LOCALLY D-OPTIMAL DESIGNS FOR HIERARCHICAL RESPONSE EXPERTIMENTS

Mingyao Ai[1], Zhiqiang Ye[1], Jun Yu[2]

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

School of Mathematics and Statistics, Beijing Institute of Technology[2]

Supplementary Material

S1 Proofs

To proof Theorem 1, we begin with two technical lemmas. Let $\boldsymbol{\gamma} \in \mathbb{R}^p$ denote the parameter vector, i.e., $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^T = (\boldsymbol{\beta}^T, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_{J-1}^T)^T$. The first lemma gives the Fisher information matrix for Model (2.2) under an exact design. The second lemma calculates $\partial\pi(\boldsymbol{x})/\partial\boldsymbol{\gamma}^T$, which is an essential part of Theorem 1.

For an exact design

$$\xi_{\text{exact}} = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m \\ n_1 & \cdots & n_m \end{pmatrix},$$

the corresponding Fisher information matrix is derived in the following lemma.

Lemma S1. *Suppose Assumptions 1 and 2 hold, the Fisher information* *matrix for Model* (2.2) *under the exact design* ξ_{exact} *can be written as*

$$M(\xi_{\text{exact}}) = \sum_{i=1}^{m} n_i M_i,$$

where $M_i = (m_{i_{st}})_{1\le s,t\le p}$ *is a* $p \times p$ *matrix with*

$$m_{i_{st}} = \sum_{j=1}^{J} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t}.$$

Proof of Lemma S1. For the experimental setting $\boldsymbol{x}_i$, for $i = 1, \ldots, m$, the responses $(Y_{i1}, \ldots, Y_{iJ})^T \sim \text{Multinomial}(n_i; \pi_{i1}, \ldots, \pi_{iJ})$. We know that $E(Y_{ij}) = n_i\pi_{ij}$, $E(Y_{ij}^2) = n_i(n_i-1)\pi_{ij}^2 + n_i\pi_{ij}$, and $E(Y_{is}Y_{it}) = n_i(n_i-1)\pi_{is}\pi_{it}$ when $s \neq t$.

The log-likelihood function (up to a constant) is

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^{m}\sum_{j=1}^{J} Y_{ij}\log \pi_{ij}.$$

Then the score function is

$$\frac{\partial l}{\partial \gamma_s} = \sum_{i=1}^{m}\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s}.$$

Note that $\pi_{i1} + \cdots + \pi_{iJ} = 1$, it follows that

$$E\left(\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s}\right) = \sum_{j=1}^{J} n_i \frac{\partial \pi_{ij}}{\partial \gamma_s} = n_i \frac{\partial}{\partial \gamma_s}\left(\sum_{j=1}^{J} \pi_{ij}\right) = 0,$$

for $i = 1, \ldots, m$. The Hessian matrix can be achieved through the following calculation.

$$E\frac{\partial l}{\partial \gamma_s}\frac{\partial l}{\partial \gamma_t} = E\left(\sum_{i=1}^{m}\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_t}\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{m} E\left(\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s}\right)\left(\sum_{j=1}^{J} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_t}\right) \\
&= \sum_{i=1}^{m} E\left(\sum_{j=1}^{J} \frac{Y_{ij}^2}{\pi_{ij}^2} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t} + 2 \sum_{1\le j<k\le m} \frac{Y_{ij}}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{Y_{ik}}{\pi_{ik}} \frac{\partial \pi_{ik}}{\partial \gamma_t}\right) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{n_i(n_i-1)\pi_{ij}^2 + n_i \pi_{ij}}{\pi_{ij}^2} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t} \\
&\quad + 2 \sum_{i=1}^{m} \sum_{1\le j<k\le m} \frac{n_i(n_i-1)\pi_{ij}\pi_{ik}}{\pi_{ij}\pi_{ik}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ik}}{\partial \gamma_t} \\
&= \sum_{i=1}^{m} n_i \left((n_i - 1)\left(\sum_{j=1}^{J} \frac{\partial \pi_{ij}}{\partial \gamma_s}\right)\left(\sum_{j=1}^{J} \frac{\partial \pi_{ij}}{\partial \gamma_t}\right) + \sum_{j=1}^{J} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t}\right) \\
&= \sum_{i=1}^{m} \left(n_i \sum_{j=1}^{J} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t}\right).
\end{aligned}
$$

By the definition of Fisher information matrix, we have

$$
M(\xi_{\text{exact}}) = E\left(\frac{\partial l}{\partial \boldsymbol{\gamma}}\right)\left(\frac{\partial l}{\partial \boldsymbol{\gamma}}\right)^T = \sum_{i=1}^{m} n_i M_i,
$$

where $M_i = (m_{i_{st}})_{1\le s,t\le p}$ is a $p \times p$ matrix with

$$
m_{i_{st}} = \sum_{j=1}^{J} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_s} \frac{\partial \pi_{ij}}{\partial \gamma_t}.
$$

□

Remark S1. From Lemma S1, the Fisher information matrix for Model (2.2) under an approximate design

$$
\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m \\ \omega_1 & \cdots & \omega_m \end{pmatrix},
$$

can be written as

$$M(\xi) = \sum_{i=1}^{m} \omega_i M_i.$$

Recall $\delta_{\boldsymbol{x}}$ denote the single point design, then $M(\delta_{\boldsymbol{x}_i}) = M_i$, for $i = 1, \ldots, m$.

