Maximum Likelihood Estimators in Growth Curve Model with Monotone Missing Data

Ayaka Yagi*, Takashi Seo* and Yasunori Fujikoshi**

* *Department of Applied Mathematics*
Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

***Department of Mathematics*
Hiroshima University
Higashi-Hiroshima, Hiroshima 739-0046, Japan

Abstract

This paper focuses on the maximum likelihood estimators (MLEs) of the mean parameter vector and the covariance matrix in a one-sample version of the growth curve model when the dataset has a monotone missing pattern. First, a closed form is obtained for the MLE of the mean parameter vector when the covariance matrix is known. Similarly, it is obtained for the MLE of the covariance matrix when the mean parameter vector is known. The distributions of these estimators and their basic properties are also given. Then, considering that these expressions give the likelihood or determining equations, we propose an algorithm that includes an iterative procedure to obtain the MLEs when all the parameters are unknown. Further, a conventional estimator for the mean parameter vector is also proposed. Finally, a numerical example is given to illustrate our estimation procedure.

Key Words and Phrases: Growth curve model, Maximum likelihood estimator, Monotone missing data.

1 Introduction

Suppose that a single variable y is measured at p time points (different conditions) $t_1, t_2, \ldots, t_p$ on n subjects chosen at random from a group. We denote the variable y at time point t_j by y_j, and let $\boldsymbol{y} = (y_1, y_2, \ldots, y_p)'$. Let the observations $y_{i1}, y_{i2}, \ldots, y_{ip}$ of the ith subject be denoted by

$$\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})', \quad i = 1, 2, \ldots, n.$$

Then, in a growth curve model, it is assumed that for $i = 1, 2, \ldots, n$,

$$\mathrm{E}[\boldsymbol{y}_i] = \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}, \quad \mathrm{Var}[\boldsymbol{y}_i] = \boldsymbol{\Sigma}, \tag{1}$$

where $\boldsymbol{X}$ is a given $p \times q$ matrix with rank q, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$ is an unknown parameter vector, and $\boldsymbol{\Sigma}$ is an unknown positive definite. Further, it is assumed that $\boldsymbol{y}_i, i = 1, 2, \ldots, n$ are independent and normally distributed. The matrix $\boldsymbol{X}$ is called a within-design matrix. The model given by (1) is a one-sample version of the growth curve model introduced by Potthoff and Roy (1964).

We consider the case where the missing data occur at random. Kleinbaum (1973) gave some inference methods for the growth curve model. Srivastava (1985) gave the likelihood equations for some multivariate models including the growth curve model. Liski (1985) applied the EM-algorithm to the estimation problem. Kanda (1994) considered the case where the covariance matrix has some special structures. In this paper, we consider the case where the missing data are of the monotone type, that is, if y_{ij} is missing, then all variables $y_{i\ell}, \ell > j$ are also missing. Here, j may depend on the subject i. If there are k types of monotone data, then the data are called k-step monotone missing data. Such monotone missing data often appear in longitudinal studies. There have been a considerable number of works on monotone missing data in one and several multivariate normal populations. For these works, see Anderson and Olkin (1985), Jinadasa and Tracy (1992), Kanda and Fujikoshi (1998), Yagi and Seo (2017), etc. However, it seems that more advanced studies have not been conducted on the growth curve model of monotone missing data, although there are considerable number of works on the general type of missing data, as mentioned above.

This paper deals with MLEs in a one-sample growth curve model (1). Our aim is to derive more advanced results for the MLEs by assuming monotone-type missing data. The remainder of this paper is organized as follows. In Section 2, a closed form is obtained for the MLE of the mean parameter vector when the covariance matrix is known. In addition, its distribution and basic properties are given. In Section 3, a closed form is obtained for the MLE of the covariance matrix when the mean parameter vector is known, and its distribution is studied. In Section 4, we show that these results give the likelihood or determining equations to obtain their MLEs when all the parameters are unknown. Further, we propose an algorithm with an iterative procedure to obtain the MLEs. A conventional estimator for the mean parameter is also proposed. A numerical example is given to illustrate our estimation procedure in Section 5. Finally, we state our conclusion in Section 6. The proofs of Theorems 1 and 2 are given in the Appendix, focusing on a three-step monotone case.

2 MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known

In this section, we derive a closed form expression of the MLE of the mean parameter vector with a known covariance matrix for a growth curve model when the dataset has a monotone pattern of missing observations. Its distribution and basic properties are also given.

2.1 Two-step monotone missing data

For simplicity, we first consider the case $k = 2$. Suppose that we have n_1 observation vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_1}$ all of whose components have been observed. It is assumed that $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_1}$ are independently distributed as a p-dimensional normal distribution with the mean vector $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$ as in (1). Further, suppose that we have n_2 observation vectors $\boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \ldots, \boldsymbol{y}_{2n_2}$ for the first p_1 components. It is assumed that $\boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \ldots, \boldsymbol{y}_{2n_2}$ are independently distributed as a p_1-dimensional normal distribution with the mean vector $\boldsymbol{\mu}_1 = \boldsymbol{X}_1\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}_{11}$, where

$$\underset{p\times 1}{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \end{matrix}, \quad \underset{p\times q}{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \end{matrix}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Let us write

$$(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_1})' = (\boldsymbol{Y}_{11}\boldsymbol{Y}_{12}), \quad (\boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \ldots, \boldsymbol{y}_{2n_2})' = \boldsymbol{Y}_{21},$$

where $\boldsymbol{Y}_{11} : n_1 \times p_1$, $\boldsymbol{Y}_{12} : n_1 \times p_2$, $\boldsymbol{Y}_{21} : n_2 \times p_1$, and $p = p_1 + p_2$. Then, the two-step monotone missing data are expressed as

$$\boldsymbol{Y} = \left(\begin{array}{c|c} \overbrace{\boldsymbol{Y}_{11}}^{p_1} & \overbrace{\boldsymbol{Y}_{12}}^{p_2} \\ \hline \boldsymbol{Y}_{21} & * \end{array} \right) \begin{array}{l} \}n_1 \\ \}n_2 \end{array}, \tag{2}$$

where "$*$" indicates a missing part. Here, note that each row of $\boldsymbol{Y}$ is independently normal with

$$\mathrm{E}[(\boldsymbol{Y}_{11}\ \boldsymbol{Y}_{12})] = \boldsymbol{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}', \ \mathrm{E}[\boldsymbol{Y}_{21}] = \boldsymbol{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_1', \tag{3}$$

where $\boldsymbol{1}_{n_i}$ is an $n_i \times 1$ vector of 1s.

Let $L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ be the likelihood of $\boldsymbol{Y}$ in (2). Then, we have

$$\begin{aligned}
&- 2\log L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
&= (N_2 p_1 + n_1 p_2)\log(2\pi) + N_2 \log|\boldsymbol{\Sigma}_{11}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(12)1} - \boldsymbol{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(12)1} - \boldsymbol{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1') \\
&\quad + n_1 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12} - \boldsymbol{1}_{n_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})' \\
&\quad \times (\boldsymbol{Y}_{12} - \boldsymbol{1}_{n_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}),
\end{aligned}$$

where

$$N_2 = n_1 + n_2, \quad \boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \quad \underset{N_2 \times p_1}{\boldsymbol{Y}_{(12)1}} = \begin{pmatrix} \boldsymbol{Y}_{11} \\ \boldsymbol{Y}_{21} \end{pmatrix} \begin{array}{l} \}n_1 \\ \}n_2 \end{array},$$

and

$$\underset{q \times p_2}{\widetilde{\boldsymbol{X}}_2'} = \boldsymbol{X}_2' - \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.$$

Consider the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known. Note that there is a one-to-one correspondence between $\boldsymbol{\Sigma}$ and $\{\boldsymbol{\Sigma}_{11}, \ \mathcal{B}_{12}, \ \boldsymbol{\Sigma}_{22\cdot 1}\}$, where $\mathcal{B}_{12} = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. We can write

$$\begin{aligned}
&g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) \equiv -2\log L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
&= (N_2 p_1 + n_1 p_2)\log(2\pi) + N_2 \log|\boldsymbol{\Sigma}_{11}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(12)1} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(12)1} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1') \\
&\quad + n_1 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12} - \boldsymbol{Y}_{11}\mathcal{B}_{12} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2')'(\boldsymbol{Y}_{12} - \boldsymbol{Y}_{11}\mathcal{B}_{12} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2') \\
&= -2(\boldsymbol{a}_1 + \boldsymbol{a}_2)'\boldsymbol{\theta} + \boldsymbol{\theta}'(N_2\boldsymbol{A}_1 + n_1\boldsymbol{A}_2)\boldsymbol{\theta} + (\text{terms without } \boldsymbol{\theta}),
\end{aligned}$$

where

$$\begin{aligned}
&\boldsymbol{a}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_{(12)1}'\mathbf{1}_{N_2}, \quad \boldsymbol{a}_2 = \widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12} - \boldsymbol{Y}_{11}\mathcal{B}_{12})'\mathbf{1}_{n_1}, \\
&\boldsymbol{A}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1, \quad \text{and} \quad \boldsymbol{A}_2 = \widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2.
\end{aligned}$$

