Parameter Estimation for Partially Observable Systems Subject to Random Failure

Michael Jong Kim*, Viliam Makis† and Rui Jiang‡

University of Toronto, Department of Mechanical and Industrial Engineering,
Toronto, ON, M5S 3G8, Canada

Abstract

In this paper we present a parameter estimation procedure for a condition-based maintenance model under partial observations. Systems can be in a healthy or unhealthy operational state, or in a failure state. System deterioration is driven by a continuous time homogeneous Markov chain and the system state is unobservable, except the failure state. Vector information that is stochastically related to the system state is obtained through condition monitoring at equidistant sampling times. Two types of data histories are available – data histories that end with observable failure, and censored data histories that end when the system has been suspended from operation but has not failed. The state and observation processes are modeled in the hidden Markov framework and the model parameters are estimated using the EM algorithm. We show that both the pseudo likelihood function and the parameter updates in each iteration of the EM algorithm have explicit formulas. A numerical example is developed which illustrates the entire procedure.

Keywords: Condition-Based Maintenance; Failing Systems; Parameter Estimation; Hidden Markov Modeling; Multivariate Observations; EM Algorithm

2000 Mathematics Subject Classification: Primary 62P30, Secondary 62N02; 90B25

* Email: kimmi@mie.utoronto.ca
† Corresponding Author. Email: makis@mie.utoronto.ca
‡ Email: jiangrui@mie.utoronto.ca
1. Introduction

In this paper, we consider a parameter estimation problem for a partially observable system subject to random failure. System’s condition can be categorized into one of three states: a healthy or “good as new” state 0, an unhealthy or deteriorated state 1, and a failure state 2. Only the failure state is observable. The state process \( (X_t : t \in \mathbb{R}_+) \) is modeled as a continuous time homogeneous Markov chain with state space \( \{0,1,2\} \). While the system is operational, vector data that is stochastically related to the system state is obtained through condition monitoring at equidistant time points \( \Delta, 2\Delta, \ldots \). For example, in mining companies, the oil that lubricates transmission units in heavy hauler trucks is sampled at regular intervals. Spectrometric analysis is carried out which provides the levels in ppm of different metals that come from the wear of the transmission units. The data histories of metal levels constitute the observation process that gives partial information about the state of the operational transmission units. We assume that two types of data histories are available: histories that end with observable system failure, and censored data histories that end when the system has been suspended from operation but has not failed. Given any number of failure and suspension histories, our objective is to determine the maximum likelihood estimates (MLEs) of the model parameters.

In recent years, a lot of research has been done on the analysis and control of maintenance models. Neuts et al. [1] considered the analysis of failing systems governed by so-called phase type distributions. Makis et al. [2] considered a repair/replacement model for a single unit with random repair costs. Jiang et al. [3] studied a maintenance model with general repair and two types of replacement actions: failure and preventive replacement. Li and Shaked [4] analyzed an imperfect repair model also subject to preventive maintenance. Makis and Jiang [5] considered an optimal replacement problem under partial observations. The authors formulated the problem in the optimal stopping framework and were able to
determine the structure of the optimal replacement policy. Other maintenance models with partial information have been developed in the literature focusing only on optimal decision making [6-7]. Surprisingly, little research has been done on parameter estimation for partially observable systems subject to random failure. Although some research has considered estimation for partially observable systems in the hidden Markov model (HMM) framework, few researchers have considered the inclusion of failure information, which is present in almost every maintenance application. For example, Ryden [8], Douc et al. [9], Genon-Catalot and Laredo [10], and Hamilton [11] considered maximum likelihood estimation for hidden Markov models in discrete time, however the results of their papers are not applicable to maintenance systems for which system failure is observable.

Lin and Makis [12] considered an interesting maintenance model with finite-valued observations and failure information, similar to the model considered in this paper. Their objective was to derive a general recursive filter, which is important mainly for on-line re-estimation. The authors were able to express the parameter updates in each iteration of the EM algorithm in terms of the recursive filter. However, such an approach has been found to be quite computationally intensive and difficult to implement when working with real data sets.

We have found through our work with diagnostic data such as spectrometric oil data and vibration data, that it is usually sufficient to consider only two operational states – a healthy state and an unhealthy state. It will be shown that the estimation problem of the 3-state model considered in this paper can be solved by directly analyzing the structure of the pseudo likelihood function. We will show that both the pseudo likelihood function and the parameter updates in each iteration of the EM algorithm have explicit formulas. This implies that each iteration of the EM algorithm can be performed with a single computation, which leads to an extremely fast and simple estimation procedure. This computational advantage is
particularly attractive for practical applications. To our knowledge, this is the first paper that presents explicit formulas for the parameter updates in the EM algorithm for partially observable system with failure information.

The paper is organized as follows. In Section 2, we present the models of the state and observation processes. In Section 3, we discuss maximum likelihood estimation using the EM algorithm and the pseudo likelihood function. We derive an explicit expression for the pseudo likelihood function and provide update formulas for both the state and observation parameters. In Section 4, we develop a numerical example which illustrates the entire procedure and Section 5 provides concluding remarks.

2. Model Assumptions

We assume that a technical system’s condition can be categorized into one of three states: a healthy or “good as new” state (state 0), an unhealthy or deteriorated state (state 1), and a failure state (state 2). In many real world applications the state of an operational system is unobservable, and only the failure state is observable. For example, the state of an operational transmission unit in a heavy hauler truck cannot be observed without full system inspection, which is typically quite costly. However, failure of the mechanical unit is immediately observable. We model the state process $X = (X_t : t \in \mathbb{R}_+)$ as a continuous time homogeneous Markov chain with state space $S = \{0, 1, 2\}$. The system is assumed to start in a healthy state, i.e. $X_0 = 0$, and that the transition rate matrix is given by

$\Lambda = \begin{pmatrix}
-\lambda_{01} - \lambda_{02} & \lambda_{01} & \lambda_{02} \\
0 & -\lambda_{12} & \lambda_{12} \\
0 & 0 & 0
\end{pmatrix}$ \quad (1)
where $\lambda_0, \lambda_2, \lambda_0, \lambda_2 \in (0, +\infty)$ are the unknown state parameters. Let $\xi = \inf \{t \in \mathbb{R} : X_t = 2\}$ be the observable failure time of the system. Suppose at equidistant sampling times $\Delta, 2\Delta, \ldots, \Delta > 0$, vector-data $Y_{\Delta}, Y_{2\Delta}, \ldots \in \mathbb{R}^d$ is collected through condition monitoring, which gives partial information about the system state. For example, in practice, the oil that lubricates transmission units of heavy hauler trucks is sampled and spectrometric analysis is carried out. The analysis provides the levels in ppm of different metals that come from the direct wear of the transmission unit. This vector of metal particles is then taken as the observation process that gives partial information about the state of the operational transmission unit. The observations are assumed to be conditionally independent given the state of the system, and for each $n \in \mathbb{N}$, we assume that $Y_{na}$ conditional on $X_{na} = x, x = 0, 1$, has $d$-variate normal distribution $N_d(\mu_x, \Sigma_x)$ with density

$$f_{Y_{na}|X_{na}}(y \mid x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_x)}} \exp \left( -\frac{1}{2} (y - \mu_x)\Sigma_x^{-1}(y - \mu_x) \right)$$  \hspace{1cm} (2)$$

where $\mu_0, \mu_1 \in \mathbb{R}^d$ and $\Sigma_0, \Sigma_1 \in \mathbb{R}^{d \times d}$ are the unknown observation process parameters. Once the system fails, $P(Y_{na} = \eta \mid X_{na} = 2) = 1$, where $\eta \not\in \mathbb{R}^d$ represents a failure signal.

