Web Appendix of “A Group-Specific Recommender System”

Xuan Bi, Annie Qu, Junhui Wang and Xiaotong Shen*

Proof of Lemma 1

By Theorem 2.1 of Ansley and Kohn (1994), each of (3) and (4) has a unique solution, and the back-fitting algorithms for (5) and (6) can be applied in (3), and (7) and (8) can be applied in (4). This guarantees convergence to the unique solution given any initial value. Therefore, Algorithm 1 is equivalent to minimizing (3) and (4) iteratively.

Note that minimizing (3) and (4) iteratively is a special case of Algorithm MBI (Chen et al., 2012) with two blocks. Therefore, following Theorem 3.1 of Chen et al. (2012), Algorithm 1 converges to a stationary point. This completes the proof.

Proof of Lemma 2

Since $\theta_{ui} = (\mathbf{p}_u + \mathbf{s}_{v_u})'(\mathbf{q}_i + \mathbf{t}_{j_i})$ is a quadratic function of $(\mathbf{p}'_u, \mathbf{q}'_i, \mathbf{s}'_{v_u}, \mathbf{t}'_{j_i})'$, and given that $\|(\mathbf{p}'_u, \mathbf{q}'_i, \mathbf{s}'_{v_u}, \mathbf{t}'_{j_i})'\|_\infty$ and $\|(\tilde{\mathbf{p}}'_u, \tilde{\mathbf{q}}'_i, \tilde{\mathbf{s}}'_{v_u}, \tilde{\mathbf{t}}'_{j_i})'\|_\infty$ are bounded by L , there exists a constant $C_1 \geq$

*Xuan Bi is Ph.D. student, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 (E-mail: xuanbi2@illinois.edu). Annie Qu is Changjiang Visiting Professor, Yunnan University, China, and Professor, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 (E-mail: anniequ@illinois.edu). Junhui Wang is Associate Professor, Department of Mathematics, City University of Hong Kong, Hong Kong, China (E-mail: junhwang@cityu.edu.hk). Xiaotong Shen is Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: xshen@umn.edu).

0, such that

$$\|\theta_{ui} - \tilde{\theta}_{ui}\|_2 \leq C_1 \|(\mathbf{p}'_u, \mathbf{q}'_i, \mathbf{s}'_{v_u}, \mathbf{t}'_{j_i})' - (\tilde{\mathbf{p}}'_u, \tilde{\mathbf{q}}'_i, \tilde{\mathbf{s}}'_{v_u}, \tilde{\mathbf{t}}'_{j_i})'\|_2.$$

Recall that $f_{ui}(r, \gamma) = f(r_{ui}|\theta_{ui})$, then based on Assumption 1:

$$\begin{aligned} h^2(\gamma, \tilde{\gamma}) &= \frac{1}{nm} \sum_{u=1}^n \sum_{i=1}^m \int |f_{ui}^{1/2}(r, \gamma) - f_{ui}^{1/2}(r, \tilde{\gamma})|^2 d\nu(r) \\ &\leq \frac{1}{nm} \sum_{u=1}^n \sum_{i=1}^m \bar{G}^2 C_1^2 \|(\mathbf{p}'_u, \mathbf{q}'_i, \mathbf{s}'_{v_u}, \mathbf{t}'_{j_i})' - (\tilde{\mathbf{p}}'_u, \tilde{\mathbf{q}}'_i, \tilde{\mathbf{s}}'_{v_u}, \tilde{\mathbf{t}}'_{j_i})'\|_2^2 \\ &\leq \frac{1}{nm} \bar{G}^2 C_1^2 (\|\mathbf{P} - \tilde{\mathbf{P}}\|_F^2 + \|\mathbf{Q} - \tilde{\mathbf{Q}}\|_F^2 + \|\mathbf{S}_c - \tilde{\mathbf{S}}_c\|_F^2 + \|\mathbf{T}_c - \tilde{\mathbf{T}}_c\|_F^2) \\ &\leq \frac{1}{nm} \bar{G}^2 C_1^2 \{\|\mathbf{P} - \tilde{\mathbf{P}}\|_F^2 + \|\mathbf{Q} - \tilde{\mathbf{Q}}\|_F^2 + (n+m)(\|\mathbf{S} - \tilde{\mathbf{S}}\|_F^2 + \|\mathbf{T} - \tilde{\mathbf{T}}\|_F^2)\} \\ &\leq \frac{n+m}{nm} \bar{G}^2 C_1^2 \|\gamma - \tilde{\gamma}\|_2^2. \end{aligned}$$

