benefits from the following notation for the sub-matrices of $A^{-1}$, $a$ and $x$:

$$A^{-1} = \begin{bmatrix}
A^{11} & A^{12,1} & A^{12,11} & A^{12,12} & A^{12,2} & A^{12,21} & A^{12,22} & A^{12,23} \\
A^{12,1T} & A^{22,1} & A^{12,1,1} & A^{12,1,2} & \times & \times & \times & \times \\
A^{12,11T} & A^{12,1,1T} & A^{22,11} & \times & \times & \times & \times & \times \\
A^{12,12T} & A^{12,1,2T} & \times & A^{22,12} & \times & \times & \times & \times \\
A^{12,2T} & \times & \times & \times & A^{22,2} & A^{12,2,1} & A^{12,2,2} & A^{12,2,3} \\
A^{12,21T} & \times & \times & \times & A^{12,2,1T} & A^{22,21} & \times & \times \\
A^{12,22T} & \times & \times & \times & A^{12,2,2T} & \times & A^{22,22} & \times \\
A^{12,23T} & \times & \times & \times & A^{12,2,3T} & \times & \times & A^{22,23}
\end{bmatrix}, \quad (3.2)$$

$$a \equiv \begin{bmatrix} a_1 \\ a_{2,1} \\ a_{2,11} \\ a_{2,12} \\ a_{2,2} \\ a_{2,21} \\ a_{2,22} \\ a_{2,23} \end{bmatrix} \quad \text{and} \quad x \equiv \begin{bmatrix} x_1 \\ x_{2,1} \\ x_{2,11} \\ x_{2,12} \\ x_{2,2} \\ x_{2,21} \\ x_{2,22} \\ x_{2,23} \end{bmatrix}.$$

The dimensions of the partitioned vectors are as follows:

- $a_1$ and $x_1$ are $p \times 1$ vectors;
- for each $1 \leq i \leq m$, $a_{2,i}$ and $x_{2,i}$ 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.

As in Section 2, $\times$ denotes the blocks of $\boldsymbol{A}^{-1}$ that are not of interest. Using the notation exemplified by (3.2), part (II) of Definition 3.1 can be expressed as follows.

  Obtain the matrix $\boldsymbol{A}^{11}$, for each $1 \le i \le m$, obtain the matrices $\boldsymbol{A}^{12,i}$, $\boldsymbol{A}^{22,i}$ and
   for each $1 \le i \le m$, $1 \le j \le n_i$, obtain the matrices $\boldsymbol{A}^{12,ij}$, $\boldsymbol{A}^{12,i,j}$, $\boldsymbol{A}^{22,ij}$.

  Theorem 3.2 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 3.2. *Consider the three-level sparse matrix problem given by Definition 3.1, and suppose that all $\boldsymbol{A}_{22,ij}$, $1 \le i \le m$, $1 \le j \le n_i$, are invertible. For $1 \le i \le m$, define*

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

*and*

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

*and suppose that all $\boldsymbol{H}_{22,i}$, $1 \le i \le m$, are invertible. The solution to part* (II) *of Definition 3.1 is:*

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

$$\boldsymbol{A}^{12,i} = -(\boldsymbol{H}_{22,i}^{-1} \, \boldsymbol{H}_{12,i}^T \, \boldsymbol{A}^{11})^T, \quad \boldsymbol{A}^{22,i} = \boldsymbol{H}_{22,i}^{-1}(\boldsymbol{I} - \boldsymbol{H}_{12,i}^T \boldsymbol{A}^{12,i}), \quad 1 \le i \le m,$$

$$\boldsymbol{A}^{12,ij} = -\{\boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{A}_{12,ij}^T \, \boldsymbol{A}^{11} + \boldsymbol{A}_{12,i,j}^T \, \boldsymbol{A}^{12,i\,T})\}^T,$$

$$\boldsymbol{A}^{12,i,j} = -\{\boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{A}_{12,ij}^T \, \boldsymbol{A}^{12,i} + \boldsymbol{A}_{12,i,j}^T \, \boldsymbol{A}^{22,i})\}^T,$$

$$\boldsymbol{A}^{22,ij} = \boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{I} - \boldsymbol{A}_{12,ij}^T \, \boldsymbol{A}^{12,ij} - \boldsymbol{A}_{12,i,j}^T \, \boldsymbol{A}^{12,i,j}), \quad 1 \le i \le m, \, 1 \le j \le n_i.$$

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

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

*The solution to part* (I) *of Definition 3.1 is:*

$$\boldsymbol{x}_1 = \boldsymbol{A}^{11}\left( \boldsymbol{a}_1 - \sum_{i=1}^m \boldsymbol{H}_{12,i} \, \boldsymbol{H}_{22,i}^{-1} \, \boldsymbol{h}_{2,i} - \sum_{i=1}^m \sum_{j=1}^{n_i} \boldsymbol{A}_{12,ij} \, \boldsymbol{A}_{22,ij}^{-1} \, \boldsymbol{a}_{2,ij} \right),$$

$$\boldsymbol{x}_{2,i} = \boldsymbol{H}_{22,i}^{-1}(\boldsymbol{h}_{2,i} - \boldsymbol{H}_{12,i}^T \boldsymbol{x}_1), \quad 1 \le i \le m,$$

$$\boldsymbol{x}_{2,ij} = \boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{a}_{2,ij} - \boldsymbol{A}_{12,ij}^T \boldsymbol{x}_1 - \boldsymbol{A}_{12,i,j}^T \boldsymbol{x}_{2,i}), \quad 1 \le i \le m, \, 1 \le j \le n_i.$$

A proof of Theorem 3.2 is given in Appendix A.3.