The assumption of conditional independence is reasonable in practice, once appropriate data pre-processing methods have been applied. In particular, one should first fit a model which accounts for cross and autocorrelation to the data histories, and choose as the observation process the residuals of the fitted model. A suggestion was given by Wu [13] and Rahman [14] to first identify the healthy portions of the data histories using Hotelling’s $T^2$ control chart. The authors then fit a common vector autoregressive (VAR) model to the healthy data histories, and obtained the residuals of the fitted model. Using this
method on real diagnostic data coming from the spectrometric analysis of failing transmission units, we have found that the residual process obtained in this manner satisfies the conditional independence and normality assumptions stated above. These assumptions considerably simplify both the parameter estimation and subsequent optimal system control problems (see e.g. Hamilton [11] and Kim and Nelson [15] who analyzed the general case in which the assumption of conditional independence does not hold). In the subsequent sections, we will assume that such data pre-processing has already been carried out. We should also note that our model assumes that only a single vector measurement is taken at each sampling epoch. This assumption is consistent with most real-world maintenance applications. However, the analysis which follows can be easily extended when it is practical to collect more than one sampling unit.

3. Parameter Estimation Using the EM Algorithm

We begin this section by briefly reviewing the well known EM algorithm in the context of our model. The EM algorithm, first introduced into the literature by Dempster et al. [16], has been found to be well-suited for solving parameter estimation problems in the hidden Markov framework. The algorithm’s optimal convergence properties and numerical advantages have been studied by researchers such as McLachlan and Krishnan [17] and Laporice [18].

Suppose we have collected $N \in \mathbb{N}$ failure histories, which we denote as $\delta_1, \ldots, \delta_N$. Failure history $\delta_i$ is assumed to be of the form $\bar{Y}_i = (y_{i1}, \ldots, y_{iT_i})$ and $\xi_i = t_i$, where $T_i \Delta < t_i \leq (T_i + 1) \Delta$. The sampling history $\bar{Y}_i$ represents the collection of all vector data $y_{ij} \in \mathbb{R}^d$, $j \leq T_i$, that was obtained through condition monitoring until system failure at time $t_i$. Suppose further that we have collected $M \in \mathbb{N}$ suspension histories, which we denote
as $\mathcal{S}_1, \ldots, \mathcal{S}_M$. Suspension history $\mathcal{S}_j$ is assumed to be of the form $\bar{Y}_j = (y^j_1, \ldots, y^j_{T_j})$ and $\xi_j > T_j \Delta$. Let $\mathcal{O} = \{\mathcal{S}_1, \ldots, \mathcal{S}_N, \mathcal{F}_1, \ldots, \mathcal{F}_M\}$ represent all observable data and $L(\lambda, \theta \mid \mathcal{O})$ be the associated likelihood function, where $\lambda = (\lambda_0, \lambda_2, \lambda_3)$ and $\theta = (\mu_0, \Sigma_0, \Sigma_1)$ are the set of unknown state and observation parameters. Because the sample paths $(X_t : t \in \mathbb{R}_+)$ of the state process are not observable, maximizing $L(\lambda, \theta \mid \mathcal{O})$ analytically is not possible. The EM algorithm resolves this difficulty by iteratively maximizing the so-called pseudo likelihood function. More specifically, the EM algorithm works as follows. Let $\lambda_0, \theta_0$ be some initial values of the unknown parameters.

E-step. For $n \geq 0$, compute the pseudo likelihood function defined by

$$Q(\lambda, \theta \mid \lambda_n, \theta_n) := E_{\mu_0, \Sigma_0} \left( \ln L(\lambda, \theta \mid \mathcal{O}) \mid \mathcal{O} \right)$$

where $\mathcal{O} = \{\mathcal{F}_1, \ldots, \mathcal{F}_N, \mathcal{F}_1, \ldots, \mathcal{F}_M\}$ represents the complete data set, in which each failure history $\mathcal{F}_i$ and suspension history $\mathcal{S}_j$ of the observable data set $\mathcal{O}$ has been augmented with the unobservable sample path information of the state process $X$.

M-step. Choose $\lambda_{n+1}, \theta_{n+1}$ such that

$$(\lambda_{n+1}, \theta_{n+1}) = \underset{\lambda, \theta}{\text{arg max}} Q(\lambda, \theta \mid \lambda_n, \theta_n)$$

The E and M steps are repeated until the Euclidean norm $|(|\lambda_{n+1}, \theta_{n+1}) - (\lambda_n, \theta_n)| < \varepsilon$, for $\varepsilon > 0$ small. The estimates $\lambda_{n+1}$ and $\theta_{n+1}$ then approximate the maximizers of $L(\lambda, \theta \mid \mathcal{O})$. 

7
We will show in Theorems 1 and 2, that (3) admits the following decomposition $Q(\lambda, \theta | \lambda_n, \theta_n) = Q^{\text{state}}(\lambda | \lambda_n, \theta_n) + Q^{\text{obs}}(\theta | \lambda_n, \theta_n)$, where $Q^{\text{state}}$ depends only on the state parameters $\lambda = (\lambda_{01}, \lambda_{02}, \lambda_{12})$ and $Q^{\text{obs}}$ depends only on the observation parameters $\theta = (\mu_0, \mu_1, \Sigma_0, \Sigma_1)$. This implies in particular that the M-step (4) can be carried out separately for the state and observation parameters, which simplifies considerably the computational algorithm and increases the speed of computation.

3.1. Formula for the Likelihood Function

In this subsection, we are interested in deriving an explicit formula for the likelihood function $L(\lambda, \theta | c)$ in (3). Let $\tau_0 = \inf\{t \in \mathbb{R}_+ : X_t > 0\}$ be the unobservable sojourn time of the state process in healthy state 0. From (1), it is clear that there is a one-to-one correspondence between the entire sample path $(X_t : t \in \mathbb{R}_+)$ of the system state and the two random variables $\tau_0$ and $\xi$. The distributional properties of the sojourn time $\tau_0$ and failure time $\xi$ are given by the following lemma.

