Computational Complexity
of the conventional method

Construct $S_h := O(np^2)$ in both steps.
Spectral decomp. : $O(p^3)$ \& $O(\max\{np^2,p^3\})$

We can do better (assume $p < n$). Write

$X = (x_1, \ldots, x_n)$, $S_h = \frac{1}{n-1} (X - \bar{x}_h)(X - \bar{x}_h)^T$

SVD of $(X - \bar{x}_h)$ = $U \text{diag} (\sigma_1, \ldots, \sigma_p) V^T$ \(\triangleq A\)

$U = O(p)$, $V^T V = I_p$, $D = \text{diag}(\sigma_1, \ldots, \sigma_p)$

Then $S_h = \frac{1}{n-1} \text{UDV} \text{V} \text{VDU}^T = \frac{1}{n-1} \text{UD}^2 \text{U}^T$

Hence $U = (u_1, \ldots, u_p)$ contains the eigenvalues of $S_h$.

Computational complexity of $A$.

SVD of $X - \bar{x}_h$: $O(\text{min}\{p^2n,p^2\})$

Only the top $k$ eigenvectors: $O(knp)$.
Finding the right $k$

Recall that $\sum_{i=1}^{k} \lambda_i(S_{X_{n}})$, with $\lambda_1 \geq \ldots \geq \lambda_p$

the eigenvalues of $S_{X_{n}}$, is the preserved variance in the projected points.

Choosing $k$ by a Scree plot in practice, depicting the ordered eigenvalues.

Choose $k$ as "elbow" - 1.

Often, PCA is applied before further processing, because these may be ineffective for high-dim. data.
Y.1.4. The eigenvalue structure of $S_n$ in high dim.

Assume:
$$x_1, \ldots, x_n \in \mathbb{R}^p$$ independent samples of a Gaussian
r.v. $x \sim N(0, \Sigma)$. Write $X = (x_1, \ldots, x_n)$

Estimate $\Sigma$ by $S_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} XX^T$.

It holds $: S_n \to \Sigma$ a.e. (p fixed) ($n \to \infty$)

Histogram and scree plot of eigenvalues of $S_n$

For $n = 1000$, $p = 500$, $S_n$ generated by $N(0, I_p)$

Th. 4.1. (Marchenko-Pastur, 1967)

Let $X_1, \ldots, X_n \in \mathbb{R}^p$, i.i.d. r.v. with $E(X) = 0$
and $\text{Cov}(X_i) = \sigma^2 I_p$. $X = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$

$S_n = \frac{1}{n} XX^T \in \mathbb{R}^{p \times p}$, $\lambda_1, \ldots, \lambda_p$ the eigenvalues of $S_n$.

Let $p_n \to \infty$ such that $\frac{p_n}{n} \to \gamma \in (0, 1]$ ($n \to \infty$).

Then the sample distribution of $\lambda_1, \ldots, \lambda_p$
(the histogram) converges a.s. to the density

$$f_\gamma(u) = \frac{1}{2\pi \sigma^2 \gamma} \sqrt{(b-u)(u-a)}, \quad a \leq u \leq b$$

with $a = a(\gamma) = \sigma^2 (1 - \gamma^{-1})^2$, $b = b(\gamma) = \sigma^2 (1 + \gamma^{-1})^2$. 

Thinking the Future
Zukunft denken

-3-
Remark: If $y > 1$ there will be a mass point at zero.

Conclusion: even in the i.i.d. uncorrelated case there is a wide spectrum of eigenvalues.

Question: What happens, if there is a low-dim. structure in the data? Is PCA useful?

4.1.5 Spike model

Model assumptions: $X_{1n-1}, X_n \in \mathbb{R}^p$, i.i.d.

$\text{Cov}(X_i) = \Sigma = \mathbf{I}_p + \beta \nu\nu^T$ for some $\nu \in \mathbb{R}^p$, $\|\nu\| = 1$, $\beta > 0$

Interpretation: $X_i = U_i + \beta \nu^T V_i \nu$, $U_i \sim N(0, \mathbf{I}_p)$ noise, $V_i \sim N(0, 1)$ signal, indep. of $U_i$ multiplied by a fixed $\beta \nu \in \mathbb{R}^p$

Then $\text{Cov}(X_i) = \text{Cov}(U_i) + \beta \text{Cov}(V_i) \nu \nu^T$

$= \mathbf{I}_p + \beta \nu \nu^T$. 
Th. 4.2. (BBP transition, Baik, Ben Arous, Péché [2005])

Assume $X_1, \ldots, X_n \in \mathbb{R}^p$ i.i.d. $E(X_i) = 0$, $\text{Cov}(X_i) = I_p + \beta v v^T$, $\beta > 0$, $v \in \mathbb{R}^p$, $\|v\|_1 = 1$. $S_n = \frac{1}{n} X X^T$

$n, p \to \infty, \frac{p}{n} \to \delta$

If $\beta \leq \sqrt{\delta}$ then $\lambda_{\max}(S_n) \to (1 + \sqrt{\delta})^2$

and $\langle v_{\max}, v \rangle^2 \to 0$

If $\beta > \sqrt{\delta}$ then $\lambda_{\max}(S_n) \to (1 + \beta)(1 + \frac{\delta}{\beta}) > (1 + \sqrt{\delta})^2$

and $\left| \langle v_{\max}, v \rangle \right|^2 \to \frac{1 - \delta / \beta^2}{1 - \delta / \beta}$