

Article

Exploring Complex Survival Data through Frailty Modeling and Regularization

Xifen Huang ¹, Jinfeng Xu ^{1,*}  and Yunpeng Zhou ² ¹ School of Mathematics, Yunnan Normal University, Kunming 650092, China; 190004@ynnu.edu.cn² Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong, China; u3514104@connect.hku.hk

* Correspondence: 204203@ynnu.edu.cn

Abstract: This study addresses the analysis of complex multivariate survival data, where each individual may experience multiple events and a wide range of relevant covariates are available. We propose an advanced modeling approach that extends the classical shared frailty framework to account for within-subject dependence. Our model incorporates a flexible frailty distribution, encompassing well-known distributions, such as gamma, log-normal, and inverse Gaussian. To ensure accurate estimation and effective model selection, we utilize innovative regularization techniques. The proposed methodology exhibits desirable theoretical properties and has been validated through comprehensive simulation studies. Additionally, we apply the approach to real-world data from the Medical Information Mart for Intensive Care (MIMIC-III) dataset, demonstrating its practical utility in analyzing complex survival data structures.

Keywords: complex data; multivariate survival data; inverse Gaussian; log normal; general frailty model

MSC: 46N30; 62N01; 62N02



Citation: Huang, X.; Xu, J.; Zhou, Y. Exploring Complex Survival Data through Frailty Modeling and Regularization. *Mathematics* **2023**, *11*, 4440. <https://doi.org/10.3390/math11214440>

Academic Editor: Carmen Patino-Alonso

Received: 22 September 2023

Revised: 22 October 2023

Accepted: 24 October 2023

Published: 26 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, advancements in biomedicine, genomics, epidemiology, image processing, and other fields have led to a surge in high-dimensional data analysis emerging as a prominent theme in statistics. A notable example is the MIMIC-III (Medical Information Mart for Intensive Care) study, which aims to identify relevant variables for modeling and predicting two distinct periods: from ICU admission to discharge and from discharge to death. This study has generated high-dimensional, correlated multivariate failure time data which is commonly analyzed by the shared frailty models [1]. Specifically, the Cox model or relative risk model [2] with frailty is employed to incorporate dependence and assess the effects of covariates [3].

However, the observed likelihood of the Cox model with shared frailty often lacks an analytic form, except for the gamma frailty model. Consequently, parameter estimation of general frailty models is always computationally challenging. To perform variable selection for frailty models, Ref. [4] initially proposed a theory based on SCAD penalized regression. Ref. [5] developed an algorithm that utilizes Laplace transformation to handle frailty distributions in their general form. Ref. [6] applied the Laplace approximation of the full likelihood and developed an R package to implement their method. However, the methods mentioned above employ the Newton–Raphson method in their algorithms, resulting in significant computational burden when dealing with high-dimensional covariates. To avoid high-dimensional matrix inversion in the Newton–Raphson method, we try to use the MM (minorize-maximization) principle, to obtain the nonparametric maximum likelihood estimation for general frailty models, as the objective likelihood derived from the MM algorithm exhibits a monotonically increasing trend with reliable convergence properties when initialized appropriately [7].

In general, we present an MM algorithm for the Cox model with general frailties in this paper, highlighting its applicability in high-dimensional scenarios. We introduce a regularized estimation method, in which our algorithm decomposes the objective function with high-dimensional parameters into separable functions with low-dimensional parameters. This approach seamlessly integrates into the analysis of multivariate failure time data, effectively handling the challenges posed by its high-dimensional nature. LASSO is a commonly used penalty, but it can introduce notable biases in the resulting estimator. Therefore, we employ concave penalties such as SCAD [8] and MCP to conduct variable selection that gives consistent estimates.

The structure of the paper is as follows: In Section 2, we introduce the formulation of the Cox model with general frailties for high-dimensional multivariate failure time data. Section 3 presents our proposed MM algorithm for estimating model parameters. In Section 4, we introduce a regularized estimation method using the profile MM algorithm, leveraging sparse regression assumptions with high-dimensional parameters. We establish the convergence properties in Section 5. To assess the finite sample performances, we conducted simulation studies as described in Section 6. In Section 7, we illustrate the proposed methods, using the aforementioned MIMIC-III dataset. Finally, we provide further discussion in Section 8.

2. Data and Model Formulation

Assume there are J different types of events and there are n patients in the study. Denote the event time for the i -th studied subject ($i = 1, \dots, n$) and its j -th event type ($j = 1, \dots, J$) by T_{ij} , and the corresponding censoring time is denoted by C_{ij} . We use t_{ij} to denote the event time and let I_{ij} be the censoring indicator ($I_{ij} = 0$ for right-censored observation; $I_{ij} = 1$ otherwise) and $\mathbf{X}_{ij}^\top = (X_{ij1}, \dots, X_{ijp})$ be the high-dimensional covariates. The observed data consist of $Y_{obs} = \{t_{ij}, I_{ij}, \mathbf{X}_{ij}\}$. Given a noninformative censoring assumption, the Cox model with generality frailties assume that, conditional on the subject-specific frailties $\xi_i, i = 1, \dots, n, T_{ij}, j = 1, \dots, J$ are independent of one other and the conditional hazard function of T_{ij} given ξ_i and X_{ij} takes the form

$$\lambda_j(t|\mathbf{X}_{ij}, \xi_i) = \xi_i \lambda_{0j}(t) \exp\{\mathbf{X}_{ij}^\top \boldsymbol{\beta}\}, \tag{1}$$

where $\xi_i, i = 1, \dots, n$, are the individual-level frailty variables with density function $f(\xi_i|\gamma)$, which are independently and identically distributed. The baseline hazard of the j -th event is denoted by $\lambda_{0j}(\cdot)$, and $\boldsymbol{\beta}$ is a parameter vector. We denote $\Lambda_0 = (\Lambda_{01}, \dots, \Lambda_{0J})$ as the cumulative hazard functions of all events. The three components make up the model parameters, i.e., $\gamma, \boldsymbol{\beta}$, and the nonparametric components Λ_0 . To simplify the wording, we define $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \Lambda_0)$. The observed data likelihood function of model (1) can be written as

$$\begin{aligned} L(\boldsymbol{\theta}|Y_{obs}) & \\ = \prod_{i=1}^n \int_{\mathbb{W}} f(\xi_i|\gamma) \prod_{j=1}^J [\lambda_{0j}(t_{ij}) \xi_i \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta})]^{I_{ij}} \exp[-\Lambda_{0j}(t_{ij}) \xi_i \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta})] d\xi_i. & \end{aligned} \tag{2}$$

