Statistica Sinica Preprint No: SS-2020-0517	
Title	Locally D-Optimal Designs for Hierarchical Response Experiments
Manuscript ID	SS-2020-0517
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0517
Complete List of Authors	Mingyao Ai, Zhiqiang Ye and Jun Yu
Corresponding Author	Jun Yu
E-mail	yujunbeta@bit.edu.cn
Notice: Accepted version subject to English editing.	

LOCALLY D -OPTIMAL DESIGNS FOR HIERARCHICAL RESPONSE EXPERIMENTS

Mingyao Ai¹, Zhiqiang Ye¹, Jun Yu²

*LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University*¹

*School of Mathematics and Statistics, Beijing Institute of Technology*²

Abstract: Categorical responses with hierarchical structure are common in social sciences, public health, and marketing. The continuation ratio model is one of the most commonly used models in characterizing such hierarchical data. Despite the research on this model is rich, few refers to its design problem. Here we study locally D -optimal designs for the model with general link function under the partial proportional odds assumption. The necessary and sufficient conditions for the positive definiteness of the Fisher information matrix are derived, which show that a feasible design may contain less supports than the number of parameters in the model. Based on some deduced characteristics of the D -optimal criterion, an efficient algorithm is proposed to search for optimal designs. It can deal with both discrete and continuous design fields. Some numerical examples illustrate the advantages of these new designs compared with the existing ones.

Key words and phrases: Approximate design; continuation ratio model; general link functions; multinomial response.

1. Introduction

Categorical responses are common in a multitude of scientific experiments. Design of experiments with categorical responses is becoming increasingly popular in a wide variety of scientific disciplines. Some progress has been achieved in both theoretical and algorithm aspects. See Sitter and Wu (1993); Atkinson, Donev, and Tobias (2007); Yang, Zhang, and Huang (2011); Yang, Biedermann, and Tang (2013); Yang, Tong, and Mandal (2017); Lukemire, Mandal, and Wong (2019) as examples. Among the various types of categorical responses, ordinal responses with hierarchical structure are often used. Examples include ratings of preference in consumer choice experiments, tumor grade in drug testing, and animal fitness in ecology. Refer to Agresti (2006) for details.

The continuation ratio (CR) model (O’Connell, 2006) is one of the most commonly applied models, which focus on estimating probabilities of successive stages when the lower stages are reached first (Fullerton, 2009). Mathematically, the CR model estimates the conditional probabilities. For example, in an experiment on the emergence of house flies (Zocchi and Atkinson, 1999), seven sets of 500 pupae were exposed to one of several doses of radiation. Observations included the number of flies that died before the opening of the pupae (y_1), the number of flies out of the opened pupae

but died before completing emergence (y_2), and the number of flies that completely emerged (y_3). Such a response is a typically ordinal variable. The two essential ratios, i.e., the mortality of flies inside the pupae (y_1/m) and the mortality of flies before complete emergence to the total number of flies that started to emerge ($y_2/(m - y_1)$), can be described by the CR model.

In a CR model, the responses are assumed to be generated from a multinomial distribution. It can be regarded as a special case of a multivariate generalized linear model (McCullagh and Nelder, 1989). Finding optimal designs on a CR model is challenging since the model involves a wide choice of link functions, among which logit link, probit link, log-log link, Cauchit link, and complementary log-log link are five popular ones. Moreover, different parameter restriction assumptions across ordinal responses are needed to be considered. There are three kinds of assumptions routinely used in the CR model, namely, proportional odds assumption, non-proportional odds assumption, and partial proportional odds assumption. The proportional odds assumption (McCullagh, 1980) assumes that the effect of each predictor is invariant across ordinal responses. The non-proportional odds assumption allows different parameters for different categories (Agresti, 2006). Since a different explanatory variable structure

may be required to model the different stages, the non-proportional odds assumption is more commonly used in practice. The partial proportional odds assumption is proposed in Peterson and Harrell (1990), in which a subset of these explanatory variables are assumed to have proportional odds while the rest are not. Clearly, partial proportional odds assumption includes both proportional odds and non-proportional odds assumptions as special cases.

Due to the aforementioned reasons, the relevant results in the design literature for CR models are very limited. Most of them are handled on a case-by-case basis. To the best of our knowledge, the most relevant works are given in Zocchi and Atkinson (1999) and Bu et al. (2020). In Zocchi and Atkinson (1999), the optimal designs are searched out for CR models with logit link function under non-proportional odds assumption. Bu et al. (2020) generalized optimal designs for the partial proportional odds case. However, only the discrete design field is considered in this work, which cannot fully meet the real needs, especially for the dose-finding studies. It is worth mentioning that there are infinite candidate design points in the continuous design field. Therefore, the optimal designs given in Bu et al. (2020) cannot be applied directly here. When the number of grid points is not small, one will meet the computation bottleneck to search

the optimal weights on the grid points over the design field. Furthermore, finding optimal designs for CR models with general link functions has not been considered in the existing works.

To address these issues, in this work, we study D -optimal designs for CR models with general link functions under the three kinds of parameters assumption. Our main contributions to the development of the optimal design problems on CR models are three-folds. Firstly, to the limit of our knowledge, our method is the first to give optimal designs for CR models with general link functions on both discrete and continuous design fields. Secondly, analogous to Bu et al. (2020), we obtain the explicit representation of the Fisher information matrix and derive the sufficient and necessary conditions for the positive definiteness of the Fisher information matrix, which ensures the non-degeneracy of the corresponding designs. Thirdly, the proposed algorithm can deal with multivariate responses cases. This computational aspect has rarely been considered in the literature. Comprehensive simulation results show the advantages of our methods comparing with some cutting edge designs.

The rest of this paper is organized as follows. In Section 2, we review the CR models with general link functions. The Fisher information matrix is also derived. Section 3 characterizes the locally D -optimality criterion

for approximate designs and shows the sufficient and necessary conditions for positive definiteness of the Fisher information matrix. In Section 4, an algorithm is provided to search the locally D -optimal designs under both discrete and continuous design fields. Moreover, the explicit form of optimal weights is derived for some special models. Section 5 illustrates our methods through several numerical examples. Section 6 concludes this paper with discussion. All the proofs are relegated to the supplementary material.