On solving $(\partial/\partial\boldsymbol{\theta})g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) = \mathbf{0}$, we obtain the MLE of $\boldsymbol{\theta}$ as

$$\widehat{\boldsymbol{\theta}} = (N_2\boldsymbol{A}_1 + n_1\boldsymbol{A}_2)^{-1}(\boldsymbol{a}_1 + \boldsymbol{a}_2). \tag{4}$$

Then, it holds that

$$\mathrm{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}, \quad \mathrm{Var}[\widehat{\boldsymbol{\theta}}] = (N_2\boldsymbol{A}_1 + n_1\boldsymbol{A}_2)^{-1}. \tag{5}$$

The mean of $\widehat{\boldsymbol{\theta}}$ is obtained using

$$\mathrm{E}_{\boldsymbol{Y}_{11}}\Big[\mathrm{E}_{\boldsymbol{Y}_{12}|\boldsymbol{Y}_{11}}[(\boldsymbol{Y}_{12} - \boldsymbol{Y}_{11}\mathcal{B}_{12})'\mathbf{1}_{n_1}]\Big] = n_1\widetilde{\boldsymbol{X}}_2\boldsymbol{\theta}.$$

The covariance matrix of $\widehat{\boldsymbol{\theta}}$ is obtained using

$$\mathrm{E}_{\boldsymbol{Z}_{11}}[\mathrm{E}_{\boldsymbol{Z}_{12}|\boldsymbol{Z}_{11}}[(\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11}\mathcal{B}_{12})'\mathbf{1}_{n_1}\cdot\mathbf{1}_{n_1}'(\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11}\mathcal{B}_{12})]] = n_1\boldsymbol{\Sigma}_{22\cdot 1},$$

where

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & * \end{pmatrix} &= \begin{pmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & * \end{pmatrix} - \begin{pmatrix} \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_1' & \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_2' \\ \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_1' & * \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{Y}_{11} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_1' & \boldsymbol{Y}_{12} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_2' \\ \boldsymbol{Y}_{21} - \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_1' & * \end{pmatrix}
\end{aligned} \tag{6}$$

and $\boldsymbol{Z}_{(12)1} = \boldsymbol{Y}_{(12)1} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1'$. Further, the distribution of $\boldsymbol{\theta}$ is normal. These results are obtained by considering that $\boldsymbol{a}_1 \sim N_q(N_2\boldsymbol{A}_1\boldsymbol{\theta}, N_2\boldsymbol{A}_1)$, $\boldsymbol{a}_2 \sim N_q(n_1\boldsymbol{A}_2\boldsymbol{\theta}, n_1\boldsymbol{A}_2)$, and $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are independent. The normality of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ follows from the fact that these are bilinear forms of $\boldsymbol{Y}$. The independence of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ follows from the fact that the conditional distribution of $\boldsymbol{a}_2$ for a given $\boldsymbol{Y}_{11}$ or $\boldsymbol{Y}_{(12)1}$ does not depend on $\boldsymbol{a}_1$.

2.2 k-step monotone missing data ($k \geq 2$)

We consider a general monotone case. Let

$$\boldsymbol{Y} = \begin{pmatrix} \overbrace{\boldsymbol{Y}_{11}}^{p_1} & \overbrace{\boldsymbol{Y}_{12}}^{p_2} & \cdots & \overbrace{\cdots}^{p_{k-i+1}} & \cdots & \overbrace{\boldsymbol{Y}_{1,k-1}}^{p_{k-1}} & \overbrace{\boldsymbol{Y}_{1k}}^{p_k} \\ \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22} & \cdots & \cdots & \cdots & \boldsymbol{Y}_{2,k-1} & * \\ \vdots & \vdots & & \vdots & & & \vdots \\ \boldsymbol{Y}_{i1} & \boldsymbol{Y}_{i2} & \cdots & \boldsymbol{Y}_{i,k-i+1} & * & \cdots & * \\ \vdots & \vdots & & & & & \vdots \\ \boldsymbol{Y}_{k-1,1} & \boldsymbol{Y}_{k-1,2} & * & \cdots & \cdots & \cdots & * \\ \boldsymbol{Y}_{k1} & * & \cdots & \cdots & \cdots & \cdots & * \end{pmatrix} \begin{matrix} \}n_1 \\ \}n_2 \\ \\ \}n_i \\ \\ \}n_{k-1} \\ \}n_k \end{matrix}$$

and let

$$\underset{n_i \times p_{(12\ldots,k-i+1)}}{\boldsymbol{Y}_{i(12\ldots,k-i+1)}} = (\boldsymbol{Y}_{i1}\ \boldsymbol{Y}_{i2} \cdots \boldsymbol{Y}_{i,k-i+1}),\ i = 1, 2, \ldots, k.$$

The rows of $\boldsymbol{Y}_{i(12\ldots,k-i+1)}$ $(i = 1, 2, \ldots, k)$ are mutually independent and

$$\mathrm{vec}(\boldsymbol{Y}'_{i(12\ldots,k-i+1)}) \sim N_{p_{(12\ldots,k-i+1)}n_i}(\mathrm{vec}(\boldsymbol{\mu}_{(12\ldots,k-i+1)}\mathbf{1}'_{n_i}), \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}_{(12\ldots,k-i+1)(12\ldots,k-i+1)}),$$
$$i = 1, 2, \ldots, k,$$

where $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector of 1s,

$$\underset{p_{(12\ldots,k-i+1)} \times 1}{\boldsymbol{\mu}_{(12\ldots,k-i+1)}} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_{k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix},$$

$$\underset{p_{(12\ldots,k-i+1)} \times p_{(12\ldots,k-i+1)}}{\boldsymbol{\Sigma}_{(12\ldots,k-i+1)(12\ldots,k-i+1)}} = \begin{pmatrix} \overbrace{\boldsymbol{\Sigma}_{11}}^{p_1} & \overbrace{\boldsymbol{\Sigma}_{12}}^{p_2} & \cdots & \overbrace{\boldsymbol{\Sigma}_{1,k-i+1}}^{p_{k-i+1}} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2,k-i+1} \\ \vdots & & & \vdots \\ \boldsymbol{\Sigma}_{k-i+1,1} & \boldsymbol{\Sigma}_{k-i+1,2} & \cdots & \boldsymbol{\Sigma}_{k-i+1,k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix}.$$

In particular, for $i = 1$, we define $\boldsymbol{\mu}_{(12\ldots k)} = \boldsymbol{\mu}$, $\boldsymbol{\Sigma}_{(12\ldots k)(12\ldots k)} = \boldsymbol{\Sigma}$. Then, the growth curve model with k-step monotone missing data is expressed as

$$\mathrm{E}[\boldsymbol{Y}_{i(12\ldots,k-i+1)}] = \underset{n_i \times 1}{\mathbf{1}_{n_i}} \underset{1 \times q}{\boldsymbol{\theta}'} \underset{q \times p_{(12\ldots,k-i+1)}}{\boldsymbol{X}'_{(12\ldots,k-i+1)}},\ i = 1, 2, \ldots, k, \tag{7}$$

where

$$\underset{p_{(12\ldots,k-i+1)} \times q}{\boldsymbol{X}_{(12\ldots,k-i+1)}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix},\quad p_{(12\ldots,k-i+1)} = \sum_{j=1}^{k-i+1} p_j,\quad i = 1, 2, \ldots, k.$$

We note that $\boldsymbol{\mu}_{(12\ldots,k-i+1)} = \boldsymbol{X}_{(12\ldots,k-i+1)}\boldsymbol{\theta}, \quad i = 1, 2, \ldots, k$. The joint probability density function of $\boldsymbol{Y}$, i.e.,
$\left\{\boldsymbol{Y}_{1(12\ldots k)}, \boldsymbol{Y}_{2(12\ldots,k-1)}, \ldots, \boldsymbol{Y}_{i(12\ldots,k-i+1)}, \ldots, \boldsymbol{Y}_{k1}\right\}$ can be written as

$$\begin{aligned}
& f(\boldsymbol{Y}_{1(12\ldots k)}) \times f(\boldsymbol{Y}_{2(12\ldots k-1)}) \times \cdots \times f(\boldsymbol{Y}_{i(12\ldots,k-i+1)}) \times \cdots \times f(\boldsymbol{Y}_{k1}) \\
= \;& f(\boldsymbol{Y}_{11}) f(\boldsymbol{Y}_{12}|\boldsymbol{Y}_{11}) \times \cdots \times f(\boldsymbol{Y}_{1,k-i+1}|\boldsymbol{Y}_{1(12\ldots,k-i)}) \times \cdots \times f(\boldsymbol{Y}_{1k}|\boldsymbol{Y}_{1(12\ldots,k-1)}) \\
& \times f(\boldsymbol{Y}_{21}) f(\boldsymbol{Y}_{22}|\boldsymbol{Y}_{21}) \times \cdots \times f(\boldsymbol{Y}_{2,k-i+1}|\boldsymbol{Y}_{2(12\ldots,k-i)}) \times \cdots \times f(\boldsymbol{Y}_{2,k-1}|\boldsymbol{Y}_{2(12\ldots,k-2)}) \\
& \vdots \\
& \times f(\boldsymbol{Y}_{i1}) f(\boldsymbol{Y}_{i2}|\boldsymbol{Y}_{i1}) \times \cdots \times f(\boldsymbol{Y}_{ij}|\boldsymbol{Y}_{i(12\ldots,j-1)}) \times \cdots \times f(\boldsymbol{Y}_{i,k-i+1}|\boldsymbol{Y}_{i(12\ldots,k-i)}) \\
& \vdots \\
& \times f(\boldsymbol{Y}_{k1}),
\end{aligned}$$