**Lemma 1.** For each $t \in \mathbb{R}_+$, the density of $\xi$ is given by

$$f_\xi(t) = p_{01} \frac{v_0 v_1}{v_0 - v_1} \left(e^{-\nu t} - e^{-\eta t}\right) + p_{02} v_0 e^{-\eta t}$$

(5)

For all non-negative $s < t$, the conditional density of $\tau_0$ given $\xi$ is given by

$$f_{\tau_0 | \xi}(s | t) = \frac{p_{01} v_1 e^{-\nu t} e^{-(\nu - \eta) s}}{p_{01} \frac{v_1}{v_0 - v_1} (e^{-\nu t} - e^{-\eta t}) + p_{02} e^{-\eta t}}$$

(6)
and for each $t \in \mathbb{R}_+$, the conditional probability $P(\tau_0 = t \mid \xi = t)$ is given by

$$m_{\tau_0 \xi}(t \mid t) = \frac{p_{02} e^{-\nu_2 t}}{p_{01} \frac{v_1}{v_0 - v_1} (e^{-\nu_1 t} - e^{-\nu_2 t}) + p_{02} e^{-\nu_2 t}}$$

where $v_0 = \lambda_0 + \lambda_{02}$, $v_1 = \lambda_{12}$, $p_{01} = \frac{\lambda_{01}}{\lambda_0 + \lambda_{02}}$, and $p_{02} = \frac{\lambda_{02}}{\lambda_0 + \lambda_{02}}$.

**Proof.** Let $S_0 = X_{\tau_0}$ be the state of the system at time $\tau_0$. Then for each $t \in \mathbb{R}_+$,

$$P(\xi \leq t) = p_{01} P(\xi \leq t \mid S_0 = 1) + p_{02} P(\xi \leq t \mid S_0 = 2)$$

$$= p_{01} \int_{u=0}^t P(\xi \leq t \mid \tau_0 = u, S_0 = 1) v_0 e^{-\nu_1 u} du$$

$$+ p_{02} \int_{u=0}^t P(\xi \leq t \mid \tau_0 = u, S_0 = 2) v_0 e^{-\nu_2 u} du$$

$$= p_{01} \int_{u=0}^t P(X_{\tau_0} = 1) v_0 e^{-\nu_1 u} du + p_{02} \int_{u=0}^t 1 \cdot v_0 e^{-\nu_2 u} du$$

$$= p_{01} \int_{u=0}^t \left(1 - e^{-\nu_1 (t-u)}\right) v_0 e^{-\nu_1 u} du + p_{02} \left(1 - e^{-\nu_2 t}\right)$$

$$= p_{01} \left(1 - e^{-\nu_1 t}\right) - p_{01} \frac{v_0}{v_0 - v_1} e^{-\nu_1 t} + p_{01} \frac{v_0}{v_0 - v_1} e^{-\nu_2 t} + p_{02} \left(1 - e^{-\nu_2 t}\right)$$

which is differentiable in $t$ so that the density of $\xi$ is given by
\[ f_\xi(t) := \frac{dP(\xi \leq t)}{dt} \]
\[ = p_{01}v_0 e^{-\gamma t} + p_{01} \frac{v_0 v_1}{v_0 - v_1} e^{-\gamma t} - p_{01} \frac{v_0}{v_0 - v_1} e^{-\gamma t} + p_{02} v_0 e^{-\gamma t} \]
\[ = p_{01} \frac{v_0 v_1}{v_0 - v_1} (e^{-\gamma t} - e^{-\gamma t}) + p_{02} v_0 e^{-\gamma t} \]

for all \( t \in \mathbb{R}_+ \), and zero otherwise. For all non-negative \( s < t \),

\[ P(\tau_0 \leq s, \xi \\leq t) = p_{01} \int_{u=0}^{x} P(\xi \leq t \mid \tau_0 = u, S_0 = 1)v_0 e^{-\gamma u} du \]
\[ + p_{02} \int_{u=0}^{x} P(\xi \leq t \mid \tau_0 = u, S_0 = 2)v_0 e^{-\gamma u} du \]
\[ = p_{01} \int_{u=0}^{x} P(X_t = 2 \mid X_u = 1)v_0 e^{-\gamma u} du + p_{02} \int_{u=0}^{x} 1 \cdot v_0 e^{-\gamma u} du \]
\[ = p_{01} \int_{u=0}^{x} \left(1 - e^{-\gamma (t-u)}\right)v_0 e^{-\gamma u} du + p_{02} \left(1 - e^{-\gamma s}\right) \]
\[ = p_{01} \left(1 - e^{-\gamma s}\right) - p_{01} \frac{v_0}{v_0 - v_1} e^{-\gamma t} + p_{01} \frac{v_0}{v_0 - v_1} e^{-\gamma t} e^{-(\gamma_0 - \gamma_1)s} + p_{02} \left(1 - e^{-\gamma s}\right) \]

which is differentiable in both variables so that the joint density of \((\tau_0, \xi)\) for all non-negative \( s < t \) is given by

\[ f_{\tau_0,\xi}(s,t) := \frac{\partial^2 P(\tau_0 \leq s, \xi \leq t)}{\partial s \partial t} = p_{01} v_0 v_1 e^{-\gamma t} e^{-(\gamma_0 - \gamma_1)s} \]

and for all non-negative \( s < t \), we define the density function

10
\[
    f_{\tau_0|\xi}(s \mid t) := \frac{f_{\tau_0,\xi}(s, t)}{f_\xi(t)} = \frac{p_0 v_1 e^{-\eta s} e^{-(v_0 - \eta) s}}{p_0 \frac{v_1}{v_0 - v_1} (e^{-\eta s} - e^{-\eta t}) + p_0 e^{-\eta t}}
\]

For \( s = t \), we define the probability mass function

\[
    m_{\tau_0|\xi}(t \mid t) := P(\tau_0 = t \mid \xi = t) = 1 - P(\tau_0 < t \mid \xi = t) = 1 - \int_{s<\xi} f_{\tau_0|\xi}(s \mid t)ds
\]

\[
    = \frac{p_0 e^{-\eta t}}{p_0 e^{-\eta t} + p_0 \frac{v_1}{v_0 - v_1} e^{-\eta t} - p_0 \frac{v_1}{v_0 - v_1} e^{-\eta s} + p_0 e^{-\eta t}}
\]

\[
    = \frac{p_0 e^{-\eta t}}{p_0 \frac{v_1}{v_0 - v_1} (e^{-\eta s} - e^{-\eta t}) + p_0 e^{-\eta t}}
\]

which completes the proof. \( \square \)

Before we derive the formula for the likelihood function \( L(\lambda, \theta \mid \mathcal{C}) \) in the general case for \( N \) observed failure histories and \( M \) suspension histories, we first consider the case with a single failure history \( \mathcal{F} \), i.e. we have collected data \( \tilde{Y} = (y_\Delta, \ldots, y_{T\Delta}) \) and the system is known to have failed at time \( \xi = t \), where \( T\Delta < t \leq (T + 1)\Delta \). Since the observable data set \( \mathcal{D} = \{\mathcal{F}\} \) and the complete data set \( \mathcal{C} = \{\mathcal{F}\} \), we denote the likelihood function \( L(\lambda, \theta \mid \mathcal{C}) \) as \( L_{\mathcal{F}}(\lambda, \theta) \).