2. Continuation ratio model and its Fisher information matrix

2.1 The CR model

Suppose we conduct $n_1, \dots, n_m$ experiments under $m(m \geq 2)$ different experimental settings $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, respectively, where $\mathbf{x}_i \in \mathcal{X}$ for $i = 1, \dots, m$, and $\mathcal{X}$ is the design field that we are interested in. Let the possible responses be denoted by $\{1, \dots, J\}(J \geq 2)$ and y_{ij} be the number of responses equal to $j(j \in \{1, \dots, J\})$ under the experimental setting $\mathbf{x}_i$, for $i = 1, \dots, m$. Clearly, $y_{i1} + \dots + y_{iJ} = n_i$. Then $\mathbf{Y}_i = (y_{i1}, \dots, y_{iJ})^T$ can be described by the following multinomial distribution

$$p(y_{i1}, \dots, y_{iJ}) = \frac{n_i!}{y_{i1}! \dots y_{iJ}!} \pi_1(\mathbf{x}_i)^{y_{i1}} \dots \pi_J(\mathbf{x}_i)^{y_{iJ}}, \quad (2.1)$$

where $\pi_j(\mathbf{x}_i)$ denotes the probability that the response is j under experimental setting $\mathbf{x}_i$. For notational simplicity, we use π_{ij} to denote $\pi_j(\mathbf{x}_i)$ in the rest of the paper. The relationship between the explanatory variables $\mathbf{x}_i$ and the probabilities $\pi_{i1}, \dots, \pi_{iJ}$ are described through the CR model, in terms of partial proportional odds structure. To be precise, for some given link function $g(\cdot)$, the conditional distribution $\mathbf{Y}|\mathbf{x}_i$ is linked to the working model through the probabilities $\pi_{i1}, \dots, \pi_{iJ}$ by the following equation

$$g\left(\frac{\pi_{ij}}{\pi_{ij} + \dots + \pi_{iJ}}\right) = \mathbf{h}_0^T(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\theta}_j, \quad (2.2)$$

for $i = 1, \dots, m, j = 1, \dots, J-1$. The transformations $\mathbf{h}_j(\cdot), j = 0, \dots, J-1$ are known in advance, where $\mathbf{h}_0(\mathbf{x}_i) \in \mathbb{R}^{p_0}$ stands for the predictors that are common in all categories and $\mathbf{h}_j(\mathbf{x}_i) \in \mathbb{R}^{p_j}$, for $j = 1, \dots, J-1$, stands for the individual predictors belong to the j th category only. $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_j(j = 1, \dots, J-1)$ are unknown parameters. We further assume all the $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_j(j = 1, \dots, J-1)$ lie in a compact parameter space. There are $p = p_0 + \dots + p_{J-1}$ unknown parameters in total. Clearly, CR models under proportional odds and non-proportional odds assumptions are two special cases with $h_j(\mathbf{x}_i) = 1, j = 1, \dots, J-1$, and $h_0(\mathbf{x}_i) = 0$, respectively. It is worth mentioning that Model (2.2) turns into multinomial logit model (Agresti, 2006) when $g(\cdot)$ is the logit link function.

To ensure that the model is well defined, we need the following two reg-

ularity assumptions mentioned in McCullagh and Nelder (1989) throughout this paper.

Assumption 1. For any $i = 1, \dots, m; j = 1, \dots, J$, $0 < \pi_{ij} < 1$.

Assumption 2. The link function $g(\cdot)$ is differentiable and its derivative $g'(\cdot) > 0$.

Under Assumption 2, $g(\cdot)$ is an injective function, as well as $g^{-1}(\cdot)$. Thus, in Model (2.2), $\pi_{ij}/(\pi_{ij} + \dots + \pi_{iJ})$ can be represented through $g^{-1}(\mathbf{h}_0^T(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\theta}_j)$ for $i = 1, \dots, m; j = 1, \dots, J - 1$. With the fact that $\pi_{i1} + \dots + \pi_{iJ} = 1$, it is clear to see that $\pi_{i1}, \dots, \pi_{iJ}$ are completely determined by the link function $g(\cdot)$, the predictors $\mathbf{h}_0(\mathbf{x}_i), \mathbf{h}_1(\mathbf{x}_i), \dots, \mathbf{h}_{J-1}(\mathbf{x}_i)$, and the parameters $\boldsymbol{\beta}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{J-1}$.

2.2 Fisher information matrix

Following Kiefer (1974), we call a probability measure

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ \omega_1 & \cdots & \omega_m \end{pmatrix}, \quad (2.3)$$

an approximate design, where $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathcal{X}$, and $\omega_i > 0$ for $i = 1, \dots, m$, with $\omega_1 + \dots + \omega_m = 1$. Here we further assume the experimental region $\mathcal{X}$ is compact throughout this paper. To get the Fisher information matrix, we start with introducing the following notations.

For $\mathbf{x} \in \mathcal{X}$, $H(\mathbf{x})$ is a $(J-1) \times p$ matrix with

$$H(\mathbf{x}) = \begin{pmatrix} \mathbf{h}_0^T(\mathbf{x}) & \mathbf{h}_1^T(\mathbf{x}) & 0 & \cdots & 0 \\ \mathbf{h}_0^T(\mathbf{x}) & 0 & \mathbf{h}_2^T(\mathbf{x}) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{h}_0^T(\mathbf{x}) & 0 & 0 & \cdots & \mathbf{h}_{J-1}^T(\mathbf{x}) \end{pmatrix}.$$

Let $D(\mathbf{x}) = \text{diag}(\pi_1(\mathbf{x}), \dots, \pi_J(\mathbf{x}))$ with $\pi_1(\mathbf{x}), \dots, \pi_J(\mathbf{x})$ defined in Model (2.1), where $\text{diag}(\cdot)$ denotes a diagonal matrix with corresponding diagonal elements.

After defining $H(\mathbf{x})$ and $D(\mathbf{x})$, the Fisher information matrix for Model (2.2) under a design ξ is derived in the following theorem.