Generally, there is no closed-form solution of the frailty distribution's Laplace transform. Hence, we cannot find the explicit form of the marginal hazard in (2). In the following, we apply the MM principle, to deal with such an intractable case and to separate parameters when the model parameters are especially large, for fast and accurate estimation of model parameters. The MM principle is presented as follows: the minorization step first constructs a surrogate function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$, which satisfies

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \leq L(\boldsymbol{\theta}|Y_{obs}), \quad \forall \boldsymbol{\theta}, \boldsymbol{\theta}^{(k)} \in \Theta \quad \text{and} \quad Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k)}) = L(\boldsymbol{\theta}^{(k)}|Y_{obs}),$$

where $\theta^{(k)}$ denotes the current estimate of $\hat{\theta}$ in the k -th iteration. Function $Q(\cdot|\theta^{(k)})$ always lies under $L(\cdot)$ and is tangent to it at the point $\theta = \theta^{(k)}$. The maximization step then updates $\theta^{(k)}$ by $\theta^{(k+1)}$, which maximizes the surrogate function $Q(\cdot|\theta^{(k)})$ instead of $L(\cdot)$:

$$L(\theta^{(k+1)}|Y_{obs}) \geq Q(\theta^{(k+1)}|\theta^{(k)}) \geq Q(\theta^{(k)}|\theta^{(k)}) = L(\theta^{(k)}|Y_{obs}).$$

3. Likelihood Function and the Proposed Methodologies

From (2), we then formulate the log-likelihood function as

$$\ell(\theta|Y_{obs}) = \sum_{i=1}^n \log \int_{\mathbb{W}} \tau_i(\xi_i|\theta) d\xi_i, \tag{3}$$

where

$$\tau_i(\xi_i|\theta) = f(\xi_i|\gamma) \prod_{j=1}^J \left\{ [\lambda_{0j}(t_{ij})\xi_i \exp(\mathbf{X}_{ij}^T\beta)]^{I_{ij}} \exp[-\Lambda_{0j}(t_{ij})\xi_i \exp(\mathbf{X}_{ij}^T\beta)] \right\}. \tag{4}$$

Given the following weight function,

$$v_i(\xi_i|\theta^{(k)}) = \frac{\tau_i(\xi_i|\theta^{(k)})}{\int_{\mathbb{W}} \tau_i(\xi_i|\theta^{(k)}) d\xi_i},$$

then we can rewrite the objective function,

$$\ell(\theta|Y_{obs}) = \sum_{i=1}^n \log \left[\int_{\mathbb{W}} \frac{\tau_i(\xi_i|\theta)}{v_i(\xi_i|\theta^{(k)})} \cdot v_i(\xi_i|\theta^{(k)}) d\xi_i \right]. \tag{5}$$

According to the measure-theoretic form of Jensen’s inequality,

$$\Psi \left[\int_{\mathbb{X}} u(x) \cdot g(x) dx \right] \geq \int_{\mathbb{X}} \Psi[u(x)] \cdot g(x) dx, \tag{6}$$

where $\mathbb{X} \subset \mathbb{R}$, $\Psi(\cdot)$ is concave, $g(\cdot)$ is the corresponding density function on real line \mathbb{X} and $u(\cdot)$ is an arbitrary function on real line \mathbb{X} . Then, applying this inequality (6) to Equation (5), we let $v_i(\xi_i|\theta^{(k)})$ be the corresponding density function $g(\cdot)$, and the $u(\cdot)$ is defined as $\tau_i(\xi_i|\theta)/v_i(\xi_i|\theta^{(k)})$. $\Psi(\cdot)$ takes the concave log function. Thus, we have,

$$\begin{aligned} \ell(\theta|Y_{obs}) &\geq \sum_{i=1}^n \int_{\mathbb{W}} \log \left[\frac{\tau_i(\xi_i|\theta)}{v_i(\xi_i|\theta^{(k)})} \right] \cdot v_i(\xi_i|\theta^{(k)}) d\xi_i \\ &\geq \sum_{i=1}^n \int_{\mathbb{W}} \log \tau_i(\xi_i|\theta) \cdot v_i(\xi_i|\theta^{(k)}) d\xi_i + C_1, \end{aligned}$$

where C_1 is a constant. Then, substituting $\tau_i(\xi_i|\theta)$, using Equation (4), we can construct the following surrogate function for $\ell(\theta|Y_{obs})$:

$$Q_1(\theta|\theta^{(k)}) = Q_{11}(\gamma|\theta^{(k)}) + Q_{12}(\beta, \Lambda_0|\theta^{(k)}), \tag{7}$$

where

$$Q_{11}(\gamma|\theta^{(k)}) = \sum_{i=1}^n \int_{\mathbb{W}} \log[f(\xi_i|\gamma)] \cdot v_i(\xi_i|\theta^{(k)}) d\xi_i, \tag{8}$$

with parameter γ . And

$$Q_{12}(\boldsymbol{\beta}, \Lambda_0 | \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^n \sum_{j=1}^J \left[I_{ij} \log(\lambda_{0j}(t_{ij})) + I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - A_i^{(k)} \Lambda_{0j}(t_{ij}) \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right], \tag{9}$$

where $A_i^{(k)} = \int_{\mathbb{W}} \xi_i \cdot v_i(\xi_i | \boldsymbol{\theta}^{(k)}) d\xi_i, i = 1, \dots, n$. Therefore, $Q_1(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)})$ successfully separates the parameters $\gamma, \boldsymbol{\beta}$, and Λ_0 into (8) and (9), correspondingly. Then, in the second M-step, the γ will be updated by maximizing (8) numerically. However, the updating of $(\boldsymbol{\beta}, \Lambda_0)$ by maximizing (9) is challenging, due to the presence of the nonparametric components Λ_0 . Therefore, to tackle this issue, following [9], we utilize their profile estimation approach in (9), to profile out Λ_0 given $\boldsymbol{\beta}$, which results in the estimate of Λ_0 as

$$d\hat{\Lambda}_{0j}(t_{ij}) = \frac{I_{ij}}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})}, \quad j = 1, \dots, J. \tag{10}$$