### 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 $A$, with $m = 2$, $n_1 = 2$ and $n_2 = 3$, the least squares criterion has the form (2.3) with

$$
B \equiv \begin{bmatrix}
B_{11} & \dot{B}_{11} & \ddot{B}_{11} & O & O & O & O & O \\
B_{12} & \dot{B}_{12} & O & \ddot{B}_{12} & O & O & O & O \\
B_{21} & O & O & O & \dot{B}_{21} & \ddot{B}_{21} & O & O \\
B_{22} & O & O & O & \dot{B}_{22} & O & \ddot{B}_{22} & O \\
B_{23} & O & O & O & \dot{B}_{23} & O & O & \ddot{B}_{23}
\end{bmatrix}
\quad \text{and} \quad
b \equiv \begin{bmatrix}
b_{11} \\
b_{12} \\
b_{21} \\
b_{22} \\
b_{23}
\end{bmatrix}.
$$

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

$$
B \equiv [\underset{1 \le i \le m}{\text{stack}}\{\underset{1 \le j \le n_i}{\text{stack}}(B_{ij})\} \mid \underset{1 \le i \le m}{\text{blockdiag}}\{[\underset{1 \le j \le n_i}{\text{stack}}(\dot{B}_{ij}) \mid \underset{1 \le j \le n_i}{\text{blockdiag}}(\ddot{B}_{ij})]\}] \tag{3.3}
$$

and

$$
b \equiv \underset{1 \le i \le m}{\text{stack}}\{\underset{1 \le j \le n_i}{\text{stack}}(b_{ij})\}. \tag{3.4}
$$

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

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

We also assume that $B$ is full rank. Then the $x$ that minimizes (2.3) is the solution to the three-level sparse linear system with

$$
A = B^T B \quad \text{and} \quad a = B^T b. \tag{3.5}
$$

For general $m$ and $\{n_i\}_{1 \le i \le m}$ the nonzero components of $A$ and the sub-vectors of $a$ are, for $1 \le i \le m$,

$$
A_{11} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} B_{ij}^T B_{ij}, \quad a_1 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} B_{ij}^T b_{ij}, \quad A_{22,i} = \sum_{j=1}^{n_i} \dot{B}_{ij}^T \dot{B}_{ij},
$$

$$
A_{12,i} = \sum_{j=1}^{n_i} B_{ij}^T \dot{B}_{ij}, \quad a_{2,i} = \sum_{j=1}^{n_i} \dot{B}_{ij}^T b_{ij}
$$

and

$$
A_{22,ij} = \ddot{B}_{ij}^T \ddot{B}_{ij}, \quad A_{12,i,j} = \dot{B}_{ij}^T \ddot{B}_{ij}, \quad A_{12,ij} = B_{ij}^T \ddot{B}_{ij},
$$

$$
a_{2,ij} = \ddot{B}_{ij}^T b_{ij}, \quad 1 \le i \le m, 1 \le j \le n_i.
$$

The form of the matrix $B$ in (3.3) arises in statistical models containing both fixed effects and random effects with three-level structure [4]. Fuller details on this connection are given by Nolan et al. [10].

Theorem 3.3 provides a QR-decomposition enhancement of Theorem 3.2 for the least-squares-forms situation.

THEOREM 3.3. *Suppose that $A$ and $a$ admit the least squares forms defined by (3.3)–(3.5), where $B$ is full rank. 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*

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

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

        (ii) *then obtain*

$$\boldsymbol{c}_{0i} \equiv \boldsymbol{Q}_i^T \{\operatorname*{stack}_{1 \le j \le n_i}(\boldsymbol{d}_{2ij})\}, \quad \boldsymbol{c}_{1i} \equiv \textit{first } q_1 \textit{ rows of } \boldsymbol{c}_{0i},$$
$$\boldsymbol{c}_{2i} \equiv \textit{remaining rows of } \boldsymbol{c}_{0i}, \quad \boldsymbol{C}_{0i} \equiv \boldsymbol{Q}_i^T \{\operatorname*{stack}_{1 \le j \le n_i}(\boldsymbol{D}_{2ij})\},$$
$$\boldsymbol{C}_{1i} \equiv \textit{first } q_1 \textit{ rows of } \boldsymbol{C}_{0i}, \quad \boldsymbol{C}_{2i} \equiv \textit{remaining rows of } \boldsymbol{C}_{0i}.$$

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

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

(3) *For $1 \le i \le m$, the solutions are:*

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

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

$$\boldsymbol{x}_{2,ij} = \boldsymbol{R}_{ij}^{-1}(\boldsymbol{d}_{1ij} - \boldsymbol{D}_{1ij} \boldsymbol{x}_1 - \dot{\boldsymbol{D}}_{1ij} \boldsymbol{x}_{2,i}),$$
$$\boldsymbol{A}^{12,ij} = -\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij} \boldsymbol{A}^{11} + \dot{\boldsymbol{D}}_{1ij} \boldsymbol{A}^{12,iT})\}^T,$$
$$\boldsymbol{A}^{12,i,j} = -\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij} \boldsymbol{A}^{12,i} + \dot{\boldsymbol{D}}_{1ij} \boldsymbol{A}^{22,i})\}^T$$
$$\textit{and} \quad \boldsymbol{A}^{22,ij} = \boldsymbol{R}_{ij}^{-1}(\boldsymbol{R}_{ij}^{-T} - \boldsymbol{D}_{1ij} \boldsymbol{A}^{12,ij} - \dot{\boldsymbol{D}}_{1ij} \boldsymbol{A}^{12,i,j}).$$

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

$$|\boldsymbol{A}| = \Bigg[ (\textit{product of the diagonal entries of } \boldsymbol{R})$$

$$\times \prod_{i=1}^{m} \Bigg\{ (\textit{product of the diagonal entries of } \boldsymbol{R}_i)$$

$$\times \prod_{j=1}^{n_i} (\textit{product of the diagonal entries of } \boldsymbol{R}_{ij}) \Bigg\}\Bigg]^2.$$

Appendix A.4 contains a proof of Theorem 3.3.

REMARK 3.4. (1) As in Theorem 2.3, Step 1(a) of Theorem 3.3 involves determi-
nation 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 2.3 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-honoured 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 by Nolan et al. [10]. Future
extensions to higher-level situations are aided by our results and derivations.

## Appendix A.  Proofs of Theorems

### A.1.  Proof of Theorem 2.2

In the case of $m = 2$, the two-level sparse matrix linear
system problem is

$$\left[ \begin{array}{c|c|c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,2} \\ \hline \boldsymbol{A}_{12,1}^T & \boldsymbol{A}_{22,1} & \boldsymbol{O} \\ \hline \boldsymbol{A}_{12,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,2} \end{array} \right] \left[ \begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,2} \end{array} \right] = \left[ \begin{array}{c} \boldsymbol{a}_1 \\ \hline \boldsymbol{a}_{2,1} \\ \hline \boldsymbol{a}_{2,2} \end{array} \right],$$

which immediately leads to

$$\boldsymbol{A}_{11}\boldsymbol{x}_1 + \boldsymbol{A}_{12,1}\boldsymbol{x}_{2,1} + \boldsymbol{A}_{12,2}\,\boldsymbol{x}_{2,2} = \boldsymbol{a}_1$$

and

$$A_{12,i}^T x_1 + A_{22,i} x_{2,i} = a_{2,i}, \quad 1 \le i \le 2.$$

It is clear that the same pattern applies for general $m$, and we have

$$A_{11} x_1 + \sum_{i=1}^{m} A_{12,i} x_{2,i} = a_1 \tag{A.1}$$

and

$$A_{12,i}^T x_1 + A_{22,i} x_{2,i} = a_{2,i}, \quad 1 \le i \le m. \tag{A.2}$$

Conditions (A.2) immediately imply that

$$x_{2,i} = A_{22,i}^{-1}(a_{2,i} - A_{12,i}^T x_1), \quad 1 \le i \le m. \tag{A.3}$$

Substitution of (A.3) into (A.1) then leads to the solution for $x_1$ as stated in Theorem 2.2.

