Optimal Designs for Nonlinear Models with Random Block Effects

Xin Wang, Min Yang and Wei Zheng

University of Illinois at Chicago, University of Illinois at Chicago, and University of Tennessee

Supplementary Material

This material contains proof of Theorem 4.3, Theorem 4.4,and 4.5 from the main context.

S1 Appendix

Proof of Theorem 4.5 . We only give proof for $p > 0$. For $p = 0$, the proof is exactly the same with ϕ_p replaced by $-\log|M(\xi)|$. In Lemma 3.1, we proved

$$M(\alpha\xi_1 + (1-\alpha)\xi_2) \geq \alpha M(\xi_1) + (1-\alpha)M(\xi_2) \tag{S1.1}$$

By the monotonicity and convexity of $\Psi_p(M)$ (Fedorov and Hackl 1997, sec. 2.2), where $\Phi_p M) = (v^{-1}Tr(M^{-p}))^{1/p}$, we have

$$\begin{aligned}
\Phi_p(M((1-\epsilon)\xi_1 + \epsilon\xi_2)) &= \left(\frac{1}{v}Tr(M(\alpha\xi_1 + (1-\alpha)\xi_2)^{-p})\right)^{1/p} \\
&\leq \left(\frac{1}{v}Tr([\alpha M(\xi_1) + (1-\alpha)M(\xi_2)]^{-p})\right)^{1/p} \\
&\leq \alpha\left(\frac{1}{v}Tr(M(\xi_1)^{-p})\right)^{1/p} + (1-\alpha)\left(\frac{1}{v}Tr(M(\xi_2)^{-p})\right)^{1/p} \\
&= \alpha\Phi_p(M(\xi_1)) + (1-\alpha)\Phi_p(M(\xi_2))
\end{aligned} \tag{S1.2}$$

$\Phi_p(\xi)$ as function of ξ is convex.

Consider iteration t in the algorithm, since

$$\begin{aligned} x_t^* &= \arg\min_{x\in\mathcal{X}} \eta(\nu_x, \xi) \\ &= \arg\min_{x\in\mathcal{X}} \left.\frac{\partial \Phi_p(M((1-\alpha)\xi + \alpha\nu_x))}{\partial\alpha}\right|_{\alpha=0} \end{aligned} \tag{S1.3}$$

and by (S1.2) we have

$$\Phi_p(\tilde{\xi}_{\alpha,t}) \le \Phi_p(\xi_{S^{(t)}}), \forall\alpha \in [0,1] \tag{S1.4}$$

where $\tilde{\xi}_{\alpha,t} = (1-\alpha)\xi_{S^{(t)}} + \alpha\nu_{x_t^*}$, then $\Phi_p(\xi_{S^{(t+1)}}) \le \Phi_p(\tilde{\xi}_{\alpha,t})$ since $\xi_{S^{(t+1)}}$ is optimal with support set $S^{(t)} \cup x_t^*$.

Thus,

$$\Phi_p(\xi_{S^{(t+1)}}) \le \Phi_p(\xi_{S^{(t)}}) \le \Phi_p(\xi_{S^{(0)}}), \forall t \in \mathscr{N} \tag{S1.5}$$

Φ_p is a decreasing non-negative function of t, its convergence follows. Then we prove $\Phi_p(\xi_{S^{(t)}})$ actually converge to $\Phi_p(\xi^*)$.

Define $\Theta_1 = \{\Phi_p(\xi) \le 2\Phi_p(\xi_{S^{(0)}})\}$. It is obvious that $\xi_{S^{(t)}} \in \Theta_1, \forall t$, since Φ_p is decreasing in t. For any $\alpha \in [0, 1/2]$, $M(\tilde{\xi}_{t,\alpha}) \ge (1-\alpha)M(\xi_{S^{(t)}}) + \alpha M(\nu_{x_t^*}) \ge 0.5M(\xi_{S^{(t)}})$, thus $\Phi_p(\xi_{\alpha,t}) \le 2\Phi_p(\xi_{S^{(t)}}) \le 2\Phi_p(\xi_{S^{(0)}})$, $\xi_{\alpha,t} \in \Theta_1$. M_ξ is nonsingular for any $\xi \in \Theta_1$, thus $\Phi_p(\alpha\xi_1 + (1-\alpha)\xi_2)$ is infinitely differentiable with respect to α for any $\alpha \in [0, 1/2]$. So there exist $K < \infty$, such that

$$\sup\left\{\frac{\partial^2\Phi_p(\alpha\xi_1 + (1-\alpha)\xi_2)}{\partial\alpha^2} : \xi_1, \xi_2 \in \Theta_1, \alpha \in [0, 1/2]\right\} = K \tag{S1.6}$$

