Supplement to "Localized Functional Principal Component Analysis"

A. EXPLORING DIFFERENT SPARSITY LEVELS IN MORTALITY DATA

We explore the sensitivity to a and found that the resulting functions change gradually as a increases. Figure 5 shows the estimated $\phi_j(t)$ for the mortality data with different a. As a increases from 0 to 0.5, we see that the first eigenfunction, shown in the first column, becomes more and more localized to mid 1990s; The third function (later becomes the second function due to the trade-off between localization and the ability of explaining variance) becomes more concentrated at around early 1980s. We also explore the stability of the results while varying a in the neighborhood of 0.3. As shown in the right panel of Figure 5, the results are quite stable.

B. ADDITIONAL DATA EXAMPLE: BERKELEY GROWTH DATA

The smooth nature of growth curves has been explored in various previous statistical analyses, including functional data analysis approaches. We apply the proposed LFPCA method to the Berkeley growth data (Tuddenham and Snyder, 1954). These data contain height measurements for 54 girls, with 31 measurements taken between ages 1 year and 18 years. A sample covariance matrix is computed based on equally spaced measurements at every half year from interpolated curves and then the proposed algorithm is used to solve problem (5). The solution path along different levels of localization is investigated. The estimated eigenfunctions without localization penalty are visualized in the top row of Figure 6, and the estimated localized eigenfunctions are given in the bottom row of Figure 6. The localization level is chosen to maintain $rFVE = 70\%$ in (12). The total number of components $k = 2$ is chosen to explain total variation of 85%. The first estimated localized basis function, explaining 70.1% of the variation, indicates a variational mode in girls' growth around age twelve, which obviously matches the well known pubertal growth spurt. The second esti-

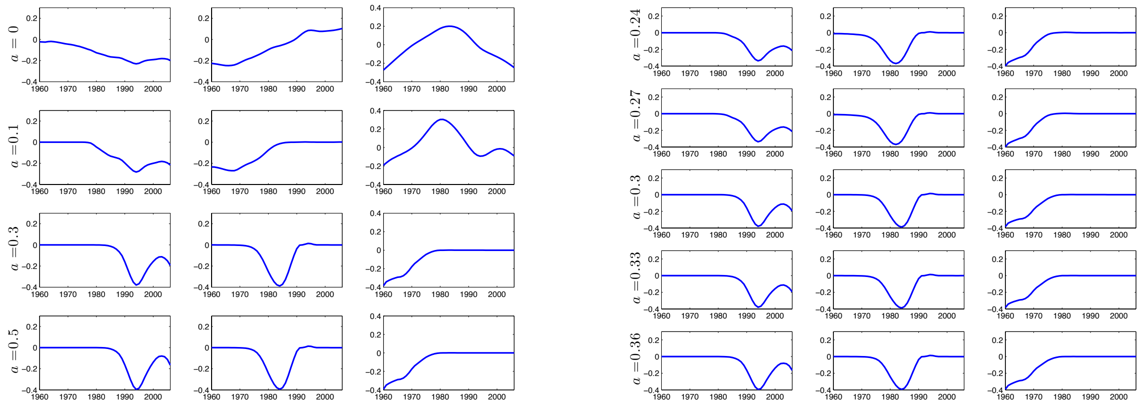

Figure 5: The estimated $\phi_j(t)$, $j = 1, 2, 3$ for the mortality data with different values of a.

mated localized basis function, explaining 18.1% of the variation, is localized around ages five and six, which remarkably matches the mid-growth spurt previously studied by many researchers (Gasser et al., 1985; Sheehy et al., 1999). The mid-growth spurt is a growth phenomenon during early childhood, expressed by a mild transitory acceleration of growth velocity between years five and eight. The individual variations in timings, durations and intensities of mid-growth spurt are of great interest and some hypotheses have been proposed for the explanation of individual differences (Mühl et al., 1991). This particular "mode of variation" is not obvious in standard FPCA. The proposed LFPCA method finds a balance between interpretability (localization) and amounts of variance explained.