For the matrix inverse derivation, we again start with the $m = 2$ case, and note that

$$\left[ \begin{array}{c|c|c} A_{11} & A_{12,1} & A_{12,2} \\ \hline A_{12,1}^T & A_{22,1} & O \\ \hline A_{12,2}^T & O & A_{22,2} \end{array} \right] \left[ \begin{array}{c|c|c} A^{11} & A^{12,1} & A^{12,2} \\ \hline A^{12,1T} & A^{22,1} & \times \\ \hline A^{12,2T} & \times & A^{22,2} \end{array} \right] = \left[ \begin{array}{c|c|c} I & O & O \\ \hline O & I & O \\ \hline O & O & I \end{array} \right].$$

Observing the pattern from the $m = 2$ case and then extending to general $m$, we obtain the system of equations

$$A_{11} A^{11} + \sum_{i=1}^{m} A_{12,i} A^{12,iT} = I, \tag{A.4}$$

$$A_{12,i}^T A^{12,i} + A_{22,i} A^{22,i} = I, \quad 1 \le i \le m, \tag{A.5}$$

$$A_{12,i}^T A^{11} + A_{22,i} A^{12,iT} = O, \quad 1 \le i \le m. \tag{A.6}$$

From (A.6)

$$A^{12,iT} = -A_{22,i}^{-1} A_{12,i}^T A^{11}, \quad 1 \le i \le m. \tag{A.7}$$

Substitution of (A.7) into (A.4) gives

$$A_{11} A^{11} - \sum_{i=1}^{m} A_{12,i} A_{22,i}^{-1} A_{12,i}^T A^{11} = I,$$

which implies that

$$A^{11} = \left( A_{11} - \sum_{i=1}^{m} A_{12,i} A_{22,i}^{-1} A_{12,i}^T \right)^{-1}.$$

Substitution of (A.7) into (A.5) gives

$$-A_{12,i}^T A^{11} A_{12,i} A_{22,i}^{-1} + A_{22,i} A^{22,i} = I, \quad 1 \le i \le m,$$

implying that

$$A^{22,i} = A_{22,i}^{-1}(I + A_{12,i}^T A^{11} A_{12,i} A_{22,i}^{-1}) = A_{22,i}^{-1}(I - A_{12,i}^T A^{12,i}), \quad 1 \le i \le m.$$

For the $|A|$ result we first prove the following lemma.

LEMMA A.1. *Let $M$ be a symmetric invertible matrix with sub-block partitioning according to the notation*

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \quad and \quad M^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{12T} & M^{22} \end{bmatrix}.$$

*Then*

$$|M| = |(M^{11})^{-1}||M_{22}|.$$

PROOF. Lemma A.1 is a direct consequence of [5, Theorem 13.3.8] concerning the determinant of a matrix with $2 \times 2$ sub-block partitioning. □

From Lemma A.1 we have

$$|A| = |(A^{11})^{-1}|| \operatorname*{blockdiag}_{1 \le i \le m}(A_{22,i})| = |(A^{11})^{-1}| \prod_{i=1}^{m} |A_{22,i}|.$$

### A.2. Proof of Theorem 2.3

We first note the following simplification:

$$B_i^T B_i = B_i^T Q_i Q_i^T B_i = C_{0i}^T C_{0i} = C_{1i}^T C_{1i} + C_{2i}^T C_{2i}, \quad 1 \le i \le m,$$

where the first equality holds by the orthogonality of $Q_i$, and the second and third equalities hold by Step 1(b) of Theorem 2.3. A similar sequence of steps can be used to show that

$$B_i^T \dot{B}_i = C_{1i}^T R_i, \quad \dot{B}_i^T \dot{B}_i = R_i^T R_i,$$
$$B_i^T b_i = C_{1i}^T c_{1i} + C_{2i}^T c_{2i} \quad \text{and} \quad \dot{B}_i^T b_i = R_i^T c_{1i}, \quad 1 \le i \le m.$$

These simplifications allow us to represent the nonzero components of $A$ and the sub-vectors of $a$ as

$$A_{11} = \sum_{i=1}^{m}(C_{1i}^T C_{1i} + C_{2i}^T C_{2i}), \quad a_1 = \sum_{i=1}^{m}(C_{1i}^T c_{1i} + C_{2i}^T c_{2i}),$$
$$A_{12,i} = C_{1i}^T R_i, \quad A_{22,i} = R_i^T R_i, \quad \text{and} \quad a_{2,i} = R_i^T c_{1i}, \quad 1 \le i \le m.$$

The derivation of the inverse matrix problem is as follows:

$$A^{11} = \left(A_{11} - \sum_{i=1}^{m} A_{12,i} A_{22,i}^{-1} A_{12,i}^T\right)^{-1}$$
$$= \left(\sum_{i=1}^{m}(C_{1i}^T C_{1i} + C_{2i}^T C_{2i}) - \sum_{i=1}^{m} C_{1i}^T R_i (R_i^T R_i)^{-1} R_i^T C_{1i}\right)^{-1}$$

$$= \left( \sum_{i=1}^{m} \boldsymbol{C}_{2i}^T \boldsymbol{C}_{2i} \right)^{-1} = [\{\operatorname*{stack}_{1 \le i \le m}(\boldsymbol{C}_{2i})\}^T \{\operatorname*{stack}_{1 \le i \le m}(\boldsymbol{C}_{2i})\}]^{-1}$$

$$= \left( \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix}^T \boldsymbol{Q}^T \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix} \right)^{-1} = \boldsymbol{R}^{-1} \boldsymbol{R}^{-T},$$

where the fifth equality holds by Step 2 of Theorem 2.3, and we have used the orthogonality of $\boldsymbol{Q}$ for the sixth equality. The other components of $\boldsymbol{A}^{-1}$ are found by simply substituting the above simplifications into Theorem 2.2.