Since \( \tau_0 \) and \( \xi \) are sufficient for characterizing the sample paths of the state process, equations (5) – (7) imply that the likelihood function \( L_{\mathcal{F}}(\lambda, \theta) \) is given by
\[
L_{\tau}(\lambda, \theta) = \begin{cases} 
g_{Y_{\tau, \lambda}}(\bar{y} \mid t, \tau_0) f_{\tau_0}(\tau_0 \mid t) f_{\xi}(t), & \text{if } \tau_0 < t 
g_{Y_{\tau, \lambda}}(\bar{y} \mid t, t) m_{\tau_0}(t \mid t) f_{\xi}(t), & \text{if } \tau_0 = t 
\end{cases}
\] (8)

where \( g_{Y_{\tau, \lambda}}(y \mid t, s) \) is the conditional density of the observation process \( \bar{Y} = (Y_A, \ldots, Y_{\tau_0}) \) given \( \xi = t \) and \( \tau_0 = s \leq t \), which can be expressed in an explicit form.

For any \( s \in ((k-1)\Delta, k\Delta], k = 1, \ldots, T \), equation (2) implies that \( g_{Y_{\tau, \lambda}}(y \mid t, s) \) is given by

\[
g_{Y_{\tau, \lambda}}(\bar{y} \mid t, s) = g_{Y_{\tau, \lambda}}(\bar{y} \mid t, k\Delta) 
= \frac{1}{\sqrt{(2\pi)^T \det^{-1}(\Sigma_0) \det^{-1}(\Sigma_{k+1})}} \exp \left( \begin{array}{c} -\frac{1}{2} \sum_{n=1}^{k+1} (y_{n\Delta} - \mu_0)' \Sigma^{-1}_0 (y_{n\Delta} - \mu_0) \\
-\frac{1}{2} \sum_{n=k}^{T} (y_{n\Delta} - \mu_1)' \Sigma^{-1}_1 (y_{n\Delta} - \mu_1) \end{array} \right)
\] (9)

and for any \( s > T\Delta, g_{Y_{\tau, \lambda}}(y \mid t, s) \) is given by

\[
g_{Y_{\tau, \lambda}}(\bar{y} \mid t, s) = g_{Y_{\tau, \lambda}}(\bar{y} \mid t, t) 
= \frac{1}{\sqrt{(2\pi)^T \det^T(\Sigma_0)}} \exp \left( -\frac{1}{2} \sum_{n=1}^{T} (y_{n\Delta} - \mu_0)' \Sigma^{-1}_0 (y_{n\Delta} - \mu_0) \right)
\] (10)

We next consider the case where we have observed only a single suspension history \( \bar{S} \), i.e. we have collected data \( \bar{Y} = (Y_A, \ldots, Y_{\tau_0}) \) and stopped observing the operating system at time \( T\Delta \). Since the observable data set \( \partial = \{\bar{S}\} \) and the complete data set \( C = \{\bar{S}\} \), in this case we denote the likelihood function \( L(\lambda, \theta \mid C) \) as \( L_{\tau}(\lambda, \theta) \). For each \( s, t \in \mathbb{R}_+ \), it is not difficult to see that the conditional reliability function of \( \xi \) given \( \tau_0 \) is given by
\( h(t \mid s) := P(\xi > t \mid \tau_0 = s) = \begin{cases} p_0 e^{-\lambda(t-s)}, & t \geq s \\ 1, & t < s \end{cases} \)  
(11)

Furthermore, it is well-known that the density function of the unobservable sojourn time \( \tau_0 \) is given by

\[
 f_{\tau_0}(s) = \begin{cases} \nu_0 e^{-\nu_0 s}, & s \geq 0 \\ 0, & s < 0 \end{cases} \]  
(12)

Then equations (9) – (12) imply that the likelihood function \( L_{\xi}(\lambda, \theta) \) is given by

\[
 L_{\xi}(\lambda, \theta) = g_{\xi, \tau_0}(y \mid t, \tau_0) h(t \mid \tau_0) f_{\tau_0}(\tau_0) \]  
(13)

Thus, for the general case in which we have observed \( N \) independent failure histories \( \delta_1, \ldots, \delta_N \) and \( M \) independent suspension histories \( \tau_1, \ldots, \tau_M \), the likelihood function is given by

\[
 L(\lambda, \theta \mid \mathcal{C}) = \prod_{i=1}^{N} L_{\xi_i}(\lambda, \theta) \prod_{j=1}^{M} L_{\tau_j}(\lambda, \theta) \]  
(14)

where the likelihood functions for the individual failure and suspension histories are given by equations (8) and (13), respectively.
3.2. Formula for the Pseudo Likelihood

In this subsection, we are interested in carrying out the E-step of the EM algorithm, i.e. deriving the pseudo likelihood by taking the expectation of the likelihood function given by (14). As in the previous subsection, we first analyze the case in which we have observed only a single failure history \( \mathcal{F} \) of the form \( \hat{Y} = (y_1, \ldots, y_{T\Delta})' \) and \( \xi = t \), where \( T\Delta < t \leq (T+1)\Delta \). Thus, for any fixed estimates \( \hat{\lambda}, \hat{\theta} \) of the state and observations parameters, we are interested in deriving the formula for the pseudo likelihood function \( Q_\mathcal{F}(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_\mathcal{F}(\lambda, \theta | \mathcal{F}) \right) \), where the likelihood function \( L_\mathcal{F}(\lambda, \theta) \) is given in (8).

To simplify notation, for the remainder of the paper we denote vectors \( \lambda = (\lambda_{01}, \lambda_{02}, \lambda_{12})' \) and \( g = (g_{Y \xi, r_0}(y | t, \Delta), \ldots, g_{Y \xi, r_0}(y | t, T\Delta), g_{Y \xi, r_0}(y | t, t))' \). Also, for any vector \( v = (v_1, \ldots, v_n)' \), we denote \( \ln v = (\ln v_1, \ldots, \ln v_n)' \). The inner product \( \langle v, w \rangle := v'w \).

**Theorem 1.** Given a single failure history \( \mathcal{F} \), the pseudo likelihood function has the following decomposition

\[
Q_\mathcal{F}(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = Q_{\mathcal{F}}^{\text{state}}(\lambda | \hat{\lambda}, \hat{\theta}) + Q_{\mathcal{F}}^{\text{obs}}(\theta | \hat{\lambda}, \hat{\theta})
\]

where

\[
Q_{\mathcal{F}}^{\text{state}}(\lambda | \hat{\lambda}, \hat{\theta}) = \langle \hat{a}, \lambda \rangle + \langle \hat{b}, \ln \lambda \rangle
\]

\[
Q_{\mathcal{F}}^{\text{obs}}(\theta | \hat{\lambda}, \hat{\theta}) = \langle \hat{c}, \ln g \rangle
\]

for some vectors \( \hat{a}, \hat{b}, \text{ and } \hat{c} \) that depend only on the fixed estimates \( \hat{\lambda}, \hat{\theta} \).
Proof. Using equations (5) – (7) of Lemma 1 and the formula for the likelihood function $L_\tau(\lambda, \theta)$ given by (8),

\[
Q_\beta(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_\tau(\lambda, \theta) | \beta \right) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_\tau(\lambda, \theta) | \tilde{Y} = \tilde{y}, \tilde{\xi} = t \right)
\]