We shall show that

$$\lim_{t\to\infty} \Phi_p(\xi_{S^{(t)}}) = \Phi_p(\xi^*), \tag{S1.7}$$

where ξ^* is optimal design. Otherwise, by Φ_p's monotocity, $\exists\delta > 0$ such that $\Phi_p(\xi_{S^{(t)}}) - \Phi_p(\xi^*) > \delta, \forall t$. By (S1.2), $\forall\epsilon \in [0,1]$, we have $\Phi_p((1-\epsilon)\xi_{S^{(t)}} + \epsilon\xi^*) \le (1-\epsilon)\Phi_p(\xi_{S^{(t)}}) + \epsilon\Phi_p(\xi^*)$

$$
\begin{aligned}
d\Phi_p(\xi_{S(t)},\alpha,\xi^*) = & \left.\frac{\partial \Phi_p((1-\alpha)\xi_{S(t)}+\alpha\xi^*)}{\partial \alpha}\right|_{\alpha=0} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}(\Phi_p(\epsilon\xi^*+(1-\epsilon)\xi_{S(t)})-\Phi_p(\xi_{S(t)})) \\
&\leq \Phi_p(\xi^*)-\Phi_p(\xi_{S(t)}) < -\delta
\end{aligned} \tag{S1.8}
$$

By definition of x_t^*, we have $\eta(\nu_{x_t^*},\xi_{S(t)}) > \int \eta(x,\xi_{S(t)})\xi^* dx$ and thus

$$
d\Phi_p(\xi_{S(t)},\alpha,\nu_{x_t^*}) = \left.\frac{\partial \Phi_p((1-\alpha)\xi_{S(t)}+\alpha\nu_{x_t^*})}{\partial \alpha}\right|_{\alpha=0} \geq \delta \tag{S1.9}
$$

Expand $\Phi_p(\tilde{\xi}_{\alpha,t})$ to a Taylor series in α and apply (S1.6) and (S1.9), we can show that

$$
\begin{aligned}
&\Phi_p(\tilde{\xi}_{\alpha,t}) \\
=&\Phi_p(\xi_{S(t)}) - d\Phi_p(\xi_{S(t)},\alpha,\nu_{x_t^*})\alpha \\
&+\frac{1}{2}\alpha^2\left.\frac{\partial^2\Phi_p(\alpha\xi_1+(1-\alpha)\xi_2)}{\partial\alpha^2}\right|_{\alpha=\alpha'} \\
&\leq \Phi_p(\xi_{S(t)}) - \delta\alpha + \frac{1}{2}K\alpha^2
\end{aligned} \tag{S1.10}
$$

Let $\alpha = \frac{\delta}{K}$, by $\Phi_p(\xi_{S(t+1)}) \leq \Phi_p(\tilde{\xi}_{\alpha,t})$ we can derive that for all $t \geq 0$ we have

$$
\Phi_p(\xi_{S(t+1)}) - \Phi_p(\xi_{S(t)}) \leq -\delta^2/2K \tag{S1.11}
$$

which is contrary to $\Phi_p \geq 0$ if let $t \to \infty$. Similar arguments can be applied to the case when $K \leq \delta$, in which we let $\alpha = 1$. $\square$

Proof of Theorem 4.4. Under Model (2.5), the information matrix under within-group design ξ can be written as

$$
M_\xi = M_\xi(\theta) = \left(\frac{\partial F}{\partial \theta}\right)^T V^{-1}\frac{\partial F}{\partial \theta}
$$

where $V^{-1} = c_1 W - c_2 W J_n W$, with $W = diag\,(w_1, ...w_n)$. The covariance matrix for the maximum likelihood estimator of θ can be written as M_ξ^{-1}. Here we consider all designs such that M_ξ is nonsingular.