Let $\partial \boldsymbol{\pi}(\boldsymbol{x})/\partial \boldsymbol{\gamma}^T$ denote a $J \times p$ matrix, whose (j,k)th entry is $\partial \pi_j(\boldsymbol{x})/\partial \gamma_k$, where $\boldsymbol{x} \in \mathcal{X}$ is a design point. We have the following lemma.

Lemma S2. *For Model* (2.2),

$$\frac{\partial \boldsymbol{\pi}(\boldsymbol{x})}{\partial \boldsymbol{\gamma}^T} = G(\boldsymbol{x}) H(\boldsymbol{x}), \tag{S1.1}$$

where $G(\boldsymbol{x})$ is defined in Section A.1 and $H(\boldsymbol{x})$ is defined in Section 2.2.

Proof of Lemma S2. To be convenience, let $e_j(\boldsymbol{x}) = g^{-1}(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x})\boldsymbol{\theta}_j)$, for $j = 1, \ldots, J-1$. We first show the following equation

$$\frac{\partial \pi_j(\boldsymbol{x})}{\partial \beta_s} = h_{0s}(\boldsymbol{x}) \sum_{k=1}^{j} g_{jk}(\boldsymbol{x}), \tag{S1.2}$$

holds, for $s = 1, \ldots, p_0$ and $j = 1, \ldots, J$. For each s, we prove Equation (S1.2) holds, for $j = 1, \ldots, J-1$, by induction.

(i) When $j = 1$, it follows that

$$\frac{\partial \pi_1(\boldsymbol{x})}{\partial \beta_s} = h_{0s}(\boldsymbol{x}) g_{11}(\boldsymbol{x}),$$

by the fact $\pi_1(\boldsymbol{x}) = e_1(\boldsymbol{x})$, which implies Equation (S1.2) holds for $j = 1$.

(ii) Suppose Equation (S1.2) holds for $2, \ldots, j-1$ $(j < J)$, by

$$\pi_j(\boldsymbol{x}) = e_j(\boldsymbol{x}) \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right),$$

we have

$$\begin{aligned}
\frac{\partial \pi_j(\boldsymbol{x})}{\partial \beta_s} &= \frac{\partial e_j(\boldsymbol{x})}{\partial \beta_s} \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right) - e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} \frac{\partial \pi_k(\boldsymbol{x})}{\partial \beta_s} \\
&= h_{0s}(\boldsymbol{x})(g^{-1})'(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x})\boldsymbol{\theta}_j) \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right) \\
&\quad - e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} \left(h_{0s}(\boldsymbol{x}) \sum_{l=1}^{k} g_{kl}(\boldsymbol{x})\right) \\
&= h_{0s}(\boldsymbol{x}) g_{jj}(\boldsymbol{x}) + h_{0s}(\boldsymbol{x}) \sum_{k=1}^{j-1} \left(-e_j(\boldsymbol{x}) \sum_{l=1}^{k} g_{kl}(\boldsymbol{x})\right) \\
&= h_{0s}(\boldsymbol{x}) g_{jj}(\boldsymbol{x}) + h_{0s}(\boldsymbol{x}) \sum_{k=1}^{j-1} \left(-e_j(\boldsymbol{x}) \sum_{l=1}^{j-1} g_{kl}(\boldsymbol{x})\right) \\
&= h_{0s}(\boldsymbol{x}) g_{jj}(\boldsymbol{x}) + h_{0s}(\boldsymbol{x}) \sum_{l=1}^{j-1} \left(-e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} g_{kl}(\boldsymbol{x})\right) \\
&= h_{0s}(\boldsymbol{x}) g_{jj}(\boldsymbol{x}) + h_{0s}(\boldsymbol{x}) \sum_{l=1}^{j-1} g_{jl}(\boldsymbol{x}) \\
&= h_{0s}(\boldsymbol{x}) \sum_{l=1}^{j} g_{jl}(\boldsymbol{x}),
\end{aligned}$$

which implies Equation (S1.2) holds for j.

As for the case $j = J$, utilizing the fact $\pi_1(\boldsymbol{x}) + \cdots + \pi_J(\boldsymbol{x}) = 1$ and the facts that have been proved in (i) and (ii), we have

$$\begin{aligned}
\frac{\partial \pi_J(\boldsymbol{x})}{\partial \beta_s} &= -\sum_{j=1}^{J-1} \frac{\partial \pi_j(\boldsymbol{x})}{\partial \beta_s} \\
&= -\sum_{j=1}^{J-1} \left(h_{0s}(\boldsymbol{x}) \sum_{k=1}^{j} g_{jk}(\boldsymbol{x}) \right) \\
&= h_{0s}(\boldsymbol{x}) \sum_{j=1}^{J-1} \left(-\sum_{k=1}^{J-1} g_{jk}(\boldsymbol{x}) \right) \\
&= h_{0s}(\boldsymbol{x}) \sum_{k=1}^{J-1} \left(-\sum_{j=1}^{J-1} g_{jk}(\boldsymbol{x}) \right) \\
&= h_{0s}(\boldsymbol{x}) \sum_{k=1}^{J-1} g_{Jk}(\boldsymbol{x}),
\end{aligned}$$

then Equation (S1.2) holds for $j = J$. Therefore, Equation (S1.2) holds, for $s = 1, \ldots, p_0$ and $j = 1, \ldots, J$.