The second-to-last inequality results from the fact that

$$\begin{aligned} \|\mathbf{S}_c - \tilde{\mathbf{S}}_c\|_F^2 &= \sum_{v=1}^N |V_v| \cdot \|\mathbf{s}_v - \tilde{\mathbf{s}}_v\|_2^2 \\ &\leq \max_{v=1, \dots, N} \{|V_v|\} \|\mathbf{S} - \tilde{\mathbf{S}}\|_F^2 \\ &\leq (n+m) \|\mathbf{S} - \tilde{\mathbf{S}}\|_F^2, \end{aligned}$$

and similarly $\|\mathbf{T}_c - \tilde{\mathbf{T}}_c\|_F^2 \leq (n+m) \|\mathbf{T} - \tilde{\mathbf{T}}\|_F^2$. The last inequality results from the fact that

$$\|\gamma - \tilde{\gamma}\|_2^2 = \|\mathbf{P} - \tilde{\mathbf{P}}\|_F^2 + \|\mathbf{Q} - \tilde{\mathbf{Q}}\|_F^2 + \|\mathbf{S} - \tilde{\mathbf{S}}\|_F^2 + \|\mathbf{T} - \tilde{\mathbf{T}}\|_F^2.$$

Define $d_0 = \bar{G}C_1$, and the result then follows. This completes the proof.

Proof of Theorem 1

We first verify the condition of Lemma 2.1 of Ossiander (1987). Based on Lemma 2,

$$\begin{aligned}
& \left\{ \frac{1}{nm} \sum_{u=1}^n \sum_{i=1}^m E(\sup_{\hat{\gamma} \in B_d(\gamma)} |f_{ui}^{1/2}(r, \hat{\gamma}) - f_{ui}^{1/2}(r, \gamma)|^2) \right\}^{1/2} \\
&= \left\{ \frac{1}{nm} \sum_{u=1}^n \sum_{i=1}^m \int \sup_{\hat{\gamma} \in B_d(\gamma)} |f_{ui}^{1/2}(r, \hat{\gamma}) - f_{ui}^{1/2}(r, \gamma)|^2 d\nu(r) \right\}^{1/2} \\
&\leq \left\{ \frac{n+m}{nm} \bar{G}^2 C_1^2 \sup_{\hat{\gamma} \in B_d(\gamma)} \|\hat{\gamma} - \gamma\|_2^2 \right\}^{1/2} \\
&\leq \sqrt{\frac{n+m}{nm}} d_0 d \\
&:= g(d)
\end{aligned}$$

Hence for $u > 0$,

$$H^B(u, \mathcal{S}(k), \rho) \leq H(g^{-1}(u/2), \mathcal{S}(k), \rho),$$

where H^B is the metric entropy of $\mathcal{S}(k)$ with bracketing of $f^{1/2}$, H is the ordinary metric entropy of $\mathcal{S}(k)$, and ρ is the L_2 -norm.

Next we provide an upper bound for $H(g^{-1}(u/2), \mathcal{S}(k), \rho)$. Since $g^{-1}(u/2) = \frac{\sqrt{nm}}{2d_0\sqrt{n+m}}u$, $N \leq n$, $M \leq m$, and $\|\gamma\|_\infty \leq L$, we have

$$\begin{aligned}
0 &\leq H^B(u, \mathcal{S}(k), \rho) \\
&\leq H(g^{-1}(u/2), \mathcal{S}(k), \rho) \\
&\leq \log \left[\max \left\{ \left(\frac{L\sqrt{(n+m+N+M)K}}{\frac{\sqrt{nm}}{2d_0\sqrt{n+m}}u} \right)^{(n+m+N+M)K}, 1 \right\} \right] \\
&\leq \max \left\{ (n+m+N+M)K \log \left(\frac{2\sqrt{2K}d_0L(n+m)}{\sqrt{nm}u} \right), 0 \right\} \\
&= \max \left\{ (n+m+N+M)K \log \left(\frac{\sqrt{K}C(n+m)}{\sqrt{nm}u} \right), 0 \right\}
\end{aligned}$$

for $u \geq \epsilon_{|\Omega|}^2$ and $C = 2\sqrt{2}d_0L$.

