Optimal designs for full and partial likelihood information - with application to survival models

Maria Konstantinou, Stefanie Biedermann and Alan Kimber

Abstract

Time-to-event data are often modelled through Cox's proportional hazards model for which inference is based on the partial likelihood function. We derive a general expression for the asymptotic covariance matrix of Cox's partial likelihood estimator for the covariate coefficients. Our approach is illustrated through an application to the special case of only one covariate, for which we construct minimum variance designs for different censoring mechanisms and both binary and interval design spaces. We compare these designs with the corresponding ones found using the full likelihood approach and demonstrate that the latter designs are highly efficient also for partial likelihood estimation.

Key words:

right-censoring, Cox's model, optimal design, full likelihood, partial likelihood

1. Introduction

Survival experiments are widely used in areas such as medicine, biostatistics, agriculture and engineering, producing data on the time until the occurrence of a particular event of interest. Such data known as time-to-event or survival data, usually feature censoring, which occurs when the event is not observed for some of the subjects in the experiment, and are often modelled through nonlinear survival models.

Currently there is little guidance on how to plan experiments involving potentially censored data. Most of the available results concern accelerated life testing, see, for example, Pascual and Montepiedra (2003), Zhang and Meeker (2006), Wu, Lin and Chen (2006) or McGree and Eccleston (2010). Becker, McDonald and Khoo (1989) use geometrical arguments and empirical values to construct D-optimal designs for proportional hazards models with one or two parameters. López-Fidalgo, Rivas-López and Del Campo (2009) consider a two-parameter exponential regression model and find D-optimal designs conditional on arrival time. Finally, Konstantinou, Biedermann and Kimber (2014) find D- and c-optimal designs for a large class of two-parameter models, including survival models with different censoring mechanisms. The results in all these papers arise from the use of the full likelihood approach.

In practice, researchers often prefer Cox's proportional hazards model (Cox (1972)) to parametric models because fewer assumptions are required and because of the simple interpretation of the regression coefficients in terms of hazard ratios. In particular, Cox's model satisfies the proportional hazards assumption of constant hazard ratio over time. However, the baseline hazard function, and hence the probability distribution of the times-to-event, is not specified and therefore inference on the covariate coefficients is based on the partial likelihood method developed by Cox (1972) which does not require knowledge of the baseline hazard.

Andersen and Gill (1982) formulate the Cox model in a counting process set-up, and provide analytical results for the asymptotic properties of the estimators from this model. However, there are only two papers in the literature so far on optimal designs for the model. Kalish and Harrington (1988) find optimal designs for the special case when two treatments are available, that is, for a binary design space. They investigate empirically the loss of efficiency when equal numbers of patients are allocated to each treatment (the balanced design). López-Fidalgo and Rivas-López (2014) use approximations to obtain an information matrix based on the partial likelihood for a binary design space. In their application, they find optimal designs based on this approximation which they then compare with the optimal designs for the full likelihood approach. However, the model they use in their full likelihood analysis does not correspond to the nominal model assumed for the partial likelihood situation.

We find a closed-form expression for the asymptotic covariance matrix in the Cox model, and provide a necessary condition for the optimality of a design that can be used to screen out non-optimal designs. In our applications, we consider both a binary and an interval design space and find the optimal designs numerically by optimising a complicated objective function that involves an integral over time. Comparisons with the optimal designs found using the full likelihood for the same underlying nominal models (Konstantinou, Biedermann, and Kimber (2014)) show that the latter designs are highly efficient for partial likelihood estimation. This suggests that the readily available optimal designs for a suitable parametric model can be used in practice, even though partial likelihood estimation is to be used. We further extend a result by Kalish and Harrington (1988) to interval design spaces, where we show that for Type-I censoring the optimal designs do not depend on the shape of the hazard function. Hence, the optimal designs found in Konstantinou, Biedermann, and Kimber (2014) for constant hazard functions are near optimal for the Cox model regardless of the true underlying hazard function.

This article is organised as follows. In Section 2, we briefly describe the type of data observed in survival studies, and define approximate designs. In Section 3, we derive the optimality criterion to be used, and find a necessary condition for the optimality of a design. Then in Section 4 we find optimal designs for various censoring scenarios, and compare our results with those by

Kalish and Harrington (1988). In Section 5 we compare the optimal designs for the Cox model with optimal designs for the corresponding parametric model, show why these are similar and give a simple illustration. Finally some conclusions and recommendations are given in Section 6.

2. Background

Let $T_1, \ldots, T_n$ be the independent times-to-event of the n subjects in the experiment with $t_1, \ldots, t_n$ their corresponding values and $[0, c]$ be the predetermined period of the experiment. Throughout this article we focus on right-censoring that occurs when the time until the occurrence of the event of interest is above a certain value called the censoring time, but it is unknown by how much.

We consider two different censoring mechanisms that result in right-censored data, Type-I and random censoring. Type-I censoring corresponds to the case where all the subjects enter the experiment at the same time and so the censoring time is common for all the subjects and is equal to the duration of the experiment c. We observe $Y_j = \min\{T_j, c\}$, $j = 1, \ldots, n$, and times-to-event greater than c are therefore right-censored. When subject j enters the experiment at a random time in $[0, c]$, independent of the time-to-event, the censoring time C_j for this subject is also random. This scenario corresponds to random censoring where we observe $Y_j = \min\{T_j, C_j\}$, $j = 1, \ldots, n$. In the following the distribution of the time of entry for each subject is assumed to be uniform.

The data, Y_j, $j = 1, \ldots, n$, may depend on several covariates held in a vector $\boldsymbol{x}$, which can be controlled by the experimenter. The aim of designing an experiment is to choose those settings of the covariates which ensure the most precise estimation of the model parameters of interest. This is formulated through an optimal experimental design. We consider approximate designs of the form

$$\xi = \begin{Bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_m \\ \omega_1 & \ldots & \omega_m \end{Bmatrix}, \;\; 0 < \omega_i \leq 1, \;\; \sum_{i=1}^{m} \omega_i = 1,$$

where the support points $\boldsymbol{x}_i \in \mathcal{X}$, $i = 1, \ldots, m$, $m \leq n$ are the distinct experimental conditions in the design and the weights ω_i represent the proportion of observations to be taken at the corresponding support point. In what follows, we will state the design problem for the general situation of possibly more than one covariate. Our applications will then focus on the one covariate case where $\mathcal{X}$ is either binary, that is $\mathcal{X} = \{0, 1\}$, corresponding, for example, to two treatments, or an interval, that is $\mathcal{X} = [u, v]$, corresponding, for example, to different drug doses.