Now we turn to prove the following equation,

$$\frac{\partial \pi_j(\boldsymbol{x})}{\partial \theta_{uv}} = h_{uv}(\boldsymbol{x}) g_{ju}(\boldsymbol{x}), \tag{S1.3}$$

for $u = 1, \ldots, J-1$, $v = 1, \ldots, p_u$, and $j = 1, \ldots, J$. Similarly, for each u, v, we prove Equation (S1.3) holds for $j = 1, \ldots, J-1$, by induction.

(1) When $j = 1$, then $\pi_1(\boldsymbol{x}) = e_1(\boldsymbol{x})$, we have

$$\frac{\partial \pi_1(\boldsymbol{x})}{\partial \theta_{1r}} = h_{1r}(\boldsymbol{x}) g_{11}(\boldsymbol{x}), \frac{\partial \pi_1(\boldsymbol{x})}{\partial \theta_{uv}} = 0 = h_{uv}(\boldsymbol{x}) g_{1u}(\boldsymbol{x}),$$

for $r = 1, \ldots, p_1$, $u = 2, \ldots, J-1$, and $v = 1, \ldots, p_u$, which implies Equation (S1.3) holds for $j = 1$.

(2) Suppose Equation (S1.3) holds for $2, \ldots, j-1 (j < J)$. For $u = 1, \ldots, j-1$ and $v = 1, \ldots, p_u$, it follows that

$$\begin{aligned}\frac{\partial \pi_j(\boldsymbol{x})}{\partial \theta_{uv}} &= \frac{\partial e_j(\boldsymbol{x})}{\partial \theta_{uv}} \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right) - e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} \frac{\partial \pi_k(\boldsymbol{x})}{\partial \theta_{uv}} \\ &= -e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} \frac{\partial \pi_k(\boldsymbol{x})}{\partial \theta_{uv}} \\ &= -e_j(\boldsymbol{x}) h_{uv}(\boldsymbol{x}) \sum_{k=1}^{j-1} g_{ku}(\boldsymbol{x}) \\ &= h_{uv}(\boldsymbol{x}) g_{ju}(\boldsymbol{x}).\end{aligned}$$

Note that for $v = 1, \ldots, p_j$, it holds that

$$\begin{aligned}\frac{\partial \pi_j(\boldsymbol{x})}{\partial \theta_{jv}} &= \frac{\partial e_j(\boldsymbol{x})}{\partial \theta_{jv}} \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right) - e_j(\boldsymbol{x}) \sum_{k=1}^{j-1} \frac{\partial \pi_k(\boldsymbol{x})}{\partial \theta_{jv}} \\ &= h_{jv}(\boldsymbol{x}) (g^{-1})'(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x})\boldsymbol{\theta}_j) \left(1 - \sum_{k=1}^{j-1} \pi_k(\boldsymbol{x})\right) \\ &= h_{jv}(\boldsymbol{x}) g_{jj}(\boldsymbol{x}).\end{aligned}$$

By the definition of $\pi_j(\boldsymbol{x})$ and $G(\boldsymbol{x})$, the following equation holds

$$\frac{\partial \pi_j(\boldsymbol{x})}{\partial \theta_{uv}} = 0 = h_{uv}(\boldsymbol{x}) g_{ju}(\boldsymbol{x}),$$

for $u = j+1, \ldots, J-1$ and $v = 1, \ldots, p_u$.

Combining the aforementioned three equations, Equation (S1.3) holds for j.

When $j = J$, utilizing the fact $\pi_1(\boldsymbol{x}) + \cdots + \pi_J(\boldsymbol{x}) = 1$ and the facts that have been proved in (a) and (b), we have

$$\begin{aligned}
\frac{\partial \pi_J(\boldsymbol{x})}{\partial \pi_{uv}} &= -\sum_{j=1}^{J-1} \frac{\partial \pi_j(\boldsymbol{x})}{\partial \pi_{uv}} \\
&= -\sum_{j=1}^{J-1} h_{uv}(\boldsymbol{x}) g_{ju}(\boldsymbol{x}) \\
&= h_{uv}(\boldsymbol{x}) \left(-\sum_{j=1}^{J-1} g_{ju}(\boldsymbol{x}) \right) \\
&= h_{uv}(\boldsymbol{x}) g_{Ju}(\boldsymbol{x}),
\end{aligned}$$

which implies Equation (S1.3) holds for $j = J$. Thus Equation (S1.3) holds for $u = 1, \ldots, J-1$, $v = 1, \ldots, p_u$, and $j = 1, \ldots, J$. Based on Equations (S1.2) and (S1.3), Lemma S2 is proved. □