where

$$f(\boldsymbol{Y}_{i1}) = (2\pi)^{-\frac{1}{2}n_i p_1} |\boldsymbol{\Sigma}_{11}|^{-\frac{1}{2}n_i} \operatorname{etr}\left\{-\frac{1}{2}(\boldsymbol{Y}_{i1} - \mathbf{1}_{n_i}\boldsymbol{\theta}'\boldsymbol{X}_1')\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{i1} - \mathbf{1}_{n_i}\boldsymbol{\theta}'\boldsymbol{X}_1')'\right\}, \qquad i = 1, 2, \ldots, k,$$

and the conditional density of $\boldsymbol{Y}_{ij}$ for a given $\boldsymbol{Y}_{i(12\cdots,j-1)}$ can be written as

$$\begin{aligned}
& f(\boldsymbol{Y}_{ij}|\boldsymbol{Y}_{i(12\ldots,j-1)}) \\
& = (2\pi)^{-\frac{1}{2}n_i p_j} |\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}|^{-\frac{1}{2}n_i} \\
& \quad \times \operatorname{etr}\Big[-\frac{1}{2}\left\{\boldsymbol{Y}_{ij} - \mathbf{1}_{n_i}\boldsymbol{\theta}'(\boldsymbol{X}_j' - \boldsymbol{X}_{(12\ldots,j-1)}'\mathcal{B}_{(12\ldots,j-1)j}) - \boldsymbol{Y}_{i(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\right\} \\
& \qquad \times \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\left\{\boldsymbol{Y}_{ij} - \mathbf{1}_{n_i}\boldsymbol{\theta}'(\boldsymbol{X}_j' - \boldsymbol{X}_{(12\ldots,j-1)}'\mathcal{B}_{(12\ldots,j-1)j}) - \boldsymbol{Y}_{i(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\right\}'\Big], \\
& \qquad\qquad i = 1, 2, \ldots, k; j = 2, 3, \ldots, k-i+1,
\end{aligned}$$

where

$$\begin{aligned}
& \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1} = \boldsymbol{\Sigma}_{jj} - \boldsymbol{\Sigma}_{(12\ldots,j-1)j}\boldsymbol{\Sigma}_{(12\ldots,j-1)(12\ldots,j-1)}^{-1}\boldsymbol{\Sigma}_{(12\ldots,j-1)j}, \\
& \mathcal{B}_{(12\ldots,j-1)j} = \boldsymbol{\Sigma}_{(12\ldots,j-1)(12\ldots,j-1)}^{-1}\boldsymbol{\Sigma}_{(12\ldots,j-1)j}, \quad \underset{p_{(12\ldots,j-1)}\times p_j}{\boldsymbol{\Sigma}_{(12\ldots,j-1)j}} = \begin{pmatrix} \boldsymbol{\Sigma}_{1j} \\ \boldsymbol{\Sigma}_{2j} \\ \vdots \\ \boldsymbol{\Sigma}_{j-1,j} \end{pmatrix}.
\end{aligned}$$

Note that there is a one-to-one correspondence between $\boldsymbol{\Sigma}$ and $\{\boldsymbol{\Sigma}_{11}, \mathcal{B}_{(12\ldots,j-1)j}, \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}, j = 2, 3, \ldots, k\}$. Using the above decomposition of the density of $\boldsymbol{Y}$, the likelihood $L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ can be expressed as

$$
\begin{aligned}
& g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{(12\ldots,j-1)j}, \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1},\ j = 2, 3, \ldots, k) \equiv -2 \log L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
&= \sum_{j=1}^{k} N_{k-j+1} p_j \log(2\pi) + N_k \log |\boldsymbol{\Sigma}_{11}| + \sum_{j=2}^{k} N_{k-j+1} \log |\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}| \\
&\quad + \operatorname{tr} \boldsymbol{\Sigma}_{11}^{-1} (\boldsymbol{Y}_{(12\ldots k)1} - \mathbf{1}_{N_k} \boldsymbol{\theta}' \boldsymbol{X}_1')' (\boldsymbol{Y}_{(12\ldots k)1} - \mathbf{1}_{N_k} \boldsymbol{\theta}' \boldsymbol{X}_1') \\
&\quad + \sum_{j=2}^{k} \operatorname{tr} \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1} \\
&\qquad \times \left\{ \boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)} \mathcal{B}_{(12\ldots,j-1)j} - \mathbf{1}_{N_{k-j+1}} \boldsymbol{\theta}' \widetilde{\boldsymbol{X}}_j' \right\}' \\
&\qquad \times \left\{ \boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)} \mathcal{B}_{(12\ldots,j-1)j} - \mathbf{1}_{N_{k-j+1}} \boldsymbol{\theta}' \widetilde{\boldsymbol{X}}_j' \right\},
\end{aligned} \tag{8}
$$

where

$$
\widetilde{\boldsymbol{X}}_j' = \boldsymbol{X}_j' - \boldsymbol{X}_{(12\ldots,j-1)}' \mathcal{B}_{(12\ldots,j-1)j}, \quad j = 2, 3, \ldots, k, \quad \text{and} \quad N_i = \sum_{j=1}^{i} n_j, \quad i = 1, 2, \ldots, k.
$$

Using the above notations, we get the following results for the MLE of $\boldsymbol{\theta}$ and its distribution when $\boldsymbol{\Sigma}$ is given:

<u>Theorem 1</u> *Suppose that we have the growth curve model (7) with a k-step monotone pattern of missing observations. Then, the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known is given by*

$$
\widehat{\boldsymbol{\theta}} = \boldsymbol{M}^{-1} \sum_{j=1}^{k} \boldsymbol{a}_j, \tag{9}
$$

where

$$
\boldsymbol{M} = \sum_{j=1}^{k} N_{k-j+1} \boldsymbol{A}_j, \quad \boldsymbol{a}_1 = \boldsymbol{X}_1' \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{Y}_{(12\ldots k)1}' \mathbf{1}_{N_k}, \quad \boldsymbol{A}_1 = \boldsymbol{X}_1' \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{X}_1,
$$

and for $j = 2, \ldots, k$,

$$
\begin{aligned}
\boldsymbol{a}_j &= \widetilde{\boldsymbol{X}}_j' \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1} \left(\boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{Y}_{(12\ldots,k-j+1)(1\ldots,j-1)} \mathcal{B}_{(12\ldots,j-1)j} \right)' \mathbf{1}_{N_{k-j+1}}, \\
\boldsymbol{A}_j &= \widetilde{\boldsymbol{X}}_j' \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1} \widetilde{\boldsymbol{X}}_j.
\end{aligned}
$$

Further, it holds that $\widehat{\boldsymbol{\theta}}$ *is normal with*

$$\mathrm{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta} \quad \text{and} \quad \mathrm{Var}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{M}^{-1}.$$

Note that we use the suffix $(1) = 1$. For example, $\mathcal{B}_{(1)2} = \mathcal{B}_{12}$, $\boldsymbol{X}_{(1)} = \boldsymbol{X}_1$, and $\boldsymbol{Z}_{(1)2} = \boldsymbol{Z}_{12}$. The proof of Theorem 1, which is given in Appendix, is based on the following Lemma.

Lemma 1 *Let* $\boldsymbol{a}_j$, $j = 1, 2, \ldots, k$ *be the random vector defined in* (9). *Then,*

(1) $\boldsymbol{a}_j \sim N_q(N_{k-j+1}\boldsymbol{A}_j\boldsymbol{\theta}, N_{k-j+1}\boldsymbol{A}_j)$, $j = 1, 2, \ldots, k$,

(2) $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ *are independent,*

where

$$\boldsymbol{A}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1,\ \boldsymbol{A}_j = \widetilde{\boldsymbol{X}}_j'\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\widetilde{\boldsymbol{X}}_j,\quad \widetilde{\boldsymbol{X}}_j = \boldsymbol{X}_j - \mathcal{B}_{(12\ldots,j-1)j}'\boldsymbol{X}_{(12\ldots,j-1)},$$
$$j = 2, 3, \ldots, k.$$

The proof of Lemma 1 is given in Appendix 1.

3 MLE of $\boldsymbol{\Sigma}$ when $\boldsymbol{\theta}$ is known

In this section, we consider the MLE of $\boldsymbol{\Sigma}$ when $\boldsymbol{\theta}$ is known, with monotone missing data in the growth curve model. The notation in Section 2 is used.

