Supplementary Material

**LOCALLY $D$-OPTIMAL DESIGNS**

**FOR HIERARCHICAL RESPONSE EXPERIMENTS**

Mingyao Ai$^1$, Zhiqiang Ye$^1$, Jun Yu$^2$

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

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

**Supplementary Material**

**S1 Proofs**

To proof Theorem 1 we begin with two technical lemmas. Let $\gamma \in \mathbb{R}^p$ denote the parameter vector, i.e., $\gamma = (\gamma_1, \ldots, \gamma_p)^T = (\beta^T, \theta_1^T, \ldots, \theta_{J-1}^T)^T$.

The first lemma gives the Fisher information matrix for Model (2.2) under an exact design. The second lemma calculates $\partial \pi(x) / \partial \gamma^T$, which is an essential part of Theorem 1.

For an exact design

$$\xi_{\text{exact}} = \begin{pmatrix} x_1 & \cdots & x_m \\ n_1 & \cdots & n_m \end{pmatrix},$$

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 $\xi_{exact}$ can be written as

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

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

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

**Proof of Lemma S1** For the experimental setting $x_i$, for $i = 1, \ldots, m$, the responses $(Y_{1i}, \ldots, Y_{Ji})^T \sim \text{Multinomial}(n_i; \pi_{i1}, \ldots, \pi_{iJ})$. 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

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

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

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

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

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

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

\[ \sum_{i=1}^{m} E \left( \sum_{j=1}^{J} \frac{Y_{ij}^2 \partial \pi_{ij} \partial \pi_{ij}}{\pi_{ij}^2 \partial \gamma_s \partial \gamma_t} + 2 \sum_{1 \leq j < k \leq m} \frac{Y_{ij} \partial \pi_{ij} Y_{ik} \partial \pi_{ik}}{\pi_{ij} \partial \gamma_s \pi_{ik} \partial \gamma_t} \right) \]

\[ \sum_{i=1}^{m} \sum_{j=1}^{J} n_i(n_i - 1) \pi_{ij}^2 + n_i \pi_{ij} \partial \pi_{ij} \partial \pi_{ij} \]

\[ + 2 \sum_{i=1}^{m} \sum_{1 \leq j < k \leq m} n_i(n_i - 1) \pi_{ij} \pi_{ik} \partial \pi_{ij} \partial \pi_{ik} \]

\[ = \sum_{i=1}^{m} n_i \left( n_i - 1 \right) \left( \sum_{j=1}^{J} \partial \pi_{ij} \right) \left( \sum_{j=1}^{J} \partial \pi_{ij} \right) + \sum_{j=1}^{J} \frac{1}{\pi_{ij} \partial \gamma_s \partial \gamma_t} \]

By the definition of Fisher information matrix, we have

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

where \( M_i = (m_{ist})_{1 \leq s, t \leq p} \) is a \( p \times p \) matrix with

\[ m_{ist} = \sum_{j=1}^{J} \frac{1}{\pi_{ij} \partial \gamma_s \partial \gamma_t}. \]

\[ \square \]

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

\[ \xi = \begin{pmatrix} x_1 & \cdots & x_m \\ \omega_1 & \cdots & \omega_m \end{pmatrix}, \]
can be written as

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

Recall \( \delta_x \) denote the single point design, then \( M(\delta_{x_i}) = M_i, \) for \( i = 1, \ldots, m. \)

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

**Lemma S2.** For Model (2.2),

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

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

**Proof of Lemma S2.** To be convenience, let \( e_j(x) = g^{-1}(h_0^T(x)\beta + h_j^T(x)\theta_j), \)
for \( j = 1, \ldots, J-1. \) We first show the following equation

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

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

\[ \frac{\partial \pi_1(x)}{\partial \beta_s} = h_{0s}(x)g_{11}(x), \]
by the fact $\pi_1(x) = e_1(x)$, which implies Equation (S1.2) holds for $j = 1$.

