Received: 15 May 2022 | Revised: 5 September 2022 | Accepted: 14 October 2022

DOI: 10.1002/qre.3228

SPECIAL ISSUE ARTICLE

WILEY

Bayesian D-optimal design for life testing with censoring

Dustin Taylor[1] | **Steven E. Rigdon**[2] | **Rong Pan**[1] | **Douglas C. Montgomery**[1]

[1]School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, USA

[2]Department of Epidemiology and Biostatistics, Saint Louis University, Saint Louis, Missouri, USA

Correspondence
Dustin Taylor, School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, USA.
Email: dtaylo57@asu.edu

Present address
Dustin Taylor, Ira A. Fulton Schools of Engineering, P.O. Box 879309, Tempe, AZ 85287-9309, USA

[Correction added on 28 July 2023, after first publication: The Article category was corrected from RESEARCH ARTICLE to SPECIAL ISSUE ARTICLE.]

Abstract

The assumption of normality is usually tied to the design and analysis of an experimental study. However, when dealing with lifetime testing and censoring at fixed time intervals, we can no longer assume that the outcomes will be normally distributed. This generally requires the use of optimal design techniques to construct the test plan for specific distribution of interest. Optimal designs in this situation depend on the parameters of the distribution, which are generally unknown a priori. A Bayesian approach can be used by placing a prior distribution on the parameters, thereby leading to an appropriate selection of experimental design. This, along with the model and number of predictors, can be used to derive the D-optimal design for an allowed number of experimental runs. This paper explores using this Bayesian approach on various lifetime regression models to select appropriate D-optimal designs in regular and irregular design regions.

KEYWORDS

Bayesian design, coordinate exchange algorithm, irregular design region, life testing, optimal design

1 | INTRODUCTION

Suppose we are planning an experiment where the outcome is a lifetime and is subject to censoring at a fixed time. The settings of the predictor variable are constrained to be in some region R of the design space. Given a fixed sample size n (which is equivalent to the number of experimental runs), how should the levels of the predictor variables be selected in each experimental run so as to provide the greatest information from the experiment? This problem of optimal experimental design has been studied thoroughly when the outcome can be assumed to be normally distributed (see, e.g., Refs.[1–5]). But, for lifetimes, the normal distribution is usually unreasonable; in addition, life testing experiments must deal with censoring, that is, the termination of testing before all units have failed.

For the normal theory case, the situation is simplified in that the optimal design is often dependent on the model, but not on the parameters in the model. For example, an optimal design for a first-order model in two variables is independent of the intercept β_0 and the slope parameters, β_1 and β_2. It is also independent of the error variance σ^2. When we step outside of the normal-linear model, the situation becomes more complicated. For instance, assuming that the lifetime has a Weibull distribution (a useful and widely applicable reliability model for life test planning), the optimal design depends on the value of the shape parameter. This creates a catch: an experimenter is running the experiment so as to learn about the parameters of the assumed distribution, but only to realize that, to obtain the desirable optimal experimental design, the values of these parameters must be known beforehand.

Dustin Taylor, Steven E. Rigdon, Rong Pan, and Douglas C. Montgomery contributed equally to this study.

Roy and Pradhan[6] consider designs for life-time plans by developing a D-optimal Bayesian censoring scheme focusing on the estimation precision of a particular lifetime quantile. They propose a generic algorithm to obtain the optimal censoring schemes under two different scenarios for the Weibull and log-normal models but do not address predictor variables as we do in our paper. Zhobi et al.[7] develop an algorithm for finding D-optimal designs when the response is a lifetime following a Weibull distribution. While censoring is taken into account, they assume a fixed and known shape parameter for the Weibull distribution, which is a strong assumption. Our approach described here assumes that the shape parameter κ is unknown and must be estimated from the data.

The Bayesian approach can be applied to this situation where the optimal design depends on unknown parameters. The Bayesian approach involves selecting a prior distribution to reflect our knowledge about parameters before data are collected, and then after data are collected, applying Bayes theorem to give the updated probability distribution of the parameters given the data. It forces us to quantify our belief about parameters in terms of their uncertainties. This knowledge can then be used to select a robust experimental design. Additional details in the construction of Bayesian models have been explored in Refs.[8–11].

Suppose that the lifetime of the ith run is T_i and the vector of predictor variables is $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^t$. We assume a model of the form

$$T_i|\boldsymbol{x}_i \sim \text{WEIBULL}(\theta = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \kappa)) \tag{1}$$

This is equivalent to assuming that the logarithm of lifetime has a smallest extreme value (SEV) distribution with location parameter μ and scale parameter σ. If we let $Y_i = \log T_i$, then we are effectively assuming

$$Y_i|\boldsymbol{x_i} \sim \text{SEV}(\mu = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \sigma = 1/\kappa). \tag{2}$$

The method for finding D-optimal designs that we propose in this article can be applied to models of any order. We illustrate the method in Section 5, for situations of up to a two predictors and a full second-order model. The most general model that we illustrate is

$$Y_i|\boldsymbol{x_i} \sim \text{SEV}(\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2, \sigma = 1/\kappa). \tag{3}$$

Thus, the parameters of the SEV are related to the parameters of the Weibull as follows:

$$\begin{aligned} \mu &= \log\theta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \\ \sigma &= 1/\kappa. \end{aligned} \tag{4}$$

Note that if the predictors $x_1, \cdots, x_p$ are environmental stress variables, which accelerate a test unit's failure process, then such reliability tests are called accelerated life tests (ALTs). In an ALT, the value ranges of parameters $\beta_1, \cdots, \beta_p$ are typically restrictive, because higher stress levels are supposed to shorten a test unit's lifetime. In this paper, however, we do not limit our models to ALT models. These predictors, for example, can be product design variables or manufacturing process variables thus their effect coefficients would not be constrained to a certain range.

As with most life testing experiments, we assume that the test can be terminated before all units have failed. Here we assume that all survival observations are censored at the same time t_c. On the log scale, this is equivalent to censoring at time $y_c = \log t_c$. Denote δ as an indicator variable for censoring,

$$\delta_i = \begin{cases} 1, & \text{observation } i \text{ is a failure} \\ 0, & \text{observation } i \text{ is censored.} \end{cases} \tag{5}$$

The problem to be addressed is how the levels for the predictor variables should be selected so as to maximize the information that we can gain from the experiment. If the Weibull distribution is assumed for lifetimes, or equivalently, the SEV distribution for log lifetimes, then the optimal design depends on the values of the parameters in these distribution models. As we mentioned previously, the normal-linear model avoids this dilemma, but for a design of a life testing experiment, this problem is unavoidable.

There have been different approaches taken to circumvent this dilemma. Park and Yum[12] determined various pre-estimates for these parameters and then found a separate test design for each individual pre-estimate. This produces

different designs based on predetermined values. Others have focused on using the log normal distribution and the distribution transformation methods that prevent the parameter-dependent problem when determining the test points.[13] Also, some researchers used predetermined stress levels based on single-stress tests, previous tests of similar units, prior knowledge on life–stress relationship, or equipment capacities.[14–18]

Our way forward is to take the Bayesian approach and express our beliefs about parameters in their prior distributions. If we assume an SEV distribution for log lifetimes as in Equation (2), then the parameter vector is $\theta = (\beta, \sigma)$. Let $p(\theta)$ denote the prior distribution for θ used for the purpose of designing the experiment. Hong et al.[19] suggest that it may be reasonable to have separate priors for the design stage and the analysis stage. For now, $p(\theta)$ represents the prior at the design stage.

Suppose ξ is a design of the form

$$\xi = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{6}$$

where x_{ij} is the ith level for variable j. Note that this is different from the model matrix, which depends on the assumed model. For example, if a second-order model with intercept is assumed, then the first column of the design matrix would consist entirely of ones, and there would also be columns for each x_j^2 and for all combinations of $x_{j_1}x_{j_2}$. The model matrix can always be constructed from the design matrix ξ by augmenting the matrix with the appropriate columns given ξ and the model. The point is that, to find an optimal design, we need to only determine the best ξ, which embodies the settings of all predictors. Once this is determined, the usual model matrix can be obtained. Additional information on Bayesian designs can be found in Chaloner and Verdinelli[20] and Anirban.[21]

Let $c(t, \xi)$ denote some design criterion for the case when we observe data (t) obtained from applying design ξ. In our case, we take $c(t, \xi)$ to be the determinant of the estimated covariance matrix for the estimate of θ. We can use the preposterior expectation of $c(t, \xi)$

$$C(\xi) = E_{t|\xi}(c(t, \xi)) \tag{7}$$

$$= \int_{\Theta} c(\hat{\theta}, \xi) p(\hat{\theta}) d\hat{\theta} \tag{8}$$

$$\approx \int_{\Theta} c(\theta, \xi) p(\theta) d\theta. \tag{9}$$

In going from Equations (7) to (8) we have assumed that the criterion c depends on the data t only through the estimate $\hat{\theta}$. The approximation from (8) to (9) was suggested by Hong et al.[19]; they describe the approximation as "The predictive distribution of $\hat{\theta}$ is a convolution of the test planning prior distribution of θ and the distribution arising from the estimation of θ from the data and will approach the test planning prior distribution as the sample size increases." The last expression has the interpretation as the criterion function c averaged across the joint prior distribution for θ.