Let Ω_i be a matrix whose (i,i)th element is 1 and (n,n)th element is -1, and 0 otherwise. We have

$$
\begin{aligned}
V_{i-} &= \frac{\partial V^{-1}}{\partial w_i} = c_1\Omega_i - c_2\Omega_i J_n W - c_2 W J_n \Omega_i,\\
V_{ij-} &= \frac{\partial^2 V^{-1}}{\partial w_i \partial w_j} = -c_2\Omega_i J_n \Omega_j - c_2 \Omega_j J_n \Omega_i,\\
M_\xi^i &= \frac{\partial M_\xi(\theta)}{\partial w_i} = \left(\frac{\partial F}{\partial \theta}\right)^T \frac{\partial V^{-1}}{\partial w_i}\frac{\partial F}{\partial \theta} = \left(\frac{\partial F}{\partial \theta}\right)^T V_{i-}\frac{\partial F}{\partial \theta}, \text{ and}\\
M_\xi^{ij} &= \frac{\partial^2 M_\xi(\theta)}{\partial w_i \partial w_j} = \left(\frac{\partial F}{\partial \theta}\right)^T \frac{\partial^2 V^{-1}}{\partial w_i \partial w_j}\frac{\partial F}{\partial \theta} = \left(\frac{\partial F}{\partial \theta}\right)^T V_{ij-}\frac{\partial F}{\partial \theta}.
\end{aligned}
$$

Thus by Lemma 15.10.5 of Harville (1997), for $i = 1, ..., n-1$, we have

$$
\begin{aligned}
\frac{\partial \Sigma_\xi(\theta)}{\partial w_i} &= \frac{\partial M_\xi^{-1}(\theta)}{\partial w_i} = -M_\xi^{-1}\frac{\partial M_\xi(\theta)}{\partial w_i}M_\xi^{-1} = -M_\xi^{-1}M_\xi^i M_\xi^{-1} \text{ and}\\
\frac{\partial^2 \Sigma_\xi(\theta)}{\partial w_i \partial w_j} &= M_\xi^{-1}(M_\xi^j M_\xi^{-1} M_\xi^i - M_\xi^{ij} + M_\xi^i M_\xi^{-1} M_\xi^j)M_\xi^{-1}.
\end{aligned}
$$

Notice that

$$
\frac{\partial^2 Tr(\Sigma_\xi(\theta))}{\partial w_i \partial w_j} = Tr\left(\frac{\partial^2 \Sigma_\xi(\theta)}{\partial w_i \partial w_j}\right), \forall i, j = 1, ...n-1.
$$

Thus the (i,j)th element of $H(w)$, the Hessian matrix of $Tr\Sigma_\xi(\theta)$, can be written as

$$
\begin{aligned}
H(w)[i,j] =& Tr(M_\xi^{-1}(M_\xi^j M_\xi^{-1} M_\xi^i - M_\xi^{ij} + M_\xi^i M_\xi^{-1} M_\xi^j)M_\xi^{-1})\\
=& 2Tr(M_\xi^{-1}M_\xi^i M_\xi^{-1} M_\xi^j M_\xi^{-1}) + Tr(-M_\xi^{-1}M_\xi^{ij}M_\xi^{-1})\\
=& 2Tr(M_\xi^{-1}M_\xi^i M_\xi^{-1} M_\xi^j M_\xi^{-1})\\
&+ Tr(-M_\xi^{-1}\left(\frac{\partial F}{\partial \theta}\right)^T(-c_2\Omega_i J_n \Omega_j - c_2\Omega_j J_n \Omega_i)\frac{\partial F}{\partial \theta}M_\xi^{-1})\\
=& 2Tr(M_\xi^{-1}M_\xi^i M_\xi^{-1} M_\xi^j M_\xi^{-1}) + 2c_2 Tr(M_\xi^{-1}\left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n \Omega_j \frac{\partial F}{\partial \theta}M_\xi^{-1}).
\end{aligned}
$$

$H(w)$ can be written as $H(w) = H_1(w) + c_2 H_2(w)$, where $c_2 = \dfrac{\rho}{[1+(k-1)\rho](1-\rho)} > 0$. This means as long as both $H_1(w)$ and $H_2(w)$ are both nonnegative definite, $H(w)$ will be nonnegative definite.

