Supervised Matrix Factorization for Survival Analysis: Objective Function & Parameter Update Algorithm

Andrew Walther

January 15, 2026

1 Introduction

This derivation builds off of the original likelihood & estimation work included in "Supervised Matrix Factorization for Survival Analysis: Likelihood & Estimation". In this document, we focus on providing more detail surrounding the final objective function for the joint matrix factorization and survival analysis components as well as a comprehensive algorithm for parameter updates of the Loadings matrix (L), Factor matrix (F), and survival regression coefficient vector (β).

In the following sections, we present a comprehensive derivation of the Supervised Bayesian Matrix Factorization model with Cox Survival Integration. This derivation includes a 2nd-order Taylor Expansion of the survival term and an explicit Block Coordinate Descent algorithm for parameter estimate updates.

2 Derivation of the Objective Function

We aim to maximize the posterior probability of the parameters L (loadings), F (factors), and β (Cox Survival regression coefficients) given the observed data Y and survival information (t, δ). Note that this objective function is equivalent to minimizing the Negative Log-Posterior.

2A Taylor Expansion of the Survival Term

The Cox partial log-likelihood, $\ell_{cox}(\eta)$, is a function of the linear predictor $\eta = L\beta$. We approximate it with a 2nd-Order Taylor Series Expansion around the current estimate $\hat{\eta}^{(b-1)}$ which is obtained from the previous iteration with index $(b-1)$.

This Taylor series expansion is:

$$\ell_{cox}(\eta) \approx \ell_{cox}\left(\hat{\eta}^{(b-1)}\right)^T \mathbf{u}^{(b-1)} + \frac{1}{2}\left(\eta - \hat{\eta}^{(b-1)}\right)^T \mathbf{H}^{(b-1)} \left(\eta - \hat{\eta}^{(b-1)}\right) \tag{1}$$

where:

- $\mathbf{u}^{(b-1)}$ is the gradient/score vector

$$u_i = \frac{\partial \ell(\eta)}{\partial \eta_i} = \delta_i - \sum_{j:i\in R(t_j)} \delta_j \left(\frac{\exp(\eta_i)}{\sum_{k\in R(t_j)} \exp(\eta_k)}\right) \tag{2}$$

  which is used to compute the "step direction" for the pseudo-response $\mathbf{z} = \hat{\eta} + \mathbf{u}/\mathrm{diag}(-\mathbf{H})$

- $\mathbf{H}^{(b-1)}$ is the hessian/information matrix

$$H_{ii} = \frac{\partial^2 \ell(\eta)}{\partial \eta_i^2} = -\sum_{j:i\in R(t_j)} \delta_j \pi_{ij}(1-\pi_{ij}) \tag{3}$$

which is used to calculate the "weights" (W) that determine how much influence subject i has on the update, $W = \text{diag}(-\mathbf{H})$

2B Transformation to Weighted Least Squares

Define the objective function for the survival component as the Negative log-likelihood, $J_{surv}(\eta) = -\ell_{cox}(\eta)$. Substituting the Taylor expansion and defining the weight matrix $W^{(b-1)} = -\text{diag}\left(\mathbf{H}^{(b-1)}\right)$, yields:

$$J_{surv}(\eta) \approx -\left(\eta - \hat{\eta}^{(b-1)}\right)^T \mathbf{u}^{(b-1)} + \frac{1}{2}\left(\eta - \hat{\eta}^{(b-1)}\right)^T W^{(b-1)} \left(\eta - \hat{\eta}^{(b-1)}\right) \tag{4}$$

Completing the square allows this expression to be transformed into a quadratic form. The pseudo-response (working data) $\mathbf{z}$ is defined as:

$$\mathbf{z}^{(b-1)} = \hat{\eta}^{(b-1)} + \left(W^{(b-1)}\right)^{-1} \mathbf{u}^{(b-1)} \tag{5}$$

so we have the simplified quadratic surrogate expression:

$$J_{surv}(\eta) \propto \frac{1}{2}\left(\mathbf{z}^{(b-1)} - L\beta\right)^T W^{(b-1)} \left(\mathbf{z}^{(b-1)} - L\beta\right) \tag{6}$$

2C Final Objective Function ($\mathcal{Q}$)

We obtain the final comprehensive objective function by combining the Gaussian likelihood for the Genomics matrix factorization component, the survival quadratic surrogate, and the L2 Ridge regularization priors and minimize for iteration b:

$$\mathcal{Q}(L, F, \beta) = \left[\frac{\tau_y}{2}\|Y - LF^T\|_F^2\right] + \left[\frac{1}{2}\left(\mathbf{z}^{(b-1)} - L\beta\right)^T W^{(b-1)} \left(\mathbf{z}^{(b-1)} - L\beta\right)\right] + \left[\frac{\tau_L}{2}\|L\|_F^2 + \frac{\tau_F}{2}\|F\|_F^2 + \frac{\tau_\beta}{2}\|\beta\|_2^2\right] \tag{7}$$

where

$$\frac{\tau_y}{2}\|Y - LF^T\|_F^2 \tag{8}$$

is the matrix factorization genomics data fit

$$\frac{1}{2}\left(\mathbf{z}^{(b-1)} - L\beta\right)^T W^{(b-1)} \left(\mathbf{z}^{(b-1)} - L\beta\right) \tag{9}$$

is the survival fit with a 2nd order approximation, and

$$\frac{\tau_L}{2}\|L\|_F^2 + \frac{\tau_F}{2}\|F\|_F^2 + \frac{\tau_\beta}{2}\|\beta\|_2^2 \tag{10}$$

are the L2 regularization priors.