Proof of Theorem 1. Combing the results in Lemmas S1 and S2, it follows that

$$\begin{aligned}
M(\xi) &= \sum_{i=1}^{m} \omega_i M_i \\
&= \sum_{i=1}^{m} \omega_i \left(\frac{\partial \boldsymbol{\pi}(\boldsymbol{x}_i)}{\partial \boldsymbol{\gamma}^T} \right)^T D^{-1}(\boldsymbol{x}_i) \left(\frac{\partial \boldsymbol{\pi}(\boldsymbol{x}_i)}{\partial \boldsymbol{\gamma}^T} \right) \\
&= \sum_{i=1}^{m} \omega_i H^T(\boldsymbol{x}_i) G^T(\boldsymbol{x}_i) D^{-1}(\boldsymbol{x}_i) G(\boldsymbol{x}_i) H(\boldsymbol{x}_i),
\end{aligned}$$

which completes the proof. □

Proof of Theorem 2. Let $\widetilde{H} = (H^T(\boldsymbol{x}_1), \ldots, H^T(\boldsymbol{x}_m))$, and

$$\widetilde{W} = \mathrm{diag}(\omega_1 G^T(\boldsymbol{x}_1)D^{-1}(\boldsymbol{x}_1)G(\boldsymbol{x}_1), \ldots, \omega_m G^T(\boldsymbol{x}_m)D^{-1}(\boldsymbol{x}_m)G(\boldsymbol{x}_m)).$$

According to Theorem 1, the Fisher information matrix can be written as $M(\xi) = \widetilde{H}\widetilde{W}\widetilde{H}^T$. Since $\pi_j(\boldsymbol{x}_i) > 0$, for $j = 1, \ldots, J$, $G(\boldsymbol{x}_i)$ has full column rank (see Appendix A.1), and $\omega_i > 0$, for $i = 1, \ldots, m$, $\widetilde{W}$ is positive definite. Therefore, $M(\xi)$ is positive definite if and only if $\widetilde{H}$ has full row rank. □

Proof of Corollary 1. After some elementary column transformations for the matrix $(H^T(\boldsymbol{x}_1), \ldots, H^T(\boldsymbol{x}_m))$, we obtain a new matrix

$$H_{new} = \begin{pmatrix} H_0 & H_0 & H_0 & \cdots & H_0 \\ H_1 & 0 & 0 & \cdots & 0 \\ 0 & H_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & H_{J-1} \end{pmatrix}.$$

In order to keep H_{new} full row rank, $H_0, \ldots, H_{J-1}$ are full row rank, thus $m \geq p_j$, for $j = 0, \ldots, J-1$.

Suppose $\cap_{j=0}^{J-1} C(H_j^T) \neq \{\mathbf{0}\}$, without loss of generality, we assume that the first row of H_0 lies in $\cap_{j=1}^{J-1} C(H_j^T)$. Therefore, the first row of H_0 can be represented by the linear combination of the rows of $H_1, \ldots, H_{J-1}$, respectively. Thus, the first row in H_{new} can be represented by the last $p-1$ rows, which contradicts the fact that H_{new} is full row rank.

Recall $r = \dim\left(\cap_{j=1}^{J-1} C(H_j^T)\right)$, utilizing the fact that $\cap_{j=0}^{J-1} C(H_j^T) = \{\mathbf{0}\}$, the rank of the matrix $(H_0^T, \ldots, H_{J-1}^T)$ is at least p_0+r. Thus $m \geq p_0+r$. $\square$

Proof of Theorem 3. As mentioned in Remark S1, $M_i = M(\delta_{\boldsymbol{x}_i})$. The information matrix under the design ξ is

$$M(\xi) = \sum_{i=1}^{m} \omega_i M(\delta_{\boldsymbol{x}_i}).$$

Using the same argument in Theorem 2 of Yang et al. (2017), it can be shown that $|M(\xi)|$ is a polynomial function of $(\omega_1, \ldots, \omega_m)$.

Now we will show that the coefficients calculated in Equation (3.1) are zero in conditions (1) or (2).

(1) For the first scenario, recall $M(\delta_{\boldsymbol{x}_i}) = H^T(\boldsymbol{x}_i)G^T(\boldsymbol{x}_i)D^{-1}(\boldsymbol{x}_i)G(\boldsymbol{x}_i)H(\boldsymbol{x}_i)$. The rank of $M(\delta_{\boldsymbol{x}_i})$ is less than or equal to the rank of $G(\boldsymbol{x}_i)$, i.e., $J-1$, for $i = 1, \ldots, m$. Since $\max_{1\leq i\leq m} \alpha_i \geq J$, without loss of generality, we assume $\alpha_1 \geq J$. Then for any $\tau \in \Delta_{\alpha_1,\ldots,\alpha_m}$, there are at least J rows of M_τ which are the same with the corresponding rows of $M(\delta_{\boldsymbol{x}_1})$, then $|M_\tau| = 0$, which implies $c_{\alpha_1,\ldots,\alpha_m} = 0$ according to Equation (3.1).

(2) For the second scenario, let $\bar{H} = (H^T(\boldsymbol{x}_1)G^T(\boldsymbol{x}_1), \ldots, H^T(\boldsymbol{x}_m)G^T(\boldsymbol{x}_m))$, and $\bar{W} = \mathrm{diag}(\omega_1 D^{-1}(\boldsymbol{x}_1), \ldots, \omega_m D^{-1}(\boldsymbol{x}_m))$, then $M(\xi) = \bar{H}\bar{W}\bar{H}^T$.