3. Cox's model and optimality criterion

When the risk of the event occurring at a particular time t depends on the values of a set of covariates Cox's proportional hazards model is specified by the hazard function

$$h(t, \boldsymbol{x}_j) = h_0(t) e^{\boldsymbol{\beta}^T \boldsymbol{x}_j} \qquad (t > 0), \tag{1}$$

where $\boldsymbol{x}_j$ is the value of the covariate vector for the jth subject, $\boldsymbol{\beta}$ is the vector of coefficients that need to be estimated and $h_0(t)$ is the baseline hazard function which remains unspecified.

Suppose that data are available for n subjects with corresponding observations denoted by $y_1, \ldots, y_n$ and that δ_j, $j = 1, \ldots, n$, is an indicator function which is equal to zero if the jth observation y_j is right-censored and unity otherwise. The partial likelihood function for model (1) is (Cox (1972))

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_j}}{\sum_{l \in R(y_j)} e^{\boldsymbol{\beta}^T \boldsymbol{x}_l}} \right\}^{\delta_j}, \tag{2}$$

where $R(y_j)$ is called the risk-set at time y_j and contains the indices of those subjects for which neither the event nor censoring have occurred at a time just prior to y_j.

The asymptotic variance of the maximum partial likelihood estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{PL}$, is the inverse of $E\left[-\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right]$. López-Fidalgo and Rivas-López (2014) approximate this expectation for one covariate, and maximise the resulting expression in order to find optimal designs. They therefore add an extra layer of approximation to the optimality criterion, in addition to the fact that the information matrix in itself approximates the inverse of the covariance matrix.

We work directly with the asymptotic covariance matrix, which we derive from Andersen and Gill (1982), who showed that under some regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta})$ converges in distribution to $\mathcal{N}(\mathbf{0}, \Sigma^{-1})$ as $n \to \infty$. Here, $\mathbf{0}$ is the zero vector of appropriate length, and for an approximate design ξ, the inverse, Σ, of the asymptotic covariance matrix is given by

$$\Sigma = \Sigma(\xi) = \sum_{i=2}^{m} \sum_{q<i} \omega_i \omega_q e^{\boldsymbol{\beta}^T (\boldsymbol{x}_i + \boldsymbol{x}_q)} (\boldsymbol{x}_i - \boldsymbol{x}_q)(\boldsymbol{x}_i - \boldsymbol{x}_q)^T \int_0^\infty \frac{\pi_i(y) \pi_q(y) h_0(y)}{\sum_{l=1}^{m} \omega_l \pi_l(y) e^{\boldsymbol{\beta}^T \boldsymbol{x}_l}} \, dy, \tag{3}$$

where $\pi_i(y)$, $i = 1, \ldots, m$, is the probability that a subject with covariate vector $\boldsymbol{x}_i$ is at risk at time y, that is, neither the event nor censoring have occurred for that subject by time y.

An optimal design for model (1) minimises the asymptotic covariance matrix or equivalently maximises $\Sigma(\xi)$ with respect to the design ξ. Thus, a design ξ^* is Σ-optimal for estimating β if

$$\xi^* = \arg\min_{\xi} \Sigma^{-1}(\xi) = \arg\max_{\xi} \Sigma(\xi).$$

We note that the optimal design will depend on the value of the β-parameter and therefore will be a locally optimal design.

It is clear from the asymptotic distribution that the bias of the estimator $\hat{\boldsymbol{\beta}}_{PL}$ is of order $o(n^{-1/2})$. Hence the variance will dominate the mean squared error for large n, thus justifying our choice of optimality criterion, which is solely based on the asymptotic covariance matrix.

For illustration purposes, in what follows, we consider the special case of only one covariate. This situation is often encountered in clinical trials where patients are randomised to different treatments or doses of a treatment. Similarly, in life testing in reliability studies there is usually just one covariate to be selected by the experimenter.

Proposition 1 gives a necessary condition for the optimality of a design ξ^* and is proven in the appendix. Unlike the general equivalence theorem for c-optimality (Atkinson, Donev and Tobias (2007)), this condition is not sufficient, since the criterion function, $\Sigma(\xi)$, is not concave. However, it can be used to discard candidate designs that do not satisfy this condition since they are non-optimal.

Proposition 1. *Let $\mathcal{H}$ be the class of all one-point designs where the support point is in the design space $\mathcal{X} = [u, v]$, and let $\eta = \{x; 1\} \in \mathcal{H}$. If a design ξ^* on $\mathcal{X}$ with support points $\{x_1, \ldots, x_m\}$ and corresponding weights $\{\omega_1, \ldots, \omega_m\}$ is optimal for estimating β via the partial likelihood method, the inequality*

$$d(\xi^*, \eta) \leq 0$$

holds for all $\eta \in \mathcal{H}$, with equality in the one-point designs $\xi_i = \{x_i; 1\}$, $i = 1, \ldots, m$, generated by the support points of ξ^*. Here $d(\xi^*, \eta)$ is the Fréchet derivative of the criterion function at ξ^* in direction of the one-point design η, and is given by

$$\begin{aligned} d(\xi^*, \eta) \;=\; & -\sum_{i=2}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l}} dy \\ & -\sum_{i=2}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)\pi_x(y) e^{\beta x}}{(\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l})^2} dy \\ & +\sum_{q=1}^{m} \omega_q e^{\beta(x + x_q)} (x - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_x(y)\pi_q(y)}{\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l}} dy, \end{aligned}$$

where $\pi_x(y)$ is the probability of being at risk at time y for covariate-value x.

Example. To fix ideas, consider the scenario with one explanatory variable, parameter value $e^\beta = 0.03$ and no censoring. In this situation, the best two-point design does not satisfy the condition in Proposition 1; see Figure 1. The best three-point design satisfies the condition. This does not prove optimality, but numerical search showed that the criterion value could not be improved by using four-point designs, so in this case there is strong evidence that the best three-point design is optimal.