Since M_ξ is nonnegative definite, its inverse M_ξ^{-1} is also nonnegative definite. Thus M_ξ^{-1} can be written as $M_\xi^{-1} = (M_\xi^{-1})^{1/2}(M_\xi^{-1})^{1/2}$, and similarly $J_n = J_n^{1/2} J_n^{1/2}$. Let $A_i = M_\xi^{-1} M_\xi^i (M_\xi^{-1})^{1/2}$. By Proposition 1 in the appendix of Stufken and Yang (2012), it follows that $H_1(w)$ is nonnegative definite. $H_2(w)$ can be proved to be nonnegative definite by the similar way. Thus $H(w) = H_1(w) + c_2 H_2(w)$ is nonnegative definite.

Therefore, $Tr(\Sigma_\xi(\theta))$ attains its minimum at any of the critical points or at the boundary.

□

Proof of Theorem 4.3. The covariance matrix for the maximum likelihood estimator of θ has the same format as that of Theorem 4.4. Here we also consider all designs such that M_ξ is nonsingular. It is equivalent to show that $\log|\Sigma_\xi(\theta)|$ is minimized at the critical points or at a point on the boundary. It suffices to show that the Hessian matrix of $\log|\Sigma_\xi(\theta)|$ is nonnegative definite. The (i,j)th entry of the Hessian matrix can be written as

$$\begin{aligned} H(w)_D[i,j] &= \frac{\partial^2 \log|\Sigma_\xi(\theta)|}{\partial w_i \partial w_j} \\ &= Tr\left(\Sigma_\xi^{-1}(\theta)\frac{\partial^2 \Sigma_\xi(\theta)}{\partial w_i \partial w_j} - \Sigma_\xi^{-1}(\theta)\frac{\partial \Sigma_\xi(\theta)}{\partial w_i}\Sigma_\xi^{-1}(\theta)\frac{\partial \Sigma_\xi(\theta)}{\partial w_j}\right). \end{aligned}$$

Similar to the proof of Theorem 4.4, we have

$$\begin{aligned} Tr\left(\Sigma_\xi^{-1}(\theta)\frac{\partial^2 \Sigma_\xi(\theta)}{\partial w_i \partial w_j}\right) =& Tr(\Sigma_\xi^{-1/2}(\theta) M_\xi^{-1} M_\xi^j M_\xi^{-1} M_\xi^i M_\xi^{-1} \Sigma_\xi^{-1/2}(\theta) \\ &+ \Sigma_\xi^{-1/2}(\theta) M_\xi^{-1} M_\xi^i M_\xi^{-1} M_\xi^j M_\xi^{-1} \Sigma_\xi^{-1/2}(\theta) \\ &+ 2c_2 \Sigma_\xi^{-1/2}(\theta) M_\xi^{-1} \left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n \Omega_j \left(\frac{\partial F}{\partial \theta}\right) M_\xi^{-1} \Sigma_\xi^{-1/2}(\theta)) \\ =& 2Tr\left(\Sigma_\xi^{-1/2}(\theta) M_\xi^{-1} M_\xi^j M_\xi^{-1} M_\xi^i M_\xi^{-1} \Sigma_\xi^{-1/2}(\theta)\right) \\ &+ 2c_2 Tr\left(\Sigma_\xi^{-1/2}(\theta) M_\xi^{-1} \left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n \Omega_j \left(\frac{\partial F}{\partial \theta}\right) M_\xi^{-1} \Sigma_\xi^{-1/2}(\theta)\right) \end{aligned}$$

and

$$Tr(\Sigma_\xi^{-1}(\theta)\frac{\partial \Sigma_\xi(\theta)}{\partial w_i}\Sigma_\xi^{-1}(\theta)\frac{\partial \Sigma_\xi(\theta)}{\partial w_j}) = Tr(\Sigma_\xi^{-1}(\theta)M_\xi^{-1}M_\xi^i M_\xi^{-1}\Sigma_\xi^{-1}(\theta)M_\xi^{-1}M_\xi^j M_\xi^{-1}$$

$$= Tr(\Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}M_\xi^j M_\xi^{-1}M_\xi^i M_\xi^{-1}\Sigma_\xi^{-1/2}(\theta)).$$

