Online supplement for "Nonlinear sufficient dimension reduction for functional data"

Bing Li and Jun Song

In this Supplement we provide the proofs of the theorems, lemmas, and corollaries in the manuscript. The equation labels such as (1) and (2) are for the equations in the manuscript; equation labels such as (A1) and (A2) are for the equations in this supplement.

Proof of Lemma 1 By the definition of $\mathfrak{M}_X^0$ above Lemma 1, a member $f \in \mathfrak{M}_X$ is orthogonal to $\mathfrak{M}_X^0$ iff $f \perp \kappa_X(\cdot, x) - \mu_X$ for all $x \in \mathcal{H}_X$. That is $\langle f, \kappa_X(\cdot, x) - \mu_X \rangle_{\mathfrak{M}_X} = 0$. However, this happens iff $f(x) = Ef(X)$ for all $x \in \mathcal{H}_X$, which is equivalent to $\mathrm{var}(f(X)) = 0$, or $\Sigma_{XX} f = 0$. Thus $(\mathfrak{M}_X^0)^\perp = \ker(\Sigma_{XX})$, which implies $\mathfrak{M}_X^0 = \overline{\mathrm{ran}}\,(\Sigma_{XX})$ because Σ_{XX} is self adjoint. $\square$

Proof of Proposition 1 For any $h \in \mathfrak{M}_Y$,

$$\mathrm{cov}(f(X) - (R_{YX}f)(Y), h(Y)) = \mathrm{cov}(f(X), h(Y)) - \mathrm{cov}((R_{YX}f)(Y), h(Y)) \tag{A1}$$

The first term on the right is $\langle \Sigma_{YX} f, h \rangle_{\mathfrak{M}_Y}$. The second term on the right is

$$\langle \Sigma_{YY} \Sigma_{YY}^\dagger \Sigma_{YX} f, h \rangle_{\mathfrak{M}_X} = \langle \Sigma_{YX} f, h \rangle_{\mathfrak{M}_X}.$$

because, by definition, $\Sigma_{YY}\Sigma_{YY}^\dagger$ is the identity mapping on $\mathrm{ran}(\Sigma_{YY})$ to $\mathrm{ran}(\Sigma_{YY})$. Hence the right hand side of (A1) is 0 for all $h \in \mathfrak{M}_X$. Since $\mathfrak{M}_X$ is dense in $L_2(P_Y)$ modulo constants (Assumption 1), we have

$$\mathrm{cov}(f(X) - (R_{YX}f)(Y), h(Y)) = 0$$

for all $h \in L_2(P_Y)$, which implies $(R_{YX}f)(Y) = E(f(X)|Y) + \text{constant}$. Taking unconditional expectation on both sides, we find the constant to be $E[(R_{YX}f)(Y)] - Ef(X)$, as desired. $\square$

Proof of Theorem 2 Let $f \in \mathfrak{M}_X$ and $f \perp \Sigma_{XX}\mathfrak{S}_{Y|X}$. Then, for any $g \in \mathfrak{S}_{Y|X}$,

$$\mathrm{cov}[f(X), g(X)] = \langle f, \Sigma_{XX} g \rangle_{\mathfrak{M}_X} = 0.$$

Since $\mathfrak{M}_X$ is dense in $L_2(P_X)$ modula constants, the above equality holds for all $f \in L_2(P_X)$ that are measurable with respect to $\mathcal{G}_{Y|X}$. By Lemma 1 of Lee, Li, and Chiaromonte (2013), $E[f(X)|\mathcal{G}_{Y|X}] = 0$. Hence

$$\begin{aligned} \mathrm{var}[f(X)|Y] &= \mathrm{var}\{E[f(X)|\mathcal{G}_{Y|X}]|Y\} + E\{\mathrm{var}[f(X)|\mathcal{G}_{Y|X}]|Y\} \\ &= E\{\mathrm{var}[f(X)|\mathcal{G}_{Y|X}]|Y\} = \mathrm{var}[f(X)|\mathcal{G}_{Y|X}], \end{aligned}$$

where the last equality follows from Assumption 7. Similarly,

$$\mathrm{var}[f(X)] = \mathrm{var}\{E[f(X)|\mathcal{G}_{Y|X}]\} + E\{\mathrm{var}[f(X)|\mathcal{G}_{Y|X}]\} = \mathrm{var}[f(X)|\mathcal{G}_{Y|X}].$$