3 Explicit Parameter Update Algorithm

For an algorithm for parameter estimate updates, we proceed with block coordinate descent. We iterate on the index b, updating parameters based on the existing values from the previous iteration index $b-1$. For $b = 0$, set initial values for the following quantities:

- Initialize $L^{(0)}, F^{(0)}$ via matrix factorization
- Initialize $\beta^{(0)} = \mathbf{0}$
- Specify hyperparameters: $\tau_y, \tau_L, \tau_F, \tau_\beta$

Then iterate over the index $b = 1, 2, \cdots$ until convergence for the following steps (1-5):

3A 1) Update Survival Approximation (z, W)

For this step, we condition on $L^{(b-1)}$ and $\beta^{(b-1)}$. To update the survival pseudo-response (z), proceed with the following:

- **Compute Linear Predictor:**
$$\hat{\eta}^{(b-1)} = L^{(b-1)}\beta^{(b-1)} \tag{11}$$

- Compute Gradient $(\mathbf{u})$:

  For each subject i, sum over all event times t_j where subject i is at risk $(t_i \geq t_j)$:
$$u_i^{(b-1)} = \delta_i - \sum_{j:i\in R(t_j)} \delta_j \frac{\exp(\hat{\eta}_i^{(b-1)})}{\sum_{k\in R(t_j)} \exp(\hat{\eta}_k^{(b-1)})} \tag{12}$$

- Compute the Hessian Diagonal (H) and Weights (W):
$$H_{ii}^{(b-1)} = -\sum_{j:i\in R(t_j)} \delta_j \left[\frac{\exp(\hat{\eta}_i^{(b-1)})}{\sum_{k\in R(t_j)} e^{\hat{\eta}_k}} - \left(\frac{\exp(\hat{\eta}_i^{(b-1)})}{\sum_{k\in R(t_j)} e^{\hat{\eta}_k}} \right)^2 \right] \tag{13}$$
$$W^{(b-1)} = \operatorname{diag}(-H^{(b-1)}) \tag{14}$$

- Compute the survival Pseudo-Response $(\mathbf{z})$:
$$\mathbf{z}^{(b-1)} = \hat{\eta}^{(b-1)} + (W^{(b-1)})^{-1}\mathbf{u}^{(b-1)} \tag{15}$$

3B 2) Update Cox Coefficients (β)

For this step, we condition on $L^{(b-1)}, z^{(b-1)}$, and $W^{(b-1)}$. Then update β by weighted ridge regression of $\mathbf{z}$ on L such that:
$$\beta^{(b)} = \left((L^{(b-1)})^T W^{(b-1)} L^{(b-1)} + \tau_\beta I \right)^{-1} (L^{(b-1)})^T W^{(b-1)} \mathbf{z}^{(b-1)} \tag{16}$$

3C 3) Supervised: Update Latent Factors (L)

For this step, we condition on $F^{(b-1)}, \beta^{(b)}, z^{(b-1)}$ and $W^{(b-1)}$. Fuse the Genomics MF data (Y) and survival information (z). Then update row-wise for each subject i. Let Y_i be the i-th row of Y with length p. We then treat Y as a column vector for the update scheme $(Y_{i.})^T$. Then for each subject $i = 1, \cdots, n$:

- Posterior Precision (a $K \times K$ matrix):
$$A_i^{(b)} = \underbrace{\tau_y (F^{(b-1)})^T F^{(b-1)}}_{\text{Genomics Matrix Factorization}} + \underbrace{W_{ii}^{(b-1)} \beta^{(b)} (\beta^{(b)})^T}_{\text{Survival}} + \underbrace{\tau_L I}_{\text{Prior}} \tag{17}$$

- Posterior Mean Weighted Sum (a $K \times 1$ vector):
$$\mathbf{v}_i^{(b)} = \underbrace{\tau_y (F^{(b-1)})^T (Y_{i.})^T}_{\text{Genomics}} + \underbrace{W_{ii}^{(b-1)} z_i^{(b-1)} \beta^{(b)}}_{\text{Survival}} \tag{18}$$

- Yields the update L_i:
$$(L_i^{(b)})^T = (A_i^{(b)})^{-1} \mathbf{v}_i^{(b)} \tag{19}$$

3D 4) Update Loadings (F)

For this step, we condition on $L^{(b)}$ and the update depends only on Y. Update row-wise for each gene j. Let $Y_{\cdot j}$ be the j-th column of Y with length n. Then for each gene $j = 1, \cdots, p$:

- Posterior Precision:
$$B^{(b)} = \tau_y \left(L^{(b)}\right)^T L^{(b)} + \tau_F I \tag{20}$$
- Each Update F_j:
$$\left(F_j^{(b)}\right)^T = (B^{(b)})^{-1} \left(\tau_y (L^{(b)})^T Y_{\cdot j}\right) \tag{21}$$

This yields a full update $F^{(b)}$ over all rows j.

3E 5) Assess Convergence Criteria

Compute the relative change in the objective function $\mathcal{Q}$ between the current and previous iterations:

$$\Delta = \frac{|\mathcal{Q}^{(b)} - \mathcal{Q}^{(b-1)}|}{|\mathcal{Q}^{(b-1)}|} \tag{22}$$

Then consider if $\Delta < \epsilon$ (for some small value), stop and consider the parameter values at b. Otherwise, set $b \leftarrow b + 1$ and continue iterating over steps 1-5.