C. PROOFS

Proof of Lemma 3.1. Because $\mathcal{D}_\Pi$ is a compact set, we know that $B = \mathcal{P}_{\mathcal{D}_\Pi}(A)$ exists and is unique. Let $G = U^T B U$, then we have $G \in \mathcal{F}^1$ and $B = UGU^T$. Note that G minimizes $\|A - UGU^T\|_F^2$ over $\mathcal{F}^1$ and

$$\begin{aligned}\|A - UGU^T\|_F^2 =& \|A\|_F^2 - 2\langle A, UGU^T\rangle + \|UGU^T\|_F^2 \\ =& \|A\|_F^2 - 2\langle U^T A U, G\rangle + \|G\|_F^2 \\ =& \|A\|_F^2 - \|U^T A U\|_F^2 + \|U^T A U - G\|_F^2 .\end{aligned}$$

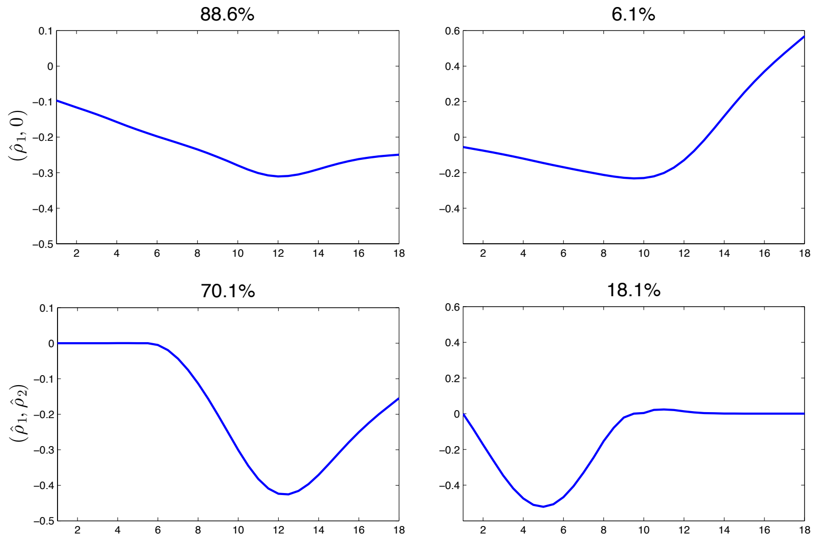

Figure 6: Top Row: Estimated eigenfunctions for the growth data, $\widehat{\rho}_1$ is chosen by 5-fold cross validation; Bottom Row: Estimated orthogonal basis functions, $\widehat{\rho}_2$ is chosen to maintain rFVE at 70%, and the number of components $k = 2$ is chosen to explain at least 85% of the total variance.

Therefore, G is the projection of $U^T AU$ onto $\mathcal{F}^1$ and by Lemma 4.1 of Vu et al. (2013) we have

$$G = \sum_{i=1}^{p-d} \gamma_i^+(\theta)\eta_i\eta_i^T$$

with γ_i, η_i, and θ specified in the theorem. □

Proof of Theorem 3.2. For any given $\tau > 0$, define the augmented Lagrangian of (8) as

$$L_\tau(H, Z, Y) = \mathbb{I}_{\mathcal{D}_\Pi}(H) - \langle S, H\rangle + \rho_2\|Z\|_{1,1} + \langle Y, H - Z\rangle + \frac{\tau}{2}\|H - Z\|_F^2 .$$

The update steps in Algorithm 1 now reads, letting $W^{(r)} = \tau Y^{(r)}$,

$$\begin{aligned} H^{(r)} &= \arg\min_H L_\tau(H, Z^{(r-1)}, W^{(r-1)}) , \\ Z^{(r)} &= \arg\min_Z L_\tau(H^{(r)}, Z, W^{(r)}) , \\ Y^{(r)} &= Y^{(r-1)} + \tau(H - Z) . \end{aligned}$$

It is obvious that $\mathbb{I}_{\mathcal{D}_\Pi}(H) - \langle S, H\rangle$ and $\rho_2\|Z\|_{1,1}$ are closed, proper, and convex functions. Here we say a function f is closed, proper and convex if $\{(x,t) : f(x) \leq t\}$ is a closed non-empty convex set (Boyd et al. (2011), Section 3.2).

By strong duality, we can find a primal-dual pair of $L_0(H, Z, Y)$, denoted as (H^{**}, Z^{**}, Y^{**}). It then follows from the primal and dual optimality that (H^{**}, Z^{**}, Y^{**}) is a saddle point of L_0 and hence by Section 3.2.1 of Boyd et al. (2011), we have

$$Z^{(r)} \to Z^* \quad \text{and} \quad H^{(r)} - Z^{(r)} \to 0, \quad \text{as} \quad t \to \infty\,,$$

where (H^*, Z^*) is an optimal primal variable for L_0. □

Proof of consistency result. To prove Theorem 4.1, we need some additional lemmas and notation as follows. The proof of lemmas are given after the proof of Theorem 4.1.