\[
= \left\{ \int_{s < t} \ln \left( g_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, s) f_{\eta, t} (s | t) f_{\xi, t} (t) \right) \hat{g}_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, s) \hat{f}_{\eta, t} (s | t) ds \right.
+ \ln \left( g_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, t) m_{\tau, t} (t | t) f_{\xi, t} (t) \right) \hat{g}_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, t) \hat{m}_{\tau, t} (t | t)
\left. \int_{u < t} \hat{g}_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, u) \hat{f}_{\eta, t} (u | t) du + \hat{g}_{\tilde{y}, \tilde{\xi}, \tau, \theta} (\tilde{y} | \tilde{t}, t) \hat{m}_{\tau, t} (t | t) \right) \right)
\]

where the notation $\hat{g}_{\tilde{y}, \tilde{\xi}, \tau, \theta}, \hat{f}_{\eta, t}, \hat{m}_{\tau, t}$ is used to signify that the functions $g_{\tilde{y}, \tilde{\xi}, \tau, \theta}, f_{\eta, t}, m_{\tau, t}$ are parameterized by fixed estimates $\hat{\lambda}, \hat{\theta}$. Since $g_{\tilde{y}, \tilde{\xi}, \tau, \theta}$ defined in (9) and (10) depends only on observation parameters $\theta = (\mu, \mu, \Sigma, \Sigma)$, and $f_{\xi}, f_{\eta, t}$, and $m_{\tau, t}$ depend only on state parameters $\lambda = (\lambda, \lambda, \lambda)$, the equation above can be decomposed into two terms $Q_\beta(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = Q^\text{state}_\beta(\lambda | \hat{\lambda}, \hat{\theta}) + Q^\text{obs}_\beta(\theta | \hat{\lambda}, \hat{\theta})$. Substituting equations (5) – (7) of Lemma 1, the first term $Q^\text{state}_\beta$ simplifies to
\[ Q^\text{state}_{\theta}(\lambda | \hat{\lambda}, \hat{\theta}) = \frac{\left\{ \int_{s<t} \ln \left( f_{\tau_{t_s}}(s \mid t) f_{v_{s}}(t) \right) \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, s) \hat{f}_{v_{t_s}}(s \mid t) ds \right.}{\left. + \ln \left( m_{\tau_{t_s}}(t \mid t) f_{v_{t_s}}(t) \right) \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, t) \hat{m}_{v_{t_s}}(t \mid t) \right\}} \]

\[ = \frac{\left\{ \int_{s<t} \ln \left( \hat{\lambda}_{21} e^{-\lambda_{21} t} e^{-(\lambda_{10} + \lambda_{01} - \hat{\lambda}_{12}) y} \right) \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, s) \hat{f}_{v_{t_s}}(s \mid t) ds \right.}{\left. + \ln \left( \hat{\lambda}_{22} e^{-\lambda_{22} t} e^{-(\lambda_{10} + \lambda_{02} - \hat{\lambda}_{21}) y} \right) \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, t) \hat{m}_{v_{t_s}}(t \mid t) \right\}} \]

\[ =: \hat{a}_{01} \lambda_{01} + \hat{a}_{02} \lambda_{02} + \hat{a}_{12} \lambda_{12} + \hat{b}_{01} \ln \lambda_{01} + \hat{b}_{02} \ln \lambda_{02} + \hat{b}_{12} \ln \lambda_{12} \]

where constants that depend only on fixed parameter estimates \( \hat{\lambda}, \hat{\theta} \) are given by

\[ \hat{a}_{01} = \hat{a}_{02} = -\frac{\hat{p}_{01} \hat{e}_2 e^{-\hat{\theta}_t}}{d} \langle \hat{e}_2, \hat{g} \rangle - \frac{\hat{p}_{02} e^{-\hat{\theta}_t}}{d} \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, t) \]
\[ \hat{a}_{12} = \frac{\hat{p}_{01} \hat{e}_2 e^{-\hat{\theta}_t}}{d} (\langle \hat{e}_2, \hat{g} \rangle - t \langle \hat{e}_1, \hat{g} \rangle) \]
\[ \hat{b}_{01} = \hat{b}_{12} = \frac{\hat{p}_{02} e^{-\hat{\theta}_t}}{d} \langle \hat{e}_1, \hat{g} \rangle \]
\[ \hat{b}_{02} = \frac{\hat{p}_{02} e^{-\hat{\theta}_t}}{d} \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, t) \]
\[ \hat{d} = \hat{p}_{01} \hat{e}_1 e^{-\hat{\theta}_t} \langle \hat{e}_1, \hat{g} \rangle + \hat{p}_{02} e^{-\hat{\theta}_t} \hat{g}_{Y_{t_s}, \tau_0}(y \mid t, t) \]

and vectors \( \hat{e}_1 = (\hat{e}_1^1, \ldots, \hat{e}_1^r, \hat{e}_1^r)' \) and \( \hat{e}_2 = (\hat{e}_2^1, \ldots, \hat{e}_2^r, \hat{e}_2^r)' \) are defined by
\[
\hat{e}_1^k = \int_{(k-1)\Delta}^{k\Delta} e^{-(\hat{b}_0-\hat{b}_1)u} du = \frac{e^{-(\hat{b}_0-\hat{b}_1)k\Delta} - e^{-(\hat{b}_0-\hat{b}_1)(k-1)\Delta}}{\hat{b}_0 - \hat{b}_1}, \quad k = 1, \ldots, T
\]
\[
\hat{e}_1^t = \int_{0}^{\Delta} e^{-(\hat{b}_0-\hat{b}_1)u} du = \frac{e^{-(\hat{b}_0-\hat{b}_1)\Delta} - e^{-(\hat{b}_0-\hat{b}_1)t}}{\hat{b}_0 - \hat{b}_1}.
\]
\[
\hat{e}_2^k = \int_{(k-1)\Delta}^{k\Delta} u e^{-(\hat{b}_0-\hat{b}_1)u} du = \frac{\hat{e}_1^k - k\Delta e^{-(\hat{b}_0-\hat{b}_1)k\Delta} + (k-1)\Delta e^{-(\hat{b}_0-\hat{b}_1)(k-1)\Delta}}{\hat{b}_0 - \hat{b}_1}, \quad k = 1, \ldots, T
\]
\[
\hat{e}_2^t = \int_{0}^{\Delta} u e^{-(\hat{b}_0-\hat{b}_1)u} du = \frac{\hat{e}_1^t - te^{-(\hat{b}_0-\hat{b}_1)t} + T\Delta e^{-(\hat{b}_0-\hat{b}_1)\Delta}}{\hat{b}_0 - \hat{b}_1},
\]

(18)

