STATISTICS IN TRANSITION new series, December 2019
Vol. 20, No. 4, pp. 59–70, DOI 10.21307/stattrans-2019-034
Submitted – 03.10.2017; Paper ready for publication – 06.02.2019

LINEAR CHOLESKY DECOMPOSITION OF COVARIANCE MATRICES IN MIXED MODELS WITH CORRELATED RANDOM EFFECTS

Anasu Rabe[1], **D. K. Shangodoyin**[2], **K.Thaga**[3]

ABSTRACT

Modelling the covariance matrix in linear mixed models provides an additional advantage in making inference about subject-specific effects, particularly in the analysis of repeated measurement data, where time-ordering of the responses induces significant correlation. Some difficulties encountered in these modelling procedures include high dimensionality and statistical interpretability of parameters, positive definiteness constraint and violation of model assumptions. One key assumption in linear mixed models is that random errors and random effects are independent, and its violation leads to biased and inefficient parameter estimates. To minimize these drawbacks, we developed a procedure that accounts for correlations induced by violation of this key assumption. In recent literature, variants of Cholesky decomposition were employed to circumvent the positive definiteness constraint, with parsimony achieved by joint modelling of mean and covariance parameters using covariates. In this article, we developed a linear Cholesky decomposition of the random effects covariance matrix, providing a framework for inference that accounts for correlations induced by covariate(s) shared by both fixed and random effects design matrices, a circumstance leading to lack of independence between random errors and random effects. The proposed decomposition is particularly useful in parameter estimation using the maximum likelihood and restricted/residual maximum likelihood procedures.

Key words: correlated random effects, covariance matrix, linear Cholesky decomposition, linear mixed models.

1. Introduction

Linear mixed models are a class of models (Laird and Ware, 1982) that provide parameter estimates (inference) for population (fixed effects) and subject-specific (random effects) characteristics via separate covariance structures.
Let $Y_i = (y_{i1}, \ldots, y_{in_i})^T$ be $n_i \times 1$ vector of responses measured on the i^{th} subject

[1]Department of Mathematics and Computer Science, Umaru Musa 'Yaradua University, PMB 2218, Dutsinma Road, Katsina, Katsina State, Nigeria. E-mail: anasu.rabe@umyu.edu.ng
[2]Department of Statistics, University of Botswana, Corner of Notwane and Mobuto Road, Pvt Bag 00706, Gaborone, Botswana. E-mail: dkshangodoyin@mopipi.ub.bw
[3]Department of Statistics, University of Botswana, Corner of Notwane and Mobuto Road, Pvt Bag 00706, Gaborone, Botswana. E-mail: kthaga@mopipi.ub.bw

$(i = 1, \ldots, m)$ from a total of $n = \sum_{i=1}^{m} n_i$ measurements. A linear mixed model (Laird and Ware, 1982) for the i^{th} subject is represented by:

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i \tag{1.1}$$

where X_i is $n_i \times p$ design matrix for the $p \times 1$ vector of fixed-effects regression coefficients β, Z_i is $n_i \times q$ design matrix for the $q \times 1$ vector of random effects γ_i and ε_i is $n_i \times 1$ vector of error terms. For model (1.1), we assume that:

i. Error terms ε_i are independent within i^{th} subject and are normally distributed with zero mean and $n_i \times n_i$ covariance matrix Σ_i: $E(\varepsilon_i) = 0$ and $\varepsilon_i \sim N(0, \Sigma_i)$.

ii. The random effects γ_i are independent and normally distributed with mean zero and $q_i \times q_i$ covariance matrix Δ_i. X and Z share no covariate(s) so that γ_i and ε_i are independent: $E(\gamma_i) = 0$, $\gamma_i \sim N(0, \Delta_i)$ and $Cov(\gamma_i, \varepsilon_i) = 0$.

iii. The response variable y_i is normally distributed with mean $X_i\beta$ and covariance matrix V_i: $y_i \sim N(X_i\beta, V_i)$, where $V_i = Z_i\Delta_i Z_i^T + \Sigma_i$.