The derivation of the solution to the linear system is:

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

$$= \boldsymbol{R}^{-1} \boldsymbol{R}^{-T} \left\{ \sum_{i=1}^{m} (\boldsymbol{C}_{1i}^T \boldsymbol{c}_{1i} + \boldsymbol{C}_{2i}^T \boldsymbol{c}_{2i}) - \sum_{i=1}^{m} \boldsymbol{C}_{1i}^T \boldsymbol{R}_i (\boldsymbol{R}_i^T \boldsymbol{R}_i)^{-1} \boldsymbol{R}_i^T \boldsymbol{c}_{1i} \right\}$$

$$= \boldsymbol{R}^{-1} \boldsymbol{R}^{-T} \sum_{i=1}^{m} \boldsymbol{C}_{2i}^T \boldsymbol{c}_{2i} = \boldsymbol{R}^{-1} \boldsymbol{R}^{-T} \{\operatorname*{stack}_{1 \le i \le m}(\boldsymbol{C}_{2i})\}^T \{\operatorname*{stack}_{1 \le i \le m}(\boldsymbol{c}_{2i})\}$$

$$= \boldsymbol{R}^{-1} \boldsymbol{R}^{-T} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix}^T \boldsymbol{Q}^T \{\operatorname*{stack}_{1 \le i \le m}(\boldsymbol{c}_{2i})\}$$

$$= \boldsymbol{R}^{-1} \boldsymbol{R}^{-T} \boldsymbol{R}^T \boldsymbol{c}$$

$$= \boldsymbol{R}^{-1} \boldsymbol{c},$$

where we have used Step 2 of Theorem 2.3 for the fifth and sixth equalities. The other sub-vectors of $\boldsymbol{x}$ are found by simply substituting the above simplifications into Theorem 2.2.

For the $|\boldsymbol{A}|$ result, the $|\boldsymbol{A}|$ expression from Theorem 2.2 implies that

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

$$= |\boldsymbol{R}^T \boldsymbol{R}| \prod_{i=1}^{m} |\dot{\boldsymbol{B}}_i^T \dot{\boldsymbol{B}}_i|$$

$$= |\boldsymbol{R}^T \boldsymbol{R}| \prod_{i=1}^{m} |\boldsymbol{R}_i^T \boldsymbol{R}_i|. \tag{A.8}$$

Then

$$|\boldsymbol{R}^T \boldsymbol{R}| = |\boldsymbol{R}^T||\boldsymbol{R}| = (|\boldsymbol{R}|)^2 = (\text{product of the diagonal entries of } \boldsymbol{R})^2, \tag{A.9}$$

where we have used the result $|\boldsymbol{M}^T| = |\boldsymbol{M}|$ for any square matrix $\boldsymbol{M}$, and the fact that the determinant of an upper-triangular matrix is the product of its diagonal entries

[5, Lemma 13.1.1]. Replacement of $\boldsymbol{R}$ with $\boldsymbol{R}_i$ in (A.9) and substitution into (A.8) leads to the stated result for $|\boldsymbol{A}|$.

**A.3. Proof of Theorem 3.2** In the case of $m = 2$, $n_1 = 2$ and $n_2 = 3$, the three-level sparse matrix linear system problem is

$$
\begin{bmatrix}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,11} & \boldsymbol{A}_{12,12} & \boldsymbol{A}_{12,2} & \boldsymbol{A}_{12,21} & \boldsymbol{A}_{12,22} & \boldsymbol{A}_{12,23} \\
\boldsymbol{A}_{12,1}^T & \boldsymbol{A}_{22,1} & \boldsymbol{A}_{12,1,1} & \boldsymbol{A}_{12,1,2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,11}^T & \boldsymbol{A}_{12,1,1}^T & \boldsymbol{A}_{22,11} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,12}^T & \boldsymbol{A}_{12,1,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,12} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,2}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,2} & \boldsymbol{A}_{12,2,1} & \boldsymbol{A}_{12,2,2} & \boldsymbol{A}_{12,2,3} \\
\boldsymbol{A}_{12,21}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,1}^T & \boldsymbol{A}_{22,21} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,22}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,22} & \boldsymbol{O} \\
\boldsymbol{A}_{12,23}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,3}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,23}
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{x}_1 \\ \boldsymbol{x}_{2,1} \\ \boldsymbol{x}_{2,11} \\ \boldsymbol{x}_{2,12} \\ \boldsymbol{x}_{2,2} \\ \boldsymbol{x}_{2,21} \\ \boldsymbol{x}_{2,22} \\ \boldsymbol{x}_{2,23}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{a}_1 \\ \boldsymbol{a}_{2,1} \\ \boldsymbol{a}_{2,11} \\ \boldsymbol{a}_{2,12} \\ \boldsymbol{a}_{2,2} \\ \boldsymbol{a}_{2,21} \\ \boldsymbol{a}_{2,22} \\ \boldsymbol{a}_{2,23}
\end{bmatrix}.
$$

For arbitrary values of $m$ and $\{n_i\}_{1 \le i \le m}$, we immediately obtain the following set of equations:

$$
\boldsymbol{A}_{11}\boldsymbol{x}_1 + \sum_{i=1}^{m} \boldsymbol{A}_{12,i}\boldsymbol{x}_{2,i} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{A}_{12,ij}\boldsymbol{x}_{2,ij} = \boldsymbol{a}_1 \tag{A.10}
$$

$$
\boldsymbol{A}_{12,i}^T\boldsymbol{x}_1 + \boldsymbol{A}_{22,i}\boldsymbol{x}_{2,i} + \sum_{j=1}^{n_i} \boldsymbol{A}_{12,i,j}\boldsymbol{x}_{2,ij} = \boldsymbol{a}_{2,i}, \quad 1 \le i \le m \tag{A.11}
$$

$$
\boldsymbol{A}_{12,ij}^T\boldsymbol{x}_1 + \boldsymbol{A}_{12,i,j}^T\boldsymbol{x}_{2,i} + \boldsymbol{A}_{22,ij}\boldsymbol{x}_{2,ij} = \boldsymbol{a}_{2,ij}, \quad 1 \le i \le m, 1 \le j \le n_i. \tag{A.12}
$$