By substituting (10) into (9), we furthermore have

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^J \left[I_{ij} \log(\lambda_{0j}(t_{ij})) + I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - A_i^{(k)} \Lambda_{0j}(t_{ij}) \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^J \left[I_{ij} \log\left(\frac{I_{ij}}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})}\right) + I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} \right. \\ & \quad \left. - \sum_{s=1}^n \frac{I(t_{sj} \leq t_{ij})}{\sum_{r=1}^n I(t_{rj} \geq t_{sj}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})} A_i^{(k)} \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^J \left\{ I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - I_{ij} \log \left[\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}) \right] \right. \\ & \quad \left. + I_{ij} \log I_{ij} \right\} - \sum_{j=1}^J \sum_{s=1}^n \sum_{i=1}^n \frac{I(t_{sj} \leq t_{ij}) A_i^{(k)} \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta})}{\sum_{r=1}^n I(t_{rj} \geq t_{sj}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})} \\ &= \sum_{i=1}^n \sum_{j=1}^J \left\{ I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - I_{ij} \log \left[\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}) \right] \right\} + C_2, \end{aligned}$$

where C_2 is a constant. Therefore,

$$Q_{13}(\boldsymbol{\beta} | \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^n \sum_{j=1}^J \left\{ I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - I_{ij} \log \left[\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}) \right] \right\}, \tag{11}$$

which includes parameter $\boldsymbol{\beta}$ only. It is obvious that $Q_{13}(\boldsymbol{\beta} | \boldsymbol{\theta}^{(k)})$ is of the same form as the Cox model’s log partial likelihood, where the modified Cox regression technique can be applied to obtain the updated estimates of $\boldsymbol{\beta}$, but this procedure will apply Newton’s method, which involves matrix inversion, which is computationally inefficient with large number of covariates. As the MM algorithm helps to separate the estimation of parameters, here we treat (11) as a new objective function and construct the minorizing function by reformulating the high-dimensional parameter estimation problem into a collection of low-dimensional optimizations. Here, we use the hyperplane inequality, which is generally applied in the MM algorithm [10]:

$$-\log(x) \geq -\log(x_0) - \frac{x - x_0}{x_0}. \tag{12}$$

In inequality (12), we take

$$x = \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}), \quad x_0 = \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)}).$$

Then,

$$\begin{aligned} Q_{13}(\boldsymbol{\beta}|\boldsymbol{\theta}^{(k)}) &\geq \sum_{i=1}^n \sum_{j=1}^J \left\{ I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - I_{ij} \left(\log \left[\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)}) \right] \right. \right. \\ &\quad \left. \left. - \frac{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}) - \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)})}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)})} \right) \right\} \\ &\geq \sum_{i=1}^n \sum_{j=1}^J \left[I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \frac{I_{ij} \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)})} \right] + C_3, \end{aligned}$$

with constant C_3 . Then we have the following surrogate function:

$$Q_{14}(\boldsymbol{\beta}|\boldsymbol{\theta}^{(k)}) = \sum_{i=1}^n \sum_{j=1}^J \left[I_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \frac{I_{ij} \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta})}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)})} \right]. \tag{13}$$

For the finite form of Jensen’s inequality:

$$\Psi \left[\sum x_i \cdot \omega_i \right] \geq \sum \Psi [x_i] \omega_i, \tag{14}$$

where ω_i are positive weights with $\sum \omega_i = 1$ and $\Psi(\cdot)$ is concave. As in [11], the concave function $-\exp(\cdot)$ can then adopt inequality (14), and a part of $Q_{14}(\boldsymbol{\beta}|\boldsymbol{\theta}^{(k)})$ can be rewritten as

$$\mathbf{X}_{rj}^\top \boldsymbol{\beta} = \sum_{p=1}^q \delta_{prj} [\delta_{prj}^{-1} X_{prj} (\beta_p - \beta_p^{(k)}) + \mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)}],$$

where $\delta_{prj} = |X_{prj}| / \sum_{p=1}^q |X_{prj}|$ is the weight in Jensen’s inequality. Thus,

$$-\exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}) \geq -\sum_{p=1}^q \delta_{prj} \exp [\delta_{prj}^{-1} X_{prj} (\beta_p - \beta_p^{(k)}) + \mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)}]. \tag{15}$$

Then, substituting (15) into $Q_{14}(\boldsymbol{\beta}|\boldsymbol{\theta}^{(k)})$, the minorizing function can be obtained as

$$Q_{15}(\beta_1, \dots, \beta_q|\boldsymbol{\theta}^{(k)}) \triangleq \sum_{p=1}^q Q_{15p}(\beta_p|\boldsymbol{\theta}^{(k)}), \tag{16}$$

where

$$\begin{aligned} Q_{15p}(\beta_p|\boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^n \sum_{j=1}^J \left\{ I_{ij} X_{pij} \beta_p \right. \\ &\quad \left. - \frac{I_{ij} \sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \delta_{prj} \exp [\delta_{prj}^{-1} X_{prj} (\beta_p - \beta_p^{(k)}) + \mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)}]}{\sum_{r=1}^n I(t_{rj} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rj}^\top \boldsymbol{\beta}^{(k)})} \right\}, \end{aligned} \tag{17}$$

for $p = 1, \dots, q$. In the first M-step of the profile MM algorithm, the final minorizing function using the profiled method for the objective log-likelihood $\ell(\boldsymbol{\theta}|Y_{obs})$ is

$$Q_{pro}(\boldsymbol{\gamma}, \boldsymbol{\beta}|\boldsymbol{\theta}^{(k)}) = Q_{11}(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(k)}) + \sum_{p=1}^q Q_{15p}(\beta_p|\boldsymbol{\theta}^{(k)}), \tag{18}$$

with the update of $d\Lambda_0$ by (10). From (18), maximizing the original objective function can be transferred into the maximization for a collection of $q + 1$ univariate functions, as the $\boldsymbol{\gamma}$ from $Q_{11}(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(k)})$ is one-dimensional in most cases. Thus, the next M-step is conducted by optimizing $q + 1$ univariate objective functions separately without the inefficient matrix inversion. We can see that there exist two integrals, $\int_{\mathbb{W}} \xi_i \cdot v_i(\xi_i|\boldsymbol{\theta}^{(k)}) d\xi_i$ and $\int_{\mathbb{W}} \log[f(\xi_i|\boldsymbol{\gamma})] \cdot v_i(\xi_i|\boldsymbol{\theta}^{(k)}) d\xi_i$. While our model is designed to address general frailty forms, the improper integrals presented above may not be computationally retrievable in extremely sparse and high-dimensional cases. However, as long as these integrals can be numerically calculated, our MM algorithm will always converge. Then, we propose the following estimation procedure.