Comparing the last two equations we see that $\mathrm{var}[f(X)|Y] = \mathrm{var}[f(X)]$, which implies $f \in \ker\{[\Sigma_{XX} - V_{XX|Y}(Y)]\}$. Hence

$$\langle f, Sf\rangle_{\mathfrak{M}_X} = E\langle f, [\Sigma_{XX} - V_{XX|Y}(Y)]^2\rangle_{\mathfrak{M}_X} = 0,$$

which implies $f \in \ker S$. Thus we have proved $[\Sigma_{XX}\mathfrak{S}_{Y|X}]^\perp \subseteq \ker[ES$, which is equivalent to the assertion of the theorem. □

Proof of Lemma 2 Because $f_1 f_2 = \sum_{i=1}^n [f_1 f_2]_i \kappa_X(\cdot, X_i)$, we have

$$(f_1(X_1), \ldots, f_1(X_n))^\intercal \odot (f_2(X_1), \ldots, f_2(X_n))^\intercal = K_X[f_1 f_2],$$

where the left hand side can be expressed as $K_X[f_1] \odot K_X[f_2]$. Multiply both sides by K_X^{-1} from the left to complete the proof. □

Proof of Lemma 3 By (18),

$$\mathrm{cov}_n(f_1, f_2|y) = [E_{YX}(f_1 f_2)]^\intercal b_Y(y) - [f_1]^\intercal [E_{YX}]^\intercal b_Y(y) b_Y^\intercal(y) [E_{YX}][f_2].$$

The first term on the right-hand side is

$$\begin{aligned}[E_{YX}(f_1 f_2)]^\intercal b_Y(y) = [f_1 f_2]^\intercal [E_{YX}]^\intercal b_Y(y) &= (K_X[f_1] \odot K_X[f_2])^\intercal K_X^{-1} [E_{YX}]^\intercal b_Y(y) \\ &= [f_1]^\intercal K_X \mathrm{diag}(K_X^{-1}[E_{YX}]^\intercal b_Y(y)) K_X [f_2],\end{aligned}$$

whence follows the desired equality (19). □

Proof of Theorem 4 Let $f_1, f_2 \in \mathfrak{M}_X^0$. Then, by Lemma 3,

$$\langle f_1, V_{XX|Y}(y) f_2\rangle_{\mathfrak{M}_X} = [f_1]^\intercal Q A(y) Q [f_2].$$

In the meantime,

$$\langle f_1, V_{XX|Y}(y) f_2\rangle_{\mathfrak{M}_X} = [f_1]^\intercal K_X [V_{XX|Y}(y)][f_2].$$

Because $\mathrm{ran}(V_{XX|Y}(y)) \subseteq \mathfrak{M}_X^0$, we have $[V_{XX|Y}(y)] = Q[V_{XX|Y}(y)]$. Also, $[f_1] = Q[f_1]$ and $[f_2] = Q[f_2]$. Hence the right hand side above is $[f_1]^\intercal Q K_X Q [V_{XX|Y}(y)][f_2]$. Thus we have

$$Q K_X Q [V_{XX|Y}(y)] = Q A(y) Q,$$

which implies the desired equality. □

Proof of Theorem 5 *1.* Let

$$\hat{A} = (\hat{\Sigma}_{XX} + \epsilon_n I)^{-1},\ A_n = (\Sigma_{XX} + \epsilon_n I)^{-1},\ A = \Sigma_{XX}^{\dagger},\ \hat{B} = \hat{\Sigma}_{XY},\ B = \Sigma_{XY}.$$

Then $\hat{M} - M = \hat{A}\hat{B}\hat{B}^*\hat{A}^* - ABB^*A^*$. Because

$$\hat{A}\hat{B}\hat{B}^*\hat{A}^* - ABB^*A^* = \hat{A}\hat{B}(\hat{B}^*\hat{A}^* - B^*A^*) + (\hat{A}\hat{B} - AB)B^*A^*,$$

we have

$$\begin{aligned}\|\hat{A}\hat{B}\hat{B}^*\hat{A}^* - ABB^*A^*\|_{\mathrm{OP}} &\le \|(\hat{A}\hat{B} - AB)\hat{B}^*\hat{A}^*\|_{\mathrm{OP}} + \|(\hat{A}\hat{B} - AB)B^*A^*\|_{\mathrm{OP}} \\ &\le \|\hat{A}\hat{B} - AB\|_{\mathrm{OP}}\,\|AB + \hat{A}\hat{B}\|_{\mathrm{OP}}\end{aligned}$$