As mentioned above, many articles have explored the cases of normal outcomes, single censoring levels or regular shaped grid. Others have explored outcomes with a Weibull distribution with presumed lifetime model. This article expands upon these ideas by exploring and deriving optimal designs for lifetime testing when the outcomes are assumed to have a Weibull distribution, multiple censoring levels and irregularly shaped design regions. We explore the case where the priors have discrete uniform supports and apply the coordinate exchange algorithm to produce the best Bayesian D-optimal design. We go further by taking into account different censoring levels, as well as irregularly shaped design grids, and explore how these may affect the optimal design.

The rest of the article is organized as follows. Section 2 discusses various optimality criteria that might be used to design optimal experiments for life testing. Then, Section 3 discusses how we approximate the posterior distribution, which is required for the design criterion suggested here. Section 4 discusses how we apply a coordinate exchange algorithm to determine the optimal design. Some examples are discussed in Section 5 and conclusions provided in Section 6. For a review of accelerated test models, we direct the reader to Escobar and Meeker.[22]

2 | OPTIMALITY CRITERIA

Lindley[23] suggested that an optimality criterion should involve the expected gain in the Shannon information. This is equivalent to maximizing the expected Kullback–Leibler distance between the prior and posterior distributions. Thus, the objective is to maximize

$$U(\xi) = \int_{\mathcal{Y}} \int_{\Theta} \log \frac{p(\boldsymbol{\theta}|\boldsymbol{y}, \xi)}{p(\boldsymbol{\theta})} p(\boldsymbol{y}, \boldsymbol{\theta}|\xi) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{y} \tag{10}$$

where $\boldsymbol{y}$ is the observed data. This is approach is explored more thoroughly in Xu and Tang.[24]

Another approach for log-location-scale families, to include the SEV, is to minimize the volume of a posterior credible region, averaged over the prior distribution. A large sample approximation to the posterior covariance matrix was given by Hong et al.[19]; they give the approximation (using our notation)

$$V(\boldsymbol{\theta}|\boldsymbol{y}, \xi) \approx \left[S^{-1} + \hat{I}_{\boldsymbol{\theta}}(\xi)\right]^{-1} \tag{11}$$

where S is the variance–covariance matrix of prior distribution and $\hat{I}_{\boldsymbol{\theta}}(\xi)$ is the information matrix evaluated at $\hat{\boldsymbol{\theta}}$, which is the estimator of $\boldsymbol{\theta}$.

Following the reasoning in Hong et al.,[19] we can derive an approximation to the expected value of a function h of the covariance matrix:

$$\begin{aligned} U_h(\xi) &\approx E\left(h\left(\left[S^{-1} + \widehat{I}_{\boldsymbol{\theta}}(\xi)\right]^{-1}\right)\right) \\ &\approx \int h\left(\left[S^{-1} + \widehat{I}_{\boldsymbol{\theta}}(\xi)\right]^{-1}\right) p_0(\widehat{\boldsymbol{\theta}}) d\widehat{\boldsymbol{\theta}} \\ &\approx \int h\left(\left[S^{-1} + I_{\boldsymbol{\theta}}(\xi)\right]^{-1}\right) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned} \tag{12}$$

Here p_0 is the PDF (probability density function) of the estimator $\hat{\theta}$ and $p(\boldsymbol{\theta})$ is the prior PDF for $\boldsymbol{\theta}$ with $\hat{\theta}$ approaching the prior as the sample sizes increase.

If the objective is to minimize the determinant of the large-sample approximation to the posterior covariance matrix, then the utility of design ξ becomes

$$U_D(\xi) = \int_{\Theta} \det\left(S^{-1} + \widehat{I}_{\boldsymbol{\theta}}(\xi)\right)^{-1} p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{13}$$

We would then want a design that minimizes $U_D(\xi)$. This is similar to the approach taken by DuMouchel and Jones.[25]

If we take a prior across a discrete set of points in the parameter space, the D-criterion becomes

$$U_D(\xi) = \sum_{\beta_0} \sum_{\beta_0} \cdots \sum_{\beta_m} \sum_{\kappa} \det\left(S^{-1} + \widehat{I}_{(\beta_0, \beta_1, \cdots, \beta_m, \kappa)}(\xi)\right)^{-1} p(\beta_0, \beta_1, \cdots, \beta_m, \kappa). \tag{14}$$

The computations are more stable if we use the reciprocal $1/U_D(\xi)$ as the criterion, which we then try to maximize. Although any discrete prior would work, for our examples, we take a discrete uniform prior across all values in the prior support; in effect, this takes the harmonic average of the Bayesian D criterion at each point in the prior parameter space. Additional Bayesian utility functions for parameter estimation are discussed in Ryan et al.[26]

3 | SELECTION OF PRIOR DISTRIBUTIONS

In order to design a life testing experiment, we must express prior distributions for the model parameters. We will put these parameters into three categories:

1. the shape parameter κ,
2. the "intercept" parameter β_0, and
3. the "slope" parameters $\beta_1, \beta_2, \ldots, \beta_m$.

The shape parameter κ is perhaps the easiest parameter for which to specify a prior. In most life testing experiments, the hazard function is increasing but it is often concave down. Since the Weibull's hazard function is

$$h(t) = \frac{\kappa}{\theta}\left(\frac{t}{\theta}\right)^{\kappa-1} t > 0, \tag{15}$$

the increasing/concave down presumption implies that the exponent on t should be between 0 and 1, that is

$$0 \leq \kappa - 1 \leq 1. \tag{16}$$

Thus,

$$1 \leq \kappa \leq 2. \tag{17}$$

A prior with support on [1,2] could, therefore, be appropriate. To be safe, the user might extend the support to an interval such as [0.8,2.2].

Next, consider the intercept parameter β_0. The mean of the Weibull distribution is

$$E(T) = \theta\Gamma\left(1 + \frac{1}{\kappa}\right). \tag{18}$$

The expression $\Gamma(1 + \frac{1}{\kappa})$ is decreasing for κ belonging to the interval [1,2] having the value of 1 when $\kappa = 1$ and 0.8862 at $\kappa = 2$. (This function decreases until reaching approximately 0.8856 when $\kappa \approx 2.17$; it then increases monotonically toward 1 as $\kappa \to \infty$.) Thus, the expected lifetime is slightly below the value of θ, with the factor being no less than 0.8862 when κ is in the interval [1,2]. At the center of a design, where all coded predictors are set to 0, $\theta = \exp(\beta_0)$ is a slight overestimate of the mean. The parameter θ is often called the characteristic life, and it gives the reliability (or survival probability) as

$$S(\theta) = \exp(-(\theta/\theta)^{\kappa}) = \exp(-1) \approx 0.3679. \tag{19}$$

In other words, approximately 36.79% of all units will survive past the characteristic life regardless of the value of κ. If we suggested a wide interval for θ, say

$$100 \leq \theta \leq 1000; \tag{20}$$

that is, a tenfold difference in the possible values of θ, then

$$100 \leq \exp(\beta_0) \leq 1000 \tag{21}$$

$$4.6 \leq \beta_0 \leq 6.9. \tag{22}$$

The parameter β_0 is a scale parameter for the Weibull distribution, so its value does not affect the optimal design, but it does affect the amount of censoring, that is, the value y_c, as described below. In the examples that follow, we will take the support for the prior for β_0 to be the interval $[-1, 1]$, which leads to the prior for θ having support [0.3679,2.7183] at the center of the design.

The slope parameters $\beta_1, \beta_2, \ldots$ are the most difficult to assess. We will consider the case where we have two predictors and a full second-order model, so that

$$\log\theta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2. \tag{23}$$

Suppose that as we move from the center design point (0,0) to the one of the corners, say (1,0) the mean lifetime is doubled, which implies that θ is doubled. Then at the corner point (1,0),

$$\begin{aligned}\theta &= \exp(\beta_0 + \beta_1 \times 1 + \beta_2 \times 2 + \beta_{11} \times 1 + \beta_{22} \times 0^2 + \beta_{12} \times 1 \times 0) \\ &= \exp(\beta_0 + \beta_1 + \beta_{11}) \\ &= \exp(\beta_0)\exp(\beta_1 + \beta_{11}). \end{aligned} \tag{24}$$

The expected lifetime at (1,0) is then

$$E(T|x_1 = 1, x_2 = 0) = \exp(\beta_0 + \beta_1 + \beta_{11})\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{25}$$

$$= \quad \exp(\beta_0)\exp(\beta_1 + \beta_{11})\Gamma\left(1 + \tfrac{1}{\kappa}\right). \tag{26}$$

At (0,0), the expected lifetime is

$$E(T|x_1 = 1, x_2 = 0) = \exp(\beta_0)\Gamma\left(1 + \frac{1}{\kappa}\right). \tag{27}$$

If we believe, the expected lifetime increases by a factor of F as we move from (0,0) to (1,0), we must have

$$\exp(\beta_0)\exp(\beta_1 + \beta_{11})\Gamma\left(1 + \frac{1}{\kappa}\right) = F\exp(\beta_0)\Gamma\left(1 + \frac{1}{\kappa}\right). \tag{28}$$

which implies that

$$F = \exp(\beta_1 + \beta_{11}), \tag{29}$$

or, equivalently,

$$\beta_1 + \beta_{11} = \log F. \tag{30}$$

This gives us information we could use to assess our prior information about the sum $\beta_1 + \beta_{11}$, but to assess our prior information about each term individually, we must make some assumptions about the relative sizes of β_1 and β_{11}. If we assume that they are roughly equal, then

$$2\beta_1 = \log F \tag{31}$$

so

$$\beta_1 = \frac{1}{2}\log F. \tag{32}$$

In ALTs, a 10 fold increase in lifetime expectancy from the low to high setting of the accelerating factor is not uncommon. Taking $F = 10$ yields

$$\beta_1 = \frac{1}{2}\log 10 \approx 1.15. \tag{33}$$

Thus, priors for β_1 and β_{11} over the interval $[-1.15, 1.15]$ might be appropriate. Similarly, we see that priors over $[-1.15, 1.15]$ might be appropriate for β_2, β_{22}, and β_{12}. Of course, more diffuse priors can be used to achieve a more objective prior.