Let $I_j = ((j-1)/p, j/p]$ for $j = 2, ..., p$ and $I_1 = [0, 1/p]$. We define $\phi_j^*(t) = \phi_j(t_i)$ for $t \in I_i$, $u_j^* = p^{-1/2}(\phi_j(t_1), \phi_j(t_2), ..., \phi_j(t_p))^T$, and $u_j = u_j^*/\|u_j^*\|_2$. Let $\widetilde{\Gamma} : [0,1]^2 \mapsto [0, \infty)$ be such that $\widetilde{\Gamma}(s,t) = \Gamma(t_i, t_j)$, if $s \in I_i, t \in I_j$. Define the discretized and diagonal-shifted covariance matrix Σ by $\Sigma(l, l') = \Gamma(t_l, t_{l'}) + a\mathbf{1}(l = l')$. Let $\widetilde{\phi}_j$ be eigenfunctions of $\widetilde{\Gamma}$ and v_j be eigenvectors of Σ. Then $p^{-1/2}\widetilde{\phi}_j(t)$ is the ith entry of v_j if $t \in I_i$. If we further denote the jth eigenvalue of $\widetilde{\Gamma}$ by $\widetilde{\lambda}_j$, then $(p\widetilde{\lambda}_j + a, v_j)$ is an eigenvalue-eigenvector pair of Σ.

Let $\Pi_j = \sum_{i=1}^j v_i v_i^T$, and $\Sigma_j = \Sigma - \Pi_{j-1}\Sigma\Pi_{j-1}$. Define $\Pi_0 = 0$ for convenience. For any measurable $B : [0,1]^2 \mapsto \mathbb{R}$, let $\|B\|_{\text{HS}} = [\int_{[0,1]^2} B(s,t)^2 dsdt]^{1/2}$ be the Hilbert-Schmidt norm.

Lemma C.1. *Under Assumptions (A1-A4), let $c_0 = L(2/\delta + 1)$ we have for p large enough, we have $\|v_j - u_j\|_2 \le c_0 p^{-1}$, for $1 \le j \le k$.*

Lemma C.2. *Under Assumptions (A1-A3), when p is large enough we have $\|\Sigma\|_F^2 \le c_2^2 p^2$, where $c_2^2 = a^2 + 4\|\Gamma\|_{\text{HS}}^2 + L^2$. Moreover, the gap between the jth and $(j+1)$th eigenvalues of Σ is at least $p\delta/2$ for all $1 \le j \le k$.*

The following lemma is elementary and can be found in Vu and Lei (2013).

Lemma C.3. *Let u and v be vectors of same length with unit norm. Then*

$$\frac{1}{\sqrt{2}}\|u - v\|_2 \le \|uu^T - vv^T\|_F \le \sqrt{2}\|u - v\|_2\,.$$

The same holds when u, v are functions and $\|\cdot\|_F$ is replaced by $\|\cdot\|_{\text{HS}}$.

The next lemma characterizes u_i as an approximate leading eigenvector of Σ_i. It extends Lemma 4.2 of Vu and Lei (2013).

Lemma C.4 (Approximate curvature)**.** *Let H_j be a solution to* (6)*. Then under Assumptions (A2-A4), for p large enough,*

$$\frac{p\delta}{8}\|H_j - u_j u_j^T\|_F^2 - \frac{3c_0c_2}{2} \leq \langle -\Sigma_j, H_j - u_j u_j^T\rangle\,, \quad \forall \quad 1 \leq j \leq k\,.$$

Proof of Theorem 4.1. The claim follows if we show that $\sup_{1\leq j\leq k}\|\widehat{v}_j - u_j\|_2 = o_p(1)$.

For simplicity denote $e_n := \|S-\Sigma\|_{\infty,\infty}$, which is $o_p(1)$ by assumption (A1). Let $\Pi_0 = \widehat{\Pi}_0 = 0$, $\widehat{v}_0 = 0$, $\beta_0 = \epsilon_0 = 0$. We use induction to show that there exist $(\epsilon_i, \beta_i : 0 \leq i \leq k)$ such that

$$\sup_{0\leq i\leq k}\epsilon_i = o_p(1), \quad \sup_{0\leq i\leq k}\beta_i = o_p(p)\,, \tag{A.1}$$

$$\max\{\|\widehat{v}_i - v_i\|_2, \|\widehat{\Pi}_i - \Pi_i\|_2\} \leq \epsilon_i \tag{A.2}$$

$$\rho_1\langle D, \widehat{v}_i\widehat{v}_i^T\rangle \leq \beta_i\,. \tag{A.3}$$

Obviously the claim holds for $k=0$. Now assume that the claim holds for $k = j-1$, and $j \geq 1$. We will construct $\epsilon_j = o_p(1)$ and $\beta_j = o_p(p)$ satisfying (A.2) and (A.3).

