1 Overview

Suppose $A \in \mathbb{R}^{m\times n}$ and $\tilde{A} = A + E$. Also, $U^\top AV = \begin{bmatrix} D \\ 0 \end{bmatrix}$ and $\tilde{U}^\top \tilde{A}\tilde{V} = \begin{bmatrix} \tilde{D} \\ 0 \end{bmatrix}$. There is a main theorem for bounding the distance between D and $\tilde{D}$:

Theorem 1.1 (Mirsky)**.**

$$\|\tilde{D} - D\| \leq \|E\|$$

where the norm can be any unitarily equivalent norm (e.g. $\|\cdot\|_2$ *or* $\|\cdot\|_F$*).*

2 Our purposes

Let the covariance matrix for X be

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$$

where

$$\Sigma_1 = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{12} \end{bmatrix} \in \mathbb{R}^{p_1\times p_1}$$

and $\Sigma_{1s} \in \mathbb{R}^{p_{1s}\times p_{1s}}$ for $s = 1, 2$. Let $\mathcal{S} = \{1, \ldots, p_1\}$ and $\mathcal{S}^c = \{1, \ldots, p\} \setminus \mathcal{S}$. Lastly, we define $\Sigma_{\mathcal{S}} = [\Sigma_1, 0]^\top$ to be the first p_1 columns of Σ.

The goal here is to extend the results of the SPCA paper by including the possibility in Σ_1 that we have missed some important features (hence their Σ_1 corresponds to our Σ_{11}. The model for X is then (here I'm writing/thinking about a single latent factor model, I'm presuming that complexifying that to multi-factor will be a matter of notation):

$$X_{ij} = v_i\rho_j + \sigma z_{ij}$$

where v_i, z_{ij} are all mutually independent standard normals and $\rho_j \neq 0$ iff $j \in \mathcal{S}$.

2.1 The result

As $F = \mathbb{X}^\top\mathbb{X}_1$, then[1]

$$F = \left[\sum_{i=1}^{n}(v_i\rho_j + \sigma z_{ij})(v_i\rho_k + \sigma z_{ik})\right]_{1\leq j\leq p, k\in\mathcal{S}} \tag{1}$$

$$= \begin{bmatrix} \sum_{i=1}^n(v_i\rho_j + \sigma z_{ij})(v_i\rho_k + \sigma z_{ik}) \\ \sum_{i=1}^n(\sigma z_{ij})(v_i\rho_k + \sigma z_{ik}) \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} \sum_{i=1}^n(v_i^2\rho_j\rho_k + \sigma z_{ij}v_i\rho_k + \sigma z_{ik}v_i\rho_j + \sigma^2 z_{ij}z_{ik}) \\ \sum_{i=1}^n(\sigma z_{ij}v_i\rho_k + \sigma^2 z_{ij}z_{ik}) \end{bmatrix}, \tag{3}$$

where the top block has $j \in S$ and the bottom block has $j \in \mathcal{S}^c$ (this convention will persist for the rest of the document).

[1]We will need to divide through by some function of n.

Using the result from the previous theorem, write $A = \Sigma_{\mathcal{S}}$ and $E = F - \Sigma_{\mathcal{S}}$. Hence, the singular values of $\Sigma_{\mathcal{S}}$ and F will be close if $F - \Sigma_{\mathcal{S}}$ is small.

Note the expectation of F:

$$\mathbb{E}F = \begin{bmatrix} n\rho\rho^\top + n\sigma^2 I \\ 0 \end{bmatrix}.$$

Hence,

$$E = \begin{bmatrix} F - \mathbb{E}F \end{bmatrix}. - \begin{bmatrix} n\rho\rho^\top + n\sigma^2 I \\ 0 \end{bmatrix}.$$

So, up to the $n\sigma^2 I$ factor, we have to bound the norm difference between a random matrix and it's expectation.

$$\begin{aligned} F - \mathbb{E}F - \begin{bmatrix} n\rho\rho^\top + n\sigma^2 I \\ 0 \end{bmatrix} &= \begin{bmatrix} \sum_{i=1}^n v_i^2 \rho_j \rho_k - n\rho_j\rho_k \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma \sum_{i=1}^n v_i (z_{ij}\rho_k + z_{ik}\rho_j) \\ \sigma \sum_{i=1}^n v_i z_{ij} \rho_k + \sigma^2 \sum_{i=1}^n z_{ij} z_{ik} \end{bmatrix} + \quad (4) \\ &+ \begin{bmatrix} \sigma^2 \sum_{i=1}^n z_{ij} z_{ik} - n\sigma^2 I \\ 0 \end{bmatrix} \quad (5) \\ &= (i) + (ii) + (iii). \quad (6) \end{aligned}$$

Now,

$$(i) = (\sum_{i=1}^n v_i^2 - n) \begin{bmatrix} \rho\rho^\top \\ 0 \end{bmatrix}$$

and

$$(iii) = \sigma^2 \begin{bmatrix} \sum_{i=1}^n Z_{i\mathcal{S}} Z_{i\mathcal{S}}^\top - nI \\ 0 \end{bmatrix},$$

where $Z_{i\mathcal{S}} = [z_{ij}]_{j \in \mathcal{S}}$. Similarly, $Z_{i\mathcal{S}^c} = [z_{ij}]_{j \in \mathcal{S}^c}$. Hence,

$$(ii) = \sigma \sum_{i=1}^n v_i \begin{bmatrix} Z_{i\mathcal{S}} \rho^\top \\ Z_{i\mathcal{S}^c} \rho^\top \end{bmatrix} + \sigma \sum_{i=1}^n v_i \begin{bmatrix} \rho Z_{i\mathcal{S}}^\top \\ 0 \end{bmatrix} + \sigma^2 \sum_{i=1}^n \begin{bmatrix} 0 \\ Z_{i\mathcal{S}} Z_{i\mathcal{S}^c}^\top \end{bmatrix}.$$

Now, concentration results can be used to show E is small in Frobenius norm. Presumably, we can find some results about spectral norm as well (which would probably be more useful as it would allow us to say $|d_j - \tilde{d}_j| \le \|E\|_2$ for all j).