# 1 Preliminary notation and definitions

- Let the covariance matrix for $X$ be
$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{1}$$
- $X_{ij} = \sum_{m=1}^{M} \lambda_m^{1/2} \eta_{im} \rho_{jm} + \sigma z_{ij}$ where $\|\rho_m\|_2^2 = 1$ and $\rho_m \equiv \rho_{\cdot m}$, $\langle \rho_m, \rho_{m'} \rangle = 0$ if $m \neq m'$, and $\rho_{jm} = 0$ if $j \geq p_1$.
- $\Sigma_1 = \sum_{m=1}^{M} \lambda_m \rho_m \rho_m^\top + \sigma^2 I_{p_1}$.
- $F = \mathbb{X}^\top \mathbb{X}_1 = V(F)\Lambda(F)U(F)^\top$ (note, I think this reversed order makes much more sense at we are looking at approximating $\mathbb{X}^\top\mathbb{X} = VD^2V^\top$...)
- The regression model:
$$Y_i = \beta_0 + \sum_{m=1}^{\tilde{M}} \beta_m \eta_{im} + W_i. \tag{2}$$
Here, I write $\tilde{M}$ to indicate the this may be different than $M$.
- Lastly, I haven't included any normalization by a function of $n$, which is surely necessary to get convergence. In particular, the sample covariance would be $n^{-1}\mathbb{X}^\top\mathbb{X}$, so defining $F \leftarrow n^{-1}F$ would seemingly make sense.

# 2 Proof outline/sketch

1. Show that $v_m(F)$ is close to $\rho_m$ (the PC loadings) and $\lambda_m(F)$ is close to $\lambda_m$
   (a) This is the topic of the document "convergenceSingularVectorsValues.pdf". We need show that $v_m(F)$ converges to $\rho_m$. So, perhaps, $v_m(F) = \rho_m + \delta_m$, where $\|\delta_m\|$ is small (note: we need to formalize the connection between bounded sin( canonical angles) of singular vectors and writing them in the fashion. Perhaps the asymptotic expansion is more amenable?)
2. The regression part of the procedure regresses $Y$ onto the PC scores, which are the coordinates in the PC, given by $\hat{u}_m = \mathbb{X} v_m(F) \lambda_m^{-1/2}(F)$. We need to show that these coordinates aren't too far from the coordinates created by inner product with $\rho_{m'}$:
$$\left\langle \sum_{m=1}^{M} \eta_{im} \rho_m, \rho_{m'} \right\rangle = \eta_{i,m'} \lambda_{m'} \tag{3}$$
   (a) This can be done via inserting the model for $X$ in for $\mathbb{X}$ in the definition of $\hat{u}_k$.
$$\mathbb{X} v_k(F) = \begin{bmatrix} \sum_{j=1}^{p} \left( \sum_{m=1}^{M} \lambda_m^{1/2} \eta_{1m} \rho_{jm} + \sigma z_{1j} \right) v_{jk}(F) \\ \vdots \\ \sum_{j=1}^{p} \left( \sum_{m=1}^{M} \lambda_m^{1/2} \eta_{nm} \rho_{jm} + \sigma z_{nj} \right) v_{jk}(F) \end{bmatrix} = \sum_{m=1}^{M} \lambda_m^{1/2} \rho_m^\top v_k(F) \begin{bmatrix} \eta_{1m} \\ \vdots \\ \eta_{nm} \end{bmatrix} + \sigma \begin{bmatrix} z_1^\top v_k(F) \\ \vdots \\ z_n^\top v_k(F) \end{bmatrix}. \tag{4}$$

Using the approximation: $v_k(F) = \rho_k + \delta_k$,

$$\eta_{im}\rho_m^\top v_k(F) = \eta_{im}\rho_m^\top(\rho_k + \delta_k) = \eta_{im}(\rho_m^\top\rho_k + \rho_m^\top\delta_k) = \begin{cases} \eta_{ik}(1 + \rho_k^\top\delta_k) & \text{if } k = m \\ \eta_{im}(\rho_m^\top\delta_k) & \text{if } k \neq m \end{cases} \tag{5}$$

i. Fix $k \neq m$:

$$\eta_{im}\lambda_m^{1/2}\rho_m^\top v_k(F)\lambda_k^{-1/2}(F) = \left(\frac{\lambda_m}{\lambda_k(F)}\right)\eta_{im}(\rho_m^\top\delta_k) \tag{6}$$

So, we need the ratio of eigenvalues to be bounded and then perhaps

$$|\rho_m^\top\delta_k| \leq \|\delta_k\|_2 = o(\text{some rate}). \tag{7}$$

ii. Fix $k = m$:

$$\eta_{ik}\lambda_k^{1/2}\rho_k^\top v_k(F)\lambda_k^{-1/2}(F) = \left(\frac{\lambda_k}{\lambda_k(F)}\right)\eta_{ik}(1 + \rho_k^\top\delta_k) \tag{8}$$

Now, we need the ratio of eigenvalues to go to one (implied by the perturbation bound?) and using the above bound in equation (7):

$$\left(\frac{\lambda_k}{\lambda_k(F)}\right)\eta_{ik}(1 + \rho_k^\top\delta_k) \to \eta_{ik} \tag{9}$$

(b) Combining (i) and (ii)

$$\sum_{m=1}^{M}\lambda_m^{1/2}\rho_m^\top v_k(F)\begin{bmatrix}\eta_{1m}\\ \vdots \\ \eta_{nm}\end{bmatrix} = \begin{bmatrix}\eta_{1k}\\ \vdots \\ \eta_{nk}\end{bmatrix} + o(\text{some other rate}) \tag{10}$$

(c) Lastly, we need to show that the measurement error term is bounded:

$$\sigma\begin{bmatrix}z_1^\top v_k(F)\\ \vdots \\ z_n^\top v_k(F)\end{bmatrix}.$$

This needs to be addressed with care as $z$ and $v$ are dependent.

3. We need to write down the form of the estimator: $\hat{U}_{\tilde{M}}^\top Y$. Plug in the regression model for $Y$ (equation (2)):

$$\hat{\beta}_m = \hat{u}_m^\top Y = \beta_0\hat{u}_m^\top\mathbf{1} + \sum_{m=1}^{\tilde{M}}\beta_m\hat{u}_m^\top\eta_m + \hat{u}_m^\top W = \text{(a) + (b) + (c)} \tag{11}$$

we need to write the regression model for $Y$ in terms of these estimated coordinates:

(a) Maybe we can get rid of this via a max norm bound?

$$|\hat{u}_m^\top\mathbf{1}| \leq \|\hat{u}_m\|_1\|\mathbf{1}\|_\infty = \|\hat{u}_m\|_1 \tag{12}$$

There should be something like a $n^{-1/2}$ running around. So, this would require that $\|\hat{u}_m\|_1 = o(n^{1/2})$, which isn't that likely.

(b) Apply the above results that show that $\hat{u}_m \approx \eta_m$ and hence

$$\beta_m \hat{u}_m^\top \eta_m \approx \beta_m \|\eta_m\|_2^2$$

So, if we have a $n^{-1}$ floating around, then $n^{-1}\|\eta_m\|_2^2 \to 1$ and

$$\beta_m \|\eta_m\|_2^2 \to \beta_m.$$

(c) $\hat{u}_m$ and $W$ are independent, so this can be shown to be small using a concentration bound (mean zero)