In practice, assumptions on random effects are difficult to satisfy. For example, non-normality of random effects has been proven in the literature by several authors, with Lange and Ryan (1989) providing some concrete examples. Assumption of independence between ε_i and γ_i may also not hold when research interest require that X and Z have common covariate(s). For example (Gelfand et al., 1995), in growth studies where individual profiles are centered about a population baseline curve. In such cases, individual models incorporate the baseline population covariate. Another example is the analysis of CD4 cell counts in HIV studies where all subjects are HIV-positive at baseline, but not all were diagnosed for the disease. Interest here is to develop a model that incorporates diagnosis as a baseline covariate and therefore should be incorporated in both X and Z. Also, in hierarchical mixed effects models (Pinheiro and Bates, 2000), the model structure is conditional on the random effects, making γ and ε_i inherently dependent.
Several approaches have been proposed in the literature to address these drawbacks, and procedures based on Cholesky decomposition of the random effects covariance matrix Δ_i provide additional advantage of guaranteeing the positive definiteness of the resulting factors, circumventing constraints of high dimensionality and statistical interpret-ability of the resulting parameters. We review some of these Cholesky-based procedures in the following section.

2. Variants of Cholesky decomposition

The standard Cholesky decomposition of a real, symmetric, positive definite matrix $\Sigma_{p\times p}$ is

$$\Sigma = U^T U \tag{2.1}$$

where U is an upper triangular with positive diagonal elements. The main advantage of this decomposition when used in parameter estimation procedures, such as maximum likelihood (ML) and residual/restricted maximum likelihood (REML), is that it provides an unconstrained parameterization of the parameters in Σ, circumventing the positive definiteness constraint. However, Pinheiro and Bates (1996) showed that the Cholesky factors U are not unique and the unconstrained $\frac{p(p+1)}{2}$ parameters lack a meaningful statistical interpretation with respect to the entries in Σ. To overcome these drawbacks, several classical and Bayesian approaches have been proposed in the literature.

Under the classical approach, Pourahmadi (1999) proposed the modified Cholesky decomposition (MCD) for modelling parameters of the precision matrix and developed a ML procedure (Pourahmadi, 2000) for normal generalized linear models:

$$\Sigma^{-1} = L^T D^{-1} L \tag{2.2}$$

where entries in the unit lower triangular Cholesky factor L are interpreted as negatives of autoregressive coefficients when a response variable y_t is regressed on its predecessors $y_{t-1}, \ldots, y_1$ and entries on the diagonal factor D as logarithms of their innovations.

However, despite the good performance of the proposed MCD, the procedure left a number of questions unanswered:

First, the proposed ML estimation procedure (and its restricted extensions) works well only when measurement times are identical across subjects, and hence may not be applicable to unbalanced data sets, particularly since it utilizes the sample covariance matrix S^2 as an initial value for Σ, which does not exist in unbalanced data settings (Pan and MacKenzie, 2006). In such cases, either the ML procedure is enhanced (Holan and Spinka, 2007) or some numerical optimization approach is adopted (Zimmermann et al., 1998). Second, the use of a saturated mean structure may be unnecessary (Pan and MacKenzie, 2003), except when the mean model is incorporated in searching the joint mean-covariance parameter space. Third, the regressogram, as proposed and utilized in model selection for the dependence and innovation variance, failed to capture the joint mean-covariance structure since the mean model was not included. However, Garcia et al. (2012) showed that as a data-driven graphical tool for model selection, they are powerful graphical tools in joint mean-covariance model selection for incomplete longitudinal data. Fourth, when subject-specific characterization is the focus of research interest, a linear mixed modelling (LMM) framework may be more flexible than a generalized linear modeling (GLM) framework.

These questions raised a number of issues and stimulated keen interest in modeling covariance structures under different frameworks and perspectives. Pan and Mackenzie (2003) observed that parameter estimates based on MCD are not optimal, and to address the first question, they proposed extending the procedure to unbalanced data with optimal parameter estimates achieved via joint search of the mean-covariance space. Zhang and Leng (2012) proposed a moving average