Conditions (A.12) imply that

$$
\boldsymbol{x}_{2,ij} = \boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{a}_{2,ij} - \boldsymbol{A}_{12,ij}^T\boldsymbol{x}_1 - \boldsymbol{A}_{12,i,j}^T\boldsymbol{x}_{2,i}), \quad 1 \le i \le m, 1 \le j \le n_i,
$$

which is the solution for $\boldsymbol{x}_{2,ij}$ as stated in Theorem 3.2. Substituting this result into conditions (A.11) leads to

$$
\boldsymbol{A}_{12,i}^T\boldsymbol{x}_1 + \boldsymbol{A}_{22,i}\boldsymbol{x}_{2,i} + \sum_{j=1}^{n_i} \boldsymbol{A}_{12,i,j}\boldsymbol{A}_{22,ij}^{-1}(\boldsymbol{a}_{2,ij} - \boldsymbol{A}_{12,ij}^T\boldsymbol{x}_1 - \boldsymbol{A}_{12,i,j}^T\boldsymbol{x}_{2,i}) = \boldsymbol{a}_{2,i}, \quad 1 \le i \le m.
$$

Solving this equation for $\boldsymbol{x}_{2,i}$ and using the definitions for $\boldsymbol{H}_{12,i}$ and $\boldsymbol{H}_{22,i}$ leads to the solution for $\boldsymbol{x}_{2,i}$ as stated in Theorem 3.2. Finally, substitution of the results for $\boldsymbol{x}_{2,i}$ and $\boldsymbol{x}_{2,ij}$ into condition (A.10) gives the solution for $\boldsymbol{x}_1$ stated in Theorem 3.2.

For the matrix inverse, we again illustrate the problem with $m = 2$, $n_1 = 2$ and $n_2 = 3$:

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

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

$$= \left[\begin{array}{c|c|c|c|c|c|c|c}
I & O & O & O & O & O & O & O \\
\hline
O & I & O & O & O & O & O & O \\
\hline
O & O & I & O & O & O & O & O \\
\hline
O & O & O & I & O & O & O & O \\
\hline
O & O & O & O & I & O & O & O \\
\hline
O & O & O & O & O & I & O & O \\
\hline
O & O & O & O & O & O & I & O \\
\hline
O & O & O & O & O & O & O & I
\end{array}\right].$$

Observing the pattern for the $m = 2$, $n_1 = 2$ and $n_2 = 3$ case and extending to general $m$ and $\{n_i\}_{1 \le i \le m}$, we obtain the following system of equations:

$$A_{11}A^{11} + \sum_{i=1}^{m} A_{12,i}A^{12,iT} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} A_{12,ij}A^{12,ijT} = I \tag{A.13}$$

$$A_{12,i}^T A^{12,i} + A_{22,i} A^{22,i} + \sum_{j=1}^{n_i} A_{12,i,j} A^{12,i,jT} = I, \quad 1 \le i \le m, \tag{A.14}$$

$$A_{12,ij}^T A^{12,ij} + A_{12,i,j}^T A^{12,i,j} + A_{22,ij} A^{22,ij} = I, \quad 1 \le i \le m, 1 \le j \le n_i, \tag{A.15}$$

$$A_{12,i}^T A^{11} + A_{22,i} A^{12,iT} + \sum_{j=1}^{n_i} A_{12,i,j} A^{12,ijT} = O, \quad 1 \le i \le m, \tag{A.16}$$

$$A_{12,ij}^T A^{11} + A_{12,i,j}^T A^{12,iT} + A_{22,ij} A^{12,ijT} = O, \quad 1 \le i \le m, 1 \le j \le n_i, \tag{A.17}$$

$$A_{12,ij}^T A^{12,i} + A_{12,i,j}^T A^{22,i} + A_{22,ij} A^{12,i,jT} = O, \quad 1 \le i \le m, 1 \le j \le n_i. \tag{A.18}$$

Rearranging conditions (A.17), we obtain

$$A^{12,ij} = -(A^{11} A_{12,ij} + A^{12,i} A_{12,i,j}) A_{22,ij}^{-1}, \quad 1 \le i \le m, 1 \le j \le n_i,$$

from which the result for $A^{12,ij}$ stated in Theorem 3.2 quickly follows. Rearranging conditions (A.16), we obtain

$$A^{12,i} = -\left(A^{11} A_{12,i} + \sum_{j=1}^{n_i} A^{12,ij} A_{12,i,j}^T\right) A_{22,i}^{-1}.$$

Substituting the results for each $A^{12,ij}$ leads, for $1 \le i \le m, 1 \le j \le n_i$, to

$$A^{12,i} = -\left\{A^{11} A_{12,i} - \sum_{j=1}^{n_i} (A^{11} A_{12,ij} + A^{12,i} A_{12,i,j}) A_{22,ij}^{-1} A_{12,i,j}^T\right\} A_{22,i}^{-1}.$$

Solving the above set of equations for each $A^{12,i}$ and using the definitions of $H_{12,i}$ and $H_{22,i}$, we obtain the result in Theorem 3.2. Substituting the results for each $A^{12,ij}$ and $A^{12,i}$ into (A.13), using the definition of $H_{12,i}$ and solving for $A^{11}$ leads to its stated result in Theorem 3.2. Rearranging conditions (A.18), we obtain

$$A^{12,i,j} = -(A^{12,iT} A_{12,ij} + A^{22,i} A_{12,i,j}) A_{22,ij}^{-1}, \quad 1 \le i \le m, 1 \le j \le n_i,$$

from which the result for $A^{12,i,j}$ stated in Theorem 3.2 quickly follows. Substitution of these results into the corresponding equations of condition (A.14) leads to

$$A_{12,i}^T A^{12,i} + A_{22,i} A^{22,i} - \sum_{j=1}^{n_i} A_{12,i,j} A_{22,ij}^{-1} (A_{12,ij}^T A^{12,i} + A_{12,i,j}^T A^{22,i}) = I, \quad 1 \le i \le m.$$

Solving the above set of equations for each $A^{22,i}$ and using the definitions of $H_{12,i}$ and $H_{22,i}$ in equation (3.2) of the main text, we obtain the result in Theorem 3.2. Rearrangement of (A.15) leads to

$$A^{22,ij} = A_{22,ij}^{-1} (I - A_{12,ij}^T A^{12,ij} - A_{12,i,j}^T A^{12,i,j}), \quad 1 \le i \le m, 1 \le j \le n_i.$$