By Cauchy-Binet formula (Horn and Johnson, 2012), it follows

$$c_{\alpha_1,\ldots,\alpha_m} = \sum_{(v_1,\ldots,v_p)\in\Lambda(\alpha_1,\ldots,\alpha_m)} |\bar{H}[i_1,\ldots,i_p]|^2 \prod_{k:\alpha_k>0} \prod_{l:(k-1)J<v_l\le kJ} \pi^{-1}_{k,v_l-(k-1)J},$$

where $1 \le v_1 < \cdots < v_p \le mJ$, $\Lambda(\alpha_1,\ldots,\alpha_m)$ only depends on $\alpha_1,\ldots,\alpha_m$, and $\bar{H}[i_1,\ldots,i_p]$ is the submatrix consisting of the i_1th, …, i_pth rows of $\bar{H}$. Without loss of generality, we assume $\alpha_1 \ge \cdots \ge \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m$, where $k+1 \le \max\{p_0+r, p_1, \ldots, p_{J-1}\}$. Suppose $c_{\alpha_1,\ldots,\alpha_m} \ne 0$ for some $(\alpha_1,\ldots,\alpha_m)$. Therefore, there exist $(v_1,\ldots,v_p)$ such that $\bar{H}[v_1,\ldots,v_p]$ has full rank p, and $1 \le v_1 < \cdots < v_p \le kJ$. Then $\bar{\bar{H}} = \bar{H}[1,\ldots,kJ]$ is full row rank. Let $\bar{\bar{W}} = k^{-1}\mathrm{diag}(D^{-1}(\boldsymbol{x}_1),\ldots,D^{-1}(\boldsymbol{x}_k))$. $\bar{\bar{H}}\bar{\bar{W}}\bar{\bar{H}}^T$ is positive definite. On the other hand, we can regard $\bar{\bar{H}}\bar{\bar{W}}\bar{\bar{H}}^T$ as the Fisher information matrix under uniform weighted design on the k support points, thus $k \ge \max\{p_0+r, p_1, \ldots, p_{J-1}\}$, which is a contradiction.

□

Proof of Theorem 4. Note that maximizing $|M(\xi)|$ is equivalent to maximize $\log|M(\xi)|$. Recall $\delta_{\boldsymbol{x}}$ denote the single point design. The Frechet derivate of $\log|M(\xi)|$ at ξ^* in the direction of $\delta_{\boldsymbol{x}} - \xi^*$ is

$$\lim_{\alpha\to 0} \frac{1}{\alpha}\left(\log|M((1-\alpha)\xi^* + \alpha\delta_{\boldsymbol{x}})| - \log|M(\xi^*)|\right)$$

$$
\begin{aligned}
&= \lim_{\alpha \to 0} \frac{1}{\alpha} \left(\log |M(\xi^*) + \alpha(M(\delta_{\boldsymbol{x}}) - M(\xi^*))| - \log |M(\xi^*)| \right) \\
&= \mathrm{tr}\left(M^{-1}(\xi^*)(M(\delta_{\boldsymbol{x}}) - M(\xi^*)) \right) \\
&= \mathrm{tr}\left(M^{-1}(\xi^*) M(\delta_{\boldsymbol{x}}) \right) - p \\
&= \mathrm{tr}\left(M^{-1}(\xi^*) H^T(\boldsymbol{x}) G^T(\boldsymbol{x}) D^{-1}(\boldsymbol{x}) G(\boldsymbol{x}) H(\boldsymbol{x}) \right) - p.
\end{aligned}
$$

Then the theorem is proved following Pukelsheim (2006). □

Proof of Theorem 5. Note that the set of all Fisher information matrices is a convex hull. Since the design region is compact, the corresponding set is a convex and compact subset of the linear space of symmetric matrices. By Carathéodory's Theorem (Danninger-Uchida, 2009), there exists a design ξ^* which contains only a finite number of design points that maximizes $\log |M(\xi)|$.

Since $\log |M(\xi_t)|$ is a bounded and increasing function of t, $\log |M(\xi_t)|$ converges when $t \to \infty$. We shall show that

$$
\lim_{t \to \infty} \log |M(\xi_t)| = \log |M(\xi^*)|. \tag{S1.4}
$$

If Equation (S1.4) does not hold, there exists $\zeta > 0$, by the monotonicity of $\log |M(\xi_t)|$, such that

$$
\log |M(\xi^*)| - \log |M(\xi_t)| > \zeta. \tag{S1.5}
$$

Utilizing the concavity of $\log|M(\xi)|$, we have

$$(1-\alpha)\log|M(\xi_t)| + \alpha\log|M(\xi^*)| \le \log|(1-\alpha)M(\xi_t) + \alpha M(\xi^*)|, \quad \text{(S1.6)}$$

for any $0 < \alpha \le 1$. Equation (S1.6) implies that

$$\frac{\log|(1-\alpha)M(\xi_t) + \alpha M(\xi^*)| - \log|M(\xi_t)|}{\alpha} \ge \log|M(\xi^*)| - \log|M(\xi_t)|.$$