3.1 Two-step monotone missing data

Here, we consider the minimization of $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1})$ with respect to $\boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}$, and $\boldsymbol{\Sigma}_{22\cdot 1}$. Note that $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1})$ is expressed in terms of $\boldsymbol{Z}$ in (6) as

$$\begin{aligned} g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) =& (N_2 p_1 + n_1 p_2)\log(2\pi) + N_2 \log|\boldsymbol{\Sigma}_{11}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Z}_{(12)1}'\boldsymbol{Z}_{(12)1} \\ & + n_1 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + \mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11}\mathcal{B}_{12})'(\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11}\mathcal{B}_{12}). \end{aligned}$$

The derivative of $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1})$ with respect to $\boldsymbol{\Sigma}_{11}$ is

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_{11}} g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) = N_2 \boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{Z}'_{(12)1} \boldsymbol{Z}_{(12)1} \boldsymbol{\Sigma}_{11}^{-1}.$$

Solving $(\partial/\partial \boldsymbol{\Sigma}_{11}) g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) = \boldsymbol{O}$, we obtain the MLE of $\boldsymbol{\Sigma}_{11}$ as

$$\widehat{\boldsymbol{\Sigma}}_{11} = \frac{1}{N_2} \boldsymbol{Z}'_{(12)1} \boldsymbol{Z}_{(12)1}.$$

As for $\mathcal{B}_{12}$, note that $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1})$ is expressed as

$$g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}) = \operatorname{tr} \boldsymbol{\Sigma}_{22\cdot 1}^{-1} (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \mathcal{B}_{12})' (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \mathcal{B}_{12}) + (\text{terms without } \mathcal{B}_{12}).$$

Then, from [A.2.11] of Fujikoshi et al. (2010), we have

$$\min_{\mathcal{B}_{12}} \operatorname{tr} \boldsymbol{\Sigma}_{22\cdot 1}^{-1} (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \mathcal{B}_{12})' (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \mathcal{B}_{12}) = \operatorname{tr} \boldsymbol{\Sigma}_{22\cdot 1}^{-1} (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \widehat{\mathcal{B}}_{12})' (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \widehat{\mathcal{B}}_{12}),$$

where

$$\widehat{\mathcal{B}}_{12} = (\boldsymbol{Z}'_{11} \boldsymbol{Z}_{11})^{-1} \boldsymbol{Z}'_{11} \boldsymbol{Z}_{12}.$$

Therefore, $\widehat{\mathcal{B}}_{12}$ is the MLE of $\mathcal{B}_{12}$. Finally, in the same way as that in the derivation of $\widehat{\boldsymbol{\Sigma}}_{11}$, the MLE of $\boldsymbol{\Sigma}_{22\cdot 1}$ is given as

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_{22\cdot 1} =& \frac{1}{n_1} (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \widehat{\mathcal{B}}_{12})' (\boldsymbol{Z}_{12} - \boldsymbol{Z}_{11} \widehat{\mathcal{B}}_{12}) \\ =& \frac{1}{n_1} \boldsymbol{Z}'_{12} (\boldsymbol{I}_{n_1} - \boldsymbol{P}_{\boldsymbol{Z}_{11}}) \boldsymbol{Z}_{12}, \end{aligned}$$

where $P_{\boldsymbol{Z}_{11}} = \boldsymbol{Z}_{11} (\boldsymbol{Z}'_{11} \boldsymbol{Z}_{11})^{-1} \boldsymbol{Z}'_{11}$. We can easily see that $N_2 \widehat{\boldsymbol{\Sigma}}_{11}$ and $n_1 \widehat{\boldsymbol{\Sigma}}_{22\cdot 1}$ are independently distributed as Wishart distributions $W_{p_1}(N_2 - 1, \boldsymbol{\Sigma}_{11})$ and $W_{p_2}(n_1 - p_1 - 1, \boldsymbol{\Sigma}_{22\cdot 1})$, respectively.

3.2 k-step monotone missing data

Let $L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ be the likelihood of $\boldsymbol{Y}$ with k-step monotone missing. Then, $-2 \log L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is expressed as (8). We use the following notation:

$$\begin{aligned} \boldsymbol{Z}_{(12\ldots k)1} =& \boldsymbol{Y}_{(12\ldots k)1} - \mathbf{1}_{N_k} \boldsymbol{\theta}' \boldsymbol{X}'_1, \\ \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)} =& \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)} - \mathbf{1}_{N_{k-j+1}} \boldsymbol{\theta}' \boldsymbol{X}'_{(12\ldots,j-1)}, \\ \boldsymbol{Z}_{(12\ldots,k-j+1)j} =& \boldsymbol{Y}_{(12\ldots,k-j+1)j} - \mathbf{1}_{N_{k-j+1}} \boldsymbol{\theta}' \boldsymbol{X}'_j, \ \ j = 2, 3, \ldots, k. \end{aligned} \tag{10}$$

Using the expression given by (8), we can obtain the MLE of $\mathbf{\Sigma}$ when $\boldsymbol{\theta}$ is known by using the following theorem.

Theorem 2 *Suppose that we have a k-step monotone pattern of missing observations. Then, for the growth curve model given by (7) with a known parameter $\boldsymbol{\theta}$, the MLEs of $\mathbf{\Sigma}_{11}, \mathcal{B}_{(12\ldots,j-1)j}, \mathbf{\Sigma}_{jj\cdot 12\ldots,j-1},\ j = 2, 3, \ldots, k$ are given by*

$$\widehat{\mathbf{\Sigma}}_{11} = \frac{1}{N_k} \boldsymbol{Z}'_{(12\ldots k)1} \boldsymbol{Z}_{(12\ldots k)1}$$

and for $j = 2, \ldots, k$,

$$\begin{aligned}
\widehat{\mathcal{B}}_{(12\ldots,j-1)j} =& (\boldsymbol{Z}'_{(12\ldots,k-j+1)(12\ldots,j-1)} \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)})^{-1} \\
& \times \boldsymbol{Z}'_{(12\ldots,k-j+1)(12\ldots,j-1)} \boldsymbol{Z}_{(12\ldots,k-j+1)j}, \\
\widehat{\mathbf{\Sigma}}_{jj\cdot 12\ldots,j-1} =& \frac{1}{N_{k-j+1}} \Big(\boldsymbol{Z}_{(12\ldots,k-j+1)j} - \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)} \widehat{\mathcal{B}}_{(12\ldots,j-1)j} \Big)' \\
& \times \Big(\boldsymbol{Z}_{(12\ldots,k-j+1)j} - \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)} \widehat{\mathcal{B}}_{(12\ldots,j-1)j} \Big) \\
=& \frac{1}{N_{k-j+1}} \boldsymbol{Z}'_{(12\ldots,k-j+1)j} \left(I_{N_j} - P_{\boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)}} \right) \boldsymbol{Z}_{(12\ldots,k-j+1)j},
\end{aligned}$$

where

$$\begin{aligned}
P_{\boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)}} =& \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)} (\boldsymbol{Z}'_{(12\ldots,k-j+1)(12\ldots,j-1)} \boldsymbol{Z}_{(12\ldots,k-j+1)(12\ldots,j-1)})^{-1} \\
& \times \boldsymbol{Z}'_{(12\ldots,k-j+1)(12\ldots,j-1)},\ j = 2, 3, \ldots, k.
\end{aligned}$$

Further,

$$\begin{aligned}
& N_k \widehat{\mathbf{\Sigma}}_{11} \sim W_{p_1}(N_k - 1, \mathbf{\Sigma}_{11}), \\
& N_{k-j+1} \widehat{\mathbf{\Sigma}}_{jj\cdot 12\ldots,j-1} \sim W_{p_j}(N_{k-j+1} - p_{(12\cdots,j-1)} - 1, \mathbf{\Sigma}_{jj\cdot 12\ldots,j-1}), \quad j = 2, 3, \ldots, k,
\end{aligned}$$

and they are independent.

The proof of Theorem 2 is given in Appendix 2.

4 MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$

In this section, we consider the MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ when all the parameters are unknown. In Theorems 1 and 2, we obtain closed forms $\boldsymbol{h}_1(\boldsymbol{\Sigma})$ and $\boldsymbol{h}_2(\boldsymbol{\theta})$ such that (i) $\widehat{\boldsymbol{\theta}} = \boldsymbol{h}_1(\boldsymbol{\Sigma})$ is the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known, and (ii) $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{h}_2(\boldsymbol{\theta})$ is the MLE of $\boldsymbol{\Sigma}$ when $\boldsymbol{\theta}$ is known. From our derivation, the likelihood equations of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are equivalent to the simultaneous equations given by

$$\boldsymbol{\theta} = \boldsymbol{h}_1(\boldsymbol{\Sigma}) \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{h}_2(\boldsymbol{\theta}). \tag{11}$$

Therefore, the solutions of the simultaneous equations given by (11) for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are the MLEs. It seems difficult to obtain an explicit solution of (11). Therefore, we consider numerical solutions based on the following iterative method.