Cholesky decomposition (MACD) for Σ as the inverse of precision matrix:

$$\Sigma = L^{-1}DL^{-T} \tag{2.3}$$

where the entries in L^{-1} have a moving average (MA) interpretation. It has the same advantages and limitations as MCD and only differs in its MA interpretation. Li and Pourahmadi (2013) utilized MCD in developing a procedure that circumvent the effect of violating normality assumption on random effects in linear mixed models. However, their procedure was based on the assumption that design matrices X and Z share no common covariates and may result into inefficient parameters when such assumptions are violated. More recently, Lee et al. (2017) proposed an autoregressive moving average Cholesky decomposition (ARMACD) by combining the modified Cholesky decomposition and moving average Cholesky decomposition to address high-dimensionality and positive definiteness constraints:

$$\Lambda_i \Sigma_i \Lambda_i^T = L_i D_i L_i^i \tag{2.4}$$

where Λ_i is unit lower triangular matrix with generalized autoregressive parameters (GARPs) $-\phi_{i,tj}$ at its $(t,j)^{th}$ position, L_i is unit lower triangular with generalized moving average parameters (GMAPs) $\imath_{i,tj}$ at its $(t,j)^{th}$ position $(j<t)$ and $D_i = diag(\sigma_{i1}^2,\ldots,\sigma_{in_i}^2)$ is diagonal with innovation variances (IV) σ_{ij}^2. This decomposition subsumes a wide variety of covariance structures which are more flexible and with better forecasting performance than separate higher order AR or MA models. The combination of MCD and MACD creates a unified framework with models that allow nonstationarity and heteroscedasticity in parameter estimates.
Under a Bayesian framework, Daniels and Zhao (2003) proposed modelling the random effects covariance matrix Δ_i for the i^{th} subject ($i=1,\ldots,m$) using the modified Cholesky decomposition

$$\Lambda_i \Delta_i \Lambda_i^T = D_i \tag{2.5}$$

where $D_i = diag(\sigma_{i1}^2,\ldots,\sigma_{iq}^2)$ is diagonal with innovation variances (IV) σ_i^2 and Λ_i is unit lower triangular with GARPs $-\phi_{i,tj}$ as its $(t,j)^{th}$ entry. This variant also provides the advantages of overcoming the positive definite constraint and statistical interpretation of parameters. By adopting a Bayesian approach, they gained additional flexibility in obtaining the sampling distribution of the random effects using a simple Gibbs sampler, which sample from the posterior distribution of parameters. To incorporate heterogeneity in Δ_i, the random effects were allowed to depend only on subject-specific covariates and parsimony was achieved by regressing the parameters using these covariates. However, when random effects differ on different linear combinations of these covariates, separate covariance structures need to be fit for each combination and misspecification of a structure can lead to inefficient parameter estimates. Another drawback is their assumption of statistical independence of the repeated measurement given the random effects. This assumption restricted the application of their approach in longitudinal studies. Chen and Dunson (2003)

proposed an alternative Cholesky decomposition for selecting the random effects components. Their approach factored the random effects covariance matrix Δ into

$$\Delta = D\Lambda\Lambda^T D \tag{2.6}$$

where D is a diagonal matrix with elements proportional to standard deviation of the random effects and Λ is a unit lower triangular matrix with off-diagonal elements describing correlations among the random effects. With separate factors for variance and correlation, their approach is computationally more tractable and provides some flexibility in selecting the random effects components. However, their approach is based on the assumption that components of the random errors and random effects are mutually independent. Gaskins and Daniels (2013) extended the Cholesky-based joint mean-covariance modelling to longitudinal data from several groups of subjects. They proposed a data-driven nonparametric method that simultaneously estimates the covariance matrix from each group by developing nonparametric priors using the matrix stick-breaking process. More recently, Han and Lee (2016) proposed a variant ARMACD decomposition:

$$\Lambda_i\Sigma_i\Lambda_i = C_i\Delta_b C_i^T + L_i D_i L_i^T \tag{2.7}$$

where Λ_i, L_i and D_i were as described above, Δ_b is the random effects covariance matrix and $C_i = (c_{i1}, \ldots, c_{in_i})^T$ is the random effects design matrix. This approach has all the advantages of ARMACD as proposed by Lee et al. (2017), but model parameters are obtained conditional on random effects in the linear mixed model. By allowing X and Z to have common covariate(s), Gelfand et al. (1995) proposed a modelling procedure in which parameters are hierarchically centred to account for between-level correlations, and to ensure model identification. However, the random effects covariance matrix of their proposed procedure is positive semi-definite, leading to poor convergence properties in some parameter estimates.
To account for heteroscedasticity in the random errors ε_i via modelling the variance function, Pinheiro and Bates (2000) proposed a variant Cholesky decomposition of the random errors covariance matrix Σ_i as