From Lemma A.1 in the proof of Theorem 2.2,

$$|A| = |(A^{11})^{-1}| \prod_{i=1}^{m} \begin{vmatrix} A_{22,i} & \{\operatorname*{stack}_{1 \le j \le n_i} (A_{12,i,j}^T)\}^T \\ \operatorname*{stack}_{1 \le j \le n_i} (A_{12,i,j}^T) & \text{blockdiag } 1 \le j \le n_i (A_{22,ij}) \end{vmatrix}.$$

Application of Lemma A.1 and Theorem 2.2 to the two-level sparse matrices in the last-written expression leads to

$$
\begin{vmatrix}
\mathbf{A}_{22,i} & \{\underset{1 \leq j \leq n_i}{\operatorname{stack}}(\mathbf{A}_{12,i,j}^T)\}^T \\
\underset{1 \leq j \leq n_i}{\operatorname{stack}}(\mathbf{A}_{12,i,j}^T) & \operatorname{blockdiag} 1 \leq j \leq n_i(\mathbf{A}_{22,ij})
\end{vmatrix}
$$

$$
= \left| \mathbf{A}_{22,i} - \sum_{j=1}^{n_i} \mathbf{A}_{12,i,j} \mathbf{A}_{22,ij}^{-1} \mathbf{A}_{12,i,j}^T \right| \prod_{j=1}^{n_i} |\mathbf{A}_{22,ij}|
$$

and the stated result for $|\mathbf{A}|$ immediately follows.

**A.4. Proof of Theorem 3.3** We first note the following re-expression:

$$
\mathbf{B}_{ij}^T \mathbf{B}_{ij} = \mathbf{B}_{ij}^T \mathbf{Q}_{ij} \mathbf{Q}_{ij}^T \mathbf{B}_{ij} = \begin{bmatrix} \mathbf{D}_{1ij} \\ \mathbf{D}_{2ij} \end{bmatrix}^T \begin{bmatrix} \mathbf{D}_{1ij} \\ \mathbf{D}_{2ij} \end{bmatrix} = \mathbf{D}_{1ij}^T \mathbf{D}_{1ij} + \mathbf{D}_{2ij}^T \mathbf{D}_{2ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n_i,
$$

where the first equality holds by orthogonality of $\mathbf{Q}_{ij}$, and the second equality holds by Step 1(a)(ii) of Theorem 3.3. A similar sequence of steps can be used to show that

$$
\dot{\mathbf{B}}_{ij}^T \dot{\mathbf{B}}_{ij} = \dot{\mathbf{D}}_{1ij}^T \dot{\mathbf{D}}_{1ij} + \dot{\mathbf{D}}_{2ij}^T \dot{\mathbf{D}}_{2ij}, \quad \mathbf{B}_{ij}^T \dot{\mathbf{B}}_{ij} = \mathbf{D}_{1ij}^T \dot{\mathbf{D}}_{1ij} + \mathbf{D}_{2ij}^T \dot{\mathbf{D}}_{2ij},
$$

$$
\mathbf{B}_{ij}^T \mathbf{b}_{ij} = \mathbf{D}_{1ij}^T \mathbf{d}_{1ij} + \mathbf{D}_{2ij}^T \mathbf{d}_{2ij} \quad \text{and} \quad \dot{\mathbf{B}}_{ij}^T \mathbf{b}_{ij} = \dot{\mathbf{D}}_{1ij}^T \mathbf{d}_{1ij} + \dot{\mathbf{D}}_{2ij}^T \mathbf{d}_{2ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n_i.
$$

We also have

$$
\ddot{\mathbf{B}}_{ij}^T \ddot{\mathbf{B}}_{ij} = \ddot{\mathbf{B}}_{ij}^T \mathbf{Q}_{ij} \mathbf{Q}_{ij}^T \ddot{\mathbf{B}}_{ij} = \begin{bmatrix} \mathbf{R}_{ij} \\ \mathbf{0} \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_{ij} \\ \mathbf{0} \end{bmatrix} = \mathbf{R}_{ij}^T \mathbf{R}_{ij},
$$

where the first equality holds by orthogonality of $\mathbf{Q}_{ij}$ and the second equality holds by Step 1(a)(i) of Theorem 3.3. A similar sequence of steps can be used to show that

$$
\dot{\mathbf{B}}_{ij}^T \ddot{\mathbf{B}}_{ij} = \dot{\mathbf{D}}_{1ij}^T \mathbf{R}_{ij}, \quad \mathbf{B}_{ij}^T \ddot{\mathbf{B}}_{ij} = \mathbf{D}_{1ij}^T \mathbf{R}_{ij} \quad \text{and} \quad \ddot{\mathbf{B}}_{ij}^T \mathbf{b}_{ij} = \mathbf{R}_{ij}^T \mathbf{d}_{1ij}.
$$

The above simplifications allow us to represent the nonzero components of $\mathbf{A}$ and the sub-vectors of $\mathbf{a}$ as

$$
\mathbf{A}_{11} = \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{D}_{1ij}^T \mathbf{D}_{1ij} + \mathbf{D}_{2ij}^T \mathbf{D}_{2ij}), \quad \mathbf{a}_1 = \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{D}_{1ij}^T \mathbf{d}_{1ij} + \mathbf{D}_{2ij}^T \mathbf{d}_{2ij}),
$$

$$
\mathbf{A}_{22,i} = \sum_{j=1}^{n_i} (\dot{\mathbf{D}}_{1ij}^T \dot{\mathbf{D}}_{1ij} + \dot{\mathbf{D}}_{2ij}^T \dot{\mathbf{D}}_{2ij}), \quad \mathbf{A}_{12,i} = \sum_{j=1}^{n_i} (\mathbf{D}_{1ij}^T \dot{\mathbf{D}}_{1ij} + \mathbf{D}_{2ij}^T \dot{\mathbf{D}}_{2ij}),
$$

$$
\mathbf{a}_{2,i} = \sum_{j=1}^{n_i} (\dot{\mathbf{D}}_{1ij}^T \mathbf{d}_{1ij} + \dot{\mathbf{D}}_{2ij}^T \mathbf{d}_{2ij}), \quad 1 \leq i \leq m,
$$

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

$$A_{22,ij} = \boldsymbol{R}_{ij}^T \boldsymbol{R}_{ij}, \quad A_{12,i,j} = \dot{\boldsymbol{D}}_{1ij}^T \boldsymbol{R}_{ij}, \quad A_{12,ij} = \boldsymbol{D}_{1ij}^T \boldsymbol{R}_{ij}, \quad \boldsymbol{a}_{2,ij} = \boldsymbol{R}_{ij}^T \boldsymbol{d}_{1ij}.$$