(0) As a starting matrix $\widehat{\boldsymbol{\Sigma}}^{(0)}$ for $\boldsymbol{\Sigma}$, we use the MLE of $\boldsymbol{\Sigma}$ when $\boldsymbol{\mu}$ is an unknown parameter vector, which is obtained from Jinadasa and Tracy (1992) and Kanda and Fujikoshi (1998).

(1) Using $\widehat{\boldsymbol{\Sigma}}^{(0)}$, compute $\boldsymbol{\Sigma}_{11}^{(0)}$, $\mathcal{B}_{(1\dots,j-1)j}^{(0)}$, $\boldsymbol{\Sigma}_{jj\cdot 1\dots,j-1}^{(0)}$, $j = 2,\dots,k$, and then compute $\widehat{\boldsymbol{\theta}}^{(1)} = \boldsymbol{h}_1(\widehat{\boldsymbol{\Sigma}}^{(0)})$ and $\widehat{\boldsymbol{\Sigma}}^{(1)} = \boldsymbol{h}_2(\widehat{\boldsymbol{\theta}}^{(1)})$. Set $i = 1$.

(2) For $j = 2,\dots,k$, compute $\boldsymbol{\Sigma}_{11}^{(i)}$, $\mathcal{B}_{(1\dots,j-1)j}^{(i)}$, and $\boldsymbol{\Sigma}_{jj\cdot 1\dots,j-1}^{(i)}$. Then, compute $\widehat{\boldsymbol{\theta}}^{(i+1)} = \boldsymbol{h}_1(\boldsymbol{\Sigma}^{(i)})$ and $\widehat{\boldsymbol{\Sigma}}^{(i+1)} = \boldsymbol{h}_2(\boldsymbol{\theta}^{(i+1)})$.

(3) Let $\boldsymbol{d}^{(i)} = \widehat{\boldsymbol{\theta}}^{(i+1)} - \widehat{\boldsymbol{\theta}}^{(i)}$ and $\boldsymbol{D}^{(i)} = \widehat{\boldsymbol{\Sigma}}^{(i+1)} - \widehat{\boldsymbol{\Sigma}}^{(i)}$. If $\boldsymbol{d}^{(i)\prime}\boldsymbol{d}^{(i)} < 10^{-6}$ and $\operatorname{tr} \boldsymbol{D}^{(i)}\boldsymbol{D}^{(i)} < 10^{-6}$, then we regard $\widehat{\boldsymbol{\theta}}^{(i+1)}$ and $\widehat{\boldsymbol{\Sigma}}^{(i+1)}$ as the solutions, and stop.

(4) Otherwise repeat steps (2) through (3).

As an alternative starting matrix $\boldsymbol{\Sigma}$, we may use the MLE based on the complete dataset, $\boldsymbol{Y}_{1(12\cdots k)}$, if $n - n_1$ is small. Then, $\boldsymbol{Y}_{1(12\cdots k)}$ has a growth curve model

$$\mathrm{E}[\boldsymbol{Y}_{1(12\cdots k)}] = \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}'. \tag{12}$$

The MLE of $\boldsymbol{\Sigma}$ based on (12) is given as follows. Consider the transformation from $\boldsymbol{Y}$ to $(\boldsymbol{U}\ \boldsymbol{V})$:

$$(\boldsymbol{U}\ \boldsymbol{V}) = \boldsymbol{Y}\boldsymbol{G} = \boldsymbol{Y}(\boldsymbol{G}_1\ \boldsymbol{G}_2),$$

where $\boldsymbol{G}_1 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\boldsymbol{G}_2 = \widetilde{\boldsymbol{X}}$; $\widetilde{\boldsymbol{X}}$ is a $p \times (p-q)$ matrix such that $\widetilde{\boldsymbol{X}}'\boldsymbol{X} = \boldsymbol{O}$ and $\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} = \boldsymbol{I}_{p-q}$. Let $\boldsymbol{\Psi}$ be the covariance matrix of each row of $(\boldsymbol{U}\ \boldsymbol{V})$. Then, $\widehat{\boldsymbol{\Sigma}}$ is obtained using the relation $\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{G}')^{-1}\widehat{\boldsymbol{\Psi}}\boldsymbol{G}$. Here, the MLE $\widehat{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$ based on $\boldsymbol{Y}_{1(12\cdots k)}$ can be expressed as follows (see, e.g. Fujikoshi et al. (2010, (12.4.6), Theorem 12.4.2)):

$$\begin{aligned} &\widehat{\boldsymbol{\Psi}}_{11\cdot 2} = \frac{n_1 - 1}{n_1}(\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}, \quad \widehat{\boldsymbol{\Psi}}_{22} = \frac{1}{n_1}\boldsymbol{G}_2'\boldsymbol{Y}_{1(12\cdots k)}'\boldsymbol{Y}_{1(12\cdots k)}\boldsymbol{G}_2, \\ &\widehat{\boldsymbol{\Gamma}} = (\boldsymbol{G}_2'\boldsymbol{S}\boldsymbol{G}_2)^{-1}\boldsymbol{G}_2'\boldsymbol{S}\boldsymbol{G}_1, \end{aligned} \tag{13}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}$ and $\boldsymbol{S}$ is the usual unbiased sample covariance matrix based on $\boldsymbol{Y}_{1(12\cdots k)}$.

Let $\widehat{\boldsymbol{\Sigma}}_{\mathrm{c}}$ be the MLE of $\boldsymbol{\Sigma}$ based on the complete data $\boldsymbol{Y}_{\mathrm{c}} = \boldsymbol{Y}_{1(12\cdots k)}$. Then, we have

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{c}} = (\boldsymbol{G}')^{-1}\widehat{\boldsymbol{\Psi}}_{\mathrm{c}}\boldsymbol{G},$$

where $\widehat{\boldsymbol{\Psi}}_{\mathrm{c}}$ is the $\widehat{\boldsymbol{\Psi}}$ given by (13). Let $\widehat{\boldsymbol{\theta}}_{\mathrm{c}}$ be the estimator of $\boldsymbol{\theta}$, which is obtained from the $\widehat{\boldsymbol{\theta}}$ in (9) by substituting $\widehat{\boldsymbol{\Sigma}}_{\mathrm{c}}$ to $\widehat{\boldsymbol{\Sigma}}$. Then, it is expected that such a conventional estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{c}}$ is useful, especially when $n - n_1$ is small.

5 Example

In this section, we consider the data on the ramus heights of 20 boys, which is given in Table 2 of Elston and Grizzle (1962), to illustrate the results of this paper. The ramus height has been measured in mm for each boy at 8, $8\frac{1}{2}$, 9, and $9\frac{1}{2}$ years of age. Let us assume that by discarding some data, the data will be missing completely at random and have a three-step monotone pattern as follows:

Let

$$\boldsymbol{Y}_{1(123)} = \begin{pmatrix} 47.8 & 48.8 & 49.0 & 49.7 \\ 46.4 & 47.3 & 47.7 & 48.4 \\ 46.3 & 46.8 & 47.8 & 48.5 \\ 45.1 & 45.3 & 46.1 & 47.2 \\ 47.6 & 48.5 & 48.9 & 49.3 \\ 52.5 & 53.2 & 53.3 & 53.7 \\ 51.2 & 53.0 & 54.3 & 54.5 \\ 49.8 & 50.0 & 50.3 & 52.7 \\ 48.1 & 50.8 & 52.3 & 54.4 \\ 45.0 & 47.0 & 47.3 & 48.3 \end{pmatrix}, \quad \boldsymbol{Y}_{2(12)} = \begin{pmatrix} 51.2 & 51.4 & 51.6 \\ 48.5 & 49.2 & 53.0 \\ 52.1 & 52.8 & 53.7 \\ 48.2 & 48.9 & 49.3 \\ 49.6 & 50.4 & 51.2 \end{pmatrix}, \quad \boldsymbol{Y}_{31} = \begin{pmatrix} 50.7 & 51.7 \\ 47.2 & 47.7 \\ 53.3 & 54.6 \\ 46.2 & 47.5 \\ 46.3 & 47.6 \end{pmatrix}.$$

Suppose that the growth or change is linear with respect to time. Then, $q = 2$ and we have

$$\begin{aligned} \mathrm{vec}(\boldsymbol{Y}'_{1(123)}) &\sim N_{40}(\mathrm{vec}(\boldsymbol{\mu}\mathbf{1}'_{10}), \boldsymbol{I}_{10} \otimes \boldsymbol{\Sigma}), \\ \mathrm{vec}(\boldsymbol{Y}'_{2(12)}) &\sim N_{15}(\mathrm{vec}(\boldsymbol{\mu}_{(12)}\mathbf{1}'_{5}), \boldsymbol{I}_{5} \otimes \boldsymbol{\Sigma}_{(12)(12)}), \\ \mathrm{vec}(\boldsymbol{Y}'_{31}) &\sim N_{10}(\mathrm{vec}(\boldsymbol{\mu}_{1}\mathbf{1}'_{5}), \boldsymbol{I}_{5} \otimes \boldsymbol{\Sigma}_{11}), \end{aligned}$$

where

$$\boldsymbol{\mu} = \underset{4\times 2}{\boldsymbol{X}}\ \underset{2\times 1}{\boldsymbol{\theta}}, \quad \underset{p\times q}{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \boldsymbol{X}_3 \end{pmatrix} \begin{matrix} \}_2 \\ \}_1 \\ \}_1 \end{matrix} = \begin{pmatrix} 1 & 8 \\ 1 & 8.5 \\ 1 & 9 \\ 1 & 9.5 \end{pmatrix}.$$