Theorem 1. *Suppose Assumptions 1 and 2 hold. The Fisher information matrix for Model (2.2) under the design ξ defined in (2.3) can be written as*

$$M(\xi) = \sum_{i=1}^m \omega_i H^T(\mathbf{x}_i) G^T(\mathbf{x}_i) D^{-1}(\mathbf{x}_i) G(\mathbf{x}_i) H(\mathbf{x}_i), \quad (2.4)$$

where $G(\mathbf{x}_i)$ is a $J \times (J-1)$ matrix decided by the model and experimental setting $\mathbf{x}_i$, which will be given in Appendix A.1.

Example 1. Consider the following model that was used in Zocchi and Atkinson (1999),

$$\begin{aligned} \log \left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}} \right) &= \theta_{11} + \theta_{12}x_i + \theta_{13}x_i^2, \\ \log \left(\frac{\pi_{i2}}{\pi_{i3}} \right) &= \theta_{21} + \theta_{22}x_i, \end{aligned} \quad (2.5)$$

for $i = 1, \dots, m$.

Note that the above model is a special case of Model (2.2) with three ($J = 3$) categories and logit link, under non-proportional odds assumption.

In this model, at experimental setting x_i , the matrices $H(x_i)$, $D(x_i)$ and $G(x_i)$ in the Fisher information matrix (2.4) can be expressed as

$$H(x_i) = \begin{pmatrix} 1 & x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_i \end{pmatrix},$$

$$D(x_i) = \text{diag} \left(\frac{b_{i1}}{1+b_{i1}}, \frac{b_{i2}}{(1+b_{i1})(1+b_{i2})}, \frac{1}{(1+b_{i1})(1+b_{i2})} \right),$$

$$G(x_i) = \frac{1}{(1+b_{i1})^2(1+b_{i2})^2} \begin{pmatrix} b_{i1}(1+b_{i2})^2 & 0 \\ -b_{i1}b_{i2}(1+b_{i2}) & (1+b_{i1})b_{i2} \\ -b_{i1}(1+b_{i2}) & -(1+b_{i1})b_{i2} \end{pmatrix},$$

where $b_{i1} = e^{\theta_{11} + \theta_{12}x_i + \theta_{13}x_i^2}$, $b_{i2} = e^{\theta_{21} + \theta_{22}x_i}$, for $i = 1, \dots, m$.

3. Determinant of the Fisher information matrix

The D -optimality criterion is one of the most popular design criteria in the optimal design theory, which seeks to maximize the determinant of the Fisher information matrix. This criterion results in minimizing the volume of the estimator's confidence ellipsoid. To avoid trivial solutions, the Fisher information matrix is required to be positive definite. The sufficient and

necessary condition for the positive definiteness of the Fisher information matrix is stated in the following theorem.

Theorem 2. *The Fisher information matrix $M(\xi)$ calculated in Equation (2.4) is positive definite if and only if the matrix $(H^T(\mathbf{x}_1), \dots, H^T(\mathbf{x}_m))$ has full row rank.*

As a direct conclusion, we obtain some equivalent conditions which are easy to verify.

Corollary 1. *Let $H_j = (\mathbf{h}_j(\mathbf{x}_1), \dots, \mathbf{h}_j(\mathbf{x}_m))$, for $j = 0, \dots, J-1$. The Fisher information matrix $M(\xi)$ is positive definite if and only if the following statements hold simultaneously.*

- (1) *The number of support points m satisfies $m \geq \max\{p_0 + r, p_1, \dots, p_{J-1}\}$, where $r = \dim(\cap_{j=1}^{J-1} C(H_j^T))$, $C(H_j^T)$ denotes the column space of H_j^T , and $\dim(\cdot)$ denotes the dimension of the corresponding space.*
- (2) *The m support points $\mathbf{x}_1, \dots, \mathbf{x}_m$ satisfy the following two conditions:*
 - (a) $H_0, \dots, H_{J-1}$ are full row rank;
 - (b) $\cap_{j=0}^{J-1} C(H_j^T) = \{\mathbf{0}\}$.

The same conditions are also required in Theorem 3.3 for logit link in Bu et al. (2020). From these conditions, one can see that the minimally supported designs have $\max\{p_0 + r, p_1, \dots, p_{J-1}\}$ different design points,

which is different from the designs for binary responses. It is worth mentioning that a feasible design for a CR model with different types of links may contain less experimental settings than parameters. For a special case, $H_1 = \cdots = H_{J-1}$, the minimum number of the support points is $p_0 + p_1$, which is strictly less than the number of parameters $p_0 + (J - 1)p_1$ when $J \geq 3$. Furthermore, for Model (2.2) under proportional odds assumption, i.e., $h_j(\mathbf{x}_i) = 1$, for $j = 1, \dots, J - 1$, the minimum number of the support points is $p_0 + 1$.

Example 2. Recall Model (2.5) mentioned in Example 1. It is clear to see that $p_0 = 0, p_1 = 3, p_2 = 2$. The dimension of $C(H_1^T) \cap C(H_2^T)$ is $r \leq 2$. Thus the number of support points m satisfies $m \geq \max\{p_0 + r, p_1, p_2\} = 3$. Note that when there are only three distinct design points x_1, x_2, x_3 , simple algebra shows that H_0, H_1 and H_2 are full row rank, and $C(H_0^T) \cap C(H_1^T) \cap C(H_2^T) = \{\mathbf{0}\}$. Therefore, the minimum number of support points is three according to Corollary 1.

To study the D -optimal designs and their properties, we start with a characterization of $|M(\xi)|$, where $|\cdot|$ denotes the determinant of the corresponding matrix. Theorem 3 will show that $|M(\xi)|$ is a homogeneous polynomial of $\omega_1, \dots, \omega_m$, when the support points $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ are predefined. Before formally presenting Theorem 3, we introduce some necessary

notations.

Let $\delta_{\mathbf{x}}$ denote the single point design and $M(\delta_{\mathbf{x}})$ be the corresponding Fisher information matrix. For simplicity, let τ be a map from $\{1, \dots, p\}$ to $\{1, \dots, m\}$ and M_{τ} be a $p \times p$ matrix whose k th row is same as the k th row of matrix $M(\delta_{\mathbf{x}_{\tau(k)}})$, $k = 1, \dots, p$. Define

$$\Delta(\alpha_1, \dots, \alpha_m) = \left\{ \tau \left| \sum_{k=1}^p \mathbb{I}_{\{i\}}(\tau(k)) = \alpha_i, i = 1, \dots, m \right. \right\},$$

where $\mathbb{I}_A(\cdot)$ is an indicator function defined on the set A , and $\alpha_1, \dots, \alpha_m \in \mathbb{N}$ with $\alpha_1 + \dots + \alpha_m = p$. Then the following theorem can be used to simplify the calculation of $|M(\xi)|$.