$$\Sigma_i = D_i C_i D_i \tag{2.8}$$

where D_i is diagonal, describing the variance of the random errors and C_i is triangular with all diagonal elements positive, describing the correlation of the random errors. The variance function model was proposed, conditional on the random effects γ_i, as a function of the conditional population mean response $\mu_{ij} = E(y_i|\gamma_i)$. Now, with these conditional dependencies, the assumption of independence between ε_i and γ_i no longer holds and $Cov(\varepsilon_i, \gamma_i) \neq 0$. To circumvent the consequences of this violation, they allow common covariates between X and Z, approximating μ_{ij} by the best linear unbiased predictor (BLUP)

$$\hat{\mu}_{ij} = x_{ij}^T\beta + z_{ij}^T\gamma_i \tag{2.9}$$

where x_{ij} and z_{ij} denote the j^{th} rows of X_i and Z_i, respectively.
In this article, we propose a more efficient approach to address the violation of this key assumption. We improve efficiency in inference and gain more insight into model bahaviour by modelling the correlation structure between ε_i and γ_i when $Cov(\varepsilon_i, \gamma_i) \neq 0$, and to achieve this, we propose a linear Cholesky decomposition of the random effects covariance matrix Δ_i . Our approach is based on linear transformation of inner products of functions of Cholesky factors when subjected to left-right or lu decomposition. We next discuss this transformation.

3. Linear Cholesky decomposition

The proposed linear decomposition is based on upper triangular Cholesky factor U. Let $\Delta = U^T U$ be $q \times q$ standard Cholesky decomposition with U upper triangular. We subject U to lu factorization, obtaining:

$$\Delta = [lu(U)]^T lu(U) = (\Phi + \Psi)^T (\Phi + \Psi) \tag{3.1}$$

$$\Delta^{-1} = \left[lu\left(U^{-1}\right)\right]^T \left[lu\left(U^{-1}\right)\right] = \left(\rho + \Psi^{-1}\right)^T \left(\rho + \Psi^{-1}\right) \tag{3.2}$$

where $\Phi(-\phi_{i,kj})$ and $\rho(\theta_{i,kj})$ are upper triangular with zeros on the diagonal, parameters $-\phi_{i,kj}$ and $\theta_{i,kj}$ in $(k,j)^{th}$ positions, respectively, and innovation $\Psi = diag(\psi_{11}, \ldots, \psi_{qq})$ with $\Psi^{-1} = diag(\psi_{11}^{-1}, \ldots, \psi_{qq}^{-1})$ as its inverse.
Expanding (3.1) and (3.2), we obtain

$$\Delta = (\Phi + \Psi)^T (\Phi + \Psi) = \Phi^T\Phi + \Phi^T\Psi + \Psi^T\Phi + \Psi^T\Psi \tag{3.3}$$

$$\Delta^{-1} = \left(\rho + \Psi^{-1}\right)^T \left(\rho + \Psi^{-1}\right) = \rho^T\rho + \rho^T\Psi^{-1} + \Psi^{-T}\rho + \Psi^{-T}\Psi^{-1} \tag{3.4}$$

Definition 1 *Let Δ be represented by (3.3). Linear Cholesky decomposition is defined by*

$$\Delta = \Phi^T\Phi + \rho^T\rho + \Psi^T\Psi \tag{3.5}$$

$$= \Delta_{AR} + \Delta_{MA} + \Delta_{IV} \tag{3.6}$$

with Cholesky factors Φ, ρ and Ψ, respectively, describing the correlation structure in γ_i, the correlation structure between ε_i and γ_i, and the innovation of γ_i.

Also, for the precision matrix we have:

Definition 2 *Let Δ^{-1} be represented by (3.4). Linear Cholesky decomposition is defined by*

$$\Delta^{-1} = \rho^T\rho + \Phi^T\Phi + [\Psi]^{-T}\Psi^{-1} \tag{3.7}$$

$$= \Delta_{MA} + \Delta_{AR} + \Delta_{IV}^{-1} \tag{3.8}$$

The following theorem provides the basis for linear Cholesky decomposition:

Theorem 1 *Linear Cholesky decomposition of real, symmetric positive definite* Δ *is*

$$\begin{aligned}\Delta(\Theta) &= \Phi^T\Phi + \rho^T\rho + \Psi^T\Psi \\ &= \Delta_{AR} + \Delta_{MA} + \Delta_{IV} \end{aligned} \tag{3.9}$$

where $\Theta = (-\phi_{i,j}, \theta_{i,j}, \psi_{ii})$ *with* $-\phi_{ij} = corr(\gamma_i, \gamma_j)$ *for* $i \neq j$, $\theta_{ij} = corr(\varepsilon_{ik}, \gamma_{jk})$ *and* $\psi_{ii} = log\,(diag\,[\Delta_{ii}])$.