Hence it suffices to show that $\|\hat{A}\hat{B} - AB\|_{\mathrm{OP}} = O_P(\epsilon_n^{\beta} + \epsilon_n^{-1}n^{-1/2})$. Because

$$\hat{A}\hat{B} = \hat{A}(\hat{B} - B) + (\hat{A} - A_n)B + (A_n - A)B + AB,$$

we have

$$\|\hat{A}\hat{B} - AB\|_{\mathrm{OP}} \le \|\hat{A}(\hat{B} - B)\|_{\mathrm{OP}} + \|(\hat{A} - A_n)B\|_{\mathrm{OP}} + \|(A_n - A)B\|_{\mathrm{OP}}. \tag{A2}$$

The first term on the right hand side is

$$\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}(\hat{\Sigma}_{XY} - \Sigma_{XY})\|_{\mathrm{OP}} \le \|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}\|_{\mathrm{OP}}\,\|(\hat{\Sigma}_{XY} - \Sigma_{XY})\|_{\mathrm{OP}}$$

where

$$\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}\|_{\mathrm{OP}} \le \epsilon_n^{-1}\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}(\hat{\Sigma}_{XX} + \epsilon_n I)\|_{\mathrm{OP}} = \epsilon_n^{-1}.$$

and, by Lemma 4, $\|(\hat{\Sigma}_{XY} - \Sigma_{XY})\|_{\mathrm{OP}} = O_P(n^{-1/2})$. Hence the first term on the right-hand side of (A2) is of the order $O_P(\epsilon_n^{-1}n^{-1/2})$. Because

$$\begin{aligned}(\hat{A} - A_n)B &= (\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}[(\Sigma_{XX} + \epsilon_n I) - (\hat{\Sigma}_{XX} + \epsilon_n I)](\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XY} \\ &= (\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}(\Sigma_{XX} - \hat{\Sigma}_{XX})(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XY}\end{aligned}$$

and because, by Assumption 3, $\Sigma_{XY} = \Sigma_{XX}R_{XY}$, the second term on the right-hand side of (A2) is no greater than

$$\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}\|_{\mathrm{OP}}\,\|\Sigma_{XX} - \hat{\Sigma}_{XX}\|_{\mathrm{OP}}\,\|(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XX}\|_{\mathrm{OP}}\,\|R_{XY}\|_{\mathrm{OP}}.$$

As argued before, $\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}\|_{\mathrm{OP}}$ and $\|\Sigma_{XX} - \hat{\Sigma}_{XX}\|_{\mathrm{OP}}$ are of the orders $O_P(\epsilon_n^{-1})$ and $O_P(n^{-1/2})$, respectively. Moreover,

$$\|(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XX}\|_{\mathrm{OP}} \le \|(\Sigma_{XX} + \epsilon_n I)^{-1}(\Sigma_{XX} + \epsilon_n I)\|_{\mathrm{OP}} = 1.$$

Hence the second term on the right-hand side of (A2) is of the order $O_P(\epsilon_n^{-1}n^{-1/2})$.

Finally, by Assumption 3 and the assumption $R_{XY} = \Sigma_{XX}^{\beta} S_{XY}$ for a bounded linear operator S_{XY}, we can rewrite the operator $(A_n - A)B$ on the right-hand side of (A2) as

$$\begin{aligned}(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XY} - \Sigma_{XX}^{\dagger}\Sigma_{XY} &= (\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XX} R_{XY} - R_{XY} \\ &= (\Sigma_{XX} + \epsilon_n I)^{-1}(\Sigma_{XX} + \epsilon_n I)R_{XY} - \epsilon_n(\Sigma_{XX} + \epsilon_n I)^{-1}R_{XY} - R_{XY} \\ &= -\epsilon_n(\Sigma_{XX} + \epsilon_n I)^{-1}R_{XY} \\ &= -\epsilon_n(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XX}^{\beta} S_{XY}.\end{aligned}$$