Theorem 3. *The determinant of the Fisher information matrix $M(\xi)$ is*

$$|M(\xi)| = \sum_{\substack{\alpha_1 \geq 0, \dots, \alpha_m \geq 0, \\ \alpha_1 + \dots + \alpha_m = p}} c_{\alpha_1, \dots, \alpha_m} \omega_1^{\alpha_1} \dots \omega_m^{\alpha_m},$$

where

$$c_{\alpha_1, \dots, \alpha_m} = \sum_{\tau \in \Delta(\alpha_1, \dots, \alpha_m)} |M_{\tau}|. \quad (3.6)$$

Furthermore, let $\{j_1^*, \dots, j_k^*\} = \{j | \alpha_j > 0\}$. Then the coefficients (3.6) are zero if the $(\alpha_1, \dots, \alpha_m)$ satisfies one of the following conditions.

- (1) The maximum of $\alpha_1, \dots, \alpha_m$ is equal or greater than J .
- (2) The number of positive α_j 's satisfy $k + 1 \leq \max\{p_0 + r, p_1, \dots, p_{J-1}\}$,
where r is defined in Corollary 1.

Theorem 3 reflects the fact that the determinant of the Fisher information matrices are homogeneous polynomials, which enable us to use the lift-one algorithm to search out the optimal weights when the support points are given in advance. This characteristic also holds for the cumulative link models. See Yang et al. (2017) for details.

Example 3. Consider Model (2.5) mentioned in Examples 1 and 2. In this case, the number of support points $m \geq 3$. According to Theorem 3, $|M(\xi)|$ is an order-5 homogeneous polynomial and the coefficients of $\omega_i^5, \omega_i^4\omega_j, \omega_i^3\omega_j^2, \omega_i^3\omega_j\omega_k$ are zero. Therefore,

$$\begin{aligned} |M(\xi)| = & \sum_{1 \leq i < j \leq m} \sum_{\substack{1 \leq k \leq m, \\ k \neq i, k \neq j}} c_{(ijk)} \cdot \omega_i^2 \omega_j^2 \omega_k + \sum_{i=1}^m \sum_{\substack{1 \leq j < k < l \leq m, \\ j \neq i, k \neq i, l \neq i}} c_{(ijkl)} \cdot \omega_i^2 \omega_j \omega_k \omega_l \\ & + \sum_{1 \leq h < i < j < k < l \leq m} c_{(hijkl)} \cdot \omega_h \omega_i \omega_j \omega_k \omega_l, \end{aligned}$$

with some coefficients $c_{(ijk)}, c_{(ijkl)}, c_{(hijkl)}$.

4. Locally D -optimal designs

As with other generalized linear models, the Fisher information matrix (2.4) depends not only on the design ξ but also on the unknown parameters. To solve the dependence on unknown parameters, many approaches have been proposed, such as local optimality (Chernoff, 1953), sequential procedure (Ford et al., 1989), Bayesian optimality (Chaloner and Verdinelli, 1995),

and maximin approach (Imhof, 2001). In this work, we focus on the locally optimal approach based on a “best guess” of the unknown parameters. Although the locally optimal designs may be inefficient if the initial guesses of parameters are far away from the true parameters, it can still be a benchmark for designs chosen to satisfy experimental constraints (Ford et al., 1992; Stufken and Yang, 2012).

To search the locally D -optimal designs, we begin with the following general equivalence theorem (Atkinson, Donev, and Tobias, 2007; Atkinson, Fedorov, Herzberg, and Zhang, 2014).

Theorem 4. *For a design ξ and a point $\mathbf{x} \in \mathcal{X}$, define*

$$\phi(\mathbf{x}, \xi) = \text{tr}(M^{-1}(\xi)H^T(\mathbf{x})G^T(\mathbf{x})D^{-1}(\mathbf{x})G(\mathbf{x})H(\mathbf{x})) - p,$$

where $\text{tr}(\cdot)$ denotes the trace of the corresponding matrix. A design ξ^ is D -optimal if and only if the inequality $\phi(\mathbf{x}, \xi^*) \leq 0$ holds for all $\mathbf{x} \in \mathcal{X}$.*

The equality holds if and only if $\mathbf{x}$ is one of the support points of ξ^ .*

When the design field is discrete, we only need to find the optimal weights for the finite experimental settings. The lift-one algorithm (Bu et al., 2020; Yang and Mandal, 2015; Yang et al., 2017) can be used to find the locally D -optimal designs. Details of this algorithm are in Appendix A.2. However, when the design field is continuous, we are also interested in

finding possible support points. Although the optimal designs found on a moderate number of grid points through the lift-one algorithm are possible, this method is often computationally infeasible when the number is large.

To solve this problem, we propose the following algorithm to find the locally D -optimal designs on the continuous design field $\mathcal{X}$.

Algorithm 1.

- (i) Set $t = 0$, and initialize ξ_0 such that $|M(\xi_0)| > 0$. Let $S^{(0)}$ be set of the support points of ξ_0 , and $\varepsilon(\varepsilon > 0)$ be the predefined tolerance.
- (ii) Derive the optimal weights for $S^{(t)}$ to form ξ_t , through the lift-one algorithm.
- (iii) Find $\mathbf{x}_t^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \phi(\mathbf{x}, \xi_t)$, where $\phi(\mathbf{x}, \xi_t)$ is defined in Theorem 4.
- (iv) If $\phi(\mathbf{x}_t^*, \xi_t) \leq \varepsilon$, then return ξ_t .
- (v) Otherwise, set $S^{(t+1)} = S^{(t)} \cup \{\mathbf{x}_t^*\}$ and update $t = t + 1$. Repeat Steps (ii)-(iv).