We now find the convergence rate $\epsilon_{|\Omega|}$, which is the smallest ϵ that satisfies the conditions of Theorem 1 of Shen (1998). That is,

$$\sup_{k \geq k_0} \psi_1(\epsilon, k) \leq c_2 |\Omega|^{1/2}$$

for a constant k_0 , where $\psi_1(\epsilon, k) = \int_x^{x^{1/2}} \{H^B(u, \mathcal{F}(k))\}^{1/2} du/x$ with $x = (c_1 \epsilon^2 + \lambda_{|\Omega|}(k - k_0))$, and $\mathcal{F}(k) = \{f^{1/2}(r, \gamma) : \gamma \in \mathcal{S}(k)\}$.

Note that $\psi_1 \leq 0 \leq c_2 |\Omega|^{1/2}$ when $x \geq 1$, so we only consider the case when $0 < x < 1$. Notice that $K \geq 1$ and $n + m \geq \sqrt{nm}$. Then with a sufficiently large L , we have $\max \left\{ \log \left(\frac{\sqrt{K}C(n+m)}{\sqrt{nm}u} \right), 0 \right\} = \log \left(\frac{\sqrt{K}C(n+m)}{\sqrt{nm}u} \right)$ for $u \in [x, x^{1/2}]$. Then:

$$\begin{aligned} \psi_1(\epsilon, k) &= \int_x^{x^{1/2}} \{H^B(u, \mathcal{F}(k))\}^{1/2} du/x \\ &\leq ((n + m + N + M)K)^{1/2} \int_x^{x^{1/2}} \left\{ \log \left(\frac{\sqrt{K}C(n+m)}{\sqrt{nm}} \right) - \log u \right\}^{1/2} du/x \\ &\leq ((n + m + N + M)K)^{1/2} (x^{-1/2} - 1) \left\{ \log \left(\frac{\sqrt{K}C(n+m)}{\sqrt{nm}} \right) + \log(x^{-1}) \right\}^{1/2}. \end{aligned}$$

Since $\lambda_{|\Omega|} < \frac{1}{2k} \epsilon_{|\Omega|}^2$ and $k \sim O(\sqrt{(n + m + N + M)K})$, we have $\lambda_{|\Omega|} = o(\epsilon_{|\Omega|}^2)$. Therefore, we solve

$$\begin{aligned} \sup_{k \geq k_0} \psi_1(\epsilon, k) &= \psi_1(\epsilon, k_0) \\ &\sim \sqrt{(n + m + N + M)K} \frac{1}{\epsilon_{|\Omega|}} \left\{ \log \left(\frac{\sqrt{K}(n+m)}{\epsilon_{|\Omega|}^2 \sqrt{nm}} \right) \right\}^{1/2} \\ &= c_2 |\Omega|^{1/2}. \end{aligned}$$

Then the smallest rate $\epsilon_{|\Omega|}$ is determined by

$$\frac{1}{\epsilon_{|\Omega|}} \left\{ \log \left(\frac{\sqrt{K}(n+m)}{\epsilon_{|\Omega|}^2 \sqrt{nm}} \right) \right\}^{1/2} \sim \frac{|\Omega|^{1/2}}{\sqrt{(n + m + N + M)K}}.$$

Note that $N \leq n$ and $M \leq m$, then we have

$$\epsilon_{|\Omega|} \sim \frac{\sqrt{(n+m)K}}{|\Omega|^{1/2}} \left\{ \log \left(\frac{|\Omega|}{\sqrt{nmK}} \right) \right\}^{1/2}.$$

For $\epsilon_{|\Omega|}$ and $\lambda_{|\Omega|}$, the conditions of Corollary 1 of Shen (1998) are now satisfied. The result then follows.

This completes the proof.