Life testing experiments are often done on units that are similar to previous generations of the unit, so some prior information is likely available. The prior distributions can reflect this information. When choosing a prior distribution for a parameter that is difficult to interpret directly, we can often transform the problem to one for which we can more easily interpret, select a prior for that, and then work back to the prior for the parameter. This is the trick we have employed here. The parameters β_1, β_2, ..., are difficult to interpret, so we transformed the problem to one involving the effect of a factor moving from 0 to 1. Once we assess our prior belief about this parameter, we can transform this back to obtain a prior for each of the βs.

The approach described above gives a general way of specifying reasonable upper and lower limits of parameter values for the parameters in a Weibull regression model. If no additional information is available, then we may use uniform priors for these parameters. In the Examples section that follows, we make the assumptions that the priors for the slope parameters are discrete uniform on the Cartesian product of the regions from −2 to 2. Here we have chosen priors that are somewhat more diffuse than described previously. For example, for the first-order model with one factor, we choose the prior to have equal probability assigned to each of the values in the Cartesian product given by

$$\begin{aligned} \beta_0 \in &\quad \texttt{seq}(-1, 1, \texttt{length.out} = 5) \\ \beta_1 \in &\quad \texttt{seq}(-2, 2, \texttt{length.out} = 5) \\ \kappa \in &\quad \texttt{seq}(1, 2, \texttt{length.out} = 5) \end{aligned} \tag{34}$$

Here we are using R's notation for creating an arithmetic sequence. The priors for the βs must reflect the knowledge about the characteristic lifetime and will depend on the time scale used to measure lifetimes.

For second-order models in two factors, we use `length.out=3` in order to keep the computations manageable; this yields a discrete prior over $3^7 = 2187$ distinct grid points.

We also consider prior distributions that guarantee a monotone increasing or decreasing characteristic life. For example, if the factor(s) are life-accelerating stress factors, for which it is safe to assume higher levels of stress will lead to a shorter characteristic life. This can be done by assuming that the predictor variables are coded so that *smaller* values indicate a higher stress and slope parameters are assigned priors with lower bounds greater than 0.

Obviously, more precise prior distributions for Weibull distribution parameters can be defined when we do have some historical lifetime data or testing data from the same or similar products. As an example, consider the data from a nickel super alloy fatigue test that was analyzed in Escobar and Meeker[22] and Rigdon et al.[27] The lifetime of this product was measured in kilocycles and one stress variable, called pseudostress (PS), was applied during testing. It was found that a suitable life-stress model was a second-order model such as

$$\log \theta = \beta_0 + \beta_1 \log PS + \beta_2 (\log PS)^2, \tag{35}$$

where $\log PS$ is the logarithm of pseudostress. These regression coefficients, β_0, β_1, and β_2, were estimated to be 217.6, −85.5, and 8.48, with standard errors (se) of 62.1, 26.5, and 2.8, respectively. If we transform the log of pseudostress through

$$x = -15.6893 + 7.711884 \log(PS) \tag{36}$$

then the values fall between −1 and 1. Solving for $\log(PS)$, we have

$$\log PS = \frac{x + 15.6893}{7.711884}. \tag{37}$$

The estimated response, kilocycles to failure, is then

$$\begin{aligned} \hat{y} &= 217.6 - 85.5\left(\frac{x + 15.6893}{7.711884}\right) + 8.48\left(\frac{x + 15.6893}{7.711884}\right)^2 \\ &= 78.7541 - 6.61266x + 0.142585x^2 \end{aligned} \tag{38}$$

TABLE 1 Censoring levels considered in various examples. The probability of being censored refers to the probability that an observation at the center of the design is censored if $\beta_0 = 0$ and $\kappa = 1.5$.

y_c	P(censor)	Description
2	2×10^{-9}	Negligible
1	0.0113	Light
0	0.3699	Moderate
−1	0.8000	Heavy
−2	0.9514	Very Heavy

where $-1 \leq x \leq 1$. Therefore, for deriving a Bayesian optimal test plan, the informative priors for these parameters could be formulated to be in the range of $\hat{\beta} \pm 2se(\hat{\beta})$, and these ranges can be further coded into $[-1, 1]$.

4 | COORDINATE EXCHANGE ALGORITHM

The coordinate exchange algorithm is described in Meyer and Nachtsheim.[28] This algorithm follows the cyclic ascent (or Gauss–Seidel) optimization method as described in Bazaraa et al.[29] Here we summarize the method and describe how we apply this algorithm to the problem of optimal design for life tests.

Let us begin with some design ξ_0 that describes the coordinates for the predictor variables. It differs from the model matrix in that it consists of only the design points, not the columns of all ones (for the constant term) or terms for squared or cross products. The model matrix, which can be computed from the design ξ_0, is dependent on the assumed model. We consider a rectangular grid of points across the design space and focus on one of the coordinates of one design point at a time, say, the specific design choices of the first and second predictors of one design point in a two-factor model. The design criterion (e.g., the Bayesian D-criterion given in Equation (13)) is then evaluated for all possible values of a single predictor in the specified grid when the other predictor value is fixed. Then the point that yielded the largest (or smallest, depending on the criterion) is taken and assigned as the value for the corresponding row in the design ξ. This process is then repeated across all rows, yielding the design ξ_1.

The above algorithm is then repeated on the design ξ_1, yielding ξ_2. This process is continued until consecutive designs are the same. This iterative approach allows the "restarting" of the search, but retains the good features and attributes of the obtained solution.[30] Often two iterations suffice, and nearly always five iterations are enough. Recall our second-order model with two factors, using `length.out=3`, yielded 2187 distinct grid points requiring evaluation.

When this algorithm is applied life testing designs, we have to consider the impact of censoring on optimal criterion evaluation. The Appendix gives the details on how to derive and evaluate the expected Fisher information matrix with right censored observations. In the following examples section, multiple scenarios are considered.

5 | EXAMPLES

In the subsections that follow, we look at the optimal designs for various situations, including one or two factors and a first- or second-order model. We mostly consider designs on the square $[-1, 1] \times [-1, 1]$, but also touch on irregular design regions such as a trapezoidal region and a circular region. Throughout, we apply Bayesian D-optimality as the criterion and use R to run the computations. An in-depth look into this algorithm can be found in Overstall and Woods[31] and in Overstall et al.[32]

We consider two types of priors. One is discrete uniform over $[-2, 2]$ for all of the β_i parameters, including the intercept β_0, and discrete uniform from 1 to 2 for κ. The other is a prior that guarantees an increasing expected lifetime as a function of the factors; this prior might be of interest if the factors were accelerating factors coded so that smaller values of the factor represent higher stress.

Often, the optimal design depends on the level of censoring, particularly when the factors are accelerating factors. We have categorized the levels of censoring according to the value of y_c. Table 1 shows the values we consider and the probability of censoring at the center of the design when $\beta_0 = 0$ and $\kappa = 1.5$.

5.1 | One predictor with first-order model

The simplest example to consider is that of one factor with a first-order model with negligible censoring. It might be expected that half of the runs would be allocated at the left and right endpoints of the design space, which is denoted as $[-1, 1]$. If $n = 10$, the optimal design is found to be have five points at -1 and five points at 1, which we indicate by the ordered pair (5,5); this design yields a D-utility of 1054.573. For $n = 20$ and $n = 40$, the optimal designs are (10,10) and (20,20).

Suppose, however, that the prior for β_1 is not symmetric, but rather has support only on 0 and positive values. This might be the case if the predictor variable was an accelerating variable, or for other reason, we were confident that the coefficient was nonnegative. In this case, the optimal design may try to put equal numbers of runs at the endpoints, but for odd values of n, the preference is to put the larger half at the low value. The reason for this seems to be that runs at the high end tend to be longer, so they are more likely to be censored, so that these runs yield less information. Under very heavy censoring, the optimal design for an even number of runs may put more points at the low end. For example, the optimal design for $n = 10$ is (5,5), but the optimal design for $n = 11$ is (6,5). The D-utility for (6,5) is 1474.605 whereas the D-utility for (5,6) is 1455.642. With very heavy censoring, an even numbered sample size can yield an unbalanced design; for example with $y_c = -2$ and $n = 10$, the optimal design is (6,4) with a D-utility of 50.04983. By contrast, the design (5,5) has D utility equal to 48.00066. With $n = 80$, the optimal design is (48,32).