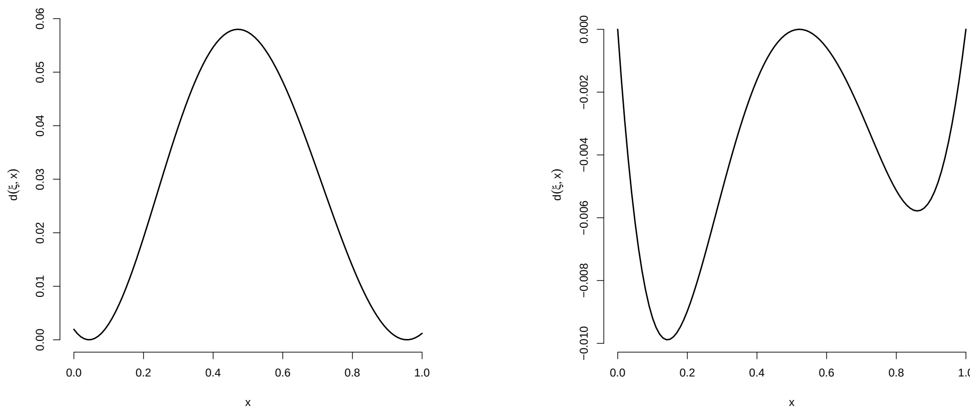

Figure 1: The function $d(\xi^*, x)$ for the best two-point design (left panel) and the best three-point design (right panel). The two-point design cannot be optimal.

We found that in most scenarios considered in this paper the best two-point design satisfied the condition from Proposition 1. It was usually the more extreme cases (large values of $|\beta|$) where three points were required. In those situations, the efficiency of the best two-point design was reasonably high (e.g. 86% for the example). For the sake of better comparison with c-optimal designs (see Section 5), we present the best two-point designs in what follows. In most cases these are optimal, and very efficient otherwise.

4. Optimal designs for partial likelihood information

In this section we find the optimal designs for model (1) considering both a binary and a continuous design space and various censoring mechanisms. We first discuss the special case of no censoring and then investigate Type-I and random censoring separately.

In the case of a binary design space $\mathcal{X} = \{0, 1\}$ the design must be supported at points $x_1 = 0$ and $x_2 = 1$. Using the results of Andersen and Gill (1982), Kalish and Harrington (1988) find the asymptotic variance of $\sqrt{n}(\hat{\beta}_{PL} - \beta)$ to be

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta}} \left[\int_0^{\infty} \frac{\pi_1(y)\pi_2(y)h_0(y)}{\omega\pi_1(y) + (1-\omega)e^{\beta}\pi_2(y)} dy \right]^{-1}, \tag{4}$$

where ω and $1-\omega$ are the weights at points 0 and 1 respectively.

For purposes of comparison with the c-optimal designs found using the full likelihood approach, for a continuous design space we consider designs with two support points x_1 and x_2 and corresponding weights ω and $1-\omega$. From (3), the asymptotic variance of $\sqrt{n}(\hat{\beta}_{PL} - \beta)$ can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2}\left[\int_0^\infty \frac{\pi_1(y)\pi_2(y)h_0(y)}{\omega e^{\beta x_1}\pi_1(y) + (1-\omega)e^{\beta x_2}\pi_2(y)}dy\right]^{-1}. \tag{5}$$

Following Kalish and Harrington (1988), the survivor function of the random variable W representing the time to censoring is given by

$$S_W(w) = \begin{cases} 1, & \text{if } 0 < w \leq c \\ 0, & \text{if } w > c \end{cases}, \qquad S_W(w) = \begin{cases} \frac{c-w}{c}, & \text{if } 0 < w \leq c \\ 0, & \text{if } w > c \end{cases}$$

for Type-I and random censoring respectively. Therefore, the probability that a subject allocated at point x_i is at risk at time y is $\pi_i(y) = S_W(y)S_i(y)$, $i = 1, \ldots, m$, where S_i is the survivor function for subjects allocated at x_i. We also use the Kalish and Harrington (1988) characterisation for the proportion of censoring as the overall probability of censoring had a balanced design been used. That is, $1 - (0.5d_1 + 0.5d_2)$, where $d_i = P(Y < W) = \int_0^\infty S_W(y)dF_i(y)$ is the probability of the event occurring and $F_i(y)$ is the distribution function of the times-to-event for subjects allocated to x_i, $i = 1, 2$.

4.1. No censoring This corresponds to $c = \infty$, that is, a study running for as long as necessary to record all event times. In this case $\pi_i(y) = S_i(y)$, $i = 1, 2$, and equations (4) and (5) can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta}}\left[\int_0^1 \frac{u^{e^{\beta}-1}}{\omega + (1-\omega)e^{\beta}u^{e^{\beta}-1}}du\right]^{-1},$$

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2}\left[\int_0^1 \frac{u^{e^{\beta x_2}-1}}{\omega e^{\beta x_1} + (1-\omega)e^{\beta x_2}u^{e^{\beta x_2}-e^{\beta x_1}}}du\right]^{-1}$$

respectively using the fact that under proportional hazards $S_i(y) = \{S_0(y)\}^{e^{\beta x_i}}$, $i = 1, 2$, and applying the transformation $u = S_0(y) = \exp\{-\int_0^y h_0(s)ds\}$. Therefore, whether a binary or an interval design space is considered, the baseline hazard does not affect the optimal choice of design.

Assuming exponential times-to-event, the optimal designs for various β-values along with the efficiency of the balanced design that allocates equal proportions of subjects at the two support points, are presented in Table 1. We note that the continuous design interval considered in these calculations is $\mathcal{X} = [0, 1]$.

We observe that for a positive value of β the optimal weight $1-\omega$ at point $x_2 = 1$ is the same as the weight ω at point $x_1 = 0$ for the corresponding negative β of equal absolute value. Moreover, for small and moderate absolute values of β, for example 0.69 and 1.39, the efficiency of the balanced design is high and decreases for larger absolute values of β ($|\beta| = 3.51$).