Let $\Sigma_j = \Sigma - \Pi_{j-1}\Sigma\Pi_{j-1}$ for $j = 1, ..., k$. By Lemma C.4 we have, for p large enough,

$$\frac{p\delta}{8}\|H_j - u_j u_j^T\|_F^2 - \frac{3c_0c_2}{2} \leq \langle -\Sigma_j, H_j - u_j u_j^T\rangle\,. \tag{A.4}$$

Next we need to control $\langle \Sigma - \Sigma_j, H_j - u_j u_j^T\rangle$ and $\langle S, H_j - u_j u_j^T\rangle$, where the first one is small because H_j and $u_j u_j^T$ are nearly orthogonal to $\Sigma - \Sigma_j$, and the second term needs to be controlled by the fact that H_j is a maximizer of (6).

For the first term $\langle \Sigma - \Sigma_j, H_j - u_j u_j^T\rangle$, by the orthogonality constraint, we have

$$\langle \Sigma - \Sigma_j, H_j\rangle \leq \lambda_1\langle \Pi_{j-1}, H_j\rangle = \lambda_1|\langle \Pi_{j-1} - \widehat{\Pi}_{j-1}, H_j\rangle| \leq \lambda_1\epsilon_{j-1} \leq c_2 p\epsilon_{j-1}\,.$$

and similarly

$$\langle \Sigma - \Sigma_j, u_j u_j^T\rangle = \langle \Sigma - \Sigma_j, u_j u_j^T - v_j v_j^T\rangle \leq \|\Sigma_{j-1}\|_F\|u_j u_j^T - v_j v_j^T\|_F \leq \sqrt{2}c_2c_0\,,$$

where the last inequality follows from Lemma C.1 and Lemma C.3, and therefore

$$\left|\langle \Sigma - \Sigma_j, H_j - u_j u_j^T \rangle\right| \leq c_2 p(\epsilon_{j-1} + \sqrt{2} c_0 p^{-1}) \,. \tag{A.5}$$

Now we turn to the term $\langle S, H_j - u_j u_j^T \rangle$. If we can show that

$$0 \leq \langle S, H_j - u_j u_j^T \rangle - \rho_1 \langle D, H_j \rangle + R_j \,, \tag{A.6}$$

for some $R_j = o_p(p)$ then we have, combining (A.4) to (A.6),

$$\begin{aligned} \frac{p\delta}{8} \|H_j - u_j u_j^2\|_F^2 \leq & \langle S - \Sigma, H_j - u_j u_j^T \rangle - \rho_1 \langle D, H_j \rangle + R_j' \\ \leq & e_n p \|H_j - u_j u_j^T\|_F - \rho_1 \langle D, H_j \rangle + R_j' \,, \end{aligned} \tag{A.7}$$

where $R_j' = c_2 p(\epsilon_{j-1} + \sqrt{2} c_0 p^{-1}) + R_j + 3c_0 c_2 / 2$. It follows that

$$\|H_j - u_j u_j^T\|_F \leq \frac{8e_n}{\delta} + \sqrt{\frac{8R_j'}{\delta p}} \,. \tag{A.8}$$

Since $\widehat{v}_j \widehat{v}_j^T$ is the closest rank one, unit norm matrix to H_j, we have

$$\begin{aligned} & \|\widehat{v}_j \widehat{v}_j^T - v_j v_j^T\|_F \leq \|\widehat{v}_j \widehat{v}_j^T - u_j u_j^T\|_F + \|u_j u_j^T - v_j v_j^T\|_F \\ \leq & 2\|H_j - u_j u_j^T\|_F + \sqrt{2} c_0 p^{-1} \leq \frac{16 e_n}{\delta} + 2\sqrt{\frac{8R_j'}{\delta p}} + \sqrt{2} c_0 p^{-1} \,, \end{aligned}$$

and

$$\begin{aligned} & \|\widehat{\Pi}_j - \Pi_j\|_F \leq \|\widehat{\Pi}_{j-1} - \Pi_{j-1}\|_F + \|\widehat{v}_j \widehat{v}_j^T - v_j v_j^T\|_F \\ \leq & \epsilon_{j-1} + \frac{16 e_n}{\delta} + 2\sqrt{\frac{8R_j'}{\delta p}} + \sqrt{2} c_0 p^{-1} =: \epsilon_j \,. \end{aligned}$$