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

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

we have

$$
\frac{\partial \pi_j(x)}{\partial \beta_s} = \frac{\partial e_j(x)}{\partial \beta_s} \left(1 - \sum_{k=1}^{j-1} \pi_k(x)\right) - e_j(x) \sum_{k=1}^{j-1} \frac{\partial \pi_k(x)}{\partial \beta_s}
$$

$$
= \pi_0(x) (g^{-1})(h_0^T(x)\beta + h_j^T(x)\theta_j) \left(1 - \sum_{k=1}^{j-1} \pi_k(x)\right)
$$

$$
- e_j(x) \sum_{k=1}^{j-1} \left(h_0(x) \sum_{l=1}^k g_{kl}(x)\right)
$$

$$
= \pi_0(x) g_{jj}(x) + \pi_0(x) \sum_{k=1}^{j-1} \left(-e_j(x) \sum_{l=1}^k g_{kl}(x)\right)
$$

$$
= \pi_0(x) g_{jj}(x) + \pi_0(x) \sum_{k=1}^{j-1} \left(-e_j(x) \sum_{l=1}^{j-1} g_{kl}(x)\right)
$$

$$
= \pi_0(x) g_{jj}(x) + \pi_0(x) \sum_{l=1}^{j-1} g_{jl}(x)
$$

$$
= \pi_0(x) \sum_{l=1}^{j} g_{jl}(x),
$$

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

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

\[
\frac{\partial \pi_j(x)}{\partial \beta_s} = -\sum_{j=1}^{J-1} \frac{\partial \pi_j(x)}{\partial \beta_s} = -\sum_{j=1}^{J-1} \left( h_{0s}(x) \sum_{k=1}^{j} g_{jk}(x) \right) = h_{0s}(x) \sum_{j=1}^{J-1} \left( -\sum_{k=1}^{J-1} g_{jk}(x) \right) = h_{0s}(x) \sum_{k=1}^{J-1} g_{Jk}(x),
\]
	hen 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,

\[
\frac{\partial \pi_j(x)}{\partial \theta_{uv}} = h_{uv}(x) g_{ju}(x), \quad (S1.3)
\]

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 \( \pi_1(x) = e_1(x) \), we have

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

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

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

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

$$
\frac{\partial \pi_j(x)}{\partial \theta_{jv}} = \frac{\partial e_j(x)}{\partial \theta_{jv}} \left(1 - \sum_{k=1}^{j-1} \pi_k(x)\right) - e_j(x) \sum_{k=1}^{j-1} \frac{\partial \pi_k(x)}{\partial \theta_{jv}} \\
= h_{jv}(x) (g^{-1})'(h_0^T(x)\beta + h_j^T(x)\theta_j) \left(1 - \sum_{k=1}^{j-1} \pi_k(x)\right) \\
= h_{jv}(x) g_{jj}(x).
$$

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

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

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

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

\begin{center}
7
\end{center}
When $j = J$, utilizing the fact $\pi_1(x) + \cdots + \pi_J(x) = 1$ and the facts that have been proved in (a) and (b), we have

$$\frac{\partial \pi_j(x)}{\partial \pi_{uv}} = -\sum_{j=1}^{J-1} \frac{\partial \pi_j(x)}{\partial \pi_{uv}}$$

$$= -\sum_{j=1}^{J-1} h_{uv}(x) g_{ju}(x)$$

$$= h_{uv}(x) \left( -\sum_{j=1}^{J-1} g_{ju}(x) \right)$$

$$= h_{uv}(x) g_{Ju}(x),$$

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. \qed

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

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

$$= \sum_{i=1}^{m} \omega_i \left( \frac{\partial \pi(x_i)}{\partial \gamma^T} \right)^T D^{-1}(x_i) \left( \frac{\partial \pi(x_i)}{\partial \gamma^T} \right)$$

$$= \sum_{i=1}^{m} \omega_i H^T(x_i) G^T(x_i) D^{-1}(x_i) G(x_i) H(x_i),$$

which completes the proof. \qed
Proof of Theorem 2. Let $\tilde{H} = (H^T(x_1), \ldots, H^T(x_m))$, and

$$\tilde{W} = \text{diag}(\omega_1 G^T(x_1)D^{-1}(x_1)G(x_1), \ldots, \omega_m G^T(x_m)D^{-1}(x_m)G(x_m)).$$