Proof 1 *Let* $\Sigma_{p\times p} = U^T U$ *be the standard Cholesky decomposition, then lu decomposition of upper triangular* U *results into ul (upper-lower) factors (see Stewart (1998), page 183):*

$$[lu\,(U)]^T\, lu\,(U) = (U^*\Psi)^T\,(U^*\Psi) \tag{3.10}$$

$$= \Psi^T U^{*T} U^*\Psi \tag{3.11}$$

with U^* *unit upper triangular and* Ψ *lower triangular (diagonal) matrices.*
Let 'b' be $p \times 1$ *suitably chosen vector such that columns of* U^* *can be sequentially extracted via repeated multiplication, forming a Krylov sequence* $b, U^*b, U^{*2}b, \ldots, U^{*(p-1)}b$. *Define as Krylov matrix*

$$\mathscr{K} = \left[b, U^*b, U^{*2}b, \ldots, U^{*(p-1)}b\right] = \langle b, U^*b\rangle \tag{3.12}$$

where $\langle .,. \rangle$ *is an inner product, with the columns forming an ordered basis whose linear combinations span the Krylov subspaces* $\mathscr{K}_1, \ldots, \mathscr{K}_{p-1} = \mathrm{K}^{(p-1)} \subseteq \mathrm{F}^p$, *where* F^p *is* $p - dimensional$ *vector field. lu decomposition described by (3.11) is nonlinear in the factors. For a linear decomposition, we have from (3.3)*

$$\Delta = \Phi^T\Phi + \underbrace{\Phi^T\Psi + \Psi^T\Phi} + \Psi^T\Psi$$

where Φ *is strictly upper triangular with dependence parameters and* Ψ *is diagonal with variance parameters. The under-braced equation is a function of inner products of the respective Cholesky factors. There are several, well-established matrix linear transformations (such as Lyapunov stability transformation, see Carlson and Datta (1979)) that can be used in obtaining meaningful interpretation of this function of inner products.*
If we let each product describe the rate of change in value of the correlation parameters θ *between* ε_i *and* γ_i *(through shared covariate(s)) as*

$$\partial f\,[\Phi(\phi)] = \left[\Phi^T(\theta)\Psi(\theta)\right]\partial\theta \tag{3.13}$$

$$\partial f\,[\Psi(\psi)] = \left[\Psi^T(\theta)\Phi(\theta)\right]\partial\theta \tag{3.14}$$

then, we obtain the correlation between ε_i *and* γ_i *(through shared covariates, with respect to parameters in* Φ*) using a direct differentiation result by De Hoog et al. (2011), as*

$$\begin{aligned}\left[\Phi^T(\theta)\Psi(\theta)+\Psi^T(\theta)\Phi(\theta)\right]\partial\theta &= \partial f\left[\Psi(\theta)\Phi(\theta)\Psi^T(\theta)\right]\\ &= \Phi^T(\theta)\Psi(\theta)\Phi(\theta)\\ &= \langle\Phi,\Psi\Phi\rangle \end{aligned} \tag{3.15}$$

The columns of $\Phi(\theta)=\left[0,U^*b,U^{*2}b,\ldots,U^{*(n-1)}b\right]$ *form linear combinations that span the same Krylov subspaces, but* $\Phi(\theta)$ *has the first column as zero vector, with zeros on its diagonal while* $U^*\left(u^*_{ij}\right)$ *is a unit triangular with the first column as a unit vector, having the first entry 1. However, congruence of* $\langle b,U^*b\rangle$ *to* $\langle\Phi(\theta),\Psi(\theta)\Phi(\theta)\rangle$ *implies congruence of* $U^*(u^*_{ij})$ *to* $\Phi(\theta)$ *and can be exploited in establishing an equivalence relation between them by changing the basis of* U^**: Define a new basis* $Z=(z_1,\ldots,z_p)\in \mathrm{K}^{(p-1)}$ *for the column space of* U^* *and let* $Z=SU^*$ *where S is invertible, then*