Let $\alpha \to 0^+$ and utilize Equation (S1.5),

$$\mathrm{tr}(M^{-1}(\xi_t)(M(\xi^*) - M(\xi_t))) > \zeta. \quad \text{(S1.7)}$$

Recall $\boldsymbol{x}_t^* = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \phi(\boldsymbol{x}, \xi_t)$, then $\phi(\boldsymbol{x}_t^*, \xi_t) \ge \phi(\boldsymbol{x}, \xi_t)$ for any $\boldsymbol{x} \in \chi$. Thus, we have

$$\phi(\boldsymbol{x}_t^*, \xi_t) \ge \int_{\boldsymbol{x}\in\chi} \phi(\boldsymbol{x}, \xi_t)\xi^*(d\boldsymbol{x}) = \mathrm{tr}(M^{-1}(\xi_t)(M(\xi^*) - M(\xi_t))).$$

Combing with Equation (S1.7), it follows that

$$\phi(\boldsymbol{x}_t^*, \xi_t) > \zeta. \quad \text{(S1.8)}$$

Let $\xi_{t+1}(\alpha) = (1-\alpha)\xi_t + \alpha\delta_{\boldsymbol{x}_t^*}$, where $0 \le \alpha \le \frac{1}{2}, t \in \mathbb{N}^*$. Since $\log|M(\xi)|$ is an increasing function and by the definition of ξ_{t+1}, it can be shown that

$$\log\left|\frac{1}{2}M(\xi_t)\right| \le \log|M(\xi_{t+1}(\alpha))| \le \log|M(\xi_{t+1})|, \quad \text{(S1.9)}$$

for any $0 \le \alpha \le \frac{1}{2}$. By the definition of ξ^*, we have

$$\log\left|\frac{1}{2}M(\xi_1)\right| \le \log|M(\xi_{t+1}(\alpha))| \le \log|M(\xi^*)|. \tag{S1.10}$$

Equation (S1.10) implies that $\log|M(\xi_{t+1}(\alpha))|$ is uniformly bounded for $0 \le \alpha \le \frac{1}{2}$ and $t \in \mathbb{N}^*$. By Theorem 3, $|M(\xi_{t+1}(\alpha))|$ is a polynomial of α, which implies that $\log|M(\xi_{t+1}(\alpha))|$ is infinitely differentiable with respect to α. Recall that both $M(\xi_t)$ and $M(\xi_{t+1}(\alpha))$ lie in a same convex and compact subset of the linear space of symmetric matrices for all t and $\alpha \in [0, \frac{1}{2}]$. Combining Equation (S1.10) with the aforementioned facts, there exists $0 < K < \infty$, such that,

$$\inf\left\{\frac{d^2\log|M(\xi_{t+1}(\alpha))|}{d\alpha^2} : \alpha \in \left[0, \frac{1}{2}\right], t \in \mathbb{N}^*\right\} = -K. \tag{S1.11}$$

Using Taylor expansion of $\log|M(\xi_{t+1}(\alpha))|$ with respect to α and applying Equations (S1.8), (S1.11), we can show that,

$$\begin{aligned}\log|M(\xi_{t+1}(\alpha))| &= \log|M(\xi_t)| + \phi(\boldsymbol{x}_t^*, \xi_t)\alpha + \frac{1}{2}\alpha^2 \left.\frac{d^2\log|M(\xi_{t+1}(\alpha))|}{d\alpha^2}\right|_{\alpha=\alpha'} \\ &\ge \log|M(\xi_t)| + \zeta\alpha - \frac{1}{2}K\alpha^2,\end{aligned}$$

where $\alpha' \in (0, \alpha)$. Combining Equation (S1.9), the following equation holds for any $0 \le \alpha \le 1/2$,

$$\log|M(\xi_{t+1})| - \log|M(\xi_t)| \ge \zeta\alpha - \frac{1}{2}K\alpha^2.$$

Now we consider the following two situations.

- If $K > 2\zeta$, let $\alpha = \frac{\zeta}{K}$, then

$$\log|M(\xi_{t+1})| - \log|M(\xi_t)| \geq \frac{\zeta^2}{2K}.$$

- If $K \leq 2\zeta$, let $\alpha = \frac{1}{2}$, then

$$\log|M(\xi_{t+1})| - \log|M(\xi_t)| \geq \frac{1}{2}\zeta - \frac{1}{8}K \geq \frac{1}{4}\zeta.$$

Note that ζ and K are finite. The two cases imply $\lim_{t\to\infty} \log|M(\xi_t)| = \infty$, which leads a contradiction. Thus, the sequence of designs $\{\xi_t\}$ converge to an optimal design that maximizes $|M(\xi)|$ as $t \to \infty$. □

Proof of Theorem 6. In this case, $H(\boldsymbol{x}) = \text{diag}\{\boldsymbol{h}_1^T(\boldsymbol{x}), \ldots, h_{J-1}^T(\boldsymbol{x})\}$ is a $(J-1) \times p_1(J-1)$ matrix. $\widetilde{H} = (H^T(\boldsymbol{x}_1), \ldots, H^T(\boldsymbol{x}_{p_1}))$ is a $p_1(J-1) \times p_1(J-1)$ matrix. For any design