4. The Algorithms

The parameter-separated surrogate functions proposed in Section 3 can cope well with sparsity-inducing penalties like SCAD or MCP. Therefore, in this section, the regularized estimation method is proposed, using the MM principle, as discussed in Section 3. Many variable selection penalties are special cases from the general form given in [8], and the likelihood function incorporated with the penalty term is written as

$$\ell^P(\boldsymbol{\theta}|Y_{obs}) = \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \Lambda_0|Y_{obs}) - n \sum_{p=1}^q P(|\beta_p|, \lambda), \tag{19}$$

where $\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \Lambda_0|Y_{obs})$ is discussed in the previous section with q -dimensional $\boldsymbol{\beta}$. λ is a non-negative tuning parameter that can also be set as λ_p in more general cases. The variable selection can be realized by shrinking some of the coefficients to zero, using a given penalty. With general frailties, the computation of MLEs is rather complicated, as the parameters involve three parts, $\boldsymbol{\gamma}, \boldsymbol{\beta}, \Lambda_0$, and it is more challenging when the number of parameters is high-dimensional, which is indeed our case. From the discussion in the previous section, our proposed profile MM algorithm separates all these parameters, which leads to an efficient and accurate estimation. This nice property of our proposed profile MM algorithm can mesh well with the different kinds of regularization penalties in (19), to produce efficient and accurate sparse estimation. Using the same profile MM strategies as in (18), as $Q_{pro}(\boldsymbol{\gamma}, \boldsymbol{\beta}|\boldsymbol{\theta}^{(k)})$ is the surrogate function for $\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \Lambda_0|Y_{obs})$, the corresponding minorization function for $\ell^P(\boldsymbol{\theta}|Y_{obs})$ can be expressed as

$$Q_{pro}(\boldsymbol{\gamma}, \boldsymbol{\beta}|\boldsymbol{\theta}^{(k)}) - n \sum_{p=1}^q P(|\beta_p|, \lambda), \tag{20}$$

where $P(\cdot, \lambda)$ is a concave function that is nondecreasing and piecewise differentiable, defined on $(0, \infty)$, Ref. [12] proposed an approximation approach for the penalty term, using quadratic functions:

$$-P(|\beta_p|, \lambda) \geq -P(|\beta_p^{(k)}|, \lambda) - \frac{[\beta_p^2 - (\beta_p^{(k)})^2]P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|}, \tag{21}$$

combining function (21) with (18):

$$\begin{aligned}
 & Q_{pro}(\gamma, \beta | \theta^{(k)}) - n \sum_{p=1}^q P(|\beta_p|, \lambda) \\
 & \geq Q_{pro}(\gamma, \beta | \theta^{(k)}) - n \sum_{p=1}^q \left(P(|\beta_p^{(k)}|, \lambda) + \frac{\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} \right) \\
 & \geq Q_{11}(\gamma | \theta^{(k)}) + \sum_{p=1}^q \left[Q_{15p}(\beta_p | \theta^{(k)}) - \frac{n\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} \right] + C_4.
 \end{aligned}$$

Therefore, the final surrogate function based on the MM algorithm for the penalized log-likelihood (19) is written as

$$Q_{pro}^P(\gamma, \beta | \theta^{(k)}, \lambda) = Q_{11}(\gamma | \theta^{(k)}) + \sum_{p=1}^q \left[Q_{15p}(\beta_p | \theta^{(k)}) - \frac{n\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} \right]. \tag{22}$$

Equation (22) decomposes the original maximizing function (19) to a sum of univariate functions that is more computationally efficient. Moreover, off-the-shelf accelerators can be applied, to improve the efficiency of the optimization problems. Then, we propose an alternative estimation procedure.

From the literature, there are different criteria for the selection of tuning parameter λ , such as the BIC (Bayesian information criterion [13]) and the GCV (generalized cross-validation [14]). In this paper, a BIC-type criterion is applied, which is defined by

$$BIC_\lambda = -2\ell(\hat{\theta}) + G_n(\hat{S} + 1) \log(n), \tag{23}$$

where $G_n = \max\{1, \log[\log(q + 1)]\}$ with q -dimensional β . \hat{S} is the degrees of freedom, and it denotes the number of estimates from $\hat{\beta}$ with nonzero values. To determine the optimal λ , we use a similar method to the one commonly used in the R package “glmnet” [15]. To determine the appropriate range of λ , we first conduct a search from \mathbb{R}^+ , using the BIC criteria. Next, we select the optimal λ , using a grid search within this determined range.

5. Theoretical Properties

The convergence properties are established in this section, for both the profile MM algorithm and the regularized MM algorithm. Let $Q(\theta | \theta^{(k)})$ be the minoring function based on the original objective $\ell(\cdot | Y_{obs})$, where θ denotes the parameters and $\theta^{(k)}$ is the k -th iteration’s estimation. The general convergence of the MM algorithm is provided by [16], as stated in Lemma 1 below. We denote $M(\theta)$ to be the maximizer of $Q(\cdot | \theta)$, given the following regularity conditions:

- I. The set \mathbb{S} of parameter θ is open \mathbb{R}^d .
- II. The objective function $\ell(\cdot | Y_{obs})$ is continuously differentiable.
- III. The set defined as $\mathbb{S}_c = \{\theta \in \mathbb{S} : \ell(\theta | Y_{obs}) \geq c\}$ is compact in \mathbb{R}^d .
- IV. The surrogate function $Q(\theta | \theta^{(k)})$ is continuously differentiable in θ and continuous in $\theta^{(k)}$.
- V. The objective function $\ell(\cdot | Y_{obs})$ ’s stationary points are isolated.
- VI. For the surrogate function $Q(\cdot | \theta^{(k)})$, there exists a unique global maximum.

