Singular value decomposition of $A \in M_{m,n}$.

I $\quad A = CR =$

$A = \begin{bmatrix} 1 & 1 \\ 1 & 1+\delta \end{bmatrix}$

$Ax = b$

$x = A^{+} b$

II $\quad A = LU = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \begin{bmatrix} * \\ 0 \end{bmatrix}$

III $\quad PA = LU = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \begin{bmatrix} * \\ 0 \; 0 \; 0 \end{bmatrix}$

IV $\quad A = QR = \begin{bmatrix} u_1 \, u_2 \cdots u_r \end{bmatrix} \begin{bmatrix} * \\ 0 \end{bmatrix}$

$\times R$

$n \begin{bmatrix} A_1 \cdots A_n \end{bmatrix}$

V $\quad S^{-1} A S = D \qquad \underline{A = SDS^{-1}} \qquad D = \begin{bmatrix} d_1 & 0 \\ 0 & d_n \end{bmatrix}$

$\in M_n$ possible when there are n-linearly independent eigenvector.

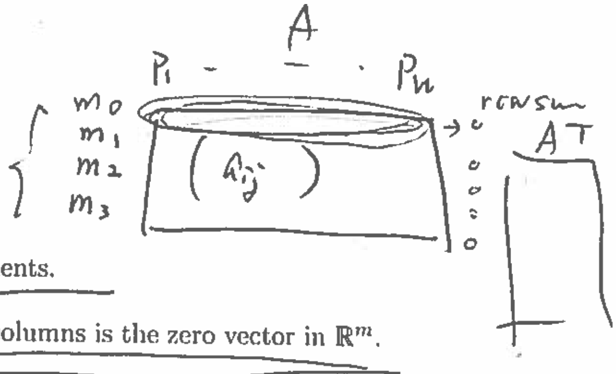
VI $\quad$ SVD of $A \in M_{m,n}$. $\quad A = \underset{\text{unitary}}{\overset{m \times m}{U}} \; \Sigma \; \overset{n \times n}{V^*}$

$\Sigma = m\left\{ \begin{bmatrix} s_1 \ldots s_r & 0 \\ 0 & 0 \end{bmatrix} \right. \qquad s_1 \geq \cdots \geq s_r > 0$

$= s_1 u_1 v_1^* + \cdots + s_r u_r v_r^*$

Best up to rank $\underline{k}$ approximation of A is

$s_1 u_1 v_1^* + \cdots + s_k u_k v_k^*$



Principal Component Analysis

- Let A be $m \times n$, each column is a sample with m measurements.
- Normalize the means to 0 for all measurements. So, sum of columns is the zero vector in $\mathbb{R}^m$.
- The variances are the diagonal entries of $AA^T \in M_m$.
- The co-variances are the off-diagonal entries of $AA^T \in M_n$.
- The sample co-variance matrix is $S = AA^T/(n-1)$.

For example, when $m = 2$, the best rank one approximation of S is $s_1 u_1 v_1^t$. So, the slop of the line is the ratio of the second entry of u_1 to the first entry of u_1.

If the rank of AA^T is low, then the two measurement is closely related, i.e., almost agree on a linear relation.

We can extend the idea to higher dimension data set recorded as $A \in M_{m,n}$. One can use k-dimensional hyperplane to approximate the data.

- The total variance is