We set the starting matrix as

$$\begin{aligned} &\widehat{\boldsymbol{\Sigma}}_{11}^{(0)} = \begin{pmatrix} 6.0135 & 5.8796 \\ 5.8796 & 6.1269 \end{pmatrix}, \quad \widehat{\mathcal{B}}_{12}^{(0)} = \begin{pmatrix} -0.1951 \\ 1.1953 \end{pmatrix}, \\ &\widehat{\boldsymbol{\Sigma}}_{22\cdot 1}^{(0)} = \begin{pmatrix} 0.7800 \end{pmatrix}, \quad \widehat{\mathcal{B}}_{(12)3}^{(0)} = \begin{pmatrix} -0.0394 \\ -0.2141 \\ 1.2259 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_{33\cdot 12}^{(0)} = \begin{pmatrix} 0.4633 \end{pmatrix}. \end{aligned}$$

Then, using the algorithm described in Section 4, we get

$$\widehat{\boldsymbol{\theta}}^{(1)} = \begin{pmatrix} 33.3995 \\ 1.9092 \end{pmatrix}, \quad \cdots, \quad \widehat{\boldsymbol{\theta}}^{(5)} = \begin{pmatrix} 33.5022 \\ 1.8962 \end{pmatrix}.$$

Finally, we obtain the MLEs of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ as follows.

$$\widehat{\boldsymbol{\mu}}^{(5)} = \begin{pmatrix} 48.6715 \\ 49.6196 \\ 50.5677 \\ 51.5158 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}^{(5)} = \begin{pmatrix} 6.0137 & 5.8795 & 5.8612 & 5.6552 \\ 5.8795 & 6.1269 & 6.1846 & 6.0114 \\ 5.8612 & 6.1846 & 7.0515 & 6.9962 \\ 5.6552 & 6.0114 & 6.9962 & 7.4318 \end{pmatrix}.$$

Throughout these steps, the values of $-2\log L(\boldsymbol{\theta},\boldsymbol{\Sigma})$ are

$$g(\widehat{\boldsymbol{\theta}}^{(1)},\widehat{\boldsymbol{\Sigma}}_{11}^{(1)},\widehat{\mathcal{B}}_{12}^{(1)},\widehat{\boldsymbol{\Sigma}}_{22\cdot1}^{(1)},\widehat{\mathcal{B}}_{(12)3}^{(1)},\widehat{\boldsymbol{\Sigma}}_{33\cdot12}^{(1)}) = 189.8262,\cdots,$$
$$g(\widehat{\boldsymbol{\theta}}^{(5)},\widehat{\boldsymbol{\Sigma}}_{11}^{(5)},\widehat{\mathcal{B}}_{12}^{(5)},\widehat{\boldsymbol{\Sigma}}_{22\cdot1}^{(5)},\widehat{\mathcal{B}}_{(12)3}^{(5)},\widehat{\boldsymbol{\Sigma}}_{33\cdot12}^{(5)}) = 189.8231.$$

In this example, we can also obtain the MLEs of $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ by using the partially complete data (10×4) as follows:

$$\widehat{\boldsymbol{\theta}}_{\rm pc} = \begin{pmatrix} 36.0103 \\ 1.4968 \end{pmatrix},\quad \widehat{\boldsymbol{\mu}}_{\rm pc} = \begin{pmatrix} 47.9844 \\ 48.7328 \\ 49.4811 \\ 50.2295 \end{pmatrix},\quad \widehat{\boldsymbol{\Sigma}}_{\rm pc} = \begin{pmatrix} 5.6796 & 5.7109 & 5.6780 & 5.5135 \\ 5.7109 & 6.4678 & 6.5778 & 6.5367 \\ 5.6780 & 6.5778 & 6.9179 & 6.9024 \\ 5.5135 & 6.5367 & 6.9024 & 7.4161 \end{pmatrix}.$$

Here, $\widehat{\boldsymbol{\Sigma}}_{\rm pc}$ is obtained from (13) by substituting the partially complete data $\boldsymbol{Y}_{1(123)}$ for $\boldsymbol{Y}_{1(12\ldots k)}$.

In this example, we can derive the MLEs based on the complete data (20×4), which is given as follows:

$$\widehat{\boldsymbol{\theta}}_{\rm c} = \begin{pmatrix} 33.7605 \\ 1.8616 \end{pmatrix},\quad \widehat{\boldsymbol{\mu}}_{\rm c} = \begin{pmatrix} 48.6534 \\ 49.5842 \\ 50.5150 \\ 51.4458 \end{pmatrix},\quad \widehat{\boldsymbol{\Sigma}}_{\rm c} = \begin{pmatrix} 6.0137 & 5.8795 & 5.4881 & 5.2718 \\ 5.8795 & 6.1269 & 5.8458 & 5.6269 \\ 5.4881 & 5.8458 & 6.5721 & 6.5988 \\ 5.2718 & 5.6269 & 6.5988 & 7.0958 \end{pmatrix}.$$

It may be noted that the estimators $\widehat{\boldsymbol{\theta}}^{(5)}$, $\widehat{\boldsymbol{\mu}}^{(5)}$, and $\widehat{\boldsymbol{\Sigma}}^{(5)}$ are considerably similar to $\widehat{\boldsymbol{\theta}}_{\rm c}$, $\widehat{\boldsymbol{\mu}}_{\rm c}$, and $\widehat{\boldsymbol{\Sigma}}_{\rm c}$, respectively. On the other hand, we can see that the estimators $\widehat{\boldsymbol{\theta}}_{\rm pc}$, $\widehat{\boldsymbol{\mu}}_{\rm pc}$, and $\widehat{\boldsymbol{\Sigma}}_{\rm pc}$ obtained by deleting the missing data are not good.

6 Conclusion

This paper provided closed forms for the MLE of the mean parameter vector when the covariance matrix is known, and the MLE of the covariance matrix when the mean parameter vector is known, based on general monotone missing data. The distributions of these MLEs were derived. Using the results, we proposed an algorithm that includes an iterative procedure to obtain the MLEs when all the parameters are unknown. These results may also be useful for considering a hypothesis testing on a growth curve model with monotone missing data, which is left as a future work. Further, the results can be extended to the multi-sample case, and we are currently investigating this problem.

7 Appendix

7.1 Proofs of Theorem 1 and Lemma 1

Since an extension to a general k-step monotone missing data is straightforward, in the following, we give proofs of Theorem 1 and Lemma 1 in the case of $k = 3$. Then, the growth curve model with three-step monotone missing data is expressed as

$$\begin{aligned} &\mathrm{E}[\underset{n_1\times p}{\boldsymbol{Y}_{1(123)}}] = \underset{n_1\times 1}{\boldsymbol{1}_{n_1}} \underset{1\times q}{\boldsymbol{\theta}'} \underset{q\times p}{\boldsymbol{X}'}, \quad \mathrm{E}[\underset{n_2\times p_{(12)}}{\boldsymbol{Y}_{2(12)}}] = \underset{n_2\times 1}{\boldsymbol{1}_{n_2}} \underset{1\times q}{\boldsymbol{\theta}'} \underset{q\times p_{(12)}}{\boldsymbol{X}'_{(12)}}, \\ &\mathrm{E}[\underset{n_3\times p_1}{\boldsymbol{Y}_{31}}] = \underset{n_3\times 1}{\boldsymbol{1}_{n_3}} \underset{1\times q}{\boldsymbol{\theta}'} \underset{q\times p_1}{\boldsymbol{X}'_1}, \end{aligned} \tag{14}$$

where

$$\boldsymbol{Y}_{1(123)} = \begin{pmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} & \boldsymbol{Y}_{13} \end{pmatrix}, \quad \boldsymbol{Y}_{2(12)} = \begin{pmatrix} \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22} \end{pmatrix},$$
$$\underset{p\times q}{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \boldsymbol{X}_3 \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \}p_3 \end{matrix} = \begin{pmatrix} \boldsymbol{X}_{(12)} \\ \boldsymbol{X}_3 \end{pmatrix} \begin{matrix} \}p_{(12)} \\ \}p_3 \end{matrix}, \quad p_{(12)} = p_1 + p_2.$$