Proof of Theorem 2

Based on Theorem 1 and Theorem 1 of Shen (1998), there exists $c_i > 0$, $i = 1, 2$, such that:

$$P^* \left(\sup_{\{\tilde{\gamma} \in \mathcal{S}(k), h(\gamma, \tilde{\gamma}) \geq \epsilon\}} \{\mathcal{L}(\gamma|R^o) - \mathcal{L}(\tilde{\gamma}|R^o)\} \geq -c_1|\Omega|\epsilon^2 \right) \leq 7 \exp(-c_2|\Omega|\epsilon^2).$$

Therefore, if there exists $\tilde{\gamma} \in \mathcal{S}(k)$ such that $h(\gamma, \tilde{\gamma}) \geq \epsilon$, then

$$P^* \left(\{\mathcal{L}(\gamma|R^o) - \mathcal{L}(\tilde{\gamma}|R^o)\} \geq -c_1|\Omega|\epsilon^2 \right) \leq 7 \exp(-c_2|\Omega|\epsilon^2).$$

We suppress the subscript, write $h_{\Theta}(\theta_{ui}, \tilde{\theta}_{ui})$ as $h_{\Theta}(\theta, \tilde{\theta})$, and write $f(r_{ui}|\theta_{ui})$ as $f(r|\theta)$.

We now lower-bound $h(\gamma, \tilde{\gamma})$ by a function of $|\theta_{ui} - \tilde{\theta}_{ui}|$:

$$\begin{aligned} h_{\Theta}^2(\theta, \tilde{\theta}) &= E \left\{ f^{1/2}(r|\theta) - f^{1/2}(r|\tilde{\theta}) \right\}^2 \\ &= \left(\int_{\{f(r|\theta) > f(r|\tilde{\theta})\}} + \int_{\{f(r|\tilde{\theta}) \geq f(r|\theta)\}} \right) \left\{ f^{1/2}(r|\theta) - f^{1/2}(r|\tilde{\theta}) \right\}^2 d\nu(r) \\ &:= I_1 + I_2, \end{aligned}$$

where $I_1 = \int_{\{f(r|\theta) > f(r|\tilde{\theta})\}} f(r|\theta) \left(1 - \exp \left[\frac{1}{2} \left\{ (\tilde{\theta} - \theta)T(r) - (A(\tilde{\theta}) - A(\theta)) \right\} \right] \right)^2 d\nu(r)$, and $I_2 = \int_{\{f(r|\tilde{\theta}) \geq f(r|\theta)\}} f(r|\tilde{\theta}) \left(1 - \exp \left[\frac{1}{2} \left\{ (\theta - \tilde{\theta})T(r) - (A(\theta) - A(\tilde{\theta})) \right\} \right] \right)^2 d\nu(r)$.

For I_1 , since $f(r|\theta) > f(r|\tilde{\theta})$, we have $Z := (\tilde{\theta} - \theta)T(r) - (A(\tilde{\theta}) - A(\theta)) \leq 0$. Since $\|\gamma\|_{\infty} \leq L$, we have θ bounded in a closed set, and hence $A'(\theta) = E_{\theta}[T(r)]$ is bounded. Let

$L_A = \sup_{\theta} |E_{\theta} T(r)|$, then

$$|A(\tilde{\theta}) - A(\theta)| \leq L_A |\tilde{\theta} - \theta|.$$

Then $-Z = |Z| \geq (|T(r)| - L_A) |\tilde{\theta} - \theta|$. That is,

$$1 - \exp\left\{-\frac{1}{2}|Z|\right\} \geq \max\left\{1 - \exp\left[\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right], 0\right\},$$

and

$$\begin{aligned} I_1 &= \int_{\{f(r|\theta) > f(r|\tilde{\theta})\}} f(r|\theta) \left(1 - \exp\left\{-\frac{1}{2}|Z|\right\}\right)^2 d\nu(r) \\ &\geq \int_{\{f(r|\theta) > f(r|\tilde{\theta})\}} f(r|\theta) \max\left\{1 - \exp\left[\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right], 0\right\}^2 d\nu(r). \end{aligned}$$

In a similar way,

$$\begin{aligned} I_2 &\geq \int_{\{f(r|\tilde{\theta}) \geq f(r|\theta)\}} f(r|\tilde{\theta}) \max\left\{1 - \exp\left[\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right], 0\right\}^2 d\nu(r) \\ &\geq \int_{\{f(r|\tilde{\theta}) \geq f(r|\theta)\}} f(r|\theta) \max\left\{1 - \exp\left[\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right], 0\right\}^2 d\nu(r). \end{aligned}$$

Notice that $1 - \exp\left\{\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right\} \geq 0$ if and only if $|T(r)| \geq L_A$. Hence,

$$\begin{aligned} h_{\Theta}^2(\theta, \tilde{\theta}) &= I_1 + I_2 \\ &\geq \int_{\{|T(r)| \geq L_A\}} f(r|\theta) \left[1 - \exp\left\{\frac{1}{2}(L_A - |T(r)|)|\tilde{\theta} - \theta|\right\}\right]^2 d\nu(r), \end{aligned}$$

which is a non-decreasing function of $|\tilde{\theta} - \theta|$.

