1. (MKB 8.2.5) Almost everybody did a lot of unnecessary work for this problem. The simple solution is that given data $X$, we can use the empirical probability distribution based on $X$ to apply the population results directly to the sample. In other words, define a pdf $\hat{F}$ by giving probability of $1/n$ to each one of the points in $X$. Now we can easily show that $\hat{F}$ has mean $\bar{x}$ and variance $S$ and thus that its principal components satisfy all the results in 8.2.1-8.2.3. But the principal components of $\hat{F}$ are trivially the same as the sample PC of $X$, and their sample moments are the same as the population moments under $\hat{F}$. Hence we are done...

2. (MKB 8.2.7) $h_i \geq 0$ is trivial, and almost as trivial is:

$$\sum_i h_i = \sum_i \sum_j f_{ij}^2 = \sum_{j=1}^k (\sum_i f_{ij}^2) = \sum_{j=1}^k 1 = k$$

To prove $h_i \leq 1$ we can imagine completing $F$ to an orthogonal matrix by adding $p - k$ appropriate columns (we have Linear Algebra results to verify we can do that). Denote that $p \cdot p$ matrix by $G$. Both $G$’s columns and its rows have norm 1, so:

$$\forall i \quad 1 = \sum_j g_{ij}^2 = \sum_{j=1}^k f_{ij}^2 + \sum_{j=k+1}^p g_{ij}^2 \geq \sum_{j=1}^k f_{ij}^2 = h_i$$

Thus we can deduce that $\Phi(h)$ is maximized by $h_1 = \ldots = h_k = 1$ at their maximal possible value and $h_{k+1} = \ldots = h_p = 0$

3. (MKB 8.4.2) This is straightforward calculations.

4. (MKB 8.7.1) This is straightforward.

5. (MKB 8.8.1)

(a) When $k = 0$ we get $\beta^* = (X'X + 0I)^{-1}X'y = \hat{\beta}_{OLS}$

(b) Let $X'X = GLG'$ be a spectral decomposition, where $L$ is diagonal with $l_1 \geq l_2 \geq \ldots \geq l_p$. We define $W = XG$, $\alpha = G'\beta$ so $W\alpha = X\beta$.

That gives us $\alpha^* = G'\beta^*$ and thus:

$$\alpha^* = G'(GLG' + kI)^{-1}X'y = G'(G(L + kI)G')^{-1}X'y = G'G(L + kI)^{-1}G'X'y = (L + kI)^{-1}W'y$$

Since we know already $\hat{\alpha} = L^{-1}W'y$, we get the desired result:

$$\alpha^* = (L + kI)^{-1}L\hat{\alpha} := D\hat{\alpha}$$

$$\alpha^*_i = l_i/(l_i + k)\hat{\alpha}_i, \quad i = 1, \ldots, p$$

Now, for $\beta^*$ we get:

$$\beta^* = G\alpha^* = GD\hat{\alpha} = GDG'\hat{\beta}$$
(c) We have the general formula \( EX^2 = Var(X) + E^2X \), so in our case:

\[
E(\beta_i^* - \beta_i)^2 = Var(\beta_i^*) + E^2(\beta_i^* - \beta_i), \quad i = 1, \ldots, p
\]

Where we are using the fact that \( \beta_i \) is a number and thus \( Var(\beta_i^* - \beta_i) = Var(\beta_i^*) \). Thus we get:

\[
\sum_{i=1}^{p} E(\beta_i^* - \beta_i)^2 = \sum_{i=1}^{p} Var(\beta_i^*) + \sum_{i=1}^{p} E^2(\beta_i^* - \beta_i) = \gamma_1(k) + \gamma_2(k)
\]

Calculating \( \gamma_1(k) \) and \( \gamma_2(k) \) explicitly:

\[
\gamma_1(k) = \text{tr}(\text{cov}(\beta^*)) = \text{tr}(GDG'\text{cov}(\hat{\beta})GD^*) = \text{tr}(GDG'GD^*\text{cov}(\hat{\beta})) = \sigma^2\text{tr}(GD^2GL^{-1}G^*) = \sigma^2\text{tr}(D^2L^{-1}) = \sigma^2 \sum_j l_j/(l_j + k)^2
\]

\[
\gamma_2(k) = (E(\beta^*) - \beta)'(E(\beta^*) - \beta) = ((GDG' - I)\beta)'((GDG' - I)\beta) = \beta'G(D - I)^2G'\beta = \alpha'(D - I)^2\alpha = \sum_j \alpha_j^2(l_j - (l_j - k))^2/(l_j - k)^2 = k^2 \sum_j \alpha_j^2/(l_j - k)^2
\]

(d)

\[
\gamma_1(k)' = -2\sigma^2 \sum_j (-2l_j)/(l_j + k)^3
\]

\[
\gamma_1(k)'|_{k=0} = -2\sigma^2 \sum_j 1/l_j^2 < 0
\]

\[
\gamma_2(k)' = 2k \sum_j \alpha_j^2/(l_j - k)^2 - 2k^2 \sum_j \alpha_j^2/(l_j - k)^3
\]

\[
\gamma_2(k)'|_{k=0} = 0
\]

Which means that for \( \epsilon > 0 \) small:

\[
\Phi(\epsilon) = \Phi(0) + (\gamma_1(k)'|_{k=0} + \gamma_2(k)'|_{k=0})\epsilon + O(\epsilon^2) < \Phi(0)
\]