Table 1: Optimal designs for binary and continuous design spaces and efficiencies, in percent, of the balanced design in the absence of censoring

optimal design	$e^{\beta}(\beta)$					
	0.03 (-3.51)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	33.3 (3.51)
$1-\omega$	0.68	0.55	0.51	0.49	0.45	0.32
efficiency	(91)	(99)	(100)	(100)	(99)	(91)
$\{x_1, x_2\}$	{0.04,0.96}	{0,1}	{0,1}	{0,1}	{0,1}	{0.1,1}
$1-\omega$	0.66	0.55	0.51	0.49	0.45	0.34
efficiency	(90)	(99)	(100)	(100)	(99)	(90)

4.2. Type-I censoring Under this censoring scheme Kalish and Harrington (1988) showed that for $\mathcal{X} = \{0,1\}$ equation (4) can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta}} \left[\int_{S_0(c)}^{1} \frac{u^{e^{\beta}-1}}{\omega + (1-\omega)e^{\beta}u^{e^{\beta}-1}} du \right]^{-1},$$

where $S_0(y) = \exp\{-\int_0^y h_0(s)ds\}$. We extend this result to the case of an interval design space. Using again that $S_i(y) = \{S_0(y)\}^{e^{\beta x_i}}$, $i = 1, 2$, and applying the transformation $u = S_0(y)$, equation (5) becomes

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2} \left[\int_{S_0(c)}^{1} \frac{u^{e^{\beta x_2}-1}}{\omega e^{\beta x_1} + (1-\omega)e^{\beta x_2}u^{e^{\beta x_2}-e^{\beta x_1}}} du \right]^{-1}.$$

In both cases $\Sigma^{-1}(\xi)$ depends on the baseline hazard only through $S_0(c)$ and hence the optimal design is independent of the shape of $h_0(t)$. Therefore, under Type-I censoring the optimal design can be found by assuming a constant baseline hazard without loss of generality. In conclusion, the optimal designs for the exponential regression model will be efficient for partial likelihood estimation whatever the baseline hazard.

Tables 2 and 3 show the optimal designs assuming the exponential regression model and the efficiency of the balanced design for $\mathcal{X} = \{0,1\}$ and $\mathcal{X} = [0,1]$ respectively. The β-values were chosen to exemplify small, moderate and large covariate effects.

For both a binary and an interval design space the symmetry of the optimal weights for equal absolute values of β is also evident for Type-I censoring. According to the sign of the parameter β, the optimal design allocates more subjects to the experimental point where the possibility of censoring is greater in order for the variance to be minimised. This will be the smaller support point x_1 when β is positive since in this case the covariate has an increasing effect on the hazard.

We found numerically that the optimal (two point) designs for the interval design space $\mathcal{X} = [0,1]$ are not necessarily unique. In particular, if an optimal (two point) design is not supported at

Table 2: Optimal weights $1-\omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and Type-I censoring

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	0.68 (92)	0.60 (97)	0.55 (99)	0.52 (100)	0.48 (100)	0.45 (99)	0.40 (97)	0.32 (92)
0.3	0.68 (92)	0.61 (96)	0.58 (98)	0.54 (99)	0.46 (99)	0.42 (98)	0.39 (96)	0.32 (92)
0.5	0.76 (80)	0.68 (88)	0.62 (95)	0.56 (99)	0.44 (99)	0.38 (95)	0.32 (88)	0.24 (80)
0.7	0.82 (71)	0.73 (83)	0.64 (93)	0.57 (98)	0.43 (98)	0.36 (93)	0.27 (83)	0.18 (71)
0.9	0.85 (68)	0.75 (80)	0.66 (91)	0.58 (97)	0.42 (97)	0.34 (91)	0.25 (80)	0.16 (68)

Table 3: Support points $\{x_1, x_2\}$, optimal weights $1-\omega$ at point x_2 and efficiency, in percent, of the balanced design under Type-I censoring for $\mathcal{X} = [0, 1]$

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	{0.04,0.96} 0.66 (90)	{0,1} 0.60 (97)	{0,1} 0.55 (99)	{0,1} 0.52 (100)	{0,1} 0.48 (100)	{0,1} 0.45 (99)	{0,1} 0.40 (97)	{0.04,0.96} 0.34 (90)
0.3	{0,0.91} 0.66 (90)	{0,1} 0.61 (96)	{0,1} 0.58 (98)	{0,1} 0.54 (99)	{0,1} 0.46 (99)	{0,1} 0.42 (98)	{0,1} 0.39 (96)	{0.09,1} 0.34 (90)
0.5	{0,0.84} 0.71 (76)	{0,1} 0.68 (88)	{0,1} 0.62 (95)	{0,1} 0.56 (99)	{0,1} 0.44 (99)	{0,1} 0.38 (95)	{0,1} 0.32 (88)	{0.16,1} 0.29 (76)
0.7	{0,0.77} 0.76 (63)	{0,1} 0.73 (83)	{0,1} 0.64 (93)	{0,1} 0.57 (98)	{0,1} 0.43 (99)	{0,1} 0.36 (93)	{0,1} 0.27 (83)	{0.23,1} 0.24 (63)
0.8	{0,0.74} 0.78 (59)	{0,1} 0.75 (80)	{0,1} 0.66 (91)	{0,1} 0.58 (97)	{0,1} 0.42 (97)	{0,1} 0.34 (91)	{0,1} 0.25 (80)	{0.26,1} 0.22 (59)

zero and one, then the support points are not unique, but their difference is. For example, if a design with support points 0 and 0.9 is optimal, then a design with support points 0.1 and 1 (and the same weights) will also be optimal.

We also note that for absolute β-values of 2.3 or more and censoring proportion of 50% or more the efficiency of the balanced design falls below 90%. This is in contrast to the result of Kalish and Harrington (1988) who only consider small $|\beta|$-values and therefore find the balanced design to have high efficiency. The lower efficiency can intuitively be explained by observing that for large $|\beta|$-values, the support points of the design move away from 0 and 1. In particular, we observed a monotonic pattern with respect to the proportion of censoring. The higher this proportion, the lower the 'threshold' value of $|\beta|$ where the support points move into the interior of the design space. For example the threshold values for $|\beta|$ are 3.22 and 2.71 for proportions of censoring of 0.3 and 0.7, respectively.