5.2 | One predictor with second-order model

Consider a second-order model in a single predictor that has the feasible region of unit interval $[-1, 1]$. The budget allows for $n = 10$ runs to be selected from the interval [0,1]. We apply the coordinate exchange algorithm to find the Bayesian D-optimal design. One prior support consists of all possible selections of

$$\begin{aligned}\beta_0 &\in \texttt{seq}(-2, 2, 0.5)\\ \beta_1 &\in \texttt{seq}(-2, 2, 0.5)\\ \beta_2 &\in \texttt{seq}(-2, 2, 0.5)\\ \kappa &\in \texttt{seq}(0.8, 2.2, 0.2)\end{aligned} \tag{39}$$

where `seq(a,b,d)` is the sequence from `a` to `b` in increments of `d`. The prior is discrete uniform over this support.

A second prior was chosen to include only those parameters that guarantee a decreasing function on the interval $[-1, 1]$. This might be reasonable when the factor is an accelerating factor for which we know that $\log\theta = \beta_0 + \beta_1 x + \beta_2 x^2$ is an increasing function from the highest stress condition ($x = -2$) to the use stress condition ($x = 1$). This is found to include those values of β_1 and β_2 that satisfy

$$-\frac{1}{2}\beta_1 \le \beta_2 \le \frac{1}{2}\beta_1 \tag{40}$$

We assumed censoring at $y_c = \log t_c = 0$ or 2, corresponding to heavy and light censoring. The parameter space is the interval $[-1, 1]$. With a second-order model, we would expect the optimal design to have at least three distinct points: one at the left end, one at the right and one near the middle. Without at least three points, the parameters of the model cannot be estimated. We ran three loops through the coordinate exchange algorithm with a grid of 10 intervals, or equivalently 11 design points between 0 and 1, yielding $\delta = 0.10$, and the last two designs were identical. The optimal designs are shown in Table 2 for various choices of a prior (symmetric vs. a prior that guarantees an increasing link function) and censoring (light vs. heavy).

In each case, the optimal design involves three points, $x = -1$, $x = 0$, and $x = 1$. With $n = 10$ runs, it is not possible to place the same number of points at each design point. The design is represented by an ordered triple, where the three numbers (i, j, k) indicate that i runs are to be placed at $x = -1$, j runs at $x = 0$, and k runs are placed at $x = 1$. The optimal design depended on the circumstance, with the extra point sometimes going at $x = -1$, sometimes at $x = 0$ and sometimes at $x = 1$. The utilities for various designs are shown in the last column of Table 2. Note that when the censoring is heavy,

TABLE 2 Optimal designs for $n = 10$ run designs over parameter space $[-1, 1]$

Prior	Censoring	Optimal design	Utility
Symmetric	Heavy ($y_c = 0$)	(4,3,3)	601.498
Symmetric	Light ($y_c = 2$)	(3,4,3)	1531.333
Guarantees increasing link	Heavy ($y_c = 0$)	(3,3,4)	464.855
Guarantees increasing link	Light ($y_c = 2$)	(3,3,4)	1439.007

TABLE 3 Designs for second-order model in one variable with a budget of $n = 10$ runs. All designs involve runs at three distinct points, $x = -1$, a point near $x = 0$, and $x = 1$. The design is represented by an ordered triple, where the three numbers indicate the number of runs at each point. The prior is either symmetric and flat for β_0, β_1, and β_2, or it is flat across the set of β_1 and β_2 that guarantee an increasing link function. Censoring is either heavy ($y_c = 0$) or light ($y_c = 2$).

Symmetric prior				**Prior guarantees increasing function**			
Censoring $y_c = 0$		**Censoring at $y_c = 2$**		**Censoring at $y_c = 0$**		**Censoring at $y_c = 2$**	
Runs at $-1, 0, 1$		Runs at $-1, 0, 1$		Runs at $-1, 0, 1$		Runs at $-1, 0, 1$	
Design	Utility	Design	Utility	Design	Utility	Design	Utility
(4,3,3)	601.4985	(4,3,3)	1526.579	(4,3,3)	444.9591	(4,3,3)	1413.522
(3,4,3)	601.3223	(3,4,3)	1531.333	(3,4,3)	462.0097	(3,4,3)	1436.669
(3,3,4)	601.4985	(3,3,4)	1526.579	(3,3,4)	464.8547	(3,3,4)	1439.007
Runs at $-1, -0.01, 1$		Runs at $-1, -0.01, 1$		Runs at $-1, -0.01, 1$		Runs at $-1, -0.01, 1$	
Design	Utility	Design	Utility	Design	Utility	Design	Utility
(4,3,3)	601.4153	(4,3,3)	1526.296	(4,3,3)	444.3329	(4,3,3)	1412.734
(3,4,3)	601.2124	(3,4,3)	1531.032	(3,4,3)	461.2848	(3,4,3)	1435.784
(3,3,4)	601.3609	(3,3,4)	1526.260	(3,3,4)	464.1806	(3,3,4)	1438.245
Runs at $-1, 0.01, 1$		Runs at $-1, 0.01, 1$		Runs at $-1, 0.01, 1$		Runs at $-1, 0.01, 1$	
Design	Utility	Design	Utility	Design	Utility	Design	Utility
(4,3,3)	601.3609	(4,3,3)	1526.260	(4,3,3)	445.3981	(4,3,3)	1413.735
(3,4,3)	601.2124	(3,4,3)	1531.032	(3,4,3)	462.5391	(3,4,3)	1436.968
(3,3,4)	601.4153	(3,3,4)	1526.260	(3,3,4)	465.3329	(3,3,4)	1439.321

with $y_c = 0$, the utilities are much smaller than when there is light censoring. This is due to less information contained in censored observations compared to uncensored ones.

The optimal design found from this algorithm is required to have runs only at the grid points. With a grid of just 11 points for each variable (121 in total), it may be that the optimal design on the continuous design space maybe slightly different from the one on the discretized grid. Since it is unlikely that moving points away from the endpoint $x = -1$ or $x = 1$ will increase a design's utility, we considered the possibility that moving the middle point to the left or right might improve the design. Table 3 shows the results of this perturbation experiment. In most cases, moving the point left or right a small amount, say 0.01, did not improve the utility of the design. For the case of a guaranteed increasing link function with heavy censoring, taking the middle point at $x = 0.01, 0.02, 0.03, 0.04$ yielded utilities of 465.3329,465.6146,465.6994,465.5874, respectively. Thus, the design with (3,3,4) points at $x = -1, x = 0.03, x = 1$ is slightly preferred over (3,3,4) points at $x = -1, x = 0, x = 1$ or $x = -1, x = 0.01, x = 1$.

5.3 | Two predictors with first-order model

With a prior for β_0, β_1, and β_2 symmetric about the origin, the optimal design for $n = 40$ puts 10 points at each of the four corners $(-1, -1), (1, -1), (-1, 1)$, and (1,1). We will denote this design as (10,10,10,10). When the prior for β_1 and β_2 has support only on nonnegative values, then runs at the high end of either x_1 or x_2 are less attractive, because they are more likely to be censored. Depending on the amount of censoring, the optimal design may put more runs at the lower left corner ($x_1 = -1$ and $x_2 = -1$). See Table 4.

TABLE 4 Optimal designs for two factors and a first-order model. The optimal designs, as determined by the coordinate exchange algorithm are in the column headed by design 1. Designs 2, 3, and 4 are other designs that are close to the optimal design (sometimes giving exactly the same utility)

		Light censoring ($y_c = 2$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
−1	−1	11	11	12	11
1	−1	10	11	10	10
−1	1	11	10	10	10
1	1	8	8	8	9
Utility		2367,228	2367,228	2359,288	2357,229
		Heavy censoring ($y_c = 0$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
−1	−1	11	11	12	11
1	−1	10	11	10	10
−1	1	11	10	10	10
1	1	8	8	8	9
Utility		844,958	844,958	844,308	826,128
		Heavier censoring ($y_c = -1$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
−1	−1	14	14	13	12
1	−1	13	12	12	11
−1	1	13	12	12	11
1	1	0	2	3	6
Utility		399,423	385,753	377,213	345,831
		Very heavy censoring ($y_c = -2$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
−1	−1	15	15	16	14
0.8	−1	12	13	12	13
−1	0.8	13	12	12	13
Utility		92,897	92,897	92,777	92,586

5.4 | Two predictors with second-order model

Now consider the case of two predictor variables with a full second-order model of the form

$$T_i \sim \text{WEIBULL}(\kappa, \theta_i) \tag{41}$$

where

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2. \tag{42}$$

If the prior for $(\beta_1, \beta_2, \beta_{12}, \beta_{11}, \beta_{22})$ is symmetric about zero and the censoring level is light, with $y_c = 2$, then the optimal design for $n = 10$ runs on a grid with 10 intervals over the domain $[-1, 1] \times [-1, 1]$ is the one shown in the top left panel of Figure 1. This design has one run at each corner, one run at the midpoint of each side, one center point run, and one additional run at one of the corners. In fact, putting two runs at any one corner and one run at every other corners yields exactly the same D criterion.