Lemma 1. Let $\theta^{(k)}, k \in \mathbb{N}$ denote a sequence from the MM algorithm:

- (i) $M(\theta)$ is continuous at $\theta^{(k)}$ if VI is satisfied.

- (ii) If I–VI are satisfied, for any initial value $\theta^{(0)}$, when $k \rightarrow \infty$, $\theta^{(k)}$ tends to stationary point θ^* . Moreover, $M(\theta^*) = \theta^*$, and the likelihood $\ell(\theta^{(k)}|Y_{obs})$ sequence strictly increases to $\ell(\theta^*|Y_{obs})$ if $\theta^{(k)} \neq \theta^*$ for all k .

Therefore, based on this Lemma, given Condition 1 below, we have the convergence of our MM algorithm.

Condition 1. Conditions for the convergence of the profile MM algorithm:

- (a) $\ell(\cdot|Y_{obs})$ is continuously differentiable.
- (b) γ and β are compact, given $\ell(\theta|Y_{obs}) \geq c$.
- (c) Stationary points for $\ell(\theta|Y_{obs})$ are isolated.

Theorem 1. If Condition 1 holds, for any initial value of $\theta^{(0)} = \{\gamma^{(0)}, \beta^{(0)}, \Lambda_0^{(0)}\}$, the profile MM algorithm (Algorithm 1) converges to $\theta^{(*)}$.

Algorithm 1 Estimating Procedure.

- S1. Provide initial values for parameters γ , β , and Λ_0 .
 - S2. Update the estimation of parameter γ , using (8).
 - S3. Update the estimation of other parameters of covariates β_p , using (17) for $p = 1, \dots, q$.
 - S4. Compute the estimation of $\Lambda_{0j}(t_{ij})$, using (10) with the previous estimated β from S3.
 - S5. Conduct iterations from S2 to S4 repeatedly till convergence.
-

Proof.

$$\begin{aligned} \ell(\theta|Y_{obs}) &= \sum_{i=1}^n \log \left[\int_{\mathbb{W}} \frac{\tau_i(\xi_i|\theta)}{v_i(\xi_i|\theta^{(k)})} \cdot v_i(\xi_i|\theta^{(k)}) d\xi_i \right] \\ &\geq \sum_{i=1}^n \int_{\mathbb{W}} \log \left[\frac{\tau_i(\xi_i|\theta)}{v_i(\xi_i|\theta^{(k)})} \right] \cdot v_i(\omega_i|\theta^{(k)}) d\xi_i \\ &= Q_1(\theta|\theta^{(k)}). \end{aligned}$$

Therefore, the initial minorizing function $Q_1(\theta|\theta^{(k)})$ satisfies the condition that

$$\ell(\theta|Y_{obs}) \geq Q_1(\theta|\theta^{(k)}), \quad \ell(\theta^{(k)}|Y_{obs}) = Q_1(\theta^{(k)}|\theta^{(k)}).$$

Then, after profiling out Λ_0 ,

$$Q_{pro}(\gamma, \beta|\theta^{(k)}) = Q_{11}(\gamma|\theta^{(k)}) + \sum_{p=1}^q Q_{15p}(\beta_p|\theta^{(k)})$$

from (18), $Q_{11}(\gamma|\theta^{(k)})$ has a unique global maximum and $Q_{15p}(\beta_p|\theta^{(k)})$ is a unimodal function with a unique global maximum that verifies Condition VI. Condition I is easily followed from the form of $\ell(\cdot|Y_{obs})$. Condition II follows by Condition 1(a) and, hence, IV is also satisfied. Also, Condition V is followed by Condition 1(c). Then, we verify Condition III. It follows from the continuity of $\ell(\theta|Y_{obs})$ that Ω_c is closed. If $d\Lambda_{0j}(t_{ij})$ is unbounded, $\ell(\theta|Y_{obs}) \rightarrow -\infty$. By contrast, $d\Lambda_{0j}(t_{ij})$ is bounded when $\ell(\theta|Y_{obs}) \geq c$. Combined with Condition 1(b), III is satisfied. Thus, by Lemma 1, the profiled MM algorithm is convergent. \square

Theorem 2. If Condition 1 holds, for any initial value of $\theta^{(0)} = \{\gamma^{(0)}, \beta^{(0)}, \Lambda_0^{(0)}\}$, the regularized profile MM algorithm (Algorithm 2) converges to $\theta^{(*)}$.

Algorithm 2 An alternative method.

- S1. Provide initial values for parameters γ , β , and Λ_0 .
 - S2. Update the estimation of parameter γ , using (8).
 - S3. Under the profile MM method, β is updated by maximizing the equation $\sum_{p=1}^q \left[Q_{15p}(\beta_p | \theta^{(k)}) - n\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda) / 2|\beta_p^{(k)}| \right]$.
 - S4. Compute the estimates of $\Lambda_{0j}(t_{ij})$, using (10) with estimated β in S3.
 - S5. Conduct iterations from S2 to S4 repeatedly till convergence.
-

Proof. Note that, after profiling out Λ_0 , the surrogate function consists of $Q_{11}(\gamma | \theta^{(k)})$ and $Q_{15p}(\beta_p | \theta^{(k)}) - \frac{n\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|}$ with the unique global maximum, which verifies condition VI. For the other conditions, I–V, the proof is similar to Theorem 1. Thus, by Lemma 1, the regularized profiled MM algorithm is convergent. \square

6. Simulation Study

Example 1. We conducted simulations for the frailty models stated below:

$$\lambda_j(t | \mathbf{X}_{ij}, \zeta_i) = \zeta_i \lambda_{0j}(t) \exp\{\mathbf{X}_{ij}^\top \beta\}, \quad \zeta_i \sim \begin{cases} \text{log-normal}(0, \gamma), & \gamma = 1/4, \\ \text{inverse Gaussian}(\gamma, \gamma^2), & \gamma = 1, \\ \text{gamma}(1/\gamma, 1/\gamma), & \gamma = 2, \end{cases}$$

with two events, $J = 2$, a sample size $n = 300$, and $\lambda_{01}(t) = 3$, $\lambda_{02}(t) = 5/(1 + 5t)$. We set the true value β as $(-2_6^\top, -1_6^\top, 1_6^\top, 2_6^\top, 3_6^\top)^\top$ with dimension $q = 30$. Covariates X_i 's were generated independently from a uniform distribution $(0, 0.5)$. The generation process ensured that the censoring rate fell between 25% and 35%.