Similarly, the second term \( Q_{\theta}^{obs} \), which is a function only of the observation parameters \( \theta \), simplifies to

\[
Q_{\theta}^{obs}(\theta | \hat{\lambda}, \hat{\theta}) = \sum_{k=1}^{T} \hat{c}_k \ln \left( g_{\hat{\nu}, \hat{\tau}, \hat{t}, \hat{z}} \right) + \hat{c}_t \ln \left( g_{\hat{\nu}, \hat{\tau}, \hat{t}} \right)
\]

where constants that depend only on \( \hat{\lambda}, \hat{\theta} \) are given by

\[
\hat{c}_k = \frac{\hat{P}_0 \hat{B}_1 e^{-\hat{b}_1} \hat{e}_1^k}{d} \hat{g}_{\hat{\nu}, \hat{\tau}, \hat{t}}(\hat{z} | t, k\Delta), \quad k = 1, \ldots, T
\]
\[
\hat{c}_t = \left( \frac{\hat{P}_0 \hat{B}_1 e^{-\hat{b}_1} \hat{e}_1^t + \hat{P}_0 e^{-\hat{b}_1}}{d} \right) \hat{g}_{\hat{\nu}, \hat{\tau}, \hat{t}}(\hat{z} | t, t)
\]

(19)

To complete the proof we put \( \hat{a} = (\hat{a}_{01}, \hat{a}_{02}, \hat{a}_{12})', \hat{b} = (\hat{b}_0, \hat{b}_0, \hat{b}_1)', \) and \( \hat{c} = (\hat{c}_1, \ldots, \hat{c}_t, \hat{c}_t)' \). □
We next analyze the case in which we have observed only a single suspension history $\mathcal{S}$ of the form $\mathbf{Y} = (y_1, \ldots, y_T)$ and $\xi > T\Delta$. That is, for any fixed estimates $\hat{\lambda}, \hat{\theta}$ of the state and observations parameters, we are interested in deriving the formula for the pseudo likelihood function $Q_s(\hat{\lambda}, \theta \mid \hat{\lambda}, \hat{\theta}) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_s(\lambda, \theta) \mid \mathcal{S} \right)$, where the likelihood function $L_s(\lambda, \theta)$ is given in equation (13).

**Theorem 2.** Given a single suspension history $\mathcal{S}$, the pseudo likelihood function has the following decomposition

$$Q_s(\hat{\lambda}, \theta \mid \hat{\lambda}, \hat{\theta}) = Q_s^{\text{state}}(\lambda \mid \hat{\lambda}, \hat{\theta}) + Q_s^{\text{obs}}(\theta \mid \hat{\lambda}, \hat{\theta})$$

(20)

where

$$Q_s^{\text{state}}(\lambda \mid \hat{\lambda}, \hat{\theta}) = \langle \hat{a}, \lambda \rangle + \hat{y}_1 \ln(\lambda_{q_1}) + \hat{y}_2 \ln(\lambda_{q_1} + \lambda_{q_2})$$

$$Q_s^{\text{obs}}(\theta \mid \hat{\lambda}, \hat{\theta}) = \langle \hat{b}, \ln g \rangle$$

(21)

for some vectors $\hat{a}, \hat{b}, \hat{y}_1, \hat{y}_2$ that depend only on the fixed estimates $\hat{\lambda}, \hat{\theta}$.

**Proof.** Using equations (11) and (12) and the formula for the likelihood function $L_s(\lambda, \theta)$ given by (13),
\[ Q_s(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_s(\lambda, \theta) | \mathcal{S} \right) \]
\[ = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L_s(\lambda, \theta) | \hat{\mathcal{Y}} = \mathcal{Y}, \xi > t \right) \]
\[ = \int \ln \left( g_{\mathcal{Y}|t,s}(y | t,s) h(t | s) f_t(s) \right) \hat{g}_{\mathcal{Y}|t,s}(y \mid t, s) \hat{h}(t \mid s) \hat{f}_t(s) ds \]
\[ \int \hat{g}_{\mathcal{Y}|t,s}(\hat{y} | t, u) \hat{h}(t \mid u) \hat{f}_t(u) du \]

Since \( h \) and \( f_t \) defined in (11) and (12) depend only on state parameters \( \lambda = (\lambda_{01}, \lambda_{02}, \lambda_{12}) \), the equation above can be decomposed into two terms \( Q_s(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = Q_s^{\text{state}}(\lambda | \hat{\lambda}, \hat{\theta}) + Q_s^{\text{obs}}(\theta | \hat{\lambda}, \hat{\theta}) \), where the first term \( Q_s^{\text{state}} \) depends only on \( \lambda \) and the second term \( Q_s^{\text{obs}} \) depends only on \( \theta \). Substituting equations (11) and (12), the first term \( Q_s^{\text{state}} \) simplifies to

\[ Q_s^{\text{state}}(\lambda | \hat{\lambda}, \hat{\theta}) = \int \ln \left( h(t \mid s) f_t(s) \right) \hat{g}_{\mathcal{Y}|t,s}(y \mid t, s) \hat{h}(t \mid s) \hat{f}_t(s) ds \]
\[ \int \hat{g}_{\mathcal{Y}|t,s}(y \mid t, u) \hat{h}(t \mid u) \hat{f}_t(u) du \]
\[ = \int \ln \left( \lambda_{01} e^{-\lambda_{01} t} e^{-(\lambda_{02} + \lambda_{12}) s} + \lambda_{02} e^{-(\lambda_{01} + \lambda_{12}) t} \right) \hat{g}_{\mathcal{Y}|t,s}(y \mid t, s) \hat{h}(t \mid s) \hat{f}_t(s) ds \]
\[ + \int \ln \left( \lambda_{01} e^{-\lambda_{01} t} e^{-(\lambda_{02} + \lambda_{12}) s} + \lambda_{02} e^{-(\lambda_{01} + \lambda_{12}) t} \right) \hat{g}_{\mathcal{Y}|t,s}(y \mid t, s) \hat{h}(t \mid s) \hat{f}_t(s) ds \]
\[ = \hat{\alpha}_{01} \hat{\lambda}_{01} + \hat{\alpha}_{02} \hat{\lambda}_{02} + \hat{\alpha}_{12} \hat{\lambda}_{12} + \hat{\gamma}_1 \ln(\lambda_{01}) + \hat{\gamma}_2 \ln(\lambda_{01} + \lambda_{02}) \]

where constants that depend only on fixed parameter estimates \( \hat{\lambda}, \hat{\theta} \) are given by
\[
\hat{a}_{01} = \hat{a}_{02} = -\hat{\lambda}_0 e^{-\delta t} \langle \hat{e}_2, \hat{g} \rangle - \frac{(t + \tau_0^{-1}) e^{-\delta t}}{\delta} \hat{g}_{\psi, \tau_0} (\hat{y} | t, t) \\
\hat{a}_{12} = \hat{\lambda}_0 e^{-\delta t} \left( \langle \hat{e}_2, \hat{g} \rangle - t \langle \hat{e}_1, \hat{g} \rangle \right) \\
\hat{\gamma}_1 = \frac{\hat{\lambda}_0 e^{-\delta t}}{\delta} \langle \hat{e}_1, \hat{g} \rangle \\
\hat{\gamma}_2 = e^{-\delta t} \hat{g}_{\psi, \tau_0} (\hat{y} | t, t) \\
\hat{\delta} = \hat{\lambda}_0 e^{-\delta t} \langle \hat{e}_1, \hat{g} \rangle + e^{-\delta t} \hat{g}_{\psi, \tau_0} (\hat{y} | t, t)
\]