To see the effect of the fineness or coarseness of the grid, we considered moving the center point around near the middle of the design region. The "center" point was chosen to be on the grid $\{-0.03, -0.02, \ldots, 0.03\} \times \{-0.03, -0.02, \ldots, 0.03\}$. The optimal design on the grid $\{-1.0, -0.8, \ldots, 1.0\} \times \{-1.0, -0.8, \ldots, 1.00\}$ yielded a D-criterion of 49243.6. The best design obtained by moving the center point across this finer grid yielded a D-criterion of 49660.78. Thus, relative to the better design (but not necessarily the *optimal* design on an even finer grid), the optimal design obtained from the coarse grid has

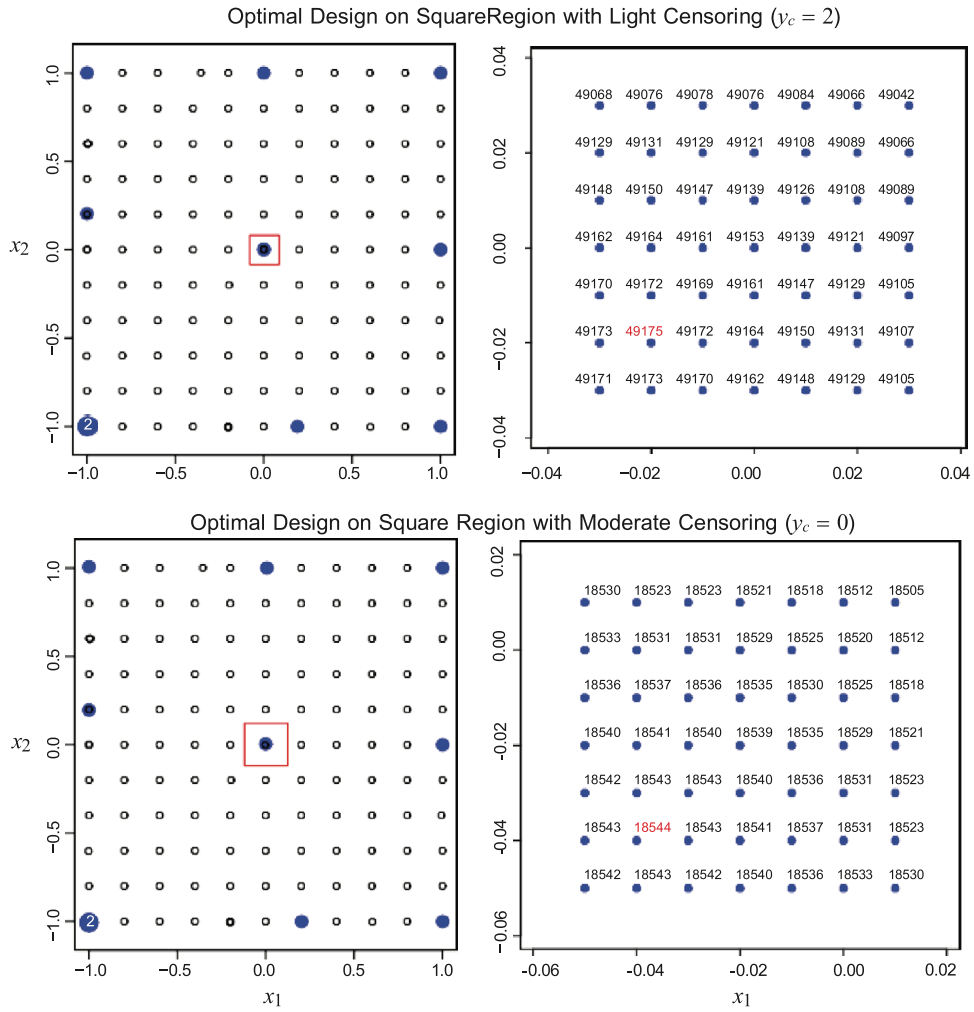


FIGURE 1 Two factors and a second-order model. The prior is symmetric about the origin and the number of runs is $n = 10$. Censoring is light ($y_c = 2$) in the top two graphs and moderate ($y_c = 0$) in the bottom two. The optimal designs on the grid with 10 intervals (width of grid points is 0.2) are shown on the left. In the right two figures, we moved the single center point around to see the effect of the fineness of the grid

a relative efficiency of $49244/49661 \approx 0.99$. Thus, it seems that a grid with $n = 10$ intervals is sufficient to obtain a design that is very close to the optimum over a finer grid.

5.5 | Two predictors with second-order model and irregular trapezoidal region

Consider now a second-order model with two predictor variables over an irregular design region. State-Ease[33] suggested an example, which is equivalent to having a design region consisting of the unit square $[0, 1] \times [0, 1]$ subject to $x_2 \geq 1/3 - x_1/3$. This region is shown in Figure 2 with a regular grid, where the lines are parallel to the coordinate axes. The problem with this grid is that boundary points are not necessarily grid points. For example, the lower left corner point is not a grid point. Since optimal designs often have experimental runs at the corners, this means that the optimal design over this grid must instead put runs at the grid point nearest to the corner.

Figure 2 further shows the result of applying the coordinate exchange algorithm where possible experimental points are constrained to be on the regular grid. Runs are placed at the two nearest grid points to the lower left corner. Also the middle point on the lower boundary occurs at $x = 0.4$, not $x = 0.5$ as we might expect.

We have found that it is better to create a grid whose lines are not necessarily parallel to the coordinate axes. Basically, the idea is to divide each of the four edges into 10 intervals, and then connect corresponding points at opposite ends. Such a grid is shown in Figure 3.

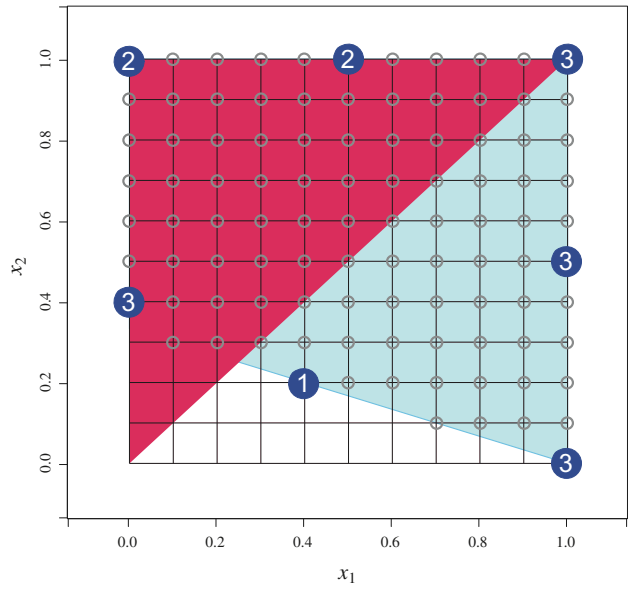


FIGURE 2 Optimal design for irregular region with nonparallel grid lines

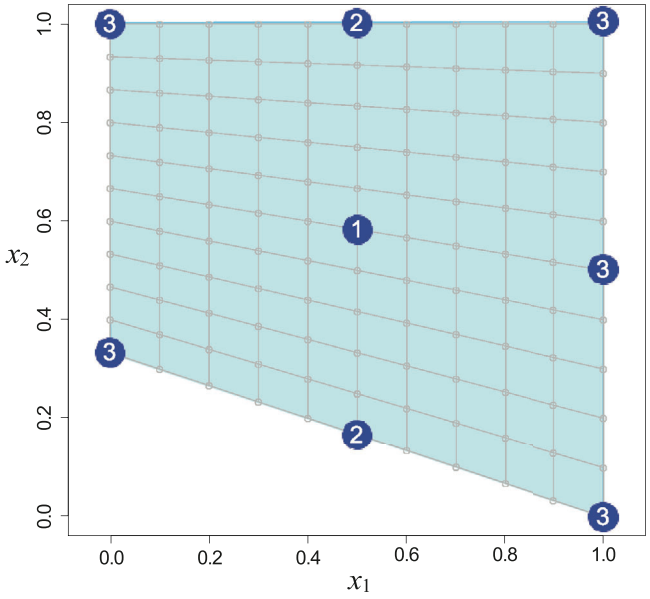


FIGURE 3 Optimal design for irregular region with grid lines parallel to the coordinate axes

5.6 | Two predictors with second-order model and circular region

Consider now a second-order model with two predictor variables over a circular design region. The mapping from a square grid to an ellipse grid can be seen below where x and y are the coordinates within a square grid over [−1,1]:

$$(x', y') = \left(x\sqrt{1 - \frac{y^2}{2}}, y\sqrt{1 - \frac{x^2}{2}} \right) \tag{43}$$