Now it remains to find β_j. Using (A.7) we have

$$\rho_1 \langle D, H_j \rangle \leq e_n p \|H_j - u_j u_j^T\|_F + R_j' \leq 2 e_n p \epsilon_j + R_j' \,.$$

On the other hand, let $\lambda_{j,1}$ be the largest eigenvalue of H_j. Then

$$\frac{\epsilon_j^2}{4} \geq \|H_j - u_j u_j^T\|_F^2 = \|H_j\|_F^2 - 2\langle H_j, u_j u_j^T \rangle + 1 \geq \lambda_{j,1}^2 - 2\lambda_{j,1} + 1 \,,$$

where we use the fact that $\|H_j\|_F^2 \ge \lambda_{j,1}^2$, and $\langle H_j, u_j u_j^T\rangle \le \lambda_{j,1}\|u_j\|_2^2$ (von Neumann trace inequality). It then follows that $\lambda_{j,1} \ge 1-\epsilon_j/2$, which implies that

$$\rho_1\langle D, \widehat{v}_j\widehat{v}_j^T\rangle \le (1-\epsilon_j/2)^{-1}\rho_1\langle D, H_j\rangle \le (1-\epsilon_j/2)^{-1}\left(2e_n p\epsilon_j + R_j'\right) =: \beta_j\,.$$

Direct verification shows that if $\max_{0\le i\le j-1}\epsilon_i = o_p(1)$, $\max_{0\le i\le j-1}\beta_j = o_p(p)$, and $R_j = o_p(p)$, then $\epsilon_j = o_p(1)$ and $\beta_j = o_p(p)$.

The rest of the proof is to show (A.6) for some $R_j = o_p(p)$. The main challenge is that u_j is not in the feasible set of problem (6) and hence $u_j u_j^T$ is not directly comparable to H_j using optimality condition of (6). To overcome this difficulty, we consider $\widetilde{u}_j$, a modified version of u_j so that (a) $\widetilde{u}_j$ is close to u_j in ℓ_2 norm; (b) $\widetilde{u}_j\widetilde{u}_j^T$ is feasible for (6); (c) $\widetilde{u}_j$ is almost as smooth as u_j.

Define $\widetilde{u}_j = (I-\widehat{\Pi}_{j-1})u_j/\|(I-\widehat{\Pi}_{j-1})u_j\|$. We first check the validity of this definition.

$$\|\widehat{\Pi}_{j-1}u_j\|_2 \le \|(\widehat{\Pi}_{j-1}-\Pi_{j-1})u_j\|_2 + \|\Pi_{j-1}v_j\|_2 + \|\Pi_{j-1}(u_j-v_j)\|_2 \le \epsilon_{j-1} + c_0 p^{-1}\,.$$

When ϵ_{j-1} is small and p large, $(I-\widehat{\Pi}_{j-1})u_j \ne 0$, and

$$\|\widetilde{u}_j - u_j\|_2 = \left\|\frac{(I-\widehat{\Pi}_{j-1})u_j}{\|(I-\widehat{\Pi}_{j-1})u_j\|_2} - \frac{u_j}{\|u_j\|_2}\right\|_2 \le 2\|(I-\widehat{\Pi}_{j-1})u_j - u_j\|_2 \le 2(\epsilon_{j-1} + c_0 p^{-1})\,, \tag{A.9}$$

where the last inequality holds when p is large and the first inequality follows from an elementary fact that, for all u, v,

$$\left\|\frac{u}{\|u\|_2} - \frac{v}{\|v\|_2}\right\|_2 \le 2\frac{\|u-v\|_2}{\max(\|u\|_2, \|v\|_2)}\,.$$

Now we establish (A.6). By feasibility of $\widetilde{u}_j\widetilde{u}_j^T$ we have

$$\begin{aligned}0 \le& \langle S, H_j - \widetilde{u}_j\widetilde{u}_j^T\rangle - \rho_1\langle D, H_j\rangle + \rho_1\langle D, \widetilde{u}_j\widetilde{u}_j^T\rangle - \rho_2(\|H_j\|_{1,1} - \|\widetilde{u}_j\widetilde{u}_j^T\|_{1,1})\\ \le& \langle S, H_j - u_j u_j^T\rangle - \rho_1\langle D, H_j\rangle + \rho_1\langle D, \widetilde{u}_j\widetilde{u}_j^T\rangle + \rho_2 p + |\langle S, \widetilde{u}_j\widetilde{u}_j^T - u_j u_j^T\rangle|\,.\end{aligned} \tag{A.10}$$