Therefore, for each θ_{ui} , and given the $\epsilon_{|\Omega|}$ in Theorem 1, there exists a $\delta_{|\Omega|}(\theta_{ui})$, such that $|\tilde{\theta}_{ui} - \theta_{ui}| > \delta_{|\Omega|}(\theta_{ui})$ implies $h_{\Theta}(\theta_{ui}, \tilde{\theta}_{ui}) \geq \epsilon_{|\Omega|}$.

Let $\delta_{|\Omega|} = \max_{1 \leq u \leq n, 1 \leq i \leq m} \sup_{\theta_{ui}} \delta_{|\Omega|}(\theta_{ui})$, then $|\tilde{\theta}_{ui} - \theta_{ui}| > \delta_{|\Omega|}$ for each (u, i) implies $h(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}) \geq \epsilon_{|\Omega|}$, and the result follows. This completes the proof.

Proof of Corollary 1

Define $\Phi = \{(u, i) : |\theta_{ui}^0 - \theta_{ui}| > \delta_{|\Omega|}\}$. Then the cardinality of Φ satisfies $|\Phi| \geq \phi nm$. From Theorem 2, for $(u, i) \in \Phi$, we have $h_{\Theta}(\theta_{ui}^0, \theta_{ui}) \geq \frac{1}{\sqrt{\phi}} \epsilon_{|\Omega|}$. Hence by the definition of γ_0 , we have:

$$h(\gamma, \gamma_0) = \left\{ \frac{1}{nm} \sum_{i=1}^m \sum_{u=1}^n h_{\Theta}^2(\theta_{ui}^0, \theta_{ui}) \right\}^{1/2} \geq \left\{ \frac{1}{nm} (\phi nm) \frac{1}{\phi} \epsilon_{|\Omega|}^2 \right\}^{1/2} = \epsilon_{|\Omega|}.$$

This completes the proof.

Proof of Corollary 2

In the proof of Corollary 1, we verify that $h(\gamma, \gamma_0) \geq \epsilon_{|\Omega|}$. Then by Lemma 2, we have:

$$\epsilon_{|\Omega|}^2 \leq h^2(\gamma, \gamma_0) \leq \frac{d_0^2(n+m)}{nm} \|\gamma - \gamma_0\|_2^2.$$

Then $\|\mathbf{S}\|_F^2 + \|\mathbf{T}\|_F^2 = \|\gamma - \gamma_0\|^2 \geq \frac{nm}{d_0^2(n+m)} \epsilon_{|\Omega|}^2$.

Meanwhile, we have each entry of $\mathbf{S}$ and $\mathbf{T}$ bounded by L , and hence $\|\mathbf{S}\|_F^2 + \|\mathbf{T}\|_F^2 \leq (N + M)KL^2$.

Therefore, $N + M \geq \frac{nm}{d_0^2 L^2 K(n+m)} \epsilon_{|\Omega|}^2$. By the rate of $\epsilon_{|\Omega|}$ provided in Theorem 1, we have:

$$O(N + M) \succeq \frac{nm}{|\Omega|} \log \left(\frac{|\Omega|}{\sqrt{nmK}} \right).$$

This completes the proof.

References

- Ansley, C. F. and Kohn, R. (1994). Convergence of the backfitting algorithm for additive models. *Journal of the Australian Mathematical Society (Series A)*, 57(3):316–329.
- Chen, B., He, S., Li, Z., and Zhang, S. (2012). Maximum block improvement and polynomial optimization. *SIAM Journal on Optimization*, 22(1):87–107.
- Ossiander, M. (1987). A central limit theorem under metric entropy with L2 bracketing. *The Annals of Probability*, 15(3):897–919.
- Shen, X. (1998). On the method of penalization. *Statistica Sinica*, 8(2):337–357.