4.3. Random censoring In the presence of random censoring the criteria functions $\Sigma^{-1}(\xi)$ for binary and interval design spaces are given by

$$\Sigma^{-1}(\xi) = \frac{c}{\omega(1-\omega)e^{\beta}} \left[\int_0^c \frac{(c-y)S_1(y)S_2(y)h_0(y)}{\omega S_1(y) + (1-\omega)e^{\beta} S_2(y)} \, dy \right]^{-1}$$

and

$$\Sigma^{-1}(\xi) = \frac{c}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2} \left[\int_0^c \frac{(c-y)S_1(y)S_2(y)h_0(y)}{\omega e^{\beta x_1} S_1(y) + (1-\omega)e^{\beta x_2} S_2(y)} \, dy \right]^{-1}$$

respectively. A transformation similar to the one used for Type-I censoring cannot be applied here. Therefore, $\Sigma^{-1}(\xi)$ and hence the optimal design does depend on the form of the underlying hazard.

For illustration purposes we compute the optimal designs for various β-values and censoring proportions again assuming a constant baseline hazard. These designs are displayed in Tables 4 and 5 for $\mathcal{X} = \{0, 1\}$ and $\mathcal{X} = [0, 1]$ respectively, along with the corresponding efficiencies of the balanced design.

As for Type-I censoring, the optimal design puts more weight at the support point where censoring is more likely. The symmetry of the optimal weights as well as of the support points for negative and positive β's of the same absolute value is also evident in both Tables 4 and 5. Overall the two censoring schemes produce similar designs which differ from the balanced design for heavy censoring and for β-values moderately far from 0.

5. Comparison of designs arising from full and partial likelihood methods

Efron (1977) compares the Fisher information for estimating β for the full and the partial likelihood methods in the same underlying model. He finds that they coincide except for an

Table 4: Optimal weights $1 - \omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and random censoring

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	0.68 (91)	0.61 (97)	0.55 (99)	0.52 (100)	0.48 (100)	0.45 (99)	0.39 (97)	0.32 (91)
0.3	0.68 (91)	0.62 (96)	0.57 (98)	0.53 (100)	0.47 (100)	0.43 (98)	0.38 (96)	0.32 (91)
0.5	0.71 (87)	0.65 (92)	0.60 (96)	0.55 (99)	0.45 (99)	0.40 (96)	0.35 (92)	0.94 (87)
0.7	0.81 (73)	0.71 (85)	0.63 (94)	0.57 (98)	0.43 (98)	0.37 (94)	0.29 (85)	0.19 (73)
0.9	0.84 (68)	0.75 (80)	0.66 (91)	0.58 (97)	0.42 (97)	0.34 (91)	0.25 (80)	0.16 (68)

Table 5: Support points $\{x_1, x_2\}$, optimal weights $1 - \omega$ at point x_2 and efficiency, in percent, of the balanced design under random censoring for $\mathcal{X} = [0, 1]$

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	{0,0.91} 0.66 (90)	{0,1} 0.61 (97)	{0,1} 0.55 (99)	{0,1} 0.52 (100)	{0,1} 0.48 (100)	{0,1} 0.45 (99)	{0,1} 0.39 (97)	{0.09,1} 0.34 (90)
0.3	{0,0.91} 0.66 (90)	{0,1} 0.62 (96)	{0,1} 0.57 (98)	{0,1} 0.53 (100)	{0,1} 0.47 (100)	{0,1} 0.43 (98)	{0,1} 0.38 (96)	{0.09,1} 0.34 (90)
0.5	{0,0.88} 0.68 (85)	{0,1} 0.65 (92)	{0,1} 0.60 (96)	{0,1} 0.55 (99)	{0,1} 0.45 (99)	{0,1} 0.40 (96)	{0,1} 0.35 (92)	{0.12,1} 0.32 (85)
0.7	{0,0.79} 0.75 (67)	{0,1} 0.71 (85)	{0,1} 0.63 (94)	{0,1} 0.57 (98)	{0,1} 0.43 (98)	{0,1} 0.37 (94)	{0,1} 0.29 (85)	{0.21,1} 0.25 (67)
0.8	{0,0.74} 0.77 (60)	{0,1} 0.75 (80)	{0,1} 0.66 (91)	{0,1} 0.58 (97)	{0,1} 0.42 (97)	{0,1} 0.34 (91)	{0,1} 0.25 (80)	{0.26,1} 0.23 (60)

extra term in the Fisher information for the full likelihood, which, however, will usually be small in practice. He therefore concludes that in most situations the partial likelihood method will be reasonably efficient.

These results suggest that the optimal designs for estimating β, which are based on the asymptotic variances and thus the Fisher information, should also be similar. In particular, we wish to find out in which situations the optimal designs for the full likelihood method, which are readily available in Konstantinou, Biedermann, and Kimber (2014), are highly efficient for estimation in the partial likelihood model. Hence finding optimal designs for the complicated criterion function $\Sigma(\xi)$ could be avoided by practitioners.

A simple explicit formula for the extra term in the Fisher information for the full likelihood could not be derived and therefore we could not work directly with the Fisher information matrix to prove the similarity of the two approaches analytically. However, we first compare the optimal designs directly for several scenarios, to identify the situations where they are similar or even coincide and then find an explanation for this phenomenon.

Throughout this section, we assume an exponential regression model with constant baseline hazard and compare the c-optimal design for estimating β in the two-parameter model readily available in Konstantinou, Biedermann, and Kimber (2014), with the Σ-optimal design for β in Cox's model. We note that López-Fidalgo and Rivas-López (2014) provide a brief comparison of such designs for a binary design space. However, they assume that $h_0(y)\exp(\alpha) = 1$, leaving them with an estimation problem for one parameter only. Hence the optimal designs they find for the parametric model are one-point designs, taking all observations at $x = 1$. This is not surprising since they completely specify the baseline hazard, implying that the hazard at $x = 0$ is known, thus not requiring any observations at $x = 0$.