and vectors \( \hat{e}_1 \) and \( \hat{e}_2 \) are defined in (18). Similarly, the second term \( Q_s^{obs} \), which is a function only of the observation parameters \( \theta \), simplifies to

\[
Q_s^{obs} (\theta | \hat{\lambda}, \hat{\theta}) = \frac{\int \ln \left( g_{\psi, \tau_0} (\hat{y} | t, s) \hat{g}_{\psi, \tau_0} (\hat{y} | t, s) \hat{h}(t | s) \hat{f}_{\tau_0} (s) ds \right)}{\int \hat{g}_{\psi, \tau_0} (\hat{y} | t, u) \hat{h}(t | u) \hat{f}_{\tau_0} (u) du} \\
= \sum_{k=1}^{T} \hat{\beta}_k \ln \left( g_{\psi, \tau_0} (\hat{y} | t, k\Delta) \right) + \hat{\beta}_c \ln \left( g_{\psi, \tau_0} (\hat{y} | t, t) \right)
\]

where constants that depend only on \( \hat{\lambda}, \hat{\theta} \) are given by

\[
\hat{\beta}_k = \frac{\hat{\lambda}_0 e^{-\delta t} \hat{e}_1^k}{\delta} \hat{g}_{\psi, \tau_0} (\hat{y} | t, k\Delta), \quad k = 1, \ldots, T \\
\hat{\beta}_c = \left( \frac{\hat{\lambda}_0 e^{-\delta t} \hat{e}_1^c + e^{-\delta t}}{\delta} \right) \hat{g}_{\psi, \tau_0} (\hat{y} | t, t)
\]

To complete the proof we put \( \hat{\alpha} = (\hat{a}_{01}, \hat{a}_{02}, \hat{a}_{12})' \) and \( \hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_T, \hat{\beta}_c)' \).
Finally, for the general case in which we have observed \( N \) independent failure histories \( \mathcal{F}_1, \ldots, \mathcal{F}_N \) and \( M \) independent suspension histories \( \mathcal{S}_1, \ldots, \mathcal{S}_M \), Theorems 1 and 2 and equation (14) imply that the pseudo likelihood function is given by

\[
Q(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = E_{\hat{\lambda}, \hat{\theta}} \left( \ln L(\lambda, \theta | \varphi) \mid \varphi \right) \\
= E_{\hat{\lambda}, \hat{\theta}} \left( \ln \left( \prod_{i=1}^N L_{\mathcal{F}_i}(\lambda_i, \theta_i) \prod_{j=1}^M L_{\mathcal{S}_j}(\lambda_j, \theta_j) \right) \mid \varphi \right) \\
= \sum_{i=1}^N E_{\hat{\lambda}_i, \hat{\theta}_i} \left( \ln \left( L_{\mathcal{F}_i}(\lambda, \theta) \right) \mid \mathcal{F}_i \right) + \sum_{j=1}^M E_{\hat{\lambda}_j, \hat{\theta}_j} \left( \ln \left( L_{\mathcal{S}_j}(\lambda, \theta) \right) \mid \mathcal{S}_j \right) \\
= \sum_{i=1}^N Q_{\mathcal{F}_i}(\lambda, \theta | \hat{\lambda}_i, \hat{\theta}_i) + \sum_{j=1}^M Q_{\mathcal{S}_j}(\lambda, \theta | \hat{\lambda}_j, \hat{\theta}_j) \tag{24}
\]

Thus, to evaluate the pseudo likelihood function for all available histories, it suffices to evaluate the pseudo likelihood function for individual failure and suspension histories separately. Equation (24) completes the E-step of the EM algorithm. In the next subsection, we solve the M-step of the EM algorithm and derive explicit parameter update formulas for the maximizers of the pseudo likelihood function defined in (24).

### 3.3. Maximization of the Pseudo Likelihood Function

In this subsection we are interested in finding maximizers of the pseudo likelihood function defined in (24). By Theorems 1 and 2, the pseudo likelihood function can be decomposed as

\[
Q(\lambda, \theta | \hat{\lambda}, \hat{\theta}) = Q_{\text{state}}(\hat{\lambda}, \hat{\theta}) + Q_{\text{obs}}(\theta | \hat{\lambda}, \hat{\theta})
\]

where \( Q_{\text{state}} \) is a function only of the state parameters \( \lambda = (\lambda_0, \lambda_1, \lambda_2) \) and \( Q_{\text{obs}} \) is a function only of the observation parameters \( \theta = (\mu_0, \mu_1, \Sigma_0, \Sigma_1) \). This means that the M-step can be carried out separately for the state and observation parameters. Using equation (24) and Theorems 1 and 2, we solve for the stationary points of the state parameters \( \hat{\lambda} = (\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2) \). After some algebra, it is not
difficult to check that there is a unique stationary point \( \mathbf{x} = (\hat{x}_{01}, \hat{x}_{02}, \hat{x}_{12}) \) of the pseudo likelihood function given explicitly by

\[
\mathbf{x}_{01}^* = -\frac{\sum_{i=1}^{N} \hat{b}_{01}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \left( \sum_{i=1}^{N} \hat{b}_{01}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \right)}{\sum_{i=1}^{N} \hat{a}_{01}^i + \sum_{j=1}^{M} \hat{\alpha}_{01}^j} + \sum_{i=1}^{N} \hat{b}_{02}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \left( \sum_{i=1}^{N} \hat{b}_{01}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \right)}{\sum_{i=1}^{N} \hat{a}_{01}^i + \sum_{j=1}^{M} \hat{\alpha}_{01}^j} + \sum_{i=1}^{N} \hat{b}_{02}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \left( \sum_{i=1}^{N} \hat{b}_{01}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \right)}{\sum_{i=1}^{N} \hat{a}_{01}^i + \sum_{j=1}^{M} \hat{\alpha}_{01}^j}
\]

\[
\mathbf{x}_{02}^* = \mathbf{x}_{01}^* \left( \sum_{i=1}^{N} \hat{b}_{02}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \right)
\]

\[
\mathbf{x}_{12}^* = -\frac{\sum_{i=1}^{N} \hat{b}_{12}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \left( \sum_{i=1}^{N} \hat{b}_{12}^i + \sum_{j=1}^{M} \hat{\gamma}_1^j \right)}{\sum_{i=1}^{N} \hat{a}_{12}^i + \sum_{j=1}^{M} \hat{\alpha}_{12}^j}
\]

where constants \( \mathbf{a}^i = (\hat{a}_{01}^i, \hat{a}_{02}^i, \hat{a}_{12}^i)' \), \( \mathbf{b}^i = (\hat{b}_{01}^i, \hat{b}_{02}^i, \hat{b}_{12}^i)' \), \( \mathbf{a}' = (\hat{a}_{01}^i, \hat{a}_{02}^i, \hat{a}_{12}^i)' \), \( \hat{\gamma}_1^j \), and \( \hat{\gamma}_2^j \) are given in (17) and (22). Similarly, using equations (9) and (10), it follows that there is a unique stationary point of the observation parameters \( \mathbf{\theta} = (\mu^*, \Sigma_0^*, \Sigma_1^*) \) given explicitly by