According to Theorem 2, the uniform weights design with the support points satisfying the conditions in Corollary 1 is a feasible initialization to start the lift-one algorithm. In step (iii), the optimal solution is easy

to achieve, when the design field $\mathcal{X}$ is regular. For example, the Nelder-Mead method (Nelder and Mead, 1965) can be applied here. It is worth mentioning that our algorithm is also suitable for the discrete design case and the maximization in step (iii) can be done through the exhaustive search.

The design achieved by Algorithm 1 will converge to an optimal design as stated in the following Theorem.

Theorem 5. *For any initial design which satisfies $|M(\xi_0)| > 0$, the sequence of designs $\{\xi_t\}$ in Algorithm 1 converges to an optimal design that maximizes $|M(\xi)|$ as $t \rightarrow \infty$.*

We consider a CR model under non-proportional odds assumption with $p_1 = \dots = p_{J-1}$ and there exist p_1 distinct experimental settings $\mathbf{x}_1, \dots, \mathbf{x}_{p_1}$ such that $H_1, \dots, H_{J-1}$ are full row rank. The following theorem shows that the optimal weights are $1/p_1$ for all the experimental settings. This can further accelerate step (ii) in our algorithm.

Theorem 6. *If $p_0 = 0, p_1 = \dots = p_{J-1}$ and there exist p_1 distinct experimental settings $\mathbf{x}_1, \dots, \mathbf{x}_{p_1}$ such that $H_1, \dots, H_{J-1}$ are full row rank, then*

the D -optimal design on $\{\mathbf{x}_1, \dots, \mathbf{x}_{p_1}\}$ is

$$\xi^* = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_{p_1} \\ \frac{1}{p_1} & \cdots & \frac{1}{p_1} \end{pmatrix}.$$

Theorem 6 shows that for non-proportional odds models with $p_1 = \dots = p_{J-1}$, uniform allocations are still D -optimal as the D -optimal designs studied in Corollary S.7 in the Supplementary Materials of Bu et al. (2020).

In general, uniform weighted designs are not always optimal. For example, consider the following CR model with two distinct experimental settings $x_1, x_2 \in \mathbb{R}$, i.e.,

$$\begin{aligned} g(\pi_{i1}) &= \theta_1 + \beta x_i, \\ g\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}}\right) &= \theta_2 + \beta x_i. \end{aligned} \tag{4.7}$$

The D -optimal design for Model (4.7) can be obtained by the following theorem.

Theorem 7. *For Model (4.7), the D -optimal design supported on $\{x_1, x_2\}$ is*

$$\xi^* = \begin{pmatrix} x_1 & x_2 \\ \frac{c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}} & \frac{c_1}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}} \end{pmatrix},$$

where

$$\begin{aligned} c_1 &= \frac{\bar{g}_{11}^2 \bar{g}_{12}^2 (\pi_{12} + \pi_{13})^2}{\pi_{11} \pi_{12} \pi_{13}} \left(\frac{\bar{g}_{21}^2}{\pi_{21} (\pi_{22} + \pi_{23})} + \frac{(\pi_{22} + \pi_{23})^2 \bar{g}_{22}^2}{\pi_{22} \pi_{23}} \right) (x_1 - x_2)^2, \\ c_2 &= \frac{\bar{g}_{21}^2 \bar{g}_{22}^2 (\pi_{22} + \pi_{23})^2}{\pi_{21} \pi_{22} \pi_{23}} \left(\frac{\bar{g}_{11}^2}{\pi_{11} (\pi_{12} + \pi_{13})} + \frac{(\pi_{12} + \pi_{13})^2 \bar{g}_{12}^2}{\pi_{12} \pi_{13}} \right) (x_1 - x_2)^2, \end{aligned}$$

and $\bar{g}_{ij} = (g^{-1})'(\theta_j + \beta x_i)$, for $i = 1, 2, j = 1, 2$.

From Theorem 7, the uniform allocation is D -optimal if and only if $c_1 = c_2$. Thus, the uniform allocations are not D -optimal in general for CR models under partial proportional odds assumption.

Now we derive the analytic solution for the following model which generalizes the model in Example 1:

$$\begin{aligned} g(\pi_{i1}) &= \theta_{11} + \theta_{12}x_i + \theta_{13}x_i^2, \\ g\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}}\right) &= \theta_{21} + \theta_{22}x_i. \end{aligned} \quad (4.8)$$

The optimal weights for the D -optimal design supported on $\{x_1, x_2, x_3\}$ are summarized in the following theorem.

Theorem 8. *For Model (4.8), define*

$$c_1 = t_1 t_2 (x_1 - x_2)^2, c_2 = t_1 t_3 (x_1 - x_3)^2, c_3 = t_2 t_3 (x_2 - x_3)^2,$$

where $t_i = ((g^{-1})'(\theta_{21} + \theta_{22}x_i))^2(\pi_{i2} + \pi_{i3})^3(\pi_{i2}\pi_{i3})^{-1}$. Without any loss of generality, $c_1 \geq c_2 \geq c_3$, then the optimal weights for the D -optimal design supported on $\{x_1, x_2, x_3\}$ satisfy following statements.

(1) If $c_1 = c_2 = c_3$, then $\omega_1 = \omega_2 = \omega_3 = 1/3$.

(2) If $c_1 = c_2 < c_3$, then $\omega_1 = \omega_2 = (-2c_1 + c_3 + \Delta_1)/\Lambda_1, \omega_3 = c_3/\Lambda_1$, where

$$\Delta_1 = \sqrt{4c_1^2 - c_1c_3 + c_3^2} \text{ and } \Lambda_1 = -4c_1 + 3c_3 + 2\Delta_1.$$

(3) If $c_1 < c_2 = c_3$, then $\omega_1 = (-c_1 + 2c_3 + \Delta_2)/\Lambda_2$ and $\omega_2 = \omega_3 = 3c_3/\Lambda_2$,

where $\Delta_2 = \sqrt{c_1^2 - c_1c_3 + 4c_3^2}$ and $\Lambda_2 = -c_1 + 8c_3 + \Delta_2$.