It is assumed that the rows of $\boldsymbol{Y}_{1(123)}$ $\boldsymbol{Y}_{2(12)}$ and $\boldsymbol{Y}_{31}$ are mutually independent and

$$\begin{aligned} &\mathrm{vec}(\boldsymbol{Y}'_{1(123)}) \sim N_{n_1p}(\mathrm{vec}(\boldsymbol{\mu}\boldsymbol{1}'_{n_1}), \boldsymbol{I}_{n_1} \otimes \boldsymbol{\Sigma}), \\ &\mathrm{vec}(\boldsymbol{Y}'_{2(12)}) \sim N_{n_2p_{(12)}}(\mathrm{vec}(\boldsymbol{\mu}_{(12)}\boldsymbol{1}'_{n_2}), \boldsymbol{I}_{n_2} \otimes \boldsymbol{\Sigma}_{(12)(12)}), \\ &\mathrm{vec}(\boldsymbol{Y}'_{31}) \sim N_{n_3p_1}(\mathrm{vec}(\boldsymbol{\mu}_1\boldsymbol{1}'_{n_3}), \boldsymbol{I}_{n_3} \otimes \boldsymbol{\Sigma}_{11}), \end{aligned}$$

where $\boldsymbol{1}_{n_i}$ is an $n_i \times 1$ vector of 1s,

$$\boldsymbol{\mu} = \left(\begin{array}{c} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \hline \boldsymbol{\mu}_3 \end{array}\right) \begin{matrix} \}p_1 \\ \}p_2 \\ \}p_3 \end{matrix} = \left(\begin{array}{c} \boldsymbol{\mu}_{(12)} \\ \hline \boldsymbol{\mu}_3 \end{array}\right),$$

and

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc|c} \overbrace{\boldsymbol{\Sigma}_{11}}^{p_1} & \overbrace{\boldsymbol{\Sigma}_{12}}^{p_2} & \overbrace{\boldsymbol{\Sigma}_{13}}^{p_3} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \hline \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{array}\right) \begin{matrix} \}p_1 \\ \}p_2 \\ \}p_3 \end{matrix} = \left(\begin{array}{cc|c} \multicolumn{2}{c|}{\boldsymbol{\Sigma}_{(12)(12)}} & \begin{matrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{matrix} \\ \hline \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{array}\right).$$

Let $L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ be the likelihood under (14). Then, we have

$$
\begin{aligned}
&-2\log L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
&= (N_3 p_1 + N_2 p_2 + N_1 p_3)\log(2\pi) + N_3 \log|\boldsymbol{\Sigma}_{11}| + N_2 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + N_1 \log|\boldsymbol{\Sigma}_{33\cdot 12}| \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(123)1} - \mathbf{1}_{N_3}\boldsymbol{\theta}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(123)1} - \mathbf{1}_{N_3}\boldsymbol{\theta}'\boldsymbol{X}_1') \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\left(\boldsymbol{Y}_{(12)2} - \boldsymbol{Y}_{(12)1}\mathcal{B}_{12} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'\right)'\left(\boldsymbol{Y}_{(12)2} - \boldsymbol{Y}_{(12)1}\mathcal{B}_{12} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'\right) \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{33\cdot 12}^{-1}\left(\boldsymbol{Y}_{13} - \boldsymbol{Y}_{1(12)}\mathcal{B}_{(12)3} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_3'\right)'\left(\boldsymbol{Y}_{13} - \boldsymbol{Y}_{1(12)}\mathcal{B}_{(12)3} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_3'\right),
\end{aligned}
$$

where

$$
\boldsymbol{Y}_{(123)1} = \begin{pmatrix} \boldsymbol{Y}_{11} \\ \boldsymbol{Y}_{21} \\ \boldsymbol{Y}_{31} \end{pmatrix}, \quad \boldsymbol{Y}_{(12)2} = \begin{pmatrix} \boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{22} \end{pmatrix}, \quad \boldsymbol{Y}_{1(12)} = \begin{pmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} \end{pmatrix},
$$

$$
\boldsymbol{\Sigma}_{33\cdot 12} = \boldsymbol{\Sigma}_{33} - \begin{pmatrix} \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} \end{pmatrix} \boldsymbol{\Sigma}_{(12)(12)}^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix}, \quad \widetilde{\boldsymbol{X}}_3' = \boldsymbol{X}_3' - \boldsymbol{X}_{(12)}'\mathcal{B}_{(12)3},
$$

$$
\mathcal{B}_{(12)3} = \boldsymbol{\Sigma}_{(12)(12)}^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix}, \quad N_i = \sum_{j=1}^{i} n_j, \quad i = 1, 2, 3.
$$

We derive the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known. Since there is a one-to-one correspondence between $\boldsymbol{\Sigma}$ and $\{\boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12}\}$,

$$
\begin{aligned}
&g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12}) \\
&\equiv -2\log L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
&= (N_3 p_1 + N_2 p_2 + N_1 p_3)\log(2\pi) + N_3 \log|\boldsymbol{\Sigma}_{11}| + N_2 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + N_1 \log|\boldsymbol{\Sigma}_{33\cdot 12}| \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(123)1} - \mathbf{1}_{N_3}\boldsymbol{\theta}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(123)1} - \mathbf{1}_{N_3}\boldsymbol{\theta}'\boldsymbol{X}_1') \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\left(\boldsymbol{Y}_{(12)2} - \boldsymbol{Y}_{(12)1}\mathcal{B}_{12} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'\right)'\left(\boldsymbol{Y}_{(12)2} - \boldsymbol{Y}_{(12)1}\mathcal{B}_{12} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'\right) \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{33\cdot 12}^{-1}\left(\boldsymbol{Y}_{13} - \boldsymbol{Y}_{1(12)}\mathcal{B}_{(12)3} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_3'\right)'\left(\boldsymbol{Y}_{13} - \boldsymbol{Y}_{1(12)}\mathcal{B}_{(12)3} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_3'\right) \\
&= -2(\boldsymbol{a}_1 + \boldsymbol{a}_2 + \boldsymbol{a}_3)'\boldsymbol{\theta} + \boldsymbol{\theta}'(N_3\boldsymbol{A}_1 + N_2\boldsymbol{A}_2 + N_1\boldsymbol{A}_3)\boldsymbol{\theta} + (\text{terms without } \boldsymbol{\theta}),
\end{aligned}
$$

where

$$\begin{aligned}
&\boldsymbol{a}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_{(123)1}'\mathbf{1}_{N_3}, \quad \boldsymbol{a}_2 = \widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{(12)2} - \boldsymbol{Y}_{(12)1}\mathcal{B}_{12})'\mathbf{1}_{N_2}, \\
&\boldsymbol{a}_3 = \widetilde{\boldsymbol{X}}_3'\boldsymbol{\Sigma}_{33\cdot 12}^{-1}(\boldsymbol{Y}_{13} - \boldsymbol{Y}_{1(12)}\mathcal{B}_{(12)3})'\mathbf{1}_{N_1}, \quad \boldsymbol{A}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1, \\
&\boldsymbol{A}_2 = \widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2, \quad \boldsymbol{A}_3 = \widetilde{\boldsymbol{X}}_3'\boldsymbol{\Sigma}_{33\cdot 12}^{-1}\widetilde{\boldsymbol{X}}_3.
\end{aligned}$$

Solving $(\partial/\partial\boldsymbol{\theta})g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12}) = \mathbf{0}$, we obtain the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known as

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{M}^{-1}(\boldsymbol{a}_1 + \boldsymbol{a}_2 + \boldsymbol{a}_3),$$

where $\boldsymbol{M} = N_3\boldsymbol{A}_1 + N_2\boldsymbol{A}_2 + N_1\boldsymbol{A}_3$. The distributional results in Theorem 1 are a direct consequence of Lemma 1. Therefore, we show Lemma 1 in the case of $k = 3$. The conditional distribution of $\boldsymbol{a}_3$ for a given $\boldsymbol{Y}_{1(12)}$ is $N_q(N_1\boldsymbol{A}_3\boldsymbol{\theta}, N_1\boldsymbol{A}_3)$, which is not dependent on $\boldsymbol{Y}_{1(12)}$. Therefore, the distribution of $\boldsymbol{a}_3$ is $N_q(N_1\boldsymbol{A}_3\boldsymbol{\theta}, N_1\boldsymbol{A}_3)$, and it is independent of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. Similarly, considering the conditional distribution of $\boldsymbol{a}_2$ for a given $\boldsymbol{Y}_{(12)1}$, we have that $\boldsymbol{a}_2 \sim N_q(N_2\boldsymbol{A}_2\boldsymbol{\theta}, N_2\boldsymbol{A}_2)$ and $\boldsymbol{a}_2$ is independent of $\boldsymbol{a}_1$. It is easy to see that the marginal distribution of $\boldsymbol{a}_1$ is $N_q(N_3\boldsymbol{A}_1\boldsymbol{\theta}, N_3\boldsymbol{A}_1)$.