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_{p_1} \\ \omega_1 & \cdots & \omega_{p_1} \end{pmatrix},$$

let $\widetilde{W} = \text{diag}\left(\omega_1 G^T(\boldsymbol{x}_1)D^{-1}(\boldsymbol{x}_1)G(\boldsymbol{x}_1), \ldots, \omega_{p_1} G^T(\boldsymbol{x}_{p_1})D^{-1}(\boldsymbol{x}_{p_1})G(\boldsymbol{x}_{p_1})\right)$. Then the determinant of $M(\xi)$ is

$$\begin{aligned}
|M(\xi)| &= |\widetilde{H}\widetilde{W}\widetilde{H}^T| \\
&= |\widetilde{H}|^2 \cdot |\widetilde{W}| \\
&= |\widetilde{H}|^2 \left(\prod_{i=1}^{p_1} \left| G^T(\boldsymbol{x}_i)D^{-1}(\boldsymbol{x}_i)G(\boldsymbol{x}_i) \right| \right) \left(\prod_{i=1}^{p_i} \omega_i \right)^{J-1}.
\end{aligned}$$

Maximizing the above expression with respect to the weights $\omega_1, \ldots, \omega_{p_1}$ under the condition $\sum_{i=1}^{p_1} \omega_i = 1$ gives $\omega_i = 1/p_1$ for all $i = 1, \ldots, p_1$, which proves this theorem. $\square$

Proof of Theorem 7. For Model (4.1),

$$H(x_i) = \begin{pmatrix} x_i & 1 & 0 \\ x_i & 0 & 1 \end{pmatrix}, G(x_i) = \begin{pmatrix} \bar{g}_{i1} & 0 \\ -\frac{\pi_{i2}}{\pi_{i2}+\pi_{i3}}\bar{g}_{i1} & (\pi_{i2}+\pi_{i3})\bar{g}_{i2} \\ -\frac{\pi_{i3}}{\pi_{i2}+\pi_{i3}}\bar{g}_{i1} & -(\pi_{i2}+\pi_{i3})\bar{g}_{i2} \end{pmatrix},$$

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

$$H^T(x_i)G^T(x_i)D^{-1}(x_i)G(x_i)H(x_i) = \begin{pmatrix} (s_i + t_i)x_i^2 & s_i x_i & t_i x_i \\ s_i x_i & s_i & 0 \\ t_i x_i & 0 & t_i \end{pmatrix},$$

where $s_i = \bar{g}_{i1}^2 \pi_{i1}^{-1}(\pi_{i2}+\pi_{i3})^{-1}$, and $t_i = (\pi_{i2}+\pi_{i3})^3 \bar{g}_{i2}^2 (\pi_{i2}\pi_{i3})^{-1}$, for $i = 1, 2$. The determinant of the Fisher information matrix can be derived as follows,

$$|M(\xi)| = \omega_1\omega_2(c_1\omega_1 + c_2\omega_2),$$

where $c_1 = (x_1 - x_2)^2 s_1 t_1 (s_2 + t_2), c_2 = (x_1 - x_2)^2 s_2 t_2 (s_1 + t_1)$. Using the facts in Corollary 2 of Yang et al. (2017), the theorem is proved. $\square$

Proof of Theorem 8. For Model (4.2), the matrices $H(x_i)$ and $G(x_i)$ have the following formula,

$$H(x_i) = \begin{pmatrix} 1 & x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_i \end{pmatrix}, G(x_i) = \begin{pmatrix} \bar{g}_{i1} & 0 \\ -\frac{\pi_{i2}}{\pi_{i2}+\pi_{i3}}\bar{g}_{i1} & (\pi_{i2}+\pi_{i3})\bar{g}_{i2} \\ -\frac{\pi_{i3}}{\pi_{i2}+\pi_{i3}}\bar{g}_{i1} & -(\pi_{i2}+\pi_{i3})\bar{g}_{i2} \end{pmatrix},$$

where $\bar{g}_{i1} = (g^{-1})'(\theta_{11}+\theta_{12}x_i+\theta_{13}x_i^2)$, $\bar{g}_{i2} = (g^{-1})'(\theta_{21}+\theta_{22}x_i)$, for $i = 1, 2, 3$. Directly calculations yield that,

$$H^T(x_i)G^T(x_i)D^{-1}(x_i)G(x_i)H(x_i) = \begin{pmatrix} s_i & s_ix_i & s_ix_i^2 & 0 & 0 \\ s_ix_i & s_ix_i^2 & s_ix_i^3 & 0 & 0 \\ s_ix_i^2 & s_ix_i^3 & s_ix_i^4 & 0 & 0 \\ 0 & 0 & 0 & t_i & t_ix_i \\ 0 & 0 & 0 & t_ix_i & t_ix_i^2 \end{pmatrix},$$

where $s_i = \bar{g}_{i1}^2\pi_{i1}^{-1}(\pi_{i2}+\pi_{i3})^{-1}$, and $t_i = (\pi_{i2}+\pi_{i3})^3\bar{g}_{i2}^2(\pi_{i2}\pi_{i3})^{-1}$, for $i = 1, 2, 3$. The determinant of the Fisher information matrix can be derived as follows,

$$|M(\xi)| = C\omega_1\omega_2\omega_3(c_1\omega_1\omega_2 + c_2\omega_1\omega_3 + c_3\omega_1\omega_2),$$

where $C = s_1s_2s_3(x_1-x_2)^2(x_1-x_3)^2(x_2-x_3)^2, c_1 = t_1t_2(x_1-x_2)^2, c_2 = t_1t_3(x_1-x_3)^2, c_3 = t_2t_3(x_2-x_3)^2$. This theorem follows by Lemma S.4 in Section S.13 and its proof in Section S.15 of Bu et al. (2020). □