We first bound $|\langle S, \widetilde{u}_j\widetilde{u}_j^T - u_j u_j^T\rangle|$:

$$|\langle S, \widetilde{u}_j\widetilde{u}_j^T - u_j u_j^T\rangle| \le \|S\|_F\|\widetilde{u}_j\widetilde{u}_j^T - u_j u_j^T\|_F \le (\|\Sigma\|_F + \|S-\Sigma\|_F)\|\widetilde{u}_j\widetilde{u}_j^T - u_j u_j^T\|_F$$

$$\leq 2\sqrt{2}\,(c_2+e_n)\,p(\epsilon_j+c_0p^{-1})\,,$$

where the last step uses (A.9), Lemma C.2, and the fact that $\|S-\Sigma\|_{\infty,\infty}=e_n$.

Now we control $\rho_1\langle D,\widetilde{u}_j\widetilde{u}_j^T\rangle$. When ϵ_{j-1} and p^{-1} are small enough such that $\|(I-\widehat{\Pi}_{j-1})u_j\|_2\geq 1/\sqrt{2}$, we have

$$\rho_1\langle D,\widetilde{u}_j\widetilde{u}_j^T\rangle=\rho_1\|\Delta\widetilde{u}_j\|_2^2\leq 2\rho_1\|\Delta(I-\widehat{\Pi}_{j-1})u_j\|_2^2$$
$$\leq 4\rho_1\left[\|\Delta u_j\|_2^2+\left(\sum_{i=1}^{j-1}\|\Delta\widehat{v}_i\|_2|\langle\widehat{v}_i,u_j\rangle|\right)^2\right]\leq 4\left[2\rho_1L^2p^{-4}+\sum_{i=1}^{j-1}\beta_i(\epsilon_{j-1}+c_0p^{-1})^2\right]\,,$$

where the first two inequalities follow from multiple applications of Cauchy-Schwartz, and the last inequality holds by definition of β_i, the smoothness of u_j, and the fact that $\sum_{i=1}^{j-1}|\langle\widehat{v}_i,u_j\rangle|^2=\|\widehat{\Pi}_{j-1}u_j\|_2^2$. As a consequence, we establish (A.6) from (A.10) with

$$R_j=2\sqrt{2}\,(c_2+e_n)\,(\epsilon_{j-1}p+c_0)+\rho_2p+8\left[\rho_1L^2p^{-4}+\sum_{i=1}^{j-1}\beta_i(\epsilon_{j-1}+c_0p^{-1})^2\right]=o_p(p)\,. \quad \square$$

Proof of Theorem 4.2. From assumption A1 and Lemma C.1 it suffices to prove that if $\|S-\Sigma\|_{\infty,\infty}=O(e_n)$ then $\sup_{1\leq j\leq k}\|\widehat{v}_j-v_j\|_2=O(e_n+\rho_2)$.

Consider the estimation procedure given by (5) for $j=1$. Let $\mathbb{B}_1$ be the collection of all $p\times p$ symmetric matrices with entries in $[-1,1]$. The optimization problem can be written in the following equivalent form.

$$\max_{H\in\mathcal{F}^1}\min_{Z\in\mathbb{B}_1}\langle S,H\rangle-\rho_2\langle Z,H\rangle\,.$$

Let H^* be any maximizer, then

$$H^*=\arg\max_{H\in\mathcal{F}^1}\langle S-\rho_2Z^*,H\rangle=\arg\max_{H\in\mathcal{F}^1}\langle\Sigma+W-\rho_2Z^*,H\rangle$$

where $W=S-\Sigma$ and Z^* is the corresponding optimal dual variable.

By Lemma C.1 the eigengap of Σ is of order at least p while the operator norm of $W-\rho_2Z^*$ is at most $p(\|W\|_{\infty,\infty}+\rho_2)$ which is $o(p)$. Thus applying standard spectral subspace perturbation theory we know that $H^*=\widehat{v}_1\widehat{v}_1^T$ where $\widehat{v}_1$ is the leading eigenvector of $\Sigma+W-\rho_2Z^*$, and satisfies for some constant c

$$\|\widehat{v}_1-v_1\|_2\leq c\|W-\rho_2Z^*\|_F/p\leq c(e_n+\rho_2)\,.$$

For $j = 2, ..., k$, we use induction. Suppose that for $j-1$ we have $\|\widehat{v}_{j-1} - v_{j-1}\|_2$ and $\|\widehat{\Pi}_{j-1} - \Pi_{j-1}\|_F$ are bounded by $O(e_n + \rho_2)$.