5.1. Numerical comparison We briefly discuss the case of no censoring for which the c-optimal design for β found using the full likelihood method is always equally supported at 0 and 1 (see Konstantinou, Biedermann, and Kimber (2014)). From Table 1 we observe that for $|\beta|$-values away from zero the two approaches do not coincide as the optimal weights for the partial likelihood method are not equal. However, the balanced design is highly efficient even for large values of $|\beta|$ making the c-optimal designs for β good alternatives to the designs found based on the partial likelihood function.

In the presence of censoring, we calculate the efficiency of the c-optimal designs found using the full likelihood function relative to the designs discussed in sections 4.2 and 4.3 by

$$eff(\xi_F^*) = \frac{\Sigma(\xi_F^*)}{\Sigma(\xi_P^*)} = \frac{\Sigma^{-1}(\xi_P^*)}{\Sigma^{-1}(\xi_F^*)},$$

where ξ_F^* and ξ_P^* are the locally optimal designs for β arising from the full and partial likelihood method respectively. The results for the two censoring schemes considered are illustrated in

Tables 6 and 7 respectively. Both the cases of $\mathcal{X} = \{0, 1\}$ and $\mathcal{X} = [0, 1]$ are examined and the efficiencies are found as functions of the proportion of censoring and the parameter of interest β. We note that for proportions of censoring above 0.5 the (rounded) efficiencies of the full likelihood designs are equal to 1 for all β-values and are therefore omitted from Tables 6 and 7.

Table 6: Efficiencies, in percent, of full likelihood designs under Type-I censoring for a binary (and a continuous) design space

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	94 (93)	98 (98)	100 (100)	100 (100)	100 (100)	100 (100)	98 (98)	94 (93)
0.3	99 (98)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	99 (98)
0.5	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Table 7: Efficiencies, in percent, of full likelihood designs under random censoring for a binary (and a continuous) design space

proportion of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	94 (92)	98 (98)	100 (100)	100 (100)	100 (100)	100 (100)	98 (98)	94 (92)
0.3	98 (97)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	98 (97)
0.5	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

We observe that the c-optimal designs found using the full likelihood function are extremely efficient under both censoring schemes, with the efficiencies under random censoring being slightly lower. Hence the c-optimal designs can be used as an efficient alternative for the Σ-optimal designs, even if the data are to be analysed through the partial likelihood approach. In particular, for heavy censoring the full likelihood designs give efficiency very close or equal to 1 even for extremely large β-values.

Moreover, by comparing the elements of Tables 6 and 7, that is, the efficiencies of the c-optimal designs, with the corresponding elements in Tables 2- 5, we find that the c-optimal designs are considerably more efficient for estimating β in the partial likelihood model than the balanced design on 0 and 1. For example, when the proportion of censoring is 0.5 and $\beta = -2.3$, the c-

optimal designs have efficiencies of 100% for Type-I and random censoring, respectively for both design spaces whereas the balanced design achieves corresponding efficiencies of 88% for Type-I censoring and 93% for random censoring again for both design spaces. This means that under Type-I censoring we require 114 subjects in a balanced design to achieve the same precision for parameter estimation as 100 subjects in a c-optimal design. For heavier censoring, the c-optimal designs are even more preferable.

5.2. Analytical results In what follows, we find an explanation for the similarities of c- and Σ-optimal designs, in particular under heavy censoring and/or small to moderate β-values. We note that here we are not trying to find the best possible approximation to the information matrix. Such an approximation is not necessary under the formulation of this paper since we have established a closed-form expression for the asymptotic covariance matrix which will always be better or at least as good as the inverse of even the actual information matrix. Therefore, improving the approximation has no extra value.

From Konstantinou, Biedermann, and Kimber (2014) we have that the asymptotic variance of ($\sqrt{n}$ times) the maximum full likelihood estimator for β for the exponential model under Type-I censoring is given by

$$Var(\hat{\beta}_{FL}) = \frac{(1-\omega)(1-e^{-ce^{\alpha+\beta x_1}}) + \omega(1-e^{-ce^{\alpha+\beta x_2}})}{\omega(1-\omega)(1-e^{-ce^{\alpha+\beta x_1}})(1-e^{-ce^{\alpha+\beta x_2}})(x_2-x_1)^2}.$$

For heavy censoring the probability of the event occurring in the time interval under consideration is small and thus we have small values of $ce^{\alpha+\beta x}$. Therefore, the above expression can be approximated by a first order Taylor expansion,

$$Var(\hat{\beta}_{FL}) \approx \frac{(1-\omega)e^{\beta x_1} + \omega e^{\beta x_2}}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2-x_1)^2},$$

using that

$$1 - e^{-ce^{\alpha+\beta x}} \approx ce^{\alpha+\beta x}.$$

Now consider the corresponding quantity for the partial likelihood model for two different treatments or drug doses x_1 and x_2. Without loss of generality we assume that among the data available for n subjects there are k distinct event times $t_1 < \ldots < t_k$. Also let r_j be the number of individuals in the risk set at time t_j, q_j of them allocated at x_2 and $r_j - q_j$ allocated at x_1. Then using equation (2) the asymptotic variance of ($\sqrt{n}$) $\hat{\beta}_{PL}$ becomes

$$Var(\hat{\beta}_{PL}) = \lim_{n\to\infty}\left[\frac{1}{n}E\left(\sum_{j=1}^{k}\frac{q_j(r_j-q_j)e^{\beta(x_1+x_2)}(x_2-x_1)^2}{[(r_j-q_j)e^{\beta x_1}+q_je^{\beta x_2}]^2}\right)\right]^{-1}. \qquad (6)$$

Let $q_j^* = q_j/r_j$ and $r_j^* = r_j/r_j = 1$, $j = 1, \ldots, k$. Then the right hand side of (6) will not change when replacing q_k and r_j with q_j^* and r_j^*, respectively. For k/n small, the proportion of observed event times is also small and this again corresponds to the case of heavy censoring. Therefore, $q_j^* \approx \omega$ and $r_j^* - q_j^* \approx 1 - \omega$ that is, the original proportion of subjects allocated at x_2 and x_1 respectively, at least for small j. Similarly, if $|\beta|$ is small, the proportion of subjects at risk in the two groups will not change much over time, and again $q_j^* \approx \omega$ in this situation. Hence in the cases of heavy censoring and/or small $|\beta|$-values the proportions of subjects at risk are approximately constant.