S2 Additional simulations results

Example S1. In this example, we demonstrate the optimal design searched out by our method. For comparison, we also report the results for the D-optimal design constructed in Bu et al. (2020) via grid-points. All the simulation settings are the same as Example 4 in the main text. For clear transparency, we drop out the points with zero weights in the following.

$$\xi_{BMY,4} = \begin{pmatrix} 0 & 66.7 & 133.3 \\ 0.206 & 0.394 & 0.400 \end{pmatrix},$$

$$\xi_{BMY,6} = \begin{pmatrix} 0 & 80.0 & 120.0 & 160.0 \\ 0.202 & 0.100 & 0.336 & 0.362 \end{pmatrix},$$

$$\xi_{BMY,10} = \begin{pmatrix} 0 & 111.1 & 155.6 \\ 0.203 & 0.398 & 0.399 \end{pmatrix},$$

$$\xi_{BMY,20} = \begin{pmatrix} 0 & 105.3 & 147.4 \\ 0.203 & 0.398 & 0.399 \end{pmatrix},$$

$$\xi_{BMY,50} = \begin{pmatrix} 0 & 102.0 & 106.1 & 146.9 & 151.0 \\ 0.203 & 0.278 & 0.120 & 0.184 & 0.215 \end{pmatrix},$$

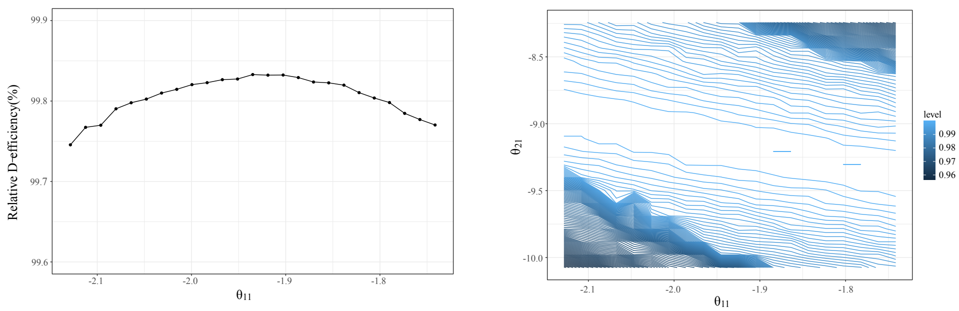
$$\xi^* = \begin{pmatrix} 0 & 101.1 & 147.8 & 149.3 \\ 0.203 & 0.397 & 0.307 & 0.093 \end{pmatrix}.$$

One can see that $\xi_{BMY,4}$, $\xi_{BMY,10}$, and $\xi_{BMY,20}$ have only three support points, which are minimally supported. While $\xi_{BMY,6}$ and $\xi_{BMY,50}$ have 4 and 5 support points, respectively. Note that the optimal design ξ^* has less support points compared with the $\xi_{BMY,50}$, which is of practical significance due to the cost of changing settings.

Example S2. Consider the situation where the pre-specified value of the parameter vector is moderately misspecified. Since all the cases have similar performance, we report the results of Model (2.5) as an example. Suppose the pre-specified value of the parameter vector for the locally optimal design fluctuates in a moderate range ($\pm 10\%$ the magnitude of the true value).

For visualization propose, we report the results for the case that only one of the five parameters is misspecified (we choose θ_{11} as an example). The results are presented in Figure S1(a). Figure S1(a) shows the relative D-efficiencies for the locally D-optimal designs under the misspecified parameter θ_{11}. Clearly, these D-optimal designs under misspecified parameters have relative D-efficiencies greater than 99.97%. When there are two parameters misspecified (we choose θ_{11}, θ_{21} as an example), we plot a contour plot in Figure S1(b). From Figure S1(b), one can see that the relative D-efficiencies are also greater than 95.0%.

To give a comprehensive result, we also consider the case that all the five parameters are misspecified, via the grid-point method. The results



(a) θ_{11} is misspecified

(b) both θ_{11} and θ_{21} are misspecified

Figure S1: Relative D-efficiencies when the parameters are misspecified.

are summarized in Table S1. The minimum efficiency is 63.6%, which is close to the efficiency of uniform design consider in Example 4. On the other hand, the 1st quartile is 94.3%, which indicates the D-optimal designs with moderately misspecified parameters are quite robust and still have satisfactory performances.

Table S1: Summary of relative D-efficiencies when all the five parameters are misspecified.

Min.	1st Quartile	Median	3rd Quartile	Max.
63.6%	94.3 %	97.6%	99.4%	100.0%

References

Bu, X., D. Majumdar, and J. Yang (2020). D-optimal designs for multinomial logistic models. *Annals of Statistics 48*(2), 983–1000.

Danninger-Uchida, G. E. (2009). Carathéodory theorem. In P. M. Floudas, Christodoulos A.and Pardalos (Ed.), *Encyclopedia of Optimization*, pp. 358–359. Springer US.

Horn, R. A. and C. R. Johnson (2012). *Matrix analysis* (2nd ed.). Cambridge university press.

Pukelsheim, F. (2006). *Optimal design of experiments*. Society for Industrial and Applied Mathematics.

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