Additional details can be found at Ref.[34]. This circular region is shown in Figure 4 with a square grid that has been transformed, or mapped, to a circle. Because of the discreteness of the grid, it is usually impossible to obtain points that are equally spaced. For example, with a budget of 20 runs, having 3 at the center leaves 17 to be placed on the circumference

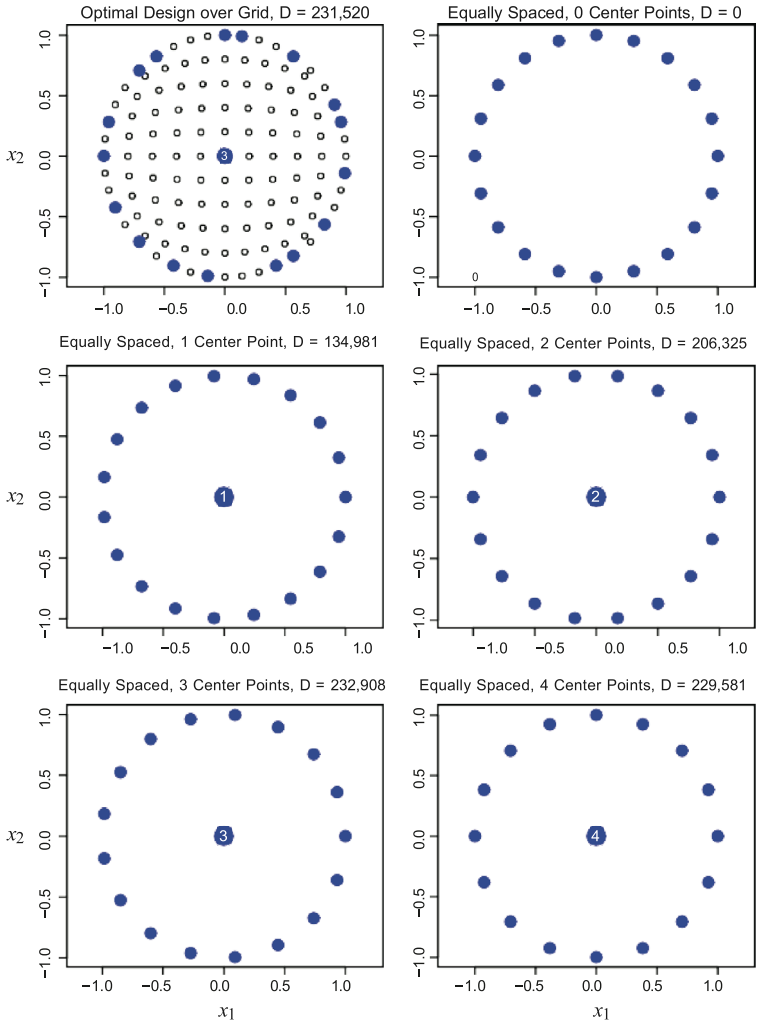


FIGURE 4 Circular design region with $n = 20$ runs for a second-order model in two factors. The prior for β_1, β_2, β_{11}, β_{22}, and β_{12} is symmetric about the origin.

of the circle. An 11×11 grid over a square yields 40 points along the perimeter of the circle. It is not possible to place these 17 points on the perimeter in a way that they are equally spaced.

The top left graph in Figure 4 shows the grid and the optimal design over this grid. The other graphs in Figure 4 show the designs with 0, 1, 2, 3, and 4 center points, leaving 20, 19, 18, 17, and 16 points, respectively, along the perimeter. Figure 4 shows the D-criterion for each of these designs. Among these designs, the one with 3 center points has the highest D-criterion. This design is slightly better ($D = 232.908$ vs. 231.520) than the one over the grid shown in the top left figure. Figure 5 further explores the grid to circle mapping for different censoring levels; light, moderate, heavy, and very heavy censoring.

6 | CONCLUSION

The Bayesian approach to experimental design offers many advantages over other approaches, the most notable of which is the ability to optimize design criteria that are functions of the posterior distribution and that can be easily tailored to the experimenter's design objective.[26] In this article, we developed an approach to perform D-optimal design for life testing models with parameter uncertainty and censoring. We further implemented our process on the 1st order, 1 predictor model through the 2nd order, 2 predictor model. "Fine" tuning is not necessarily needed past the tenth (.1) decimal for providing a high utility of the design. We then expanded the typical grid design region into a trapezoidal or spherical grid. We also varied the amount of censoring when establishing the optimal designs. That is, placing runs of an experiment at the higher end of the grid, that is, the low stress levels, would increase run times and lead to censoring, which provides less information than setting runs at the lower end of the grid.

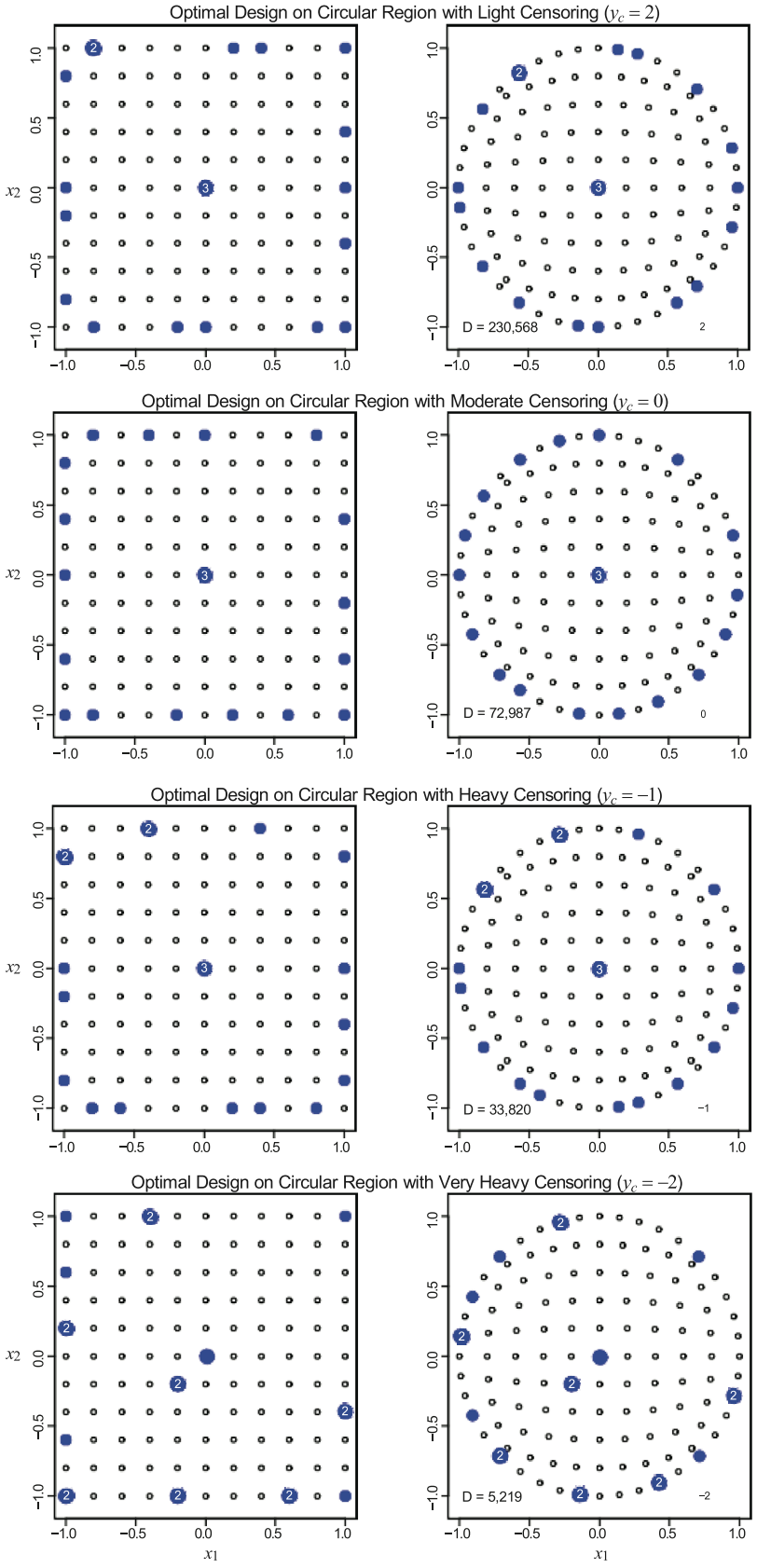


FIGURE 5 Optimal designs for $n = 20$ runs with asymmetric prior that guarantees an increasing link function in both x_1 and x_2. Censoring is light, moderate, heavy, and very heavy.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Rong Pan https://orcid.org/0000-0001-5171-8248