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

Proof of Corollary 1. After some elementary column transformations for the matrix $(H^T(x_1), \ldots, H^T(x_m))$, we obtain a new matrix

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

In order to keep $H_{\text{new}}$ 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 $\cap_{j=0}^{l-1}C(H_j^T) \neq \{0\}$, without loss of generality, we assume that the first row of $H_0$ lies in $\cap_{j=1}^{l-1}C(H_j^T)$. 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 $r = \dim \left( \bigcap_{j=1}^{J} C(H_j^T) \right)$, utilizing the fact that $\bigcap_{j=0}^{J-1} C(H_j^T) = \{0\}$, the rank of the matrix $(H_0^T, \ldots, H_{J-1}^T)$ is at least $p_0+r$. Thus $m \geq p_0+r$. □

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

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

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 $M(\delta_{x_i}) = H^T(x_i)G^T(x_i)D^{-1}(x_i)G(x_i)H(x_i)$.

The rank of $M(\delta_{x_i})$ is less than or equal to the rank of $G(x_i)$, 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 $M_\tau$ which are the same with the corresponding rows of $M(\delta_{x_1})$, then $|M_\tau| = 0$, which implies $c_{\alpha_1, \ldots, \alpha_m} = 0$ according to Equation (3.1).

(2) For the second scenario, let $\bar{H} = (H^T(x_1)G^T(x_1), \ldots, H^T(x_m)G^T(x_m))$, and $W = \text{diag}(\omega_1D^{-1}(x_1), \ldots, \omega_mD^{-1}(x_m))$, then $M(\xi) = \bar{H}W\bar{H}^T$. 10
By Cauchy-Binet formula (Horn and Johnson [2012]), it follows

\[ c_{\alpha_1, \ldots, \alpha_m} = \sum_{(v_1, \ldots, v_p) \in \Lambda(\alpha_1, \ldots, \alpha_m)} |\tilde{H}[i_1, \ldots, i_p]|^2 \prod_{k: \alpha_k > 0} \prod_{(k-1)J < u \leq kJ} \pi_{k,v,-(k-1)J}, \]

where \( 1 \leq v_1 < \cdots < v_p \leq mJ \), \( \Lambda(\alpha_1, \ldots, \alpha_m) \) only depends on \( \alpha_1, \ldots, \alpha_m \), and \( \tilde{H}[i_1, \ldots, i_p] \) is the submatrix consisting of the \( i_1 \)th, \( \ldots, i_p \)th rows of \( \tilde{H} \). Without loss of generality, we assume \( \alpha_1 \geq \cdots \geq \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m \), where \( k + 1 \leq \max\{p_0 + r, p_1, \ldots, p_{J-1}\} \).

Suppose \( c_{\alpha_1, \ldots, \alpha_m} \neq 0 \) for some \( (\alpha_1, \ldots, \alpha_m) \). Therefore, there exist \( (v_1, \ldots, v_p) \) such that \( \tilde{H}[v_1, \ldots, v_p] \) has full rank \( p \), and \( 1 \leq v_1 < \cdots < v_p \leq kJ \). Then \( \tilde{H} = \tilde{H}[1, \ldots, kJ] \) is full row rank. Let \( \tilde{W} = k^{-1} \text{diag}(D^{-1}(x_1), \ldots, D^{-1}(x_k)) \). \( \tilde{H}\tilde{W}\tilde{H}^T \) is positive definite. On the other hand, we can regard \( \tilde{H}\tilde{W}\tilde{H}^T \) as the Fisher information matrix under uniform weighted design on the \( k \) support points, thus \( k \geq \max\{p_0 + r, p_1, \ldots, p_{J-1}\} \), which is a contradiction.