$$f(Z)=Z^T\Psi Z=\langle Z,\Psi Z\rangle=\langle SU^*,\Psi SU^*\rangle$$

Now, $\langle SU^*,\Psi SU^*\rangle$ *is congruent to* $\langle\Phi(\theta),\Psi(\theta)\Phi(\theta)\rangle$ *if there exists a non-singular matrix Q such that*

$$\begin{aligned} Q\Phi(\theta) &= SU^*Q \;\Rightarrow\; \Phi(\theta) &&= Q^{-1}SU^*Q\\ Q\Psi(\theta)\Phi(\theta) &= \Psi(\theta)SU^*Q \;\Rightarrow\; \Psi(\theta)\Phi(\theta) &&= Q^T\Psi(\theta)SU^*Q \end{aligned} \tag{3.16}$$

Then, we have

$$\begin{aligned}\Phi^T(\theta)\Psi(\theta)\Phi(\theta) &= \left(Q^{-1}SU^*Q\right)^T\left(Q^T\Psi(\theta)SU^*Q\right)\\ &= Q^TU^{*T}S^TQ^{-T}Q^T\Psi(\theta)SU^*Q\\ &= (U^*Q)^TS^T\Psi(\theta)S(U^*Q)\\ &= (U^*Q)^TI(U^*Q)\\ &= (U^*Q)^T(U^*Q)\end{aligned} \tag{3.17}$$

with invertible S diagonalizing $\Psi(\theta)$ *to identity:* $S^T\Psi(\theta)S\longrightarrow I$*. The matrix* $Q=[q_1,\ldots,q_p]$ *is obtained using the Lanczos algorithm. The main advantage of the lu factorization is that columns of* U *can be reconstructed using the non-zero rows of* U^* *as coefficients and we exploit this feature in estimating MA parameters by reducing an upper Hessenberg matrix H to tridiagonal via column operations:*

$$
\begin{aligned}
U^*Q &= [U^*q_1,\ldots,U^*q_p] \\
&= [q_1,\ldots,q_p]\begin{pmatrix} h_{11} & h_{12} & h_{13} & \ldots & \ldots & h_{1p} \\ h_{21} & h_{22} & h_{23} & \ldots & \ldots & h_{2p} \\ 0 & h_{32} & h_{33} & \ldots & \ldots & h_{3p} \\ \ldots & 0 & h_{43} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & h_{(p-1)p} \\ 0 & 0 & \ldots & \ldots & h_{(p-1)p} & h_{pp} \end{pmatrix} \\
&= QH = \rho(\theta)
\end{aligned} \tag{3.18}
$$

where $\rho(\theta)$ is tridiagonal. With U^* being symmetric,

$$H = Q^{-1}U^*Q$$

is also symmetric and tridiagonal. Note that Q need not be orthogonal. Now, using the above relations, we have

$$
\begin{aligned}
\Phi^T(\theta)\Psi(\theta)\Phi(\theta) &= (U^*Q)^T(U^*Q) \\
&= \rho^T(\theta)\rho(\theta)
\end{aligned} \tag{3.19}
$$

To ensure that our approach also provides the basic advantage offered by Cholesky decompositions, we show that Δ_i is positive definite:

Theorem 2 *Let Δ_i be represented by the linear Cholesky decomposition (3.5). Then, Δ_i is positive definite.*

Proof 2 *By definition (3.5), $\Delta_i = \Phi^T\Phi + \rho^T\rho + \Psi^T\Psi$. Then, for any conformable nonzero vector x, we have*

$$
\begin{aligned}
x^T\Delta x &= x^T\left(\Phi^T\Phi + \rho^T\rho + \Psi^T\Psi\right)x \\
&= x^T\Phi^T\Phi x + x^T\rho^T\rho x + x^T\Psi^T\Psi x \\
&= (\Phi x)^T\Phi x + (\rho x)^T\rho x + (\Psi x)^T\Psi x \\
&= Y_1^TY_1 + Y_2^TY_2 + Y_3^TY_3 \\
&= \textstyle\sum_i y_{1i}^2 + \sum_j y_{2j}^2 + \sum_k y_{3k}^2 > 0
\end{aligned} \tag{3.20}
$$

where $Y_1 = \Phi x, Y_2 = \rho x, and\ Y_3 = \Psi x$. Therefore, $x^T\Delta x > 0$ and Δ is positive definite.