Furthermore, each $\boldsymbol{H}_{22,i}$ matrix takes the form

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

$$= \sum_{j=1}^{n_i} (\dot{\boldsymbol{D}}_{1ij}^T \dot{\boldsymbol{D}}_{1ij} + \dot{\boldsymbol{D}}_{2ij}^T \dot{\boldsymbol{D}}_{2ij}) - \sum_{j=1}^{n_i} \dot{\boldsymbol{D}}_{1ij}^T \boldsymbol{R}_{ij} (\boldsymbol{R}_{ij}^T \boldsymbol{R}_{ij})^{-1} \boldsymbol{R}_{ij}^T \dot{\boldsymbol{D}}_{1ij}$$

$$= \sum_{j=1}^{n_i} \dot{\boldsymbol{D}}_{2ij}^T \dot{\boldsymbol{D}}_{2ij} = \{\operatorname*{stack}_{1 \leq j \leq n_i} \dot{\boldsymbol{D}}_{2ij}\}^T \{\operatorname*{stack}_{1 \leq j \leq n_i} \dot{\boldsymbol{D}}_{2ij}\}$$

$$= \left( \boldsymbol{Q}_i \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix} \right)^T \boldsymbol{Q}_i \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{R}_i^T \boldsymbol{R}_i, \quad 1 \leq i \leq m,$$

where the fifth equality holds by Step 1(b)(i) of Theorem 3.3. Similarly,

$$\boldsymbol{H}_{12,i} = \boldsymbol{C}_{1i}^T \boldsymbol{R}_i \quad \text{and} \quad \boldsymbol{H}_{12,i} \boldsymbol{H}_{22,i}^{-1} \boldsymbol{H}_{12,i}^T = \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i}, \quad 1 \leq i \leq m.$$

We are now in a position to derive the solutions to the inverse matrix problem:

$$\boldsymbol{A}^{11} = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\boldsymbol{D}_{1ij}^T \boldsymbol{D}_{1ij} + \boldsymbol{D}_{2ij}^T \boldsymbol{D}_{2ij}) - \sum_{i=i}^{m} \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i} - \sum_{i=1}^{m} \sum_{j=1}^{n_i} \boldsymbol{D}_{1ij}^T \boldsymbol{R}_{ij} (\boldsymbol{R}_{ij}^T \boldsymbol{R}_{ij})^{-1} \boldsymbol{R}_{ij}^T \boldsymbol{D}_{1ij} \right\}^{-1}$$

$$= \left( \sum_{i=1}^{m} \sum_{j=1}^{n_i} \boldsymbol{D}_{2ij}^T \boldsymbol{D}_{2ij} - \sum_{i=i}^{m} \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i} \right)^{-1}$$

$$= \left( \sum_{i=1}^{m} [\{\operatorname*{stack}_{1 \leq j \leq n_i}(\boldsymbol{D}_{2ij})\}^T \{\operatorname*{stack}_{1 \leq j \leq n_i}(\boldsymbol{D}_{2ij})\} - \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i}] \right)^{-1}$$

$$= \left( \sum_{i=1}^{m} [\{\operatorname*{stack}_{1 \leq j \leq n_i}(\boldsymbol{D}_{2ij})\}^T \boldsymbol{Q}_i \boldsymbol{Q}_i^T \{\operatorname*{stack}_{1 \leq j \leq n_i}(\boldsymbol{D}_{2ij})\} - \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i}] \right)^{-1}$$

$$= \left( \sum_{i=1}^{m} \left[ \begin{bmatrix} \boldsymbol{C}_{1i} \\ \boldsymbol{C}_{2i} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{C}_{1i} \\ \boldsymbol{C}_{2i} \end{bmatrix} - \boldsymbol{C}_{1i}^T \boldsymbol{C}_{1i} \right] \right)^{-1} = \left( \sum_{i=1}^{m} \boldsymbol{C}_{2i}^T \boldsymbol{C}_{2i} \right)^{-1}$$

$$= [\{\operatorname*{stack}_{1 \leq i \leq m}(\boldsymbol{C}_{2i})\}^T \{\operatorname*{stack}_{1 \leq i \leq m}(\boldsymbol{C}_{2i})\}]^{-1} = \left\{ \left( \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \right)^T \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \right\}^{-1} = \left( \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \right)^{-1} = \boldsymbol{R}^{-1} \boldsymbol{R}^{-T},$$

where the fourth equality holds by orthogonality of each $\boldsymbol{Q}_i$, the fifth equality holds by Step 1(b)(ii), the eighth equality holds by Step 2 and the ninth equality holds by orthogonality of $\boldsymbol{Q}$. The other components of $\boldsymbol{A}^{-1}$ are found by simply substituting the above simplifications into Theorem 3.2, and using a similar sequence of steps.

For the linear system solution we have

$$h_{2,i} = \sum_{j=1}^{n_i}(\dot{D}_{1ij}^T d_{1ij} + \dot{D}_{2ij}^T d_{2ij}) - \sum_{j=1}^{n_i} \dot{D}_{1ij}^T R_{ij}(R_{ij}^T R_{ij})^{-1} R_{ij}^T d_{1ij} = \sum_{j=1}^{n_i} \dot{D}_{2ij}^T d_{2ij}$$

$$= \{\operatorname*{stack}_{1 \le j \le n_i}(\dot{D}_{2ij})\}^T \{\operatorname*{stack}_{1 \le j \le n_i}(d_{2ij})\}$$

$$= \begin{bmatrix} R_i \\ 0 \end{bmatrix}^T Q_i^T \{\operatorname*{stack}_{1 \le j \le n_i}(d_{2ij})\}$$

$$= \begin{bmatrix} R_i \\ 0 \end{bmatrix}^T \begin{bmatrix} c_{1i} \\ c_{2i} \end{bmatrix} = R_i^T c_{1i}, \quad 1 \le i \le m,$$

where the fourth and fifth equalities hold by Steps 1(b)(i) and 1(b)(ii), respectively. Also,

$$A_{12,ij}A_{22,ij}^{-1}a_{2,ij} = D_{1ij}^T R_{ij}(R_{ij}^T R_{ij})^{-1} R_{ij}^T d_{1ij} = D_{1ij}^T d_{1ij}, \quad 1 \le i \le m, 1 \le j \le n_i.$$