7.2 Proof of Theorem 2

Next, we consider the minimization of $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12})$ with respect to $\boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}$, $\boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}$, and $\boldsymbol{\Sigma}_{33\cdot 12}$. We define

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} & \boldsymbol{Z}_{13} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} & * \\ \boldsymbol{Z}_{31} & * & * \end{pmatrix} &= \begin{pmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} & \boldsymbol{Y}_{13} \\ \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22} & * \\ \boldsymbol{Y}_{31} & * & * \end{pmatrix} - \begin{pmatrix} \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_1' & \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_2' & \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_3' \\ \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_1' & \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_2' & * \\ \mathbf{1}_{n_3}\boldsymbol{\theta}'\boldsymbol{X}_1' & * & * \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{Y}_{11} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_1' & \boldsymbol{Y}_{12} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_2' & \boldsymbol{Y}_{13} - \mathbf{1}_{n_1}\boldsymbol{\theta}'\boldsymbol{X}_3' \\ \boldsymbol{Y}_{21} - \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_1' & \boldsymbol{Y}_{22} - \mathbf{1}_{n_2}\boldsymbol{\theta}'\boldsymbol{X}_2' & * \\ \boldsymbol{Y}_{31} - \mathbf{1}_{n_3}\boldsymbol{\theta}'\boldsymbol{X}_1' & * & * \end{pmatrix},
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{Z}_{(123)1} = \boldsymbol{Y}_{(123)1} - \mathbf{1}_{N_3}\boldsymbol{\theta}'\boldsymbol{X}_1', \quad \boldsymbol{Z}_{(12)2} = \boldsymbol{Y}_{(12)2} - \mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_2', \\
&\boldsymbol{Z}_{1(12)} = \boldsymbol{Y}_{1(12)} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\boldsymbol{X}_{(12)}'.
\end{aligned}$$

Then, we can reduce $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12})$ as follows:

$$
\begin{aligned}
&g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12}) \\
&= (N_3p_1 + N_2p_2 + N_1p_3)\log(2\pi) + N_3 \log|\boldsymbol{\Sigma}_{11}| + N_2 \log|\boldsymbol{\Sigma}_{22\cdot 1}| + N_1 \log|\boldsymbol{\Sigma}_{33\cdot 12}| \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{Z}'_{(123)1}\boldsymbol{Z}_{(123)1} + \mathrm{tr}\, \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\Big(\boldsymbol{Z}_{(12)2} - \boldsymbol{Z}_{(12)1}\mathcal{B}_{12}\Big)'\Big(\boldsymbol{Z}_{(12)2} - \boldsymbol{Z}_{(12)1}\mathcal{B}_{12}\Big) \\
&\quad + \mathrm{tr}\, \boldsymbol{\Sigma}_{33\cdot 12}^{-1}\Big(\boldsymbol{Z}_{13} - \boldsymbol{Z}_{1(12)}\mathcal{B}_{(12)3}\Big)'\Big(\boldsymbol{Z}_{13} - \boldsymbol{Z}_{1(12)}\mathcal{B}_{(12)3}\Big).
\end{aligned}
$$

As with $k = 2$, calculating the derivative of $g(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}, \mathcal{B}_{(12)3}, \boldsymbol{\Sigma}_{33\cdot 12})$ with respect to $\boldsymbol{\Sigma}_{11}, \mathcal{B}_{12}, \boldsymbol{\Sigma}_{22\cdot 1}$, $\mathcal{B}_{(12)3}$, and $\boldsymbol{\Sigma}_{33\cdot 12}$, respectively, and solving the likelihood equation, we obtain the following MLEs with the known parameter $\boldsymbol{\theta}$.

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{11} &= \frac{1}{N_3}\boldsymbol{Z}'_{(123)1}\boldsymbol{Z}_{(123)1}, \quad \widehat{\mathcal{B}}_{12} = (\boldsymbol{Z}'_{(12)1}\boldsymbol{Z}_{(12)1})^{-1}\boldsymbol{Z}'_{(12)1}\boldsymbol{Z}_{(12)2}, \\
\widehat{\boldsymbol{\Sigma}}_{22\cdot 1} &= \frac{1}{N_2}\Big(\boldsymbol{Z}_{(12)2} - \boldsymbol{Z}_{(12)1}\mathcal{B}_{12}\Big)'\Big(\boldsymbol{Z}_{(12)2} - \boldsymbol{Z}_{(12)1}\mathcal{B}_{12}\Big) \\
&= \frac{1}{N_2}\boldsymbol{Z}'_{(12)2}\Big(\boldsymbol{I}_{N_2} - \boldsymbol{P}_{\boldsymbol{Z}_{(12)1}}\Big)\boldsymbol{Z}_{(12)2}, \\
\widehat{\mathcal{B}}_{(12)3} &= (\boldsymbol{Z}'_{1(12)}\boldsymbol{Z}_{1(12)})^{-1}\boldsymbol{Z}'_{1(12)}\boldsymbol{Z}_{13}, \\
\widehat{\boldsymbol{\Sigma}}_{33\cdot 12} &= \frac{1}{N_1}\Big(\boldsymbol{Z}_{13} - \boldsymbol{Z}_{1(12)}\mathcal{B}_{(12)3}\Big)'\Big(\boldsymbol{Z}_{13} - \boldsymbol{Z}_{1(12)}\mathcal{B}_{(12)3}\Big) \\
&= \frac{1}{N_1}\boldsymbol{Z}'_{13}\Big(\boldsymbol{I}_{N_1} - \boldsymbol{P}_{\boldsymbol{Z}_{1(12)}}\Big)\boldsymbol{Z}_{13},
\end{aligned}
$$

where $P_{\boldsymbol{Z}_{(12)1}} = \boldsymbol{Z}_{(12)1}(\boldsymbol{Z}'_{(12)1}\boldsymbol{Z}_{(12)1})^{-1}\boldsymbol{Z}'_{(12)1}$ and $P_{\boldsymbol{Z}_{1(12)}} = \boldsymbol{Z}_{1(12)}(\boldsymbol{Z}'_{1(12)}\boldsymbol{Z}_{1(12)})^{-1}\ \boldsymbol{Z}'_{1(12)}$.

Consider the conditional distribution of $N_1\widehat{\boldsymbol{\Sigma}}_{33\cdot 12}$ for a given $\boldsymbol{Z}_{1(12)}$. Then,

$$
\mathrm{E}[\boldsymbol{Z}_{13}|\boldsymbol{Z}_{1(12)}] = \boldsymbol{Z}_{1(12)}\mathcal{B}_{12}.
$$

We can see that $N_1\widehat{\boldsymbol{\Sigma}}_{33\cdot 12}|\boldsymbol{Z}_{1(12)} \sim W_{p_3}(N_1 - p_{(12)} - 1, \boldsymbol{\Sigma}_{33\cdot 12})$, and hence $N_1\widehat{\boldsymbol{\Sigma}}_{33\cdot 12} \sim W_{p_3}(N_1 - p_{(12)} - 1, \boldsymbol{\Sigma}_{33\cdot 12})$. Further, $N_1\widehat{\boldsymbol{\Sigma}}_{33\cdot 12}$ is independent of $\boldsymbol{Z}_{1(12)}$. Similarly, considering the conditional distribution of $N_2\widehat{\boldsymbol{\Sigma}}_{22\cdot 1}$ for a given $\boldsymbol{Z}_{(12)1}$, we get $N_2\widehat{\boldsymbol{\Sigma}}_{22\cdot 1} \sim W_{p_2}(N_2 - p_1 - 1, \boldsymbol{\Sigma}_{22\cdot 1})$. It is easy to see that $N_3\widehat{\boldsymbol{\Sigma}}_{11} \sim W_{p_1}(N_3 - 1, \boldsymbol{\Sigma}_{11})$, and $N_3\widehat{\boldsymbol{\Sigma}}_{11}$, $N_2\widehat{\boldsymbol{\Sigma}}_{22\cdot 1}$, and $N_1\widehat{\boldsymbol{\Sigma}}_{33\cdot 12}$ are independent.

Acknowledgment

The first and second authors' research is partly supported by a Grant-in-Aid for Young Scientists (JSPS KAKENHI Grant Number JP19K20225) and a Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP17K00058), respectively. The third author's research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP16K00047), 2016-2018.

References

[1] Anderson, T. W. and Olkin, I. (1985). Maximum-likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra and Its Applications*, **70**, 147–171.

[2] Elston, R. C. and Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics*, **18**, 148-159.

[3] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Hoboken, NJ: Wiley.

[4] Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics – Theory and Methods*, **21**, 41–50.

[5] Kanda, T. (1994). Growth curve model with covariance structures. *Hiroshima Mathematical Journal*, **24**, 135–176.

[6] Kanda, T. and Fujikoshi, Y. (1998). Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. *American Journal of Mathematical and Management Sciences*, **18**, 161–190.

[7] Kleinbaum, D. G. (1973). A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, **3**, 117–124.

[8] Liski, E. P. (1985). Estimation from incomplete data in growth curves models. *Communications in Statistics – Simulation and Computation*, **14**, 13–27.

[9] Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[10] Srivastava, M. S. (1985). Multivariate data with missing observations. *Communications in Statistics – Theory and Methods*, **14**, 775–792.

[11] Yagi, A. and Seo, T. (2017). Tests for normal mean vectors with monotone incomplete data. *American Journal of Mathematical and Management Sciences*, **36**, 1–20.