(4) If $c_1 < c_2 < c_3$, then $\omega_1 = u_1/(u_1 + u_2 + 1)$, $\omega_2 = u_2/(u_1 + u_2 + 1)$, $\omega_3 =$

$1/(u_1 + u_2 + 1)$, where u_1, u_2 satisfies

$$c_3u_1u_2(u_1 - 2) + 2c_2u_1(u_1 - 1) = c_1u_2(-2u_1 + 1), \quad (4.9)$$

$$c_3u_1u_2(u_2 - 2) + 2c_1u_2(u_2 - 1) = c_2u_1(-2u_2 + 1).$$

Theorem 8 generalizes the results for logit link which is studied in Bu et al. (2020). The characteristic of such design is the same regardless of c_1, c_2, c_3 , which are decided by the link function and model parameters. The explicit solution of Equation (4.9) was given in the proof of Lemma S.4 of Bu et al. (2020), and we omit it for simplicity.

5. Simulation studies

In this section, we evaluate our method through several examples. The CR models with five commonly used link functions, listed in Table 1, are considered throughout this section. The quality of a design ξ is evaluated by the relative D -efficiency

$$\text{Eff}_D(\xi) = \left(\frac{|M(\xi)|}{|M(\xi^*)|} \right)^{1/p} \times 100\%,$$

where ξ^* is found through the Algorithm 1 in Section 4.

Table 1: Link functions considered in Section 5.

link name	link function	link name	link function
logit	$\log(\mu/(1 - \mu))$	probit	$\Phi^{-1}(\mu)$
log-log	$-\log(-\log(\mu))$	c-log-log	$\log(-\log(1 - \mu))$
Cauchit	$\tan(\pi(\mu - 1/2))$		

Here $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$ and c-log-log is the abbreviation of the terminology “complementary log-log”.

Two kinds of designs are compared with our method (ξ^*). The first is the uniform weighted design, i.e., all the given support points are assigned equal weights. The second is the design proposed by Bu et al. (2020) on discrete design field. Note that both the uniform weighted design and the design proposed by Bu et al. (2020) are available only if the design field is discrete. When the design field is continuous, a common method is to consider the grid points as its discrete levels. Let ξ_U and ξ_{BMY} denote the uniform weighted design and design proposed by Bu et al. (2020) on the grid points, respectively. Note that only the optimal design for Model (2.2) with logit link is considered in Bu et al. (2020), and hence ξ_{BMY} is not available when the link is not logit link. Thus, we denote the results as “NA” in the corresponding part.

Example 4. In the experiment on the emergence of house flies (Zocchi and Atkinson, 1999), seven sets of pupae were exposed to seven different doses of Gamma radiation, respectively. After a period of time, for each set of pupae, observations included the number of flies that died before the opening of the pupae, the number of flies out of the opened pupae but died before completing emergence, and the number of flies that completely emerged. See Table 1 in Zocchi and Atkinson (1999) for more details.

We confirm that Model (2.5) used in Zocchi and Atkinson (1999) fits the data the best among the five link functions in terms of AIC. Therefore, we only consider the optimal design for Model (2.5).

Following Zocchi and Atkinson (1999), the design field $\mathcal{X} = [0, 200]$. The locally D -optimal design ξ^* is searched out through the algorithm in Section 4 with the true parameters estimated from the original data. To show the efficiencies of our design (ξ^*) against ξ_U and ξ_{BMY} , we consider five different cases, i.e., 4, 6, 10, 20, and 50 grid points. The summarized efficiencies of ξ_U and ξ_{BMY} are listed in Table 2. The numbers in the first row of Table 2 stand for the numbers of the grid points used in the corresponding designs.

As shown in Table 2, the uniform weighted designs are far from satisfactory for all cases. The relative D -efficiency will increase as the number

Table 2: Relative D -efficiencies of our method (ξ^*) against the uniform weights designs (ξ_U) and the designs proposed by Bu et al. (2020) (ξ_{BMY}).

		4	6	10	20	50
logit	$\text{Eff}_D(\xi_U)$	60.1%	65.8%	66.5%	66.1%	65.6%
	$\text{Eff}_D(\xi_{BMY})$	75.3%	90.3%	97.3%	99.6%	99.9%

of grid points increased when the grid point search mentioned in Bu et al. (2020) is adopted. However, the corresponding design cannot achieve satisfactory efficiency in this example when only four grid points are used. In practice, the number of grid points is hard to decide. Therefore, our method is prior to the grid point search on the continuous design field.

To evaluate the computational efficiency, we report the computing time (in seconds) for ξ^* and ξ_{BMY} with 4, 6, 10, 20, and 50 grid points. For each case, we repeat 1000 times and compute the average time. All computations are carried out on a Mac Mini with an M1 chip. Table 3 shows the results, where $\xi_{BMY,k}$ indicates the design proposed by Bu et al. (2020) with k grid points. From Table 3, one can see that to achieve the optimal design (in the sense the relative efficiency is more than 99%), ξ_{BMY} needs much more time compared with our method.

Example 5. In the developmental toxicity study (Agresti, 2006), in early

Table 3: Average of computing time (in seconds) of our method (ξ^*) and the designs proposed by Bu et al. (2020) with different grid points ($\xi_{BMY,k}$).

ξ^*	$\xi_{BMY,4}$	$\xi_{BMY,6}$	$\xi_{BMY,10}$	$\xi_{BMY,20}$	$\xi_{BMY,50}$
0.198	0.0257	0.0747	0.529	3.033	22.73

pregnancy, each mouse was exposed to one of five scores for concentration level for ten days. Each fetus had three possible outcomes, dead, malformation, and normal. See Table 6.11 in Agresti (2006) for more details. Agresti (2006) applied Model (2.2) with logit link under proportional odds assumption. The model is

$$\begin{aligned} g(\pi_{i1}) &= \theta_1 + \beta x_i, \\ g\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}}\right) &= \theta_2 + \beta x_i. \end{aligned} \tag{5.10}$$

where $i = 1, \dots, m$ and $g(\cdot)$ is logit link.

The score for the concentration level is a continuous factor ranging from 0 to 500. As in Example 4, we also compare our design with ξ_U and ξ_{BMY} under the same settings.

Actually, Cauchit link is the best among the five link functions in terms of AIC. Therefore, we also consider Model (5.10) with $g(\cdot)$ being Cauchit link. The relative efficiencies of ξ_U and ξ_{BMY} are listed in Table 4.

