	Covariance estimation ($\mathbb{E}\|\hat{\Sigma}-\Sigma\|_2^2$ or canonical angle)	Eigenvalues	Eigenvectors	PCR ($\hat{Y}$)	PCR ($\hat{\beta}$)
Nystrom					
CS			CS_3		
Martinsson					

Table 1: Note that the CS results apply to AIMER.

1 Table for matrix sketching results

2 Preliminary notation and definitions

- $p = \{1, \ldots, p\}$ (it probably won't be confusing to use p as either the number of covariates or the index set, depending on context)
- $\mathcal{A} = \{$ of active covariates $\}$
- $\mathcal{S} = \{$ nonzero marginal covariance $\}$ (using $\mathcal{S}$ as it is the 'selected' model)
- $\mathcal{D} = \mathcal{A} \setminus \mathcal{S}$ (to be the difference between active and selected covariates)
- $\mathcal{T} = \{$ nonzero θ $\}$ (using $\mathcal{T}$ due to whatever)

The underlying machinery of these supervised PCA papers is a suite of estimators of the form $\hat{\Sigma}_{A,B}$, where $A, B \subseteq \{1, \ldots, p\}$. In the SPCA paper, they choose $\hat{\Sigma}_{\mathcal{S},\mathcal{S}}$. Using F is tantamount of using $\hat{\Sigma}_{p,\mathcal{S}}$. This protects somewhat against $\mathcal{S} \subset \mathcal{A}$. If we had a good estimator of $\mathcal{T}$, we would could use $\hat{\Sigma}_{\mathcal{T},\mathcal{S}}$, instead. Perhaps this estimator should in investigated as well...

3 Showing CS_3

Using the result that

$$\|\hat{v} - v\|_2^2 \leq 2\sin(\angle(\hat{v}, v))$$

we can do the following. Supposing that $\Sigma = [\tilde{\Sigma}_{\mathcal{S}}|\tilde{\Sigma}_{\mathcal{S}^c}]$, $F = V(F)D(F)U(F)^\top$, $\tilde{\Sigma} = VDU^\top$, and $\Sigma = \Theta\Lambda\Theta^\top$, then for $k \in \mathcal{S}$,

$$\|v_q(F) - \theta_q\|_2 \leq \|v_q(F) - v_q\|_2 + \|v_q - \theta_q\|_2 \leq \sqrt{2}\left(\sin(\angle(v_q(F), v_q)) + \sin(\angle(v_q, \theta_q))\right).$$

So, by Yu et al. (2015)[1], Theorem 3

$$\sin(\angle(v_q(F), v_q)) \leq 2\frac{(2d_{\max} + \|F - \tilde{\Sigma}\|_{op})\min\{\|F - \tilde{\Sigma}\|_{op}, \|F - \tilde{\Sigma}\|_F\}}{\tilde{\delta}_q},$$

where $\tilde{\delta}_q = \min\{d_{q-1} - d_q, d_q - d_{q+1}\}$. This quantity will be controlled by assumption on Σ (for instance, $d_{\max} \leq \lambda_{\max}$).

[1] http://www.statslab.cam.ac.uk/~yy366/index_files/Biometrika-2015-Yu-biomet_asv008.pdf

Now, looking at $\|F - \tilde{\Sigma}\|_F^2$ component wise for $j \in p$ and $k \in \mathcal{S}$

$$(F(j,k) - \tilde{\Sigma}(j,k))^2 = (\mathbf{x}_j^\top \mathbf{x}_k - \mathbb{E}x_j x_k)^2.$$

Here $\mathbf{x}_j$ is the j^{th} column of $\mathbb{X}$ and $x_j \sim N(0, \Sigma(j,j))$. This will be controllable via asymptotics or concentration.

There will be nonzero approximation bias if $\mathcal{D} \neq \emptyset$. Using the same result as above

$$\sin(\angle(v_q, \theta_q)) \leq 2\frac{(2\lambda_{\max} + \|\tilde{\Sigma} - \Sigma\|_{op}) \min\{\|\tilde{\Sigma} - \Sigma\|_{op}, \|\tilde{\Sigma} - \Sigma\|_F\}}{\delta_q},$$

where $\delta_q = \min\{\lambda_{q-1} - \lambda_q, \lambda_q - \lambda_{q+1}\}$. This quantity will again be controlled by assumption on Σ.

Now, looking at $\|\tilde{\Sigma} - \Sigma\|_F^2$ component wise for $j, k \in p$

$$(\tilde{\Sigma}(j,k)) - \Sigma(j,k))^2 = \begin{cases} 0 & \text{if } k \in \mathcal{S} \\ (\sum_{m=1}^M \lambda_m \theta_j \theta_k)^2 & \text{if } j \in \mathcal{A}, k \in \mathcal{D} \\ (\sum_{m=1}^M \lambda_m \theta_j \theta_k + \sigma^2)^2 & \text{if } j = k \in \mathcal{D} \\ (\sigma^2)^2 & \text{if } j = k \notin \mathcal{A} \end{cases}$$

Now, we might make some assumptions about the side of this "residual" components, due to a norm constraint on these components implying a norm constraint on the β's. So, the result might look like if

- $(\sum_{m=1}^M \lambda_m \theta_j \theta_k)^2 \leq \gamma_n$
- We can estimate σ^2 well so we consider it known (really, just to simplify things so we can just subtract off the diagonal component before hand)
- $\lambda_{\max} \leq C_\Lambda$ independent of n

Then

$$\|\tilde{\Sigma} - \Sigma\|_F^2 \leq |\mathcal{A}||\mathcal{D}|\gamma_n,$$

which implies that

$$\sin(\angle(v_q, \theta_q)) \leq 2\frac{(2\lambda_{\max} + |\mathcal{A}||\mathcal{D}|\gamma_n)|\mathcal{A}||\mathcal{D}|\gamma_n}{\delta_q},$$