Now consider the procedure (5) for j. Similarly let H^* be any solution and Z^* the corresponding optimal dual variable. We have

$$\begin{aligned} H^* =& \arg\max_{H \in \mathcal{D}_{\widehat{\Pi}_{j-1}}} \langle S - \rho_2 Z^*, H \rangle \\ =& \arg\max_{H \in \mathcal{F}^1} \langle (I - \widehat{\Pi}_{j-1})(S - \rho_2 Z^*)(I - \widehat{\Pi}_{j-1}), H \rangle \,. \end{aligned}$$

The remainder of the proof focuses on analyzing the matrix $(I - \widehat{\Pi}_{j-1})(S - \rho_2 Z^*)(I - \widehat{\Pi}_{j-1})$. In particular, we show that its leading eigenvector is close to v_j with the desired rate. We first write this matrix in four terms

$$\begin{aligned} & (I - \widehat{\Pi}_{j-1})(\Sigma + W - \rho_2 Z^*)(I - \widehat{\Pi}_{j-1}) \\ =& (I - \Pi_{j-1})\Sigma(I - \Pi_{j-1}) \\ & + (\Pi_{j-1} - \widehat{\Pi}_{j-1})\Sigma(I - \widehat{\Pi}_{j-1}) + (I - \Pi_{j-1})\Sigma(\Pi_{j-1} - \widehat{\Pi}_{j-1}) \\ & + (I - \widehat{\Pi}_{j-1})(W + \rho_2 Z^*)(I - \widehat{\Pi}_{j-1}) \\ =& T_0 + T_1 + T_2 \,. \end{aligned}$$

The main term is T_0. The leading eigenvector of T_0 is v_j with an eigengap at least $\delta p/2$ according to Lemma C.2. Next we bound T_1 and T_2. In fact we have

$$\|T_1\|_F \le 2\|\Pi_{j-1} - \widehat{\Pi}_{j-1}\|_F \|\Sigma\|_2 \le 2c_2 \|\Pi_{j-1} - \widehat{\Pi}_{j-1}\|_F p \,, \tag{A.11}$$

where c_2 is the constant in Lemma C.2 and

$$\|T_2\|_F \le \|W + \rho_2 Z^*\|_F \le (e_n + \rho_2)p \,.$$

Then we have

$$\|T_1 + T_2\|_F \le 2c_2 \|\Pi_{j-1} - \widehat{\Pi}_{j-1}\|_F p + (e_n + \rho_2)p \,. \tag{A.12}$$

When n and p are large enough, $\|T_1 + T_2\|_F$ is smaller than the gap between the first and second largest eigenvalues of T_0. Therefore, the induction completes by using Davis-Kahan

$\sin\Theta$ theorem (Bhatia, 1997, Theorem VII.3.1)

$$\|\widehat{v}_j\widehat{v}_j^T - v_j v_j^T\|_F \leq 2\|T_1 + T_2\|_F/(\delta p/2) \leq 8c_2\delta^{-1}\|\Pi_{j-1} - \widehat{\Pi}_{j-1}\|_F + 4\delta^{-1}(e_n + \rho_2)\,, \quad \text{(A.13)}$$

where c_2 is the constant given in Lemma C.2. □

Proof of Technical Lemmas

Proof of Lemma C.1. Note that $\widetilde{\Gamma}$ is a compact self-adjoint operator from $L^2(0,1)$ to $L^2(0,1)$ with eigen-decomposition $\widetilde{\Gamma}(s,t) = \sum_{j=1}^{p} \widetilde{\lambda}_j \widetilde{\phi}_j(s)\widetilde{\phi}_j(t)$.