From Table 4, we can see that the performance of ξ_U and ξ_{BMY} are

Table 4: Relative D -efficiencies of our methods against the uniform weights designs (ξ_U) and the designs proposed by Bu et al. (2020) (ξ_{BMY}).

		4	6	10	20	50
logit	$\text{Eff}_D(\xi_U)$	83.7%	84.0%	83.8%	83.3%	83.0%
	$\text{Eff}_D(\xi_{BMY})$	97.4%	99.2%	99.7%	99.9%	99.9%
Cauchit	$\text{Eff}_D(\xi_U)$	66.2%	55.5%	55.3%	53.2%	51.9%
	$\text{Eff}_D(\xi_{BMY})$	NA	NA	NA	NA	NA

similar to Example 4, in logit link case. When Cauchit link is applied, both ξ_U 's and ξ_{BMY} 's are far away from satisfaction. This is because the method in Bu et al. (2020) is only designed for models with logit link.

Example 6. In this example, we demonstrate the case that the parameters in the Fisher information matrix are not known in advance. In this case, locally optimal design can not be applied directly. To overcome such a problem, a Bayesian optimal design (Chaloner and Verdinelli, 1995) is adopted here. To be precise, we seek a design such that $E_{\gamma}(\log |M(\xi, \gamma)|)$ attains its maximum, where $\gamma = (\beta^T, \theta_1^T, \dots, \theta_{j-1}^T)^T$.

Consider the house flies experiment in Example 4. A multivariate normal prior with its mean and variance-covariance matrix estimated from the initial dataset in Zocchi and Atkinson (1999) is considered in this ex-

ample. This prior is also used as the prior for the EW D -optimal design in Bu et al. (2020). To avoid calculating the expectation of the Fisher information matrix analytically, we use its Monte Carlo version instead. More precisely, we consider the Bayesian optimal design ξ_B which maximize $\sum_{i=1}^{1000} \log |M(\xi, \gamma_i)|$, where $\gamma_1, \dots, \gamma_{1000}$ are independent and identically generated from the prior. The relative D -efficiency among the 1000 locally optimal designs are summarized in Figure 1, where ξ_L is the locally optimal design studied in Example 4, and $\xi_{U,k}$ indicates the uniform weighted design for k grid points. Clearly, Figure 1 illustrates that ξ_B and ξ_L are highly robust than the uniform weights design.

Figure 1: Summary of the relative D -efficiencies of the Bayesian optimal design (ξ_B), locally optimal design (ξ_L), and the uniform weights designs ($\xi_{U,k}$).

6. Conclusion

In this paper, we address the problem of constructing locally D -optimal designs for CR models with general link functions under partial proportional odds assumption. The Fisher information matrix of the CR model is derived, which helps us better understand the characteristics of D -optimal designs. An algorithm to search out optimal designs on continuous design fields is presented, which is much better than the grid point search. See Example 4 for details. Furthermore, our method can also be applied to discrete design fields with a simple modification. Example 5 shows that blindly using the optimal designs under logit link will be not very efficient when other links are used to characterize the data. Several numerical results show the necessity and superiority of the proposed method comparing with some state-of-arts methods.

Although the locally optimal designs may not be very efficient if the initial guesses are far away from the true parameters, they can still be served as the basis for constructing the optimal designs with respect to more robust criteria such as Bayesian optimality (Chaloner and Verdinelli, 1995). These can be regarded as possible implementations of our proposed designs.

Supplementary Material

All technical proofs and additional simulation results are included in the Supplementary Material.

Acknowledgements

The authors sincerely thank the editor, associate editor, and two referees for their valuable comments. Ai's work was supported by NSFC grants 11671019 and 12071014, SSFC grant 19ZDA121 and LMEQF. Yu's work was supported by NNSF of China grant 12001042 and Beijing Institute of Technology research fund program for young scholars.

A. Appendix

A.1 Definition of $G(\mathbf{x})$

Recall $g(\cdot)$ is a general link function, $\mathbf{x} \in \mathcal{X}$ is a design point, and $G(\mathbf{x}) = (g_{jk}(\mathbf{x}))$ in Theorem 1 is a $J \times (J - 1)$ matrix.

When $j = 1$, $g_{1k}(\mathbf{x})$ has the form

$$g_{1k}(\mathbf{x}) = \begin{cases} (g^{-1})'(\mathbf{h}_0^T(\mathbf{x})\boldsymbol{\beta} + \mathbf{h}_1^T(\mathbf{x})\boldsymbol{\theta}_1) & k = 1, \\ 0 & k = 2, \dots, J - 1. \end{cases}$$

When $j = 2, \dots, J - 1$,

$$g_{jk}(\mathbf{x}) = \begin{cases} -g^{-1}(\mathbf{h}_0^T(\mathbf{x})\boldsymbol{\beta} + \mathbf{h}_j^T(\mathbf{x})\boldsymbol{\theta}_j) \left(\sum_{l=1}^{j-1} g_{lk}(\mathbf{x}) \right) & k = 1, \dots, j-1, \\ (g^{-1})'(\mathbf{h}_0^T(\mathbf{x})\boldsymbol{\beta} + \mathbf{h}_j^T(\mathbf{x})\boldsymbol{\theta}_j) \left(1 - \sum_{l=1}^{j-1} \pi_{il} \right) & k = j, \\ 0 & k = j+1, \dots, J-1. \end{cases}$$

Finally, when $j = J$,

$$g_{Jk}(\mathbf{x}) = - \sum_{j=1}^{J-1} g_{jk}(\mathbf{x}), k = 1, \dots, J-1.$$

It is worth mentioning that $G(\mathbf{x})$ is a lower triangular matrix, i.e., $g_{jk}(\mathbf{x}) = 0$ for $k > j$. By Assumption 2, $g_{jj}(\mathbf{x}) > 0$ for $j = 1, \dots, J-1$, thus $G(\mathbf{x})$ has full column rank.