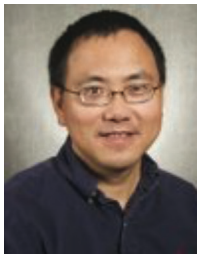
REFERENCES

1. Montgomery D. *Design and Analysis of Experiments, 10e Enhanced eText with Abridged Print Companion*. Wiley; 2019. https://books.google.com/books?id=TpK9wAEACAAJ
2. Ryan T. *Sample Size Determination and Power. Wiley Series in Probability and Statistics*. Wiley; 2013. https://books.google.com/books?id=WBSwhw-j1ZoC
3. Cobb GW. *Introduction to Design and Analysis of Experiments*. Springer; 1998.
4. Rasch D, Pilz J, Verdooren LR, Gebhardt A. *Optimal Experimental Design with R*. CRC Press; 2011.
5. Shah KR, Sinha B. *Theory of Optimal Designs*. Vol 54. Springer Science & Business Media; 2012.
6. Roy S, Pradhan B. Bayesian optimum life testing plans under progressive Type-I interval censoring scheme. *Qual Reliab Eng Int*. 2017;33(8):2727-2737.
7. Zohbi I, Wainakh M, Arafeh H. Bayesian D-optimal designs for Weibull distribution with censoring. *Qual Reliab Eng Int*. 2016;32(8):2789-2802.
8. Burghaus I, Dette H. Optimal designs for nonlinear regression models with respect to non-informative priors. *J Stat Plann Inference*. 2014;154(1):12-25.
9. Dette Hea. Maximin and Bayesian Optimal Designs for Regression Models. *Statistica Sinica*. 2007;17(2):463-480.
10. Dette H, Neugebauer HM. Bayesian D-optimal designs for exponential regression models. *J Stat Plann Inference*. 1997;60(2):331-349.
11. Chaloner K, Larntz K. Design of accelerated life testing plans under multiple stresses. *J Stat Plann Inference*. 1992;33(2):245-259.
12. Park JW, Yum BJ. Optimal design of accelerated life tests with two stresses. *Nav Res Logist*. 1996;43(6):863-884.
13. Zhang X, Yang J, Kong X. Optimal design of accelerated life tests with two stresses. *Qual Reliab Eng Int*. 2021;37(1):60-77.
14. Nasir E, Pan R. Simulation-based Bayesian optimal design for multi-factor accelerated life tests. *Commun Stat-Simul Comput*. 2017;46(2):980-993.
15. Pan R, Yang T, Seo K. Planning constant-stress accelerated life tests for acceleration model selection. *IEEE Trans Reliab*. 2015;64(4):1356-1366.
16. Zhu Y, Elsayed EA. Design of accelerated life testing plans under multiple stresses. *Nav Res Logist*. 2013;60(6):468-478.
17. Monroe EM, Pan R, Anderson-Cook CM, Montgomery DC, Borror CM. Sensitivity analysis of optimal designs for accelerated life testing. *J Qual Technol*. 2010;42(2):121-135.
18. Monroe EM, Pan R. Experimental design considerations for accelerated life tests with nonlinear constraints and censoring. *J Qual Technol*. 2008;40(4):355-367.
19. Hong Y, King C, Zhang Y, Meeker WQ. Bayesian life test planning for log-location-scale family of distributions. *J Qual Technol*. 2015;47(4):336-350.
20. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat Sci*. 1995;10(3):273-304.
21. DasGupta A. Review of optimal Bayes designs. *Handb Stat*. 1996;13:1099-1147.
22. Escobar LA, Meeker WQ. A review of accelerated test models. *Stat Sci*. 2006;21(4):552-577.
23. Lindley DV. On a measure of the information provided by an experiment. *Ann Math Stat*. 1956;27(4):986-1005.
24. Xu A, Tang Y. Reference optimality criterion for planning accelerated life testing. *J Stat Plann Inference*. 2015;167:14-26.
25. DuMouchel W, Jones B. A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics*. 1994;36(1):37-47.
26. Ryan EG, Drovandi CC, McGree JM, Pettitt AN. A review of modern computational algorithms for Bayesian optimal design. *Int Stat Rev*. 2016;84(1):128-154.
27. Rigdon SE, Pan R, Montgomery DC, Freeman LJ. *Design of Experiments for Reliability Achievement*. Wiley; 2022.
28. Meyer RK, Nachtsheim CJ. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Am Stat Assoc*. 1995;37(1):60-69.
29. Bazaraa MS, Sherali HD, Shetty CM. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons; 2013.
30. Yang Y, Guang J, Zhengqiang P, Rui G. An iterated local coordinate-exchange algorithm for constructing experimental designs for multi-dimensional constrained spaces. *J Syst Eng Electron*. 2021;32(5):1212-1220.
31. Overstall A, Woods D. The approximate coordinate exchange algorithm for Bayesian optimal design of experiments. Unpublished, 2015.
32. Overstall AM, Woods DC. Bayesian design of experiments using approximate coordinate exchange. *Technometrics*. 2017;59(4):458-470.
33. StatEase. Design-Expert > Tutorials > Optimal Design. 2020. Accessed April 30, 2022. https://www.statease.com/docs/v11/tutorials/optimal-numeric/
34. Robert E. Accessed April 30, 2022. https://www.xarg.org/2017/07/how-to-map-a-square-to-a-circle/
35. Escobar LA, Meeker WQ. Algorithm AS 292: Fisher information matrix for the extreme value, normal and logistic distributions and censored data. *J R Stat Soc Ser C (Appl Stat)*. 1994;43(3):533-540.
36. Inc. WR. Mathematica, Version 13.0.0. 2021. https://www.wolfram.com/mathematica

AUTHOR BIOGRAPHIES

Dustin Taylor is a Major in the US Air Force and an Industrial Engineering doctoral student at Arizona State University. He received his undergraduate degree in 2009 from Valdosta State University in Applied Mathematics where he coauthored "Investigation of a pest control method involving chemical and biological methods using the Lotka–Volterra model with fishing." He received his commission in October 2009 through Officer Candidate School (OCS) in Newport, RI and his Master's degree in Mathematical Sciences from the University of West Florida.

Steven E. Rigdon is a Professor of biostatistics in the Department of Epidemiology and Biostatistics, College for Public Health and Social Justice at Saint Louis University. He is also Distinguished Research Professor Emeritus at Southern Illinois University Edwardsville. He has been on the faculty at Saint Louis University since 2012, and from 1986 to 2012, he was at Southern Illinois University Edwardsville. From 1982 to 1986, he was a Senior Mathematical Statistician for Monsanto. He is the Coauthor of Calculus, 8th Edition and 9th Edition, published by Pearson, Statistical Methods for the Reliability of Repairable Systems, published by Wiley and Design of Experiments for Reliability Achievement, also published by Wiley. He has served on the editorial review board of Journal of Quality Technology and was section editor for Encyclopedia for Statistics in Quality and Reliability, published by Wiley. He is currently on the editorial board of the Journal of Quantitative Analysis in Sports and the Journal of Quality Technology. He is a Fellow of the American Statistical Association.



Rong Pan is an Associate Professor of Industrial Engineering in the School of Computing and Augmented Intelligence at Arizona State University. He received his doctoral degree in Industrial Engineering from Penn State University in 2002. His research interests include failure time data analysis, system reliability, design of experiments, multivariate statistical process control, time series analysis, and computational Bayesian methods.

Douglas C. Montgomery, Ph.D., is Regent's Professor of Industrial Engineering and ASU Foundation Professor of Engineering at Arizona State University. He was the John M. Fluke Distinguished Professor of Engineering, Director of Industrial Engineering, and Professor of Mechanical Engineering at the University of Washington in Seattle. He was a Professor of Industrial and Systems Engineering at the Georgia Institute of Technology. He holds BSIE, MS, and Ph.D. degrees from Virginia Tech. Dr Montgomery's professional interests focus on industrial statistics, including design of experiments, quality and reliability engineering, applications of linear models, and time series analysis and forecasting. Dr Montgomery is an author of 16 books, and he has authored over 280 papers that have appeared in many archival journals. He is currently one of the Chief Editors of Quality and Reliability Engineering International and is a former Editor of the Journal of Quality Technology. He is a Fellow of the American Statistical Association, an Honorary Member of the American Society for Quality, a Fellow of the Royal Statistical Society, a Fellow of the Institute of Industrial and Systems Engineers, an Elected Member of the International Statistical Institute, and an Academician of the International Association for Quality.