Hence the third term on the right-hand side of (A2) satisfies

$$\begin{aligned}\|(A_n - A)B\|_{\mathrm{OP}} &\le \epsilon_n \|(\Sigma_{XX} + \epsilon_I)^{-1+\beta}\|_{\mathrm{OP}}\,\|S_{XY}\|_{\mathrm{OP}} \\ &= \epsilon_n \epsilon_n^{\beta-1}\|(\Sigma_{XX} + \epsilon_n I)^{-1+\beta}(\epsilon_n I)^{1-\beta}\|_{\mathrm{OP}}\,\|S_{XY}\|_{\mathrm{OP}} \\ &\le \epsilon_n \epsilon_n^{\beta-1}\|(\Sigma_{XX} + \epsilon_n I)^{-1+\beta}(\Sigma_{XX} + \epsilon_n I)^{1-\beta}\|_{\mathrm{OP}}\,\|S_{XY}\|_{\mathrm{OP}} = \epsilon_n^{\beta}\|S_{XY}\|_{\mathrm{OP}},\end{aligned}$$

which is of the order $O(\epsilon_n^{\beta})$. This proves part *1*.
2. The proof is similar to part 1; so we only highlight the differences. We use the following facts: (i) if A is a bounded operator and B is Hilbert Schmidt operator and $\mathrm{ran}(A) \subseteq \mathrm{dom}(B)$, then AB is a Hilbert Schmidt operator with $\|AB\|_{\mathrm{HS}} \le \|A\|_{\mathrm{OP}}\,\|B\|_{\mathrm{HS}}$; (ii) if A is Hilbert Schmidt then so is A^* and $\|A\|_{\mathrm{HS}} = \|A^*\|_{\mathrm{HS}}$. Inequality (A2) now becomes

$$\|\hat{A}\hat{B} - AB\|_{\mathrm{HS}} \le \|\hat{A}(\hat{B} - B)\|_{\mathrm{HS}} + \|(\hat{A} - A_n)B\|_{\mathrm{HS}} + \|(A_n - A)B\|_{\mathrm{HS}}. \tag{A3}$$

The first term on the right-hand side is no more than $\|\hat{A}\|_{\mathrm{OP}}\|\hat{B} - B\|_{\mathrm{HS}}$, which is of the order $O_P(\epsilon_n^{-1}n^{-1/2})$ by Lemma 4 and the proof of part 1. The second term on the right-side of (A3) is bounded by

$$\|(\hat{\Sigma}_{XX} + \epsilon_n I)^{-1}\|_{\mathrm{OP}}\,\|\Sigma_{XX} - \hat{\Sigma}_{XX}\|_{\mathrm{OP}}\,\|(\Sigma_{XX} + \epsilon_n I)^{-1}\Sigma_{XX}\|_{\mathrm{OP}}\,\|R_{XY}\|_{\mathrm{HS}}.$$

Because $R_{XY} = \Sigma_{XX}^{\beta} S_{XY}$ and S_{XY} is Hilbert Schmidt, R_{XY} is itself Hilbert Schmidt. Hence, the above is of the order $O_P(\epsilon_n^{-1}n^{-1/2})$. Finally, the third term in (A3) is bounded by $\epsilon_n^{\beta}\|S_{YX}\|_{\mathrm{HS}}$, which, because S_{YX} is Hilbert Schmidt, is of the order $O(\epsilon_n^{\beta})$. □

Proof of Theorem 7 If $k < d$, then

$$\begin{aligned}G_n(k) - G_n(d) &= -\textstyle\sum_{i=k+1}^{d}\hat{\lambda}_i - a\hat{\lambda}_1 n^{\alpha\beta/(1+\beta)}\log(n)(k-d) \\ &= -\textstyle\sum_{i=k+1}^{d}\lambda_i + o_P(1).\end{aligned}$$

Thus $P(G_n(k) < G_n(d)) \to 1$. If $k > d$, then

$$\begin{aligned}G_n(k) - G_n(d) &= \textstyle\sum_{i=d+1}^{k}\hat{\lambda}_i + a\hat{\lambda}_1 n^{\alpha\beta/(1+\beta)}\log(n)(d-k) \\ &= 0 + O_P(n^{-\alpha\beta/(1+\beta)}) + a\hat{\lambda}_1 n^{\alpha\beta/(1+\beta)}\log(n)(d-k).\end{aligned}$$

Since the right hand side is dominated by $c_{1n}(c_{2d}-c_{2k})$, which is negative, we have $P(G_n(k) < G_n(d)) \to 1$. Consequently, with probability tending to 1, $G_n(k)$ is maximized at d. □

References

Lee, K.-Y., Li, B., and Chiaromonte, F. (2013), "A general theory for nonlinear sufficient dimension reduction: Formulation and estimation," *The Annals of Statistics*, 41, 221–249.