Appendix: Connection with the GSVD

We provide the derivation of a LongPEER estimate using the GSVD. This can be viewed as an extension of the estimation discussed by Randolph *et al.* (2012) in two ways: we allow for a general covariance matrix $\mathbf{V}$ (for $\mathbf{y}$) and we extend the penalty operator to apply across multiply-defined domains, $\mathbf{L}_0, \ldots, \mathbf{L}_D$.

After some algebra, the generalized ridge estimate in equation (3.3) for $\boldsymbol{\gamma}$ can be expressed as

$$\hat{\gamma} = -\mathbf{A}_1 \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y} + \mathbf{A}_2 \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{y}$$

where

$$\begin{aligned} \mathbf{A}_1^\top &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{W} [\mathbf{W}^\top \mathbf{V}^{-1} \mathbf{W} + \mathbf{L}^\top \mathbf{L} - \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{W}]^{-1} \\ \mathbf{A}_2 &= \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{W} + \mathbf{L}^\top \mathbf{L} - \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{W} \end{aligned}$$

When $\mathbf{X} = \mathbf{0}$ (a situation without any scalar predictors) or $\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{W} = 0$ the generalized ridge estimation of $\boldsymbol{\gamma}$ can be put into a PEER estimation framework in terms of GS vectors, as discussed below.

With $\mathbf{X} = \mathbf{0}$ or $\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{W} = \mathbf{0}$, the $\hat{\boldsymbol{\gamma}}$ reduces to $[\mathbf{W}^\top \mathbf{V}^{-1} \mathbf{W} + \mathbf{L}^\top \mathbf{L}]^{-1} \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{y}$. Moreover, in this case generalized ridge estimate of β becomes $[\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}$. Now, if we transform $\tilde{\mathbf{W}} := \mathbf{V}^{-1/2} \mathbf{W}$ and $\tilde{\mathbf{y}} := \mathbf{V}^{-1/2} \mathbf{y}$, we can rewrite $\mathbf{L}$ as

$$\mathbf{L} = \lambda_0 \text{ blockdiag} \left\{ \mathbf{L}_0, \frac{\lambda_1}{\lambda_0} \mathbf{L}_1, \cdots, \frac{\lambda_D}{\lambda_0} \mathbf{L}_D \right\} = \lambda_0 \mathbf{L}^s$$

Here, $\mathbf{L}^s$ can be interpreted as a scaled $\mathbf{L}$ where scaling is done for all the tuning parameters associated with the 'longitudinal' part of the regression function with respect to the 'baseline' tuning parameter.

Set $\tilde{p} = (D+1)p$, let m denote the number of rows in $\mathbf{L}$ and set $c = \dim[\text{Null}(\mathbf{L})]$. Further, assume that $n_\bullet \leq m \leq \tilde{p} \leq m + n_\bullet$ and the rank of the $(n_\bullet + m) \times \tilde{p}$ matrix $[\tilde{\mathbf{W}}^\top \quad (\mathbf{L}^s)^\top]^\top$ is $\tilde{p}$. The following describes the GSVD of the pair $(\tilde{\mathbf{W}}, \mathbf{L}^s)$: there exist orthogonal matrices $\mathcal{U}$ and $\mathcal{V}$, a nonsingular $\mathcal{G}$ and diagonal matrices $\mathbf{S}$ and $\mathbf{M}$ such that

$$\tilde{\mathbf{W}} = \mathcal{U} \mathcal{S} \mathcal{G}^{-1} \qquad \mathcal{S} = [\mathbf{0} \;\; \mathbf{S}] \qquad \mathbf{S} = \text{blockdiag}\{\mathbf{S}_1, \;\; \mathbf{I}_{\tilde{p}-m}\}$$

$$\mathbf{L}^s = \mathcal{V} \mathcal{M} \mathcal{G}^{-1} \qquad \mathcal{M} = [\mathbf{M} \;\; \mathbf{0}] \qquad \mathbf{M} = \text{blockdiag}\{\mathbf{I}_{\tilde{p}-n_\bullet}, \;\; \mathbf{M}_1\}$$

Submatrices $\mathbf{S}_1$ and $\mathbf{M}_1$ have $\ell = n_\bullet + m - \tilde{p}$ diagonal entries ordered as

$$\begin{array}{l} 0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_\ell < 1 \\ 0 > \mu_1 \geq \mu_2 \geq \cdots \geq \mu_\ell > 1 \end{array} \qquad \text{where,} \qquad \sigma_k^2 + \mu_k^2 = 1, \qquad k = 1, \ldots, \ell$$