4. Conclusions

We propose a linear Cholesky decomposition for estimating correlation parameters between random errors ε_i and random effects γ_i in linear mixed models when the independence assumption between the two does not hold. Our approach can be

regarded as an extension of the Pinheiro and Bates (2000) result, in which their Cholesky decomposition of Σ_i has two factors, while our decomposition of Δ_i has three factors. Application of this decomposition to parameter estimation using the maximum likelihood and restricted/residual maximum likelihood procedures is the topic of our ongoing research.

Acknowledgements

We would like to thank the editorial office and two referees for their constructive feedback, which considerably improved the original version of the paper. The first author is grateful to Umaru Musa Yar'adua University Katsina for funding his research.

REFERENCES

CARLSON, D. H., DATTA, B. N., (1979). The Lyapunov matrix equation $SA + A^*S = S^*B^*BS$. Linear Algebra and Its Applications, 28, pp. 43–52.

CHEN, Z., DUNSON, D. B., (2003). Random effects selection in linear mixed models. Biometrics, 59, pp. 762–769.

DANIELS, M. J., ZHAO, Y. D., (2003). Modeling the random effects covariance matrix in longitudinal data. Statistics in Medicine, 22(10), pp. 1631–1647.

DE HOOG, F. R., ANDERSSEN, R. S., LUKAS, M. A., (2011). Differentiation of matrix functions using triangular factorization. Mathematics of Computation, 80(275), pp. 1585–1600.

GARCIA, T. P., KOHLI, P., POURAHMADI, M., (2012). Regressogram and mean-covariance models for incomplete longitudinal data, Journal of the American Statistical Association, 66(2), pp. 85–91.

GASKINS, J. T., DANIELS, M. J., (2013). A nonparametric prior for simultaneous covariance estimation, Biometrika, 100(1), pp. 125–138.

GELFAND, A. E., SAHU, S. K., CARLIN, B. P., (1995). Efficient parameterization for normal linear mixed models, Biometrika, 82(3), pp. 479–488.

HAN, E-J., LEE, K., (2016). Dynamic linear mixed models with ARMA covariance matrix, Communications for Statistical Applications and Methods, 23(6), pp. 575–585.

HOLAN, S., SPINKA, C., (2007). Maximum likelihood estimation for joint mean-covariance models for unbalanced repeated measures data, Statistics and Probability letters, 77, pp. 319–328.

LAIRD, N. M., WARE, J. H., (1982). Random-effects models for longitudinal data, Biometrics, 38(4), pp. 963–974.

LANGE, N., RYAN, L., (1989). Assessing normality of random effects models, Annals of Statistics, 17, pp. 624–642.

LEE, K., BAEK C., DANIELS, M. J., (2017). ARMA Cholesky factor models for the covariance matrix of linear models, Computational Statistics and Data Analysis, 115, pp. 267–280.

LI, E., POURAHMADI, M., (2013). An alternative REML estimation of covariance matrices in linear mixed models, Statistics and Probability Letters, 83, pp. 1071–1077.

PAN, J-X., MACKENZIE, G., (2003). On modeling mean-covariance structures in longitudinal studies, Biometrika, 90(1), pp. 239–244.

PAN, J-X., MACKENZIE, G., (2006). Regression models for covariance structures in longitudinal studies, Statistical Modeling, 6, pp. 43–57.

PINHEIRO, J. C., BATES, D. M., (1996). Unconstrained parameterizations for variance-covariance matrices, Statistics and Computing, 6, pp. 289–296.

PINHEIRO, J. C., BATES, D. M., (2000). Mixed-effects models in S and S-Plus, Springer-Verlag, New York, pp. 205–207.

POURAHMADI, M., (1999). Joint mean-covariance models with application to longitudinal data: Unconstrained parameterization, Biometrika, 86(3), pp. 677–690.

POURAHMADI, M., (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix, Biometrika, 87(2), pp. 425–435.

STEWART, G. W., (1998). Matrix algorithms volume I: Basic decompositions, SIAM Publications, Philadelphia, U.S.A., p. 183.

ZHANG, W., LENG, C., (2012). A moving average Cholesky factor model in covariance modelling for longitudinal data, Biometrika, 99(1), pp. 141–150.

ZIMMERMANN, D. L., VINCENT, N-A., HAMMOU, E., (1998). Computational aspects of likelihood-based estimation of first order antedependence models, Journal of Statistical Computation and Simulation, 60(1), pp. 67–84.