\[
\mu_0^* = \frac{\sum_{i=1}^{N} \langle \hat{c}^i, n_1^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, n_1^j \rangle}{\sum_{i=1}^{N} \langle \hat{c}^i, d_1^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, d_1^j \rangle}, \quad \Sigma_0^* = \frac{\sum_{i=1}^{N} \langle \hat{c}^i, n_2^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, n_2^j \rangle}{\sum_{i=1}^{N} \langle \hat{c}^i, d_2^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, d_2^j \rangle},
\]

\[
\mu_1^* = \frac{\sum_{i=1}^{N} \langle \hat{c}^i, n_3^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, n_3^j \rangle}{\sum_{i=1}^{N} \langle \hat{c}^i, d_3^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, d_3^j \rangle}, \quad \Sigma_1^* = \frac{\sum_{i=1}^{N} \langle \hat{c}^i, n_4^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, n_4^i \rangle}{\sum_{i=1}^{N} \langle \hat{c}^i, d_4^i \rangle + \sum_{j=1}^{M} \langle \hat{p}^j, d_4^i \rangle},
\]

where vectors
and constants \( \hat{\mathbf{c}}^i = (\hat{c}_{i1}, \ldots, \hat{c}_{id}, \hat{c}_iy) \) and \( \hat{\beta}^j = (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{iy}^j, \hat{\beta}_{ij}^j) \) are defined in (19) and (23). This completes the M-step of the EM algorithm. In the next section, we present a numerical example which illustrates the entire estimation procedure.

4. Numerical Example

Suppose the state process \( \mathbf{X} = (X_t : t \in \mathbb{R}_+) \) follows a continuous time homogeneous Markov chain with transition rate matrix

\[
\begin{bmatrix}
-0.30 & 0.25 & 0.05 \\
0 & -0.10 & 0.10 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

We have chosen sampling interval \( \Delta = 1 \), and simulated 25 data histories (20 failure histories and 5 suspension histories) of a bivariate \( (d = 2) \) observation process \( \mathbf{Y} = (Y_n : n \in \mathbb{N}) \), which follows \( N_2(\mu_0, \Sigma_0) \) when \( X_n = 0 \), and \( N_2(\mu_i, \Sigma_i) \) when \( X_n = 1 \), where
\[ \mu_0 = \begin{pmatrix} 1.0 \\ 1.5 \end{pmatrix}, \mu_i = \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \Sigma_0 = \begin{pmatrix} 1.0 & 1.5 \\ 1.5 & 4.0 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1.5 & 3.0 \\ 3.0 & 8.0 \end{pmatrix} \]

Using update formulas (25) and (26), and the Euclidean norm stopping criterion \(|(\lambda_{n+1}, \psi_{n+1}) - (\lambda_n, \psi_n)| < 10^{-4}\), we have obtained the following results.

### Table 1. Iterations of the EM Algorithm

<table>
<thead>
<tr>
<th></th>
<th>Initial Values</th>
<th>Update 1</th>
<th>Update 2</th>
<th>Update 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\lambda}_{01})</td>
<td>0.27</td>
<td>0.27</td>
<td>0.29</td>
<td>0.30</td>
</tr>
<tr>
<td>(\hat{\lambda}_{02})</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>(\hat{\lambda}_{12})</td>
<td>0.12</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>(\hat{\mu}_0)</td>
<td>\begin{pmatrix} 0.0 \ 2.0 \end{pmatrix}</td>
<td>\begin{pmatrix} 0.9 \ 1.3 \end{pmatrix}</td>
<td>\begin{pmatrix} 0.9 \ 1.4 \end{pmatrix}</td>
<td>\begin{pmatrix} 0.9 \ 1.4 \end{pmatrix}</td>
</tr>
<tr>
<td>(\hat{\mu}_1)</td>
<td>\begin{pmatrix} 3.0 \ 3.0 \end{pmatrix}</td>
<td>\begin{pmatrix} 2.2 \ 2.8 \end{pmatrix}</td>
<td>\begin{pmatrix} 2.2 \ 2.7 \end{pmatrix}</td>
<td>\begin{pmatrix} 2.2 \ 2.9 \end{pmatrix}</td>
</tr>
<tr>
<td>(\hat{\Sigma}_0)</td>
<td>\begin{pmatrix} 1.0 &amp; 2.0 \ 2.0 &amp; 5.0 \end{pmatrix}</td>
<td>\begin{pmatrix} 0.9 &amp; 1.4 \ 1.4 &amp; 3.5 \end{pmatrix}</td>
<td>\begin{pmatrix} 0.9 &amp; 1.4 \ 1.4 &amp; 4.2 \end{pmatrix}</td>
<td>\begin{pmatrix} 1.0 &amp; 1.4 \ 1.4 &amp; 4.2 \end{pmatrix}</td>
</tr>
<tr>
<td>(\hat{\Sigma}_1)</td>
<td>\begin{pmatrix} 2.0 &amp; 2.5 \ 2.5 &amp; 7.0 \end{pmatrix}</td>
<td>\begin{pmatrix} 1.6 &amp; 3.3 \ 3.3 &amp; 8.5 \end{pmatrix}</td>
<td>\begin{pmatrix} 1.7 &amp; 3.2 \ 3.2 &amp; 8.3 \end{pmatrix}</td>
<td>\begin{pmatrix} 1.6 &amp; 3.0 \ 3.0 &amp; 8.2 \end{pmatrix}</td>
</tr>
<tr>
<td>(Q)</td>
<td>(-6.36 \times 10^4)</td>
<td>(-5.17 \times 10^4)</td>
<td>(-4.32 \times 10^4)</td>
<td>(-4.30 \times 10^4)</td>
</tr>
<tr>
<td>Time (sec)</td>
<td>–</td>
<td>10.2</td>
<td>7.4</td>
<td>8.1</td>
</tr>
</tbody>
</table>

Table 1 shows that each iteration of the EM algorithm takes around 10 seconds which is extremely fast for offline computations. Furthermore, the estimates converge rapidly in 3 iterations. For this numerical example, we have used only 25 data histories, including 5 suspension histories, and obtained very close estimates to the true simulation parameters. Clearly, better estimates of the model parameters would be expected as the number of available histories increases.
5. Conclusion

A parameter estimation problem for partially observable failing systems has been considered. System deterioration is driven by a continuous time homogeneous Markov chain and the system state is unobservable, except the failure state. Vector autoregressive information is obtained through condition monitoring at equidistant sampling times. Two types of data histories are available – data histories that end with observable failure and data histories that end when the system has been suspended from operation. The state and observation process has been modeled in the hidden Markov framework and the maximum likelihood estimates of the model parameters have been obtained using the EM algorithm. A numerical example has been developed to illustrate the estimation procedure using simulated data histories. It has been found that the procedure is both computationally efficient and converges rapidly to reasonably close parameter estimates, considering a relatively small number of data histories.

References