Now k, the number of observed events, is itself random, and its expectation can be approximated by $E(k) \approx n[(1-\omega)ce^{\alpha+\beta x_1} + \omega ce^{\alpha+\beta x_2}]$. Overall, we obtain

$$Var(\hat{\beta}_{PL}) \approx \frac{(1-\omega)e^{\beta x_1} + \omega e^{\beta x_2}}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2-x_1)^2}.$$

We conclude that the two variances, and thus the optimal designs, are approximately equal which confirms the numerical results in Tables 6 and 7.

Under random censoring $Var(\hat{\beta}_{FL})$ is given by the expression

$$\frac{(1-\omega)(1+(e^{-ce^{\alpha+\beta x_1}}-1)/ce^{\alpha+\beta x_1}) + \omega(1+(e^{-ce^{\alpha+\beta x_2}}-1)/ce^{\alpha+\beta x_2})}{\omega(1-\omega)(1+(e^{-ce^{\alpha+\beta x_1}}-1)/ce^{\alpha+\beta x_1})(1+(e^{-ce^{\alpha+\beta x_2}}-1)/ce^{\alpha+\beta x_2})(x_2-x_1)^2}.$$

Following along the same lines as for Type-I censoring, we find that for heavy censoring and/or small $|\beta|$-values

$$Var(\hat{\beta}_{FL}) \approx Var(\hat{\beta}_{PL}) \approx \frac{2((1-\omega)e^{\beta x_1} + \omega e^{\beta x_2})}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2-x_1)^2}.$$

and therefore, again the two asymptotic variances, and thus the corresponding optimal designs, are approximately equal.

5.3. Example We now give an example to illustrate the simplicity of our approach to obtaining highly efficient designs for fitting the Cox model in the case of Type-I censoring with $\mathcal{X} = \{0, 1\}$. A key result is that of Section 4.2 that allows us to use the exponential regression model results of Konstantinou, Biedermann, and Kimber (2014). Collett (2003) briefly discusses designing a survival trial for chronic active hepatitis patients. The proposed analysis will involve fitting a Cox model. Each patient will be followed up for $c = 2$ years, and it is thought that when $x = 0$, 70% of patients will survive beyond 2 years, and the corresponding figure when $x = 1$ is 82%. This corresponds to a log-hazard ratio β of around -0.6. The c-optimal design based on the full-likelihood exponential model puts weight $\sqrt{0.18}/(\sqrt{0.18}+\sqrt{0.30}) = 0.436$ on $x = 0$ and from Table 6 we can see that this design is virtually fully efficient relative to the much harder to calculate Σ-optimal design.

6. Conclusions

There is only limited guidance in the literature on efficient design of survival experiments with possibly censored data, with the majority of the available articles considering parametric models. However, in practice this type of data is often modelled through a Cox model in which case the parameters of interest are estimated using the partial likelihood method.

We have met the needs of this practical scenario by setting up a general framework for the construction of optimal designs for Cox's model, later focussing on the model with only one covariate. Our approach contains the results by Kalish and Harrington (1988) as a special case and it differs from that by López-Fidalgo and Rivas-López (2014) in that we work directly with the asymptotic covariance matrix, without adding another level of approximation.

Kalish and Harrington (1988) conclude that equal allocation to both support points will be sufficiently efficient for partial likelihood estimation under both Type-I and random censoring. However, we have found this not to be the case for large effect sizes β and/or if there is heavy censoring.

Optimal designs for partial likelihood estimation are not trivial to find, and may therefore not be popular with practitioners. We have compared these designs with the c-optimal designs for the corresponding model using the full likelihood information, and found that the optimal designs for both methods are similar, in particular for heavy censoring, which is often observed in practice. We have shown that the two asymptotic variances are indeed approximately equal under heavy censoring. Moreover, we have found that the c-optimal designs are considerably more efficient than the balanced design for estimating β using the partial likelihood approach.

We have extended a result of Kalish and Harrington (1988) to include more scenarios commonly encountered in practice. In particular, for Type-I censoring, the optimal design for partial likelihood estimation does not depend on the shape of the baseline hazard function, but only on the value of the survival function at the censoring time c. This means that the c-optimal designs found in Konstantinou, Biedermann, and Kimber (2014) for constant baseline hazard will be highly efficient for partial likelihood estimation *whatever the baseline hazard*.

We therefore recommend the use of the readily available and highly efficient c-optimal designs, also for partial likelihood estimation, since they can be used without detriment in most situations.

Optimally designed experiments are cost effective, since a smaller sample size is required to obtain estimates with a given accuracy. We hope that this work will encourage practitioners to use optimal designs thus influencing the planning of survival experiments in the future.

Acknowledgement. The authors would like to thank two reviewers whose insightful comments led to a substantial improvement of an earlier version of this paper. The first author's work has been supported in part by the Deutsche Forschungsgemeinschaft (SFB 823: Statistik

nichtlinearer dynamischer Prozesse, Teilprojekt C2).

A. Proof of Proposition 1

We first find the Fréchet derivative of the criterion function $\Sigma(\xi)$ defined in (5), for the case of one covariate, at a design ξ in the direction of another design η, where

$$\xi = \begin{Bmatrix} x_1 & \dots & x_m \\ \omega_1 & \dots & \omega_m \end{Bmatrix} \quad \text{and} \quad \eta = \begin{Bmatrix} x_{m+1} & \dots & x_l \\ \omega_{m+1} & \dots & \omega_l \end{Bmatrix}.$$