\[ \square \]

**Proof of Theorem 4** Note that maximizing \( |M(\xi)| \) is equivalent to maximizing \( \log |M(\xi)| \). Recall \( \delta_x \) denote the single point design. The Frechet derivate of \( \log |M(\xi)| \) at \( \xi^* \) in the direction of \( \delta_x - \xi^* \) is

\[ \lim_{\alpha \to 0} \frac{1}{\alpha} (\log |M((1 - \alpha)\xi^* + \alpha\delta_x)| - \log |M(\xi^*)|) \]
\[ = \lim_{\alpha \to 0} \frac{1}{\alpha} (\log |M(\xi^*) + \alpha(M(\delta_x) - M(\xi^*))| - \log |M(\xi^*)|) \]
\[ = \text{tr} \left( M^{-1}(\xi^*)(M(\delta_x) - M(\xi^*)) \right) \]
\[ = \text{tr} \left( M^{-1}(\xi^*)M(\delta_x) \right) - p \]
\[ = \text{tr} \left( M^{-1}(\xi^*)H^T(x)G^T(x)D^{-1}(x)G(x)H(x) \right) - p. \]

Then the theorem is proved following Pukelsheim (2006).

**Proof of Theorem**

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 \( \xi^* \) 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

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

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

\[ \log |M(\xi^*)| - \log |M(\xi_t)| > \zeta. \quad (S1.5) \]
Utilizing the concavity of $\log |M(\xi)|$, we have

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

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

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

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

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

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

$$\phi(x^*_t, \xi_t) \geq \int_{x \in \chi} \phi(x, \xi_t)\xi^*(dx) = \text{tr}(M^{-1}(\xi_t)(M(\xi^*) - M(\xi_t))).$$

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

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

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

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

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

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

Equation (S1.10) implies that $\log |M(\xi_{t+1}(\alpha))|$ is uniformly bounded for $0 \leq \alpha \leq \frac{1}{2}$ and $t \in \mathbb{N}^*$. By Theorem 3, $|M(\xi_{t+1}(\alpha))|$ is a polynomial of $\alpha$, which implies that $\log |M(\xi_{t+1}(\alpha))|$ is infinitely differentiable with respect to $\alpha$. 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,

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

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

$$\log |M(\xi_{t+1}(\alpha))| = \log |M(\xi_t)| + \phi(\mathbf{x}_t^*, \xi_t) \alpha + \frac{1}{2} \alpha^2 \frac{d^2 \log |M(\xi_{t+1}(\alpha))|}{d\alpha^2} \bigg|_{\alpha = \alpha'} \geq \log |M(\xi_t)| + \zeta \alpha - \frac{1}{2} K \alpha^2,$$

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

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

Now we consider the following two situations.
Supplementary Material

• 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 $\zeta$ 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 $t \to \infty$. \hfill \Box

**Proof of Theorem 6.** In this case, $H(x) = \text{diag}\{h_1^T(x), \ldots, h_{J-1}^T(x)\}$ is a $(J - 1) \times p_1(J - 1)$ matrix. $\tilde{H} = (H^T(x_1), \ldots, H^T(x_{p_1}))$ is a $p_1(J - 1) \times p_1(J - 1)$ matrix. For any design

$$ \xi = \begin{pmatrix} x_1 & \cdots & x_{p_1} \\ \omega_1 & \cdots & \omega_{p_1} \end{pmatrix}, $$

let $\tilde{W} = \text{diag} (\omega_1 G^T(x_1)D^{-1}(x_1)G(x_1), \ldots, \omega_{p_1} G^T(x_{p_1})D^{-1}(x_{p_1})G(x_{p_1}))$. Then the determinant of $M(\xi)$ is

$$ |M(\xi)| = |\tilde{H}\tilde{W}\tilde{H}^T| $$

$$ = |\tilde{H}|^2 \cdot |\tilde{W}| $$

$$ = |\tilde{H}|^2 \left( \prod_{i=1}^{p_1} |G^T(x_i)D^{-1}(x_i)G(x_i)| \right) \left( \prod_{i=1}^{p_1} \omega_i \right)^{J-1}.$$
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),

\[ H(x_i) = \begin{pmatrix} x_i & 1 & 0 \\ x_i & 0 & 1 \end{pmatrix}, \quad G(x_i) = \begin{pmatrix} g_{i1} & 0 \\ -\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}} g_{i1} & (\pi_{i2} + \pi_{i3}) g_{i2} \\ -\frac{\pi_{i3}}{\pi_{i2} + \pi_{i3}} g_{i1} & -(\pi_{i2} + \pi_{i3}) g_{i2} \end{pmatrix}, \]

where \( g_{ij} = (g^{-1})'(\theta_j + \beta x_i) \), 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 \( s_i = g_{i1}^2 \pi_{i1}^{-1}(\pi_{i2} + \pi_{i3})^{-1} \), and \( t_i = (\pi_{i2} + \pi_{i3})^3 g_{i2}^2(\pi_{i2} \pi_{i3})^{-1} \), 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 \( c_1 = \pi_{i1}^2 x_i^2 s_1 t_1(s_2 + t_2), c_2 = (x_1 - x_2)^2 s_2 t_2(s_1 + t_1) \). 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,

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

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,

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

where \( s_i = \bar{g}_{i1}^2(\pi_{i2} + \pi_{i3})^{-1} \), 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 \( C = s_1 s_2 s_3(x_1 - x_2)^2(x_1 - x_3)^2(x_2 - x_3)^2 \), \( c_1 = t_1 t_2(x_1 - x_2)^2 \), \( c_2 = t_1 t_3(x_1 - x_3)^2 \), \( c_3 = t_2 t_3(x_2 - x_3)^2 \). This theorem follows by Lemma S.4 in Section S.13 and its proof in Section S.15 of [Bu et al. (2020)](https://example.com). □
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.

\[
\begin{align*}
\xi_{BMY, 4} & = \begin{pmatrix} 0 & 66.7 & 133.3 \\ 0.206 & 0.394 & 0.400 \end{pmatrix}, \\
\xi_{BMY, 6} & = \begin{pmatrix} 0 & 80.0 & 120.0 & 160.0 \\ 0.202 & 0.100 & 0.336 & 0.362 \end{pmatrix}, \\
\xi_{BMY, 10} & = \begin{pmatrix} 0 & 111.1 & 155.6 \\ 0.203 & 0.398 & 0.399 \end{pmatrix}, \\
\xi_{BMY, 20} & = \begin{pmatrix} 0 & 105.3 & 147.4 \\ 0.203 & 0.398 & 0.399 \end{pmatrix}, \\
\xi_{BMY, 50} & = \begin{pmatrix} 0 & 102.0 & 106.1 & 146.9 & 151.0 \\ 0.203 & 0.278 & 0.120 & 0.184 & 0.215 \end{pmatrix}, \\
\xi^* & = \begin{pmatrix} 0 & 101.1 & 147.8 & 149.3 \\ 0.203 & 0.397 & 0.307 & 0.093 \end{pmatrix}.
\end{align*}
\]

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 $\xi^*$ 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 $\theta_{11}$ 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 $\theta_{11}$. Clearly, these $D$-optimal designs under misspecified parameters have relative $D$-efficiencies greater than 99.97%. When there are two parameters misspecified (we choose $\theta_{11}, \theta_{21}$ 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
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.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Quartile</th>
<th>Median</th>
<th>3rd Quartile</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>63.6%</td>
<td>94.3 %</td>
<td>97.6%</td>
<td>99.4%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
References