The Lipschitz condition on Γ implies that

$$\|\widetilde{\Gamma} - \Gamma\|_{\text{HS}}^2 := \int\int |\Gamma_p(s,t) - \Gamma(s,t)|^2 ds dt \leq \frac{L^2}{4p^2}\,. \quad \text{(A.14)}$$

By Weyl's inequality, $|\widetilde{\lambda}_j - \lambda_j| \leq \delta/2$ for large p. Let E_j and $\widetilde{E}_j$ be the projection operators onto the one-dimensional subspaces spanned by ϕ_j and $\widetilde{\phi}_j$, respectively. Then

$$\|\widetilde{\phi}_j - \phi_j\|_2 \leq \sqrt{2}\|\widetilde{E}_j - E_j\|_{\text{HS}} \leq \frac{4\|\widetilde{\Gamma} - \Gamma\|_{\text{HS}}}{\delta} \leq \frac{2L}{\delta p}\,, \quad \text{(A.15)}$$

where the first inequality follows from Lemma C.3, and the second inequality follows from the Davis-Kahan $\sin\Theta$ theorem (Chapter VII of Bhatia (1997)).

On the other hand, by assumption (A4) we have

$$\|\phi_j^* - \phi_j\|_2 \leq \|\phi_j^* - \phi_j\|_\infty \leq \frac{L}{2p}\,. \quad \text{(A.16)}$$

which, together with (A.15), implies that

$$\|v_j - u_j^*\|_2 \leq \left(\frac{2}{\delta} + \frac{1}{2}\right)\frac{L}{p}\,.$$

Also note that

$$\|u_j^* - u_j\|_2 = \big|\|u_j^*\|_2 - 1\big| \leq \|\phi_j^* - \phi_j\|_2 \leq \|\phi_j^* - \phi_j\|_\infty \leq \frac{L}{2p}\,.$$

Combining the previous two inequalities, we have

$$\|v_j - u_j\|_2 \leq \left(\frac{2}{\delta} + 1\right)\frac{L}{p} := c_0 p^{-1}\,.$$

□

Proof of Lemma C.2. The first claim follows from, letting Σ^* be the discretized Γ evaluated at the grid,

$$
\begin{aligned}
\|\Sigma\|_F^2 \leq & 2\|aI_p\|_F^2 + 2\|\Sigma^*\|_F^2 = 2a^2p + 2p^2\|\widetilde{\Gamma}\|_{\mathrm{HS}}^2 \\
\leq & 2a^2p + 2p^2\left(2\|\Gamma\|_{\mathrm{HS}}^2 + 2\|\widetilde{\Gamma} - \Gamma\|_{\mathrm{HS}}^2\right) \leq 2a^2p + 2p^2\left(2\|\Gamma\|_{\mathrm{HS}}^2 + \frac{L^2}{2p^2}\right) \\
\leq & p^2\left(2a^2p^{-1} + 4\|\Gamma\|_{\mathrm{HS}}^2 + L^2\right) \leq c_2^2p^2\,,
\end{aligned}
$$

where (A.14) is used to bound $\|\widetilde{\Gamma} - \Gamma\|_{\mathrm{HS}}$.

The second claim follows from the fact that the eigengaps of Σ are the same as those of Σ^*, and by Weyl's inequality:

$$
\widetilde{\lambda}_j - \widetilde{\lambda}_{j+1} \geq \lambda_j - \lambda_{j+1} - 2\|\widetilde{\Gamma} - \Gamma\|_{\mathrm{HS}} \geq \delta - \frac{\delta}{2} = \delta/2\,. \qquad \square
$$

Proof of Lemma C.4. Note that v_j is the leading eigenvector of Σ_j, with eigengap at least $p\delta/2$ as implied by Lemma C.2. Then we have

$$
\begin{aligned}
\|H_j - u_ju_j^T\|_F^2 \leq & 2\|H_j - v_jv_j^T\|_F^2 + 2\|v_jv_j^T - u_ju_j^T\|_F^2 \leq \frac{8}{p\delta}\langle -\Sigma, H_j - v_jv_j\rangle + 4c_0^2p^{-2} \\
\leq & \frac{8}{p\delta}\langle -\Sigma, H_j - u_ju_j^T\rangle + \frac{8}{p\delta}\|\Sigma\|_F\|u_ju_j^T - v_jv_j^T\|_F + 4c_0^2p^{-2} \\
\leq & \frac{8}{p\delta}\langle -\Sigma, H_j - u_ju_j^T\rangle + \frac{8\sqrt{2}c_0c_2}{p\delta} + 4c_0^2p^{-2} \\
\leq & \frac{8}{p\delta}\langle -\Sigma, H_j - u_ju_j^T\rangle + \frac{12c_0c_2}{p\delta}\,,
\end{aligned}
$$

where the first and third inequalities come from Cauchy-Schwartz, the second from the curvature lemma of principal subspace (Vu and Lei (2013), Lemma 4.2), and the last holds provided that p is sufficiently large. $\square$