A.2 Lift-one algorithm

When the experimental settings $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is given, lift-one algorithm gives the optimal weights $(\omega_1^*, \dots, \omega_m^*)$ for the corresponding experiment settings. According to Theorem 3, we write the determinant of the Fisher information matrix as $|M(\xi)| = f(\omega_1, \dots, \omega_m)$ for an order p homogeneous polynomial function f . Following Yang et al. (2016, 2017), define $S_+ = \{(\omega_1, \dots, \omega_m)^T | \omega_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \omega_i = 1, f(\omega_1, \dots, \omega_m) > 0\}$, and

$$f_i(z) = f\left(\frac{1-z}{1-\omega_i}\omega_1, \dots, \frac{1-z}{1-\omega_i}\omega_{i-1}, z, \frac{1-z}{1-\omega_i}\omega_{i+1}, \dots, \frac{1-z}{1-\omega_i}\omega_m\right). \quad (\text{A.11})$$

with $0 \leq z \leq 1$ and $(\omega_1, \dots, \omega_m)^T \in S_+$. Note that $f_i(z)$ in Equation (A.11) can be represented as

$$f_i(z) = \sum_{\substack{\alpha_1 \geq 0, \dots, \alpha_m \geq 0 \\ \alpha_1 + \dots + \alpha_m = p}} c_{\alpha_1, \dots, \alpha_m} z^{\alpha_i} \prod_{j=1, j \neq i}^m \left(\frac{1-z}{1-\omega_i} \omega_j \right)^{\alpha_j}.$$

Moreover, $c_{\alpha_1, \dots, \alpha_m} = 0$ when some $\alpha_i \geq J$, thus $f_i(z)$ can be further simplified as $f_i(z) = \sum_{k=0}^{J-1} c_{(k)} z^k (1-z)^{p-k}$, for some coefficients $c_{(0)}, \dots, c_{(J-1)}$. The coefficients $c_{(0)}, \dots, c_{(J-1)}$ can be calculated through solving the following J linear equations, $f_i(z_j) = \sum_{k=0}^{J-1} c_{(k)} z_j^k (1-z_j)^{p-k}$, for J distinct $z_1, \dots, z_J \in [0, 1)$, where $f_i(z_j)$ can be directly calculated through Equation (A.11).

The lift-one algorithm is described in what follows.

Algorithm 2.

- (i) Start with a uniform allocation $\omega_0 = (1/m, \dots, 1/m)^T$ and compute $f(\omega_0)$.
- (ii) Set up a random order of i going through $\{1, \dots, m\}$.
- (iii) For each i , determine $f_i(z)$.
- (iv) Use quasi-Newton algorithm to find z^* maximizing $f_i(z)$ with $0 \leq z \leq 1$. If $f_i(z^*) \leq f_i(0)$, let $z^* = 0$. Define $\omega_*^{(i)} = (\omega_1(1-z_*)/(1-$

$$\omega_i), \dots, \omega_{i-1}(1 - z_*)/(1 - \omega_i), z_*, \omega_{i+1}(1 - z_*)/(1 - \omega_i), \dots, \omega_m(1 - z_*)/(1 - \omega_i))^T.$$

(v) Replace ω_0 with $\omega_*^{(i)}$, and $f(\omega_0)$ with $f(\omega_*^{(i)})$.

(vi) Repeat (ii)-(v) until convergence, that is $f(\omega_0) = f(\omega_*^{(i)})$ for each i .

References

- Agresti, A. (2006). *An Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons.
- Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*, Volume 34. Oxford University Press.
- Atkinson, A. C., V. V. Fedorov, A. M. Herzberg, and R. Zhang (2014). Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference* 144, 81–91.
- Bu, X., D. Majumdar, and J. Yang (2020). D-optimal designs for multinomial logistic models. *Annals of Statistics* 48(2), 983–1000.
- Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical Science* 10(3), 273–304.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics* 24(4), 586–602.

-
- Ford, I., D. Titterton, and C. P. Kitsos (1989). Recent advances in nonlinear experimental design. *Technometrics* 31(1), 49–60.
- Ford, I., B. Torsney, and C. Wu (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society: Series B (Methodological)* 54(2), 569–583.
- Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological methods & research* 38(2), 306–347.
- Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics* 29(2), 561–576.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* 2(5), 849–879.
- Lukemire, J., A. Mandal, and W. K. Wong (2019). d-QPSO: A quantum-behaved particle swarm technique for finding D-optimal designs with discrete and continuous factors and a binary response. *Technometrics* 61(1), 77–87.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42(2), 109–127.
- McCullagh, P. and J. A. Nelder (1989). *Generalized linear models* (2nd ed.). Chapman and Hall/CRC.
- Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer*

Journal 7(4), 308–313.

O’Connell, A. A. (2006). *Logistic regression models for ordinal response variables* (1st ed.).
Sage.

Peterson, B. and F. E. Harrell (1990). Partial proportional odds models for ordinal response
variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 39(2),
205–217.

Sitter, R. R. and C. Wu (1993). Optimal designs for binary response experiments: Fieller, D,
and A criteria. *Scandinavian Journal of Statistics* 20(4), 329–341.

Stufken, J. and M. Yang (2012). Optimal designs for generalized linear models. In K. Hinkel-
mann (Ed.), *Design and Analysis of Experiments, Volume 3: Special Designs and Appli-
cations*. Wiley.

Yang, J. and A. Mandal (2015). D-optimal factorial designs under generalized linear models.
Communications in Statistics-Simulation and Computation 44(9), 2264–2277.

Yang, J., A. Mandal, and D. Majumdar (2016). Optimal designs for 2^k factorial experiments
with binary response. *Statistica Sinica* 26(1), 385–411.

Yang, J., L. Tong, and A. Mandal (2017). D-optimal designs with ordered categorical data.
Statistica Sinica 27(4), 1879–1902.

Yang, M., S. Biedermann, and E. Tang (2013). On optimal designs for nonlinear models: A
general and efficient algorithm. *Journal of the American Statistical Association* 108(504),

1411–1420.

Yang, M., B. Zhang, and S. Huang (2011). Optimal designs for binary response experiments with multiple variables. *Statistica Sinica* 21(3), 1415–1430.

Zocchi, S. S. and A. C. Atkinson (1999). Optimum experimental designs for multinomial logistic models. *Biometrics* 55(2), 437–444.

LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China.

E-mails: myai@math.pku.edu.cn, zqye@pku.edu.cn

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100811, China.

E-mail: yujunbeta@bit.edu.cn