How to cite this article: Taylor D, Rigdon SE, Pan R, Montgomery DC. Bayesian D-optimal design for life testing with censoring. *Qual Reliab Engng Int*. 2022;1-20. https://doi.org/10.1002/qre.3228

APPENDIX A: ALGORITHM FOR FISHER INFORMATION MATRIX

Assume that the log of the lifetimes have an SEV(μ, σ) distribution. Define

$$\begin{aligned} G(x) &= 1 - \exp(-\exp(x)), \quad -\infty < x < \infty \\ g(x) &= G'(x) = \exp(-\exp(x) + x). \end{aligned} \tag{A1}$$

Assuming that an observation is right censored at z, we can use Escobar et al.[35] to compute the entries in the Fisher information matrix as

$$\begin{aligned} f_{11}(z) &= \frac{\sigma^2}{n} E\left[-\frac{\partial^2 \ell}{\partial \mu^2}\right] = \Psi_0(z) = \int_{-\infty}^{z} g(x)\, dx \\ f_{12}(z) &= \frac{\sigma^2}{n} E\left[-\frac{\partial^2 \ell}{\partial \mu \partial \sigma}\right] = \Psi_1(z) = \int_{-\infty}^{z} (1+x) g(x)\, dx \\ f_{22}(z) &= \frac{\sigma^2}{n} E\left[-\frac{\partial^2 \ell}{\partial \sigma^2}\right] = \Psi_2(z) = \int_{-\infty}^{z} (1+x)^2 g(x)\, dx. \end{aligned} \tag{A2}$$

Using Mathematica,[36] we find that

$$\Psi_0(z) = \int_{-\infty}^{z} g(x)\, dx = 1 - e^{-e^z} \tag{A3}$$

and

$$\Psi_1(z) = \int_{-\infty}^{z} (1+x) g(x)\, dx = 1 - e^{-e^z} - \gamma - z e^{-e^z} + \mathrm{Ei}\,(-e^z) \tag{A4}$$

where γ is Euler's constant ($\gamma \approx 0.577$) and $\mathrm{Ei}(x)$ is the exponential integral function defined by

$$\mathrm{Ei}\,(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\, dt. \tag{A5}$$

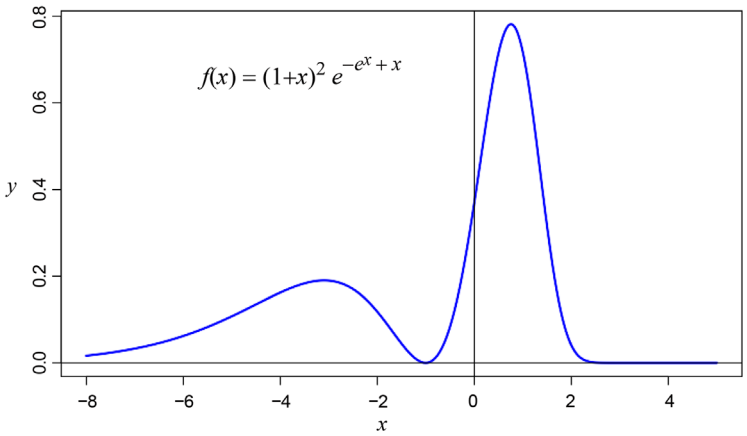
Note that if $x > 0$, the interval of integration covers the singular value $t = 0$, in which case, the integral is interpreted as the Cauchy principal value. In our case, the argument, namely $-e^z$, is always negative so there is no need to address this issue. We have

$$\mathrm{Ei}\,(-e^z) = -\int_{-(-e^z)}^{\infty} \frac{e^{-t}}{t}\, dt = -\int_{e^z}^{\infty} \frac{e^{-t}}{t}\, dt. \tag{A6}$$

To compute Ψ_2, we can write

$$\begin{aligned} \Psi_2(z) &= \int_{-\infty}^{z} (1+x)^2 g(x)\, dx \\ &= \int_{-\infty}^{0} (1+x)^2 g(x)\, dx + \int_{0}^{z} (1+x)^2 g(x)\, dx \\ &= \left[1 - e^{-1} - 2\gamma + 2\,\mathrm{Ei}(-1) + 2{}_3F_3(\{1,1,1\},\{2,2,2\},-1)\right] + \int_{0}^{z} (1+x)^2 g(x)\, dx \end{aligned} \tag{A7}$$

FIGURE A1 Integrand in the expression for $\Psi_2(z)$



The first integral on the second line can be evaluated analytically using Mathematica[36] and it is the value shown in the last line. The function ${}_3F_3$ in Equation (A7) is the generalized hypergeometric function, defined by

$$ {}_3F_3(\{a_1, a_2, z_3\}, \{b_1, b_2, b_3\}, z) = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n (a_3)_n}{(b_1)_n (b_2)_n (b_3)_n} \frac{z^n}{n!} \tag{A8} $$

where the Plochhammer function $(a)_n$ is defined by

$$ \begin{aligned} (a)_0 &= 1 \\ (a)_n &= a(a+1)(a+2)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)}. \end{aligned} \tag{A9} $$

The expression in brackets in the last line of Equation (A7) is simply a number, which can approximated numerically as 0.821346956. Thus,

$$ \Psi_2(z) \approx 0.821346956 + \int_0^z (1+x)^2 g(x)\, dx \tag{A10} $$

$$ = 0.821346956 + \int_0^z (1+x)^2 e^{-e^x + x}\, dx. \tag{A10} $$

Note that this is true regardless of whether z is greater or less than 0. Ordinarily, we would expect the censoring time to be larger than the location parameter μ, which would lead to a z that is positive, but the relationship holds nonetheless. The graph of the integrand from -5 to 5 is shown in Figure A1. Since the integral from $-\infty$ to 0 can be evaluated with virtual exactness, we need only numerically integrate from 0 to z. The integrand $(1+x)^2 e^{-e^x+x}$ is well-behaved for positive values of x, so the integral can be approximated by Gaussian quadrature. The integrand approaches zero rapidly as $x \to \infty$; The limiting value for the integral is

$$ \Psi_2(\infty) = 1 - 2\gamma + \gamma^2 + \frac{\pi^2}{6} \approx 1.82368066. \tag{A11} $$

For $z > 3.2$, $\Psi(z)$ is within 4×10^{-10} of this limiting value.

The integrand reaches a maximum near $x = 0.759313$; it may be advantageous to use Gaussian quadrature separately on the intervals $[0, 0.759313]$ and $[0.759313, z]$ for the case when $z > 0.759313$.

Consider now the regression model

$$ Y_i | x_i \sim SEV(\mu = \beta_0 + \beta_1 x_i, \sigma) \tag{A12} $$

where observation i will be censored at time y_i^c. The lifetime T_i given x_i, therefore, has a WEI($\theta = \exp(\beta_0 + \beta_1 x_i), \kappa = 1/\sigma$) distribution. The Fisher information matrix is then

$$I(\beta_0, \beta_1, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_0\left(\frac{y_i^c-(\beta_0+\beta_1 x_i)}{\sigma}\right)\begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-(\beta_0+\beta_1 x_i)}{\sigma}\right)\begin{bmatrix} 1 \\ x_i \end{bmatrix} \\ \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-(\beta_0+\beta_1 x_i)}{\sigma}\right)\begin{bmatrix} 1 & x_i \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_2\left(\frac{y_i^c-(\beta_0+\beta_1 x_i)}{\sigma}\right) \end{bmatrix} \tag{A13}$$

The first-order regression model in more than one variable is handled in a similar manner. For example, if there are two predictor variables, x_{i1} and x_{i2}, the Fisher information matrix is

$$I(\beta_0, \beta_1, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_0\left(\frac{y_i^c-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma}\right)\begin{bmatrix} 1 & x_{i1} & x_{i2} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} \\ x_{i2} & x_{i1}x_{i2} & x_{i2}^2 \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma}\right)\begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \end{bmatrix} \\ \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma}\right)\begin{bmatrix} 1 & x_{i1} & x_{i2} \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_2\left(\frac{y_i^c-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma}\right) \end{bmatrix} \tag{A14}$$

For the case where

$$\mu_i = \log\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} \tag{A15}$$

the Fisher information matrix is

$$I(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_0\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i1}^2 & x_{i2}^2 & x_{i1}x_{i2} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} & x_{i1}^3 & x_{i1}x_{i2}^2 & x_{i1}^2 x_{i2} \\ x_{i2} & x_{i1}x_{i2} & x_{i2}^2 & x_{i1}^2 x_{i2} & x_{i2}^3 & x_{i1}x_{i2}^2 \\ x_{i1}^2 & x_{i1}^3 & x_{i1}^2 x_{i2} & x_{i1}^4 & x_{i1}^2 x_{i2}^2 & x_{i1}^3 x_{i2} \\ x_{i2}^2 & x_{i1}x_{i2}^2 & x_{i2}^3 & x_{i1}^2 x_{i2}^2 & x_{i2}^4 & x_{i1}x_{i2}^3 \\ x_{i1}x_{i2} & x_{i1}^2 x_{i2} & x_{i1}x_{i2}^2 & x_{i1}^3 x_{i2} & x_{i1}x_{i2}^3 & x_{i1}^2 x_{i2}^2 \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ x_{i1}^2 \\ x_{i2}^2 \\ x_{i1}x_{i2} \end{bmatrix} \\ \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i^c-\boldsymbol{x}_i^t}{\boldsymbol{\beta}}\sigma\right)\begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i1}^2 & x_{i2}^2 & x_{i1}x_{i2} \end{bmatrix} & \frac{1}{\sigma^2}\sum_{i=1}^{n} \Psi_2\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right) \end{bmatrix} \tag{A16}$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{bmatrix} \Psi_0\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{X}^T\boldsymbol{X} & \Psi_1\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{x}_i \\ \Psi_1\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{x}_i^T & \Psi_2\left(\frac{y_i^c-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right) \end{bmatrix} \tag{A17}$$

where $\boldsymbol{X}$ is the model matrix and $\boldsymbol{x}_i$ is the i th row of the model matrix. Here Ψ_0, Ψ_1, and Ψ_2 are the functions calculated in Equations (A3), (A4), and (A10).