Then

$$(1-\varepsilon)\xi + \varepsilon\eta = \begin{Bmatrix} x_1 & \dots & x_m & x_{m+1} & \dots & x_l \\ \omega_1^* & \dots & \omega_m^* & \omega_{m+1}^* & \dots & \omega_l^* \end{Bmatrix}$$

where $\omega_i^* = (1-\varepsilon)\omega_i$ if $i \le m$ or $\omega_i^* = \varepsilon\omega_i$ if $i > m$. Let $R_1(y) = \sum_{r=1}^m \omega_r \pi_r(y) \exp(\beta x_r)$ and $R_2(y) = \sum_{r=m+1}^l \omega_r \pi_r(m) \exp(\beta x_r)$. Then

$$\begin{aligned}
&\Sigma((1-\varepsilon)\xi + \varepsilon\eta) - \Sigma(\xi) \\
=&\sum_{i=2}^{l}\sum_{q<i} \omega_i^*\omega_q^* \exp(\beta(x_i+x_q))(x_i-x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y)+\varepsilon R_2(y)}\, dy \\
-&\sum_{i=2}^{m}\sum_{q<i} \omega_i\omega_q \exp(\beta(x_i+x_q))(x_i-x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)}\, dy \\
=&\sum_{i=2}^{m}\sum_{q<i} \omega_i\omega_q \exp(\beta(x_i+x_q))(x_i-x_q)^2 \\
&\int_0^\infty h_0(y)\pi_i(y)\pi_q(y)\left[\frac{(1-\varepsilon)^2}{(1-\varepsilon)R_1(y)+\varepsilon R_2(y)} - \frac{1}{R_1(y)}\right] dy \\
+&(1-\varepsilon)\varepsilon \sum_{i=m+1}^{l}\sum_{q=1}^{m} \omega_i\omega_q \exp(\beta(x_i+x_q))(x_i-x_q)^2 \\
&\int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y)+\varepsilon R_2(y)}\, dy + O(\varepsilon^2)
\end{aligned}$$

$$\begin{aligned}
=&\sum_{i=2}^{m}\sum_{q<i} \omega_i\omega_q \exp(\beta(x_i+x_q))(x_i-x_q)^2 \\
&\int_0^\infty h_0(y)\pi_i(y)\pi_q(y)\frac{-\varepsilon(R_1(y)+R_2(y)) + O(\varepsilon^2)}{R_1(y)[(1-\varepsilon)R_1(y)+\varepsilon R_2(y)]}\, dy \\
+&\varepsilon \sum_{i=m+1}^{l}\sum_{q=1}^{m} \omega_i\omega_q \exp(\beta(x_i+x_q))(x_i-x_q)^2 \\
&\int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y)+\varepsilon R_2(y)}\, dy + O(\varepsilon^2).
\end{aligned}$$

The Fréchet derivative is therefore

$$
\begin{aligned}
d(\xi,\eta) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(\Sigma((1-\varepsilon)\xi+\varepsilon\eta)-\Sigma(\xi)) \\
&= -\sum_{i=2}^{m}\sum_{q<i}\omega_i\omega_q\exp(\beta(x_i+x_q))(x_i-x_q)^2\int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)}\,dy \\
&\quad -\sum_{i=2}^{m}\sum_{q<i}\omega_i\omega_q\exp(\beta(x_i+x_q))(x_i-x_q)^2\int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)R_2(y)}{R_1^2(y)}\,dy \\
&\quad +\sum_{i=m+1}^{l}\sum_{q=1}^{m}\omega_i\omega_q\exp(\beta(x_i+x_q))(x_i-x_q)^2\int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)}\,dt.
\end{aligned}
$$

Clearly, $d(\xi,\eta)=\sum_{i=m+1}^{l}\omega_i d(\xi,\eta_i)$, where η_i is the one-point design with support x_i and weight 1, $i=m+1,\ldots,l$. (Equivalently, it can be shown that the Gâteaux derivative is linear in its second argument.) Therefore we only need to consider directions towards one-point designs. If ξ is optimal, $\Sigma((1-\varepsilon)\xi+\varepsilon\eta_i)-\Sigma(\xi)\le 0$ for all designs $\eta_i\in\mathcal{H}$, and the inequality $d(\xi^*,\eta)\le 0$ follows with $l=k+1$ and $x_{m+1}=x$.

Now, if ξ is optimal, $\max_\eta d(\xi,\eta)=0$, and clearly $0=d(\xi,\xi)=\sum_{i=1}^{m}\omega_i d(\xi,\xi_i)$ where $\xi_i=\{x_i;1\}$, $i=1,\ldots,m$. Hence $d(\xi,\xi_i)=0$ for all $i=1,\ldots,m$.

References

Andersen, P. K., Gill, R. D., 1982. Cox's regression model for counting processes: A large sample study. The Annals of Statistics 10 (4), pp. 1100–1120.

Atkinson, A., Donev, A., Tobias, R., 2007. Optimum Experimental Designs, with SAS (Oxford Statistical Science Series), 2nd Edition. Oxford University Press, USA.

Becker, N., McDonald, B., Khoo, C., 1989. Optimal designs for fitting a proportional hazards regression model to data subject to censoring. Australian Journal of Statistics 31 (3), pp. 449–468.

Collett, D., 2003. Modelling survival data in medical research. Chapman & Hall, London, New York.

Cox, D. R., 1972. Regression models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 34 (2), pp. 187–220.

Efron, B., 1977. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 72 (359), pp. 557–565.

Kalish, L. A., Harrington, D. P., 1988. Efficiency of balanced treatment allocation for survival analysis. Biometrics 44 (3), pp. 815–821.

Konstantinou, M., Biedermann, S., Kimber, A. C., 2014. Optimal designs for two-parameter nonlinear mdoels with application to survival models. Statistica Sinica 24, 415–428.

López-Fidalgo, J., Rivas-López, M. J., 2014. Optimal experimental designs for partial likelihood information. Computational Statistics & Data Analysis 71, 856–867.

López-Fidalgo, J., Rivas-López, M. J., Del Campo, R., 2009. Optimal designs for cox regression. Statistica Neerlandica 63 (2), pp. 135–148.

McGree, J., Eccleston, J., 2010. Investigating design for survival models. Metrika 72 (3), pp. 295–311.

Pascual, F. G., Montepiedra, G., 2003. Model-robust test plans with applications in accelerated life testing. Technometrics 45 (1), pp. 47–57.

Wu, S. J., Lin, Y. P., Chen, Y. J., 2006. Planning step-stress life test with progressively type i group-censored exponential data. Statistica Neerlandica 60 (1), pp. 46–56.

Zhang, Y., Meeker, W. Q., 2006. Bayesian methods for planning accelerated life tests. Technometrics 48 (1), pp. 49–60.