Here, the columns $\{\mathbf{g}_k\}$ of $\mathcal{G}$ are the GS vectors determined by the GSVD of the pair $(\tilde{\mathbf{W}}, \mathbf{L}^s)$. Denote the columns of $\mathcal{U}$ and $\mathcal{V}$ by $\mathbf{u}_k$ and $\mathbf{v}_k$, respectively. Now, it can be shown that $[\mathbf{W}^\top\mathbf{V}^{-1}\mathbf{W}+\mathbf{L}^\top\mathbf{L}]^{-1}\mathbf{W}^\top\mathbf{V}^{-1} = [\mathbf{W}^\top\mathbf{V}^{-1}\mathbf{W}+\lambda_0^2(\mathbf{L}^s)^\top\mathbf{L}^s]^{-1}\mathbf{W}^\top\mathbf{V}^{-1} = \mathcal{G}(\mathcal{S}^\top\mathcal{S}+\lambda_0^2\mathcal{M}^\top\mathcal{M})^{-1}\mathcal{G}^\top\ \tilde{W}^\top\mathbf{V}^{-1/2}$ and consequently, $\hat{\boldsymbol{\gamma}}$ can be expressed as

$$\hat{\boldsymbol{\gamma}} = \mathcal{G}(\mathcal{S}^\top\mathcal{S}+\lambda_0^2\mathcal{M}^\top\mathcal{M})^{-1}\mathcal{S}^\top\mathcal{U}^\top\tilde{\mathbf{y}} = \sum_{k=\tilde{p}-n_\bullet+1}^{\tilde{p}-c}\frac{\sigma_k^2}{\sigma_k^2+\lambda_0^2\mu_k^2}\frac{1}{\sigma_k}\mathbf{u}_k^\top\tilde{\mathbf{y}}\mathbf{g}_k + \sum_{k=\tilde{p}-c+1}^{\tilde{p}}\mathbf{u}_k^\top\tilde{\mathbf{y}}\mathbf{g}_k$$

Further, the bias and variance can be expressed as

$$\begin{aligned} Bias[\hat{\gamma}] = (\mathbf{I}-\mathbf{W}^{\#}\mathbf{W})\boldsymbol{\gamma} \quad &= \mathcal{G}(\mathcal{S}^\top\mathcal{S}+\lambda_0^2\mathcal{M}^\top\mathcal{M})^{-1}(\lambda_0^2\mathcal{M}^\top\mathcal{M})\mathcal{G}^{-1} \\ &= \textstyle\sum_{k=1}^{\tilde{p}-n_\bullet}\mathbf{g}_k\tilde{\mathbf{g}}_k^\top\boldsymbol{\gamma} + \sum_{k=\tilde{p}-n_\bullet+1}^{\tilde{p}-c}\frac{\lambda_0^2\mu_k^2}{\sigma_k^2+\lambda_0^2\mu_k^2}\mathbf{g}_k\tilde{\mathbf{g}}_k^\top\boldsymbol{\gamma} \end{aligned}$$

$$\begin{aligned} Var[\hat{\gamma}] = \mathbf{W}^{\#}\mathbf{V}(\mathbf{W}^{\#})^\top \quad &= \mathcal{G}(\mathcal{S}^\top\mathcal{S}+\lambda_0^2\mathcal{M}^\top\mathcal{M})^{-1}\mathcal{S}^\top\mathcal{S}(\mathcal{S}^\top\mathcal{S}+\lambda_0^2\mathcal{M}^\top\mathcal{M})^{-1}\mathcal{G}^\top \\ &= \textstyle\sum_{k=\tilde{p}-n_\bullet+1}^{\tilde{p}-c}\frac{\sigma_k^2}{(\sigma_k^2+\lambda_0^2\mu_k^2)^2}\mathbf{g}_k\mathbf{g}_k^\top + \sum_{k=\tilde{p}-c+1}^{\tilde{p}}\mathbf{g}_k\mathbf{g}_k^\top \end{aligned}$$

where, $\mathbf{W}^{\#} = [\mathbf{W}^\top\mathbf{V}^{-1}\mathbf{W}+\mathbf{L}^\top\mathbf{L}]^{-1}\mathbf{W}^\top\mathbf{V}^{-1}$ and $\tilde{\mathbf{g}}_k$ denotes the kth column of $\mathcal{G}^{-T} = (\mathcal{G}^{-1})^\top = (\mathcal{G}^\top)^{-1}$. Further, we can express bias as $[\mathbf{W}^\top\mathbf{V}^{-1}\mathbf{W}+\mathbf{L}^\top\mathbf{L}]^{-1}\mathbf{L}^\top\mathbf{L}\boldsymbol{\gamma}$ which means $\hat{\boldsymbol{\gamma}}$ will be unbiased only when $\boldsymbol{\gamma} \in \text{Null}(\mathbf{L})$.

For estimates obtained using this technique, the bias and variance can be expressed in terms of generalized singular vectors, provided the assumption of $\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{W} = \mathbf{0}$ applies. In this case, one can show that $\hat{\boldsymbol{\beta}}$ is simply the generalized least squares estimate from the linear model $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}^*$, and $\hat{\boldsymbol{\gamma}}$ is the generalized ridge estimate from $\mathbf{y} = \mathbf{W}\boldsymbol{\gamma}+\boldsymbol{\epsilon}^*$ with penalty $\mathbf{L}$. That is, $\boldsymbol{\beta}$ is estimated as if $\mathbf{W}\boldsymbol{\gamma}$ were not present, and $\boldsymbol{\gamma}$ is estimated as if $\mathbf{X}\boldsymbol{\beta}$ were not present.

References

Randolph, T., Harezlak, J. and Feng, Z. (2012). Structured penalties for functional linear models – partially empirical eigenvectors for regression. *Electronic Journal of Statistics* **6,** 323–353.