We tested the performance of our proposed profile MM algorithms, using the three different frailty models provided above. For computational issue, our algorithm separated the estimation of all parameters into univariate objectives, making it easily adaptable to off-the-shelf accelerators. Utilizing simple off-the-shelf accelerators [17], we can speed up our MM algorithms. As described in [17], there are several types of accelerators that utilize quasi-Newton approximation. For our article, we chose to adopt the squared iterative (SqS1) approach. The simulation results are presented in Table 1, where (MLE) denotes the average estimation of regression and frailty parameters, and (Bias) denotes their biases. (SD) is their estimated standard deviations, estimated using empirical standard deviation (for β_i , the estimated standard deviation was $\frac{1}{200} \sum_{k=1}^{200} (\hat{\beta}_i^{(k)} - \bar{\beta}_i)$ where $\hat{\beta}_i^{(k)}$ was the estimation of β_i in k^{th} replication), (K) is the number of iterations until convergence, (T) is the total run time, in terms of seconds, and the final value for the objective function is denoted by (L). The results indicate that our algorithms exhibited fast convergence and performed well across all three frailty models. Our algorithm also demonstrated good estimation accuracy, as evidenced by the nearly unbiased results and the high significance of the estimates of β , which was observed from their estimated standard deviations in all cases.

Example 2. In this example, we conducted a simulation for multivariate frailty models:

$$\lambda_j(t | \mathbf{X}_{ij}, \zeta_i) = \zeta_i \lambda_{0j}(t) \exp\{\mathbf{X}_{ij}^\top \beta\}, \quad \zeta_i \sim \begin{cases} \text{gamma}(1/\gamma, 1/\gamma), & \gamma = 2, \\ \text{log-normal}(0, \gamma), & \gamma = 0.5, \\ \text{inverse Gaussian}(\gamma, \gamma^2), & \gamma = 1, \end{cases}$$

for $i = 1, \dots, 400$, $j = 1, 2$, with $\lambda_{01}(t) = 3$, $\lambda_{02}(t) = 5/(1 + 5t)$. Here, we set the true coefficient vector β as $(2, 3, 4, 0_{27}^\top)^\top$ with dimension $q = 30$. We generated the covariates $\mathbf{X}^\top = (X_1, \dots, X_q)$ following multivariate normal distribution. The mean was equal to zero and

the covariance was $\Sigma = \{\rho^{|r-s|}\}_{rs}$ for $r, s = 1, \dots, q$ with a first-order autoregressive form, where $\rho = 0.2$. Similar to Example 1, there existed censored observation while the censoring proportion was set to be between 5% to 15%.

In the following simulation study, we tested the correctness of the variable selection of our proposed regularized profile MM methods in a high-dimensional regression model, using MCP and SCAD penalties with high parameter sparsity (19). The first derivative of the penalty function for SCAD was $P'(|\beta|, \lambda) = \lambda\{I(|\beta| \leq \lambda) + \frac{(a\lambda - |\beta|)_+}{(a-1)\lambda}I(|\beta| > \lambda)\}$, and for MCP it was $P'(|\beta|, \lambda) = \text{sign}(\beta)(\lambda - \frac{|\beta|}{a})$. According to [4,18], they suggested taking $a = 3.7$ for SCAD and $a = 3$ for MCP. To assess the performance of the model selection method, we conducted 200 replications. Based on these replications, we calculated the probability of obtaining the true model, where all covariates with non-zero parameters were estimated to have non-zero values, while covariates with zero parameters were not selected. This measure provided an evaluation of how well the method performed in correctly selecting the model that accurately represented the underlying data. We counted the number of correctly identified zero coefficients and the non-zero parameters that were estimated to be zero in each simulation run, and the average is shown in Table 2. The column titled ‘‘Correct’’ denotes the average count of true zero coefficients with zero estimated value, and the column titled ‘‘Incorrect’’ represents the average count of parameters where the true value was not zero but had an estimated zero value. We can observe that it consistently provided correct variable selection results across three different frailty models. Furthermore, we report the average parameter estimates and their bias and the corresponding mean squared error with 200 replications in Table 3. We observe that our proposed MM profile algorithms effectively handled penalties such as MCP and SCAD, producing accurate estimation results for various frailties. In particular, the inverse Gaussian frailty model with the SCAD penalty demonstrated excellent performance. Moreover, we also found that the estimates from the SCAD penalty were consistently more accurate than the results from the MCP penalty. The results from Tables 2 and 3 highlight the efficacy of our approach in dealing with different penalty functions and its ability to yield reliable estimation results for various types of frailty models.

Table 1. Simulation results of Example 1.

	Log-Normal Frailty			Inverse Gaussian Frailty			Gamma Frailty		
K	31.2780			162.1980			64.7648		
T	4.7591			12.6620			5.1203		
L	−2702.3917			−2586.3026			−2696.2794		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
γ	0.2729	0.0229	0.4236	0.8502	−0.1498	0.3856	2.1017	0.1017	0.4097
β_1	−2.0917	−0.0917	0.3814	−2.1046	−0.1046	0.4252	−2.0687	−0.0687	0.3844
β_5	−2.0692	−0.0692	0.4076	−2.1536	−0.1536	0.4224	−2.0894	−0.0894	0.3818
β_{10}	−1.0320	−0.0320	0.3822	−1.0597	−0.0597	0.3893	−1.0610	−0.0610	0.4010
β_{15}	1.0606	0.0606	0.3928	1.0692	0.0692	0.3973	1.0348	0.0348	0.4015
β_{20}	2.0530	0.0530	0.3851	2.1129	0.1129	0.4162	2.0891	0.0891	0.3830
β_{25}	3.1770	0.1770	0.4159	3.1770	0.1770	0.4159	3.1120	0.1120	0.3890
β_{30}	3.1071	0.1071	0.3849	3.2102	0.2102	0.4277	3.1410	0.1410	0.4070

Table 2. Simulation results based on three frailty models using two types of penalties. The sample size was 400, and 200 replications were conducted in Example 2.