The final task is to derive the expressions in Theorem 3.3 for the components of $x$. For the first block we have

$$x_1 = A^{11}\Big\{\sum_{i=1}^m \sum_{j=1}^{n_i}(D_{1ij}^T d_{1ij} + D_{2ij}^T d_{2ij}) - \sum_{i=1}^m C_{1i}^T R_i^{-T} R_i^T c_{1i} - \sum_{i=1}^m \sum_{j=1}^{n_i} D_{1ij}^T d_{1ij}\Big\}$$

$$= A^{11}\Big\{\sum_{i=1}^m \Big(\sum_{j=1}^{n_i} D_{2ij}^T d_{2ij} - C_{1i}^T c_{1i}\Big)\Big\}$$

$$= A^{11}\Big(\sum_{i=1}^m [\{\operatorname*{stack}_{1 \le j \le n_i}(D_{2ij})\}^T \{\operatorname*{stack}_{1 \le j \le n_i}(d_{2ij})\} - C_{1i}^T c_{1i}]\Big)$$

$$= A^{11}\Big(\sum_{i=1}^m [\{\operatorname*{stack}_{1 \le j \le n_i}(D_{2ij})\}^T Q_i Q_i^T \{\operatorname*{stack}_{1 \le j \le n_i}(d_{2ij})\} - C_{1i}^T c_{1i}]\Big)$$

$$= A^{11}\Big(\sum_{i=1}^m \Big[\begin{bmatrix} C_{1i} \\ C_{2i} \end{bmatrix}^T \begin{bmatrix} c_{1i} \\ c_{2i} \end{bmatrix} - C_{1i}^T c_{1i}\Big]\Big) = A^{11}\Big\{\sum_{i=1}^m (C_{1i}^T c_{1i} + C_{2i}^T c_{2i} - C_{1i}^T c_{1i})\Big\}$$

$$= A^{11} \sum_{i=1}^m C_{2i}^T c_{2i} = A^{11}\{\operatorname*{stack}_{1 \le i \le m}(C_{2i})\}^T \{\operatorname*{stack}_{1 \le i \le m}(c_{2i})\}$$

$$= A^{11}\Big(Q\begin{bmatrix} R \\ 0 \end{bmatrix}\Big)^T \{\operatorname*{stack}_{1 \le i \le m}(c_{2i})\} = R^{-1}R^{-T}\begin{bmatrix} R \\ 0 \end{bmatrix}^T Q^T \{\operatorname*{stack}_{1 \le i \le m}(c_{2i})\}$$

$$= R^{-1}R^{-T}R^T c = R^{-1}c,$$

where the fourth equality holds by the orthogonality of $Q_i$, the fifth equality holds by Step 1(b)(ii) and the ninth and eleventh equalities hold by Step 2 of Theorem 3.3. We get the other sub-vectors of $x$ by a similar sequence of computations.

From Theorem 3.2,

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

From part 3 of Theorem 3.3 we have $A^{11} = R^{-1}R^{-T}$, so steps similar to those used for simplification of $|A|$ in the proof of Theorem 2.3 lead to

$$|(A^{11})^{-1}| = (\text{product of the diagonal entries of } R)^2.$$

From the above arguments involving the $H_{22,i}$ matrices,

$$\left|A_{22,i} - \sum_{j=1}^{n_i} A_{12,i,j}\, A_{22,ij}^{-1}\, A_{12,i,j}^{T}\right| = |H_{22,i}| = |R_i^T R_i| = (|R_i|)^2$$

$$= (\text{product of the diagonal entries of } R_i)^2$$

for $1 \le i \le m$. Lastly,

$$|A_{22,ij}| = |\ddot{B}_{ij}^T \ddot{B}_{ij}| = |R_{ij}^T R_{ij}| = (|R_{ij}|)^2 = (\text{product of the diagonal entries of } R_{ij})^2$$

for $1 \le i \le m,\ 1 \le j \le n_i$. The stated result for $|A|$ immediately follows.

## Acknowledgements

## References

[1] B. H. Baltagi, *Econometric analysis of panel data* (John Wiley & Sons, Chichester, 2013).
[2] G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs (eds), *Longitudinal data analysis* (Chapman & Hall/CRC, Boca Raton, FL, 2008); doi:10.1201/9781420011579.
[3] J. E. Gentle, *Matrix algebra* (Springer, New York, 2007); doi:10.1007/978-0-387-70873-7.
[4] H. Goldstein, *Multilevel statistical models*, 4th edn (John Wiley & Sons, Chichester, 2010); doi:10.1002/9780470973394.
[5] D. A. Harville, *Matrix algebra from a statistician's perspective* (Springer, New York, 2008).
[6] C. R. Henderson, "Best linear unbiased estimation and prediction under a selection model", *Biometrics* **31** (1975) 423–447; doi:10.2307/2529430.
[7] W. Hołubowski, D. Kurzyk and T. Trawiński, "A fast method for computing the inverse of symmetric block arrowhead matrices", *Appl. Math. Inf. Sci.* **9** (2015) 319–324; doi:10.12785/amis/092L06.
[8] N. T. Longford, "A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects", *Biometrika* **74** (1987) 817–827; doi:10.1093/biomet/74.4.817.
[9] C. E. McCulloch, S. R. Searle and J. M. Neuhaus, *Generalized, linear, and mixed models*, 2nd edn (John Wiley & Sons, Hoboken, NJ, 2008).
[10] T. H. Nolan, M. Menictas and M. P. Wand, *Streamlined computing for variational inference with higher level random effects*. arXiv:1903.06616v3 (2020).

[11]  J. C. Pinheiro and D. M. Bates, *Mixed-effects models in S and S-PLUS* (Springer, New York, 2000); doi:10.1007/978-1-4419-0318-1.

[12]  J. N. K. Rao and I. Molina, *Small area estimation*, 2nd edn (John Wiley & Sons, Hoboken, NJ, 2015); doi:10.1002/9781118735855.

[13]  S. Saberi Nejafi, S. A. Edalatpanah and G. A. Gravvanis, "An efficient method for computing the inverse of arrowhead matrices", *Appl. Math. Lett.* **33** (2014) 1–5; doi:10.1016/j.aml.2014.02.010.

[14]  P. S. Stanimirović, V. N. Katsikis and D. Kolundžija, "Inversion and pseudoinversion of block arrowhead matrices", *Appl. Math. Comput.* **341** (2019) 379–401; doi:10.1016/j.amc.2018.09.006.