Solving for β to estimate the B-spline bases resolves to finding the score function $\mathbf{S}(\beta)$ and Hessian matrix $\mathbf{H}(\beta)$, which are respectively the first derivative of the partial log-likelihood function with respect to β_j and the $p \times p$ matrix of second derivatives of the log-likelihood function with respect to β_j, β_k.

$$\begin{aligned}
\mathbf{S}(\beta) &= \frac{\partial}{\partial \beta_j}\phi(y, \delta|\mathbf{x}, \beta) \\
&= \delta B_j(y|\mathbf{x}) - \int_0^y B_j(u|\mathbf{x}) \exp(\alpha(u|\mathbf{x}, \beta))du, \quad 1 \leq j \leq p, \quad y \geq 0, \quad \delta \in \{0, 1\} \\
\mathbf{H}(\beta) &= \frac{\partial^2}{\partial \beta_j \partial \beta_k}\phi(y, \delta|\mathbf{x}, \beta) \\
&= -\int_0^y B_j(u|\mathbf{x})B_k(u|\mathbf{x}) \exp(\alpha(u|\mathbf{x}, \beta))du, \quad 1 \leq j, \quad k \leq p, \quad y \geq 0, \quad \delta \in \{0, 1\}.
\end{aligned}$$

This means the estimation procedure is not much different from most numerical solutions to statistical computations. In HARE, the Newton-Raphson method is used to estimate $\hat{\beta}$. The initial value for $\hat{\beta}$ is $\hat{\beta}^{(0)}$ and $\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - 2^{-\nu}[\mathbf{H}(\hat{\beta}^{(m)})]^{-1}\mathbf{S}(\hat{\beta}^{(m)})$ where ν is a step-halving constant (Lange, 2010). The iterations stop when the difference between repeated log-likelihood calculations is $< 10^{-6}$. The primary goal for HARE, then, is to determine what basis functions should be used to estimate $\alpha(t|\mathbf{x})$.

Denote the space of all acceptable basis functions B_j for HARE as $\mathcal{G}$. B_j are linear, which minimizes numerical integrations over the knot sequence of $\mathbf{t} := (t_1, t_2, \ldots, t_k)$ (Kooperberg, Stone, & Truong, 1995a) and ensures L_2 convergence in function estimation. (Stone, 1994; Kooperberg, Stone, & Truong, 1995b). Formally, the allowable spaces $G \in \mathcal{G}$ in HARE are defined as follows:

- There is only one $G \in \mathcal{G}$ with minimal dimension p_{min},
- Each $G \in \mathcal{G}$ is a linear space having dimension $p \geq p_{min}$,
- If $G \in \mathcal{G}$ has dimension $p > p_{min}$, then there is at least one subspace $G_0 \in \mathcal{G}$ of G with dimension $p - 1$,
- If $G_0 \in \mathcal{G}$ has dimension p, then there is at least one space $G \in \mathcal{G}$ with dimension $p + 1$ whose subspace is G_0.

These criteria state that HARE determines B_j and estimates its corresponding β_j in a stepwise fashion where tensor products of two basis spaces require each single factor space. Let k represent a knot along a knot sequence. Then HARE has basis functions of the form $1, (t_k - t)_+, x_m, (x_{mk} - x_m)_+, x_m x_n, (t_k - t)_+ x_m, (t_k - t)_+(x_{mk} - x_m)_+, x_m(x_{nk} - x_n)_+, (x_{mk} - x_m)_+ x_n,$ and $(x_{mk} - x_m)_+(x_{nk} - x_n)_+$, where x_m, x_n are separate covariates, t_k is a knot in time, $x_{\cdot k}$ is a knot in a covariate, and $(\cdot)_+$ represents the positive part of the function. The tensor products $x_m x_n, (t_k - t)_+ x_m, (t_k - t)_+(x_{mk} - x_m)_+, x_m(x_{nk} - x_n)_+, (x_{mk} - x_m)_+ x_n,$ and $(x_{mk} - x_m)_+(x_{nk} - x_n)_+$ are allowable only if the factor basis functions are included. For example, if x_1, x_2, and $(7 - t)_+$ are included in the estimate of $\alpha(t|\mathbf{x})$, but x_3 is not, then $x_1 x_2, x_1(7 - t)_+,$ and $x_2(7 - t)_+$ are allowable, but $x_3(7 - t)_+$ is not.

HARE begins the partitioning procedure of determining $G \in \mathcal{G}$ with the constant space $G_{min} = 1$. The process adds new spaces $G \in \mathcal{G}$ where each $(p-1)$-dimensional space G_0 is replaced by a p-dimensional space G that includes G_0 as a subspace. When determining the new G allowable space, candidate basis functions B_j include linear covariates, a new knot in time, a new knot in a covariate, and a tensor product of two existing basis functions from G_0. G is determined from finding the candidate basis function that maximizes the Rao statistic $R = \mathbf{S}(\hat{\beta}_p^{(0)})/\sqrt{\mathbf{I}^{-1}(\hat{\beta}^{(0)})_{pp}}$, where $\hat{\beta}^{(0)}$ is the maximum likelihood estimate of β corresponding to space G, β_p is the coefficient for the basis function needed to go from G_0 to G, $\mathbf{S}(\cdot)$ is the score function, and $\mathbf{I}^{-1}(\cdot)$ is the observed Fisher information matrix. The addition of new G allowable spaces follows this algorithm:

- Calculate Rao statistic for all spaces obtained from G_0 by adding a basis function $B_{l0}(x_l) = x_l$ to G_0
- Calculate Rao for all allowable spaces obtained from G_0 by adding a basis function to G_0 comprising a tensor product of two tensor functions in G_0
- Calculate Rao statistic for a space obtained from G_0 by adding a basis function constructed by adding a new knot in t
- Calculate Rao statistic for a space obtained from G_0 by adding a basis function constructed by adding a new knot in covariate m
- Select space G that maximizes the absolute value of the Rao statistic.

After each space G is determined, the BIC (Schwarz, 1978) is stored for model selection procedures. Candidate basis functions are no longer added when either (a) the number of basis functions included is $\min(6n^{1/5}, n/4, 50)$, (b) the change in the maximized log-likelihood function is $< \frac{1}{2}(P-p) - \frac{1}{2}$ where P is the number of basis functions and p is the dimensionality of G, or (c) the algorithm yields no possible new basis function.

Following the addition phase of $G \in \mathcal{G}$ is the deletion phase, which carries out the candidate basis function algorithm above with two features changed: (i) the deletion phase goes from space G to space G_0, and (ii) Wald statistic $\hat{\beta}_p/SE(\hat{\beta}_p)$ is used instead of the Rao statistic. This latter modification is needed because the Rao statistic is based on the maximum likelihood estimate in G_0 whereas Wald is based on G space (Kooperberg, Stone, & Truong; 1995a). After each model has been estimated through the iterative addition and deletion phases of G construction, the model that minimizes BIC is the final estimate of the conditional log-hazard function $\alpha(t|\mathbf{x}, \beta)$.

References

Kooperberg, C., Stone, C. J., & Truong, Y. K. (1995a). Hazard regression. *Journal of the American Statistical Association*, 90(429), 78-94.

Kooperberg, C., Stone, C. J., & Truong, Y. K. (1995b). The L2 rate of convergence for hazard regression. *Scandinavian Journal of Statistics*, 22(2), 143-157.

Lange, K. *Numerical Analysis for Statisticians*. Springer, New York, USA, second edition, 2010.