Therefore

$$\begin{aligned}
&\frac{\partial^2\left(\log|\Sigma_\xi(\theta)|\right)}{\partial w_i \partial w_j}\\
=&Tr\left(\Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}M_\xi^i M_\xi^{-1}M_\xi^j M_\xi^{-1}\Sigma_\xi^{-1/2}(\theta)\right)\\
&+ 2c_2 Tr\left(\Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}\left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n \Omega_j \left(\frac{\partial F}{\partial \theta}\right) M_\xi^{-1}\Sigma_\xi^{-1/2}(\theta)\right).
\end{aligned}$$

Let $H(w)_D = H(w)_{D1} + 2c_2 H(w)_{D2}$, where

$$\begin{aligned}
H(w)_{D1}[i,j] &= Tr\left(\Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}M_\xi^i M_\xi^{-1}M_\xi^j M_\xi^{-1}\Sigma_\xi^{-1/2}(\theta)\right) \text{ and}\\
H(w)_{D2}[i,j] &= Tr\left(\Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}\left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n \Omega_j \left(\frac{\partial F}{\partial \theta}\right) M_\xi^{-1}\Sigma_\xi^{-1/2}(\theta)\right).
\end{aligned} \tag{S1.12}$$

Let $A_i = \Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}M_\xi^i(M_\xi^{-1})^{1/2}$, by Proposition 1 in the appendix of Stufken and Yang (2012), it follows that $H(w)_{D1}$ is nonnegative definite. Similarly, let $A_i = \Sigma_\xi^{-1/2}(\theta)M_\xi^{-1}\left(\frac{\partial F}{\partial \theta}\right)^T \Omega_i J_n^{1/2}$, we can show that $H(w)_{D2}$ is also nonnegative definite. Thus $H(w)_D$ is nonnegative definite. Consequently, $|\Sigma_\xi(\theta)|$ is minimized at any of the critical points or at a point on the boundary. $\square$

Proof of Theorem 4.3 Remark. The (i,j)th entry of the corresponding Hessian matrix can be written as

$$\begin{aligned}
&\frac{\partial^2\left(\log|\Sigma_\xi(\eta(\theta))|\right)}{\partial w_i \partial w_j}\\
=&Tr\left(\Sigma^{-1/2}\left(\frac{\partial \eta}{\partial \theta}\right) M_\xi^- \left\{M_\xi^i (M_\xi^-)^{1/2} P^\perp\left((\partial\eta/\partial\theta)(M_\xi^-)^{1/2}\right)(M_\xi^-)^{1/2} M_\xi^j\right\} M_\xi^- \left(\frac{\partial \eta}{\partial \theta}\right)^T \Sigma^{-1/2}\right)\\
&+ Tr\left(\Sigma^{-1/2}\left(\frac{\partial \eta}{\partial \theta}\right) M_\xi^- M_\xi^i M_\xi^- M_\xi^j M_\xi \left(\frac{\partial \eta}{\partial \theta}\right)^T \Sigma^{-1/2}\right)
\end{aligned}$$

where $P^{\perp}\left((\partial\eta/\partial\theta)(M_\xi^-)^{1/2}\right) = I_n - (M_\xi^-)^{1/2}\left(\frac{\partial\eta}{\partial\theta}\right)^T \Sigma_\xi^{-1}(\eta(\theta))\left(\frac{\partial\eta}{\partial\theta}\right)(M_\xi^-)^{1/2}$ is projection matrix onto the complement of column space of $\frac{\partial\eta}{\partial\theta}(M_\xi^-)^{1/2}$.

Let

$$A_i = \Sigma_\xi(\theta)^{-1/2}\left(\frac{\partial\eta}{\partial\theta}\right) M_\xi^- M_\xi^i (M_\xi^-)^{1/2}$$

and

$$A_i = \Sigma_\xi(\theta)^{-1/2}\left(\frac{\partial\eta}{\partial\theta}\right) M_\xi^- M_\xi^i (M_\xi^-)^{1/2} P^{\perp}\left((\partial\eta/\partial\theta)(M_\xi^-)^{1/2}\right)$$

respectively, by Proposition 1 in the appendix of Stufken and Yang (2012), it follows that the first part and the second part of the Hessian matrix are nonnegative definite respectively. Thus the conclusion follows. $\square$