Frailty	Sparsity Penalties	P (Selecting the True Model)	Zeros	
			Correct	Incorrect
Gamma	MCP ($a = 3$)	1	27	0
	SCAD ($a = 3.7$)	1	27	0

Table 2. Cont.

Frailty	Sparsity Penalties	P (Selecting the True Model)	Zeros	
			Correct	Incorrect
Log-Normal	MCP ($a = 3$)	1	27	0
	SCAD ($a = 3.7$)	1	27	0
Inverse Gaussian	MCP ($a = 3$)	1	27	0
	SCAD ($a = 3.7$)	1	27	0

Table 3. The estimated parameters' MLE, Bias, and SD (standard error). The sample size equaled 400 and the replications number was 200 in Example 2.

Frailty	Par.	MCP			SCAD		
		MLE	Bias	SD	MLE	Bias	SD
Gamma	γ	2.1012	-0.1012	0.3106	2.0678	-0.0687	0.3223
	β_1	2.0891	-0.0891	0.2438	1.9984	0.0016	0.2081
	β_2	3.1120	-0.1120	0.3517	2.9893	0.0107	0.2729
	β_3	4.1410	-0.1410	0.3459	4.0175	-0.0175	0.2920
Log-Normal	γ	0.4641	0.0359	0.1182	0.4696	0.0304	0.1101
	β_1	2.0563	-0.0563	0.1831	2.0282	-0.0282	0.1566
	β_2	2.9962	0.0038	0.3565	2.9705	0.0295	0.3080
	β_3	3.9756	0.0244	0.2931	4.0010	-0.0010	0.2230
Inverse Gaussian	γ	0.8999	0.1001	0.2024	1.0193	-0.0193	0.1717
	β_1	2.0562	-0.0562	0.2240	2.0305	-0.0305	0.1978
	β_2	3.0888	-0.0888	0.2591	3.0052	-0.0052	0.2548
	β_3	4.0915	-0.0915	0.1816	4.0055	-0.0055	0.1794

7. Real Data Application

MIMIC-III [19] is a dataset that includes information about patients who have been admitted to a critical care unit for various diseases. In this study, we specifically focused on a subset of 8516 patients who were admitted due to respiratory disease. The dataset included 82 covariates, which consisted of clinical variables such as mean blood pressure, heart rate, height, and other features, like whether the patient was transferred from another hospital. Two events were considered in our analysis. The first event was the time from admission to discharge from the ICU, and the second event was the time from discharge to death.

Three frailty models—namely gamma, log-normal, and inverse Gaussian—were fitted, using the regularized profile MM estimation method, with both MCP and SCAD penalties. The results are given in Table 4. The result obtained using MCP and SCAD penalties were similar, as they both selected the same set of significant variables, using the BIC criterion. For both the log-normal and the inverse Gaussian models, a total of 11 significant covariates were selected. These covariates were primarily clinical variables, as indicated by Tables 5 and 6. Among the 11 clinical variables that were considered significant, more than half of them were numerical variables, such as height. These numerical variables were considered to be more relevant to the two events being studied. On the other hand, other types of variables, such as categorical variables, may have had less of a direct relationship with the events. In addition to the clinical variables, it was observed that patients with booked admission (ELECTIVE) had a significant positive impact on the two events being studied. This suggests that patients who have planned admissions may have better outcomes compared to those who are admitted in emergency situations. On the other hand, variables related to patients' personal information, such as marital status, had smaller effects on the two events. This indicates that these personal factors may have had less influence on the outcomes being studied. For the Gamma frailty model, as presented by Table 7, more significant covariates were found: for example, the ethnicity was considered to be significant. It is reasonable that different frailty models will generate

different estimates. It is worth noting that the Gamma frailty model selected too many sparse features, including ethnicity, which may not have been preferable in this particular case. However, regardless of the choice of frailty, there were seven key features that were consistently selected by all models. This suggests that our model can consistently select those features which play a vital role in reflecting the length of a patient’s stay in the ICU.

Table 4. The minimum BIC scores (BIC) and number of significant variables from the three frailty models.

Penalty	BIC		No. of Significant Variables
		Log-Normal Frailty	
MCP	160,361		11
SCAD	160,361		11
		Inverse Gaussian Frailty	
MCP	175,223		11
SCAD	175,127		11
		Gamma Frailty	
MCP	160,684		31
SCAD	160,690		31

Table 5. Parameter estimates for covariates using the inverse Gaussian frailty model.

The Inverse Gaussian Frailty Model	
	Clinical
Fraction of inspired oxygen	0.221
Glucose	−0.128
Heart Rate	−0.069
Height	−0.173
Oxygen saturation	0.109
Respiratory rate	−0.147
Elective	0.486
Emergency room admit (Admission location)	−0.152
Phys referral/normal deli (Admission location)	0.220
	Personal Information
Engl (Language)	0.156
Married (Marriage status)	0.079

Table 6. Parameter estimates for covariates, using the log-normal frailty model.

The Log-Normal Frailty Model	
	Clinical
Fraction of inspired oxygen	0.208
Glucose	−0.120
Heart Rate	−0.065
Height	−0.165
Oxygen saturation	0.098
Respiratory rate	−0.139
Elective	0.433
Emergency room admit (Admission location)	−0.138
Phys referral/normal deli (Admission location)	0.229
	Personal Information
Engl (Language)	0.134
Married (Marriage status)	0.070

Table 7. Parameter estimates for covariates, using the gamma frailty model.

The Gamma Frailty Model	
Clinical	
Fraction of inspired oxygen	0.246
Glucose	−0.048
Height	−0.162
Respiratory rate	−0.141
Elective	0.533
Phys referral/normal deli (Admission location)	0.365
TRANSFER FROM OTHER HEALT	0.833
Personal Information	
Engl (Language)	0.220
Separated (Marriage status)	0.355
Unitarian universalist	0.338
Ethnicity	
Asian	−0.317
Hispanic/Latino–Puerto Rican	0.373
Multi-race ethnicity	0.262
Asian–other	−0.265
Hispanic/Latino–Colombian	1.790
Hispanic/Latino–Dominican	0.465
Middle eastern	0.306
Hispanic/Latino–Cuban	0.664
Asian–Asian Indian	0.664
White–eastern European	0.362
White–Brazilian	0.888
Portuguese	−0.315
Hispanic/Latino–Mexican	−0.345
Asian–Japanese	−1.110
Hispanic/Latino–Salvadoran	0.808
American Indian/Alaska native	−0.841
Federally recognized tribe	−0.841
Asian–Filipino	−0.465
Asian–Korean	0.263
Hispanic/Latino–Guatemalan	−0.221
American Indian/Alaska native	−0.426
Asian–Cambodian	0.240

8. Discussion

We have developed an innovative algorithm for parameter estimation in the analysis of complex multivariate failure time data with high-dimensional covariates. This type of data presents challenges, due to its intricate nature and the presence of multiple correlated survival outcomes. Estimating frailty models, which incorporate random effects to capture unobserved heterogeneity, typically requires nonparametric maximum likelihood estimation, due to the unknown baseline hazard function. Most existing researches apply the gamma frailty model because it provides an explicit formula for parameter estimation in each iteration, reducing the computational burden. By contrast, our method offers the advantage of being applicable to general frailty models and providing efficient estimation results, even in high-dimensional cases. This is achieved through the incorporation of regularization methods.

Our proposed algorithm efficiently addresses these challenges by accurately estimating the high-dimensional parameters. Our method avoids the Laplace approximation and the traditional Newton–Raphson method, which can result in inaccurate and inefficient estimation when dealing with high-dimensional data. A significant contribution of our method is the utilization of a decomposition approach, which splits the minorizing function into a collection of univariate functions. This approach improves computational efficiency and ensures robust parameter estimation. In high-dimensional cases, the traditional EM

algorithm for parameter estimation of frailty models can lead to significant computational costs. In practical applications involving various clinical and genetic covariates, our algorithm enables the estimation procedure to be completed within a reasonable timeframe, thus facilitating effective identification of key features.

Importantly, our algorithm can be applied to various settings and complexities of multivariate survival data without relying on specific regularization techniques. In addition to the multiple event data modeled in this article, our algorithm can be further extended, to handle data with different structures of correlation, such as recurrent event data and clustered survival data with general frailties. The package “frailtyMMpen” can be downloaded from CRAN, providing convenient support for setting up various types of survival data and regularization methods in the model. By leveraging this algorithm, researchers and practitioners can effectively analyze and gain insights from real-world applications of multivariate survival data. Understanding the relationships between covariates and survival outcomes is crucial for informed decision-making and predictive modeling in these scenarios.

Author Contributions: Data curation, X.H. and Y.Z.; Formal analysis, Y.Z. and X.H.; Investigation, X.H. and Y.Z.; Methodology, X.H. and Y.Z.; Project administration, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151. [[CrossRef](#)]
2. David, C.R. Regression models and life-tables (with discussion). *J. R. Stat. Soc. B* **1972**, *34*, 187220.
3. Andersen, P.K.; Klein, J.P.; Knudsen, K.M.; Tabanera y Palacios, R. Estimation of variance in Cox’s regression model with shared gamma frailties. *Biometrics* **1997**, *53*, 1475–1484. [[CrossRef](#)] [[PubMed](#)]
4. Fan, J.; Li, R. Variable selection for Cox’s proportional hazards model and frailty model. *Ann. Stat.* **2002**, *30*, 74–99. [[CrossRef](#)]
5. Androulakis, E.; Koukouvinos, C.; Vonta, F. Estimation and variable selection via frailty models with penalized likelihood. *Stat. Med.* **2012**, *31*, 2223–2239. [[CrossRef](#)] [[PubMed](#)]
6. Groll, A.; Hastie, T.; Tutz, G. Selection of effects in Cox frailty models by regularization methods. *Biometrics* **2017**, *73*, 846–856. [[CrossRef](#)] [[PubMed](#)]
7. Becker, M.P.; Yang, I.; Lange, K. EM algorithms without missing data. *Stat. Methods Med. Res.* **1997**, *6*, 38–54. [[CrossRef](#)] [[PubMed](#)]
8. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
9. Johansen, S. An extension of Cox’s regression model. *Int. Stat. Rev.* **1983**, *51*, 165–174. [[CrossRef](#)]
10. Hunter, D.R.; Lange, K. A tutorial on MM algorithms. *Am. Stat.* **2004**, *58*, 30–37. [[CrossRef](#)]
11. Ding, J.; Tian, G.L.; Yuen, K.C. A new MM algorithm for constrained estimation in the proportional hazards model. *Comput. Stat. Data Anal.* **2015**, *84*, 135–151. [[CrossRef](#)]
12. Hunter, D.R.; Li, R. Variable selection using MM algorithms. *Ann. Stat.* **2005**, *33*, 1617. [[CrossRef](#)] [[PubMed](#)]
13. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
14. Craven, P.; Wahba, G. Smoothing noisy data with spline functions. *Numer. Math.* **1978**, *31*, 377–403. [[CrossRef](#)]
15. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)] [[PubMed](#)]
16. Vaida, F. Parameter convergence for EM and MM algorithms. *Stat. Sin.* **2005**, *15*, 831–840.
17. Zhou, H.; Alexander, D.; Lange, K. A quasi-Newton acceleration for high-dimensional optimization algorithms. *Stat. Comput.* **2011**, *21*, 261–273. [[CrossRef](#)] [[PubMed](#)]
18. Ma, S.; Huang, J. A concave pairwise fusion approach to subgroup analysis. *J. Am. Stat. Assoc.* **2017**, *112*, 410–423. [[CrossRef](#)]
19. Johnson, A.E.; Pollard, T.J.; Shen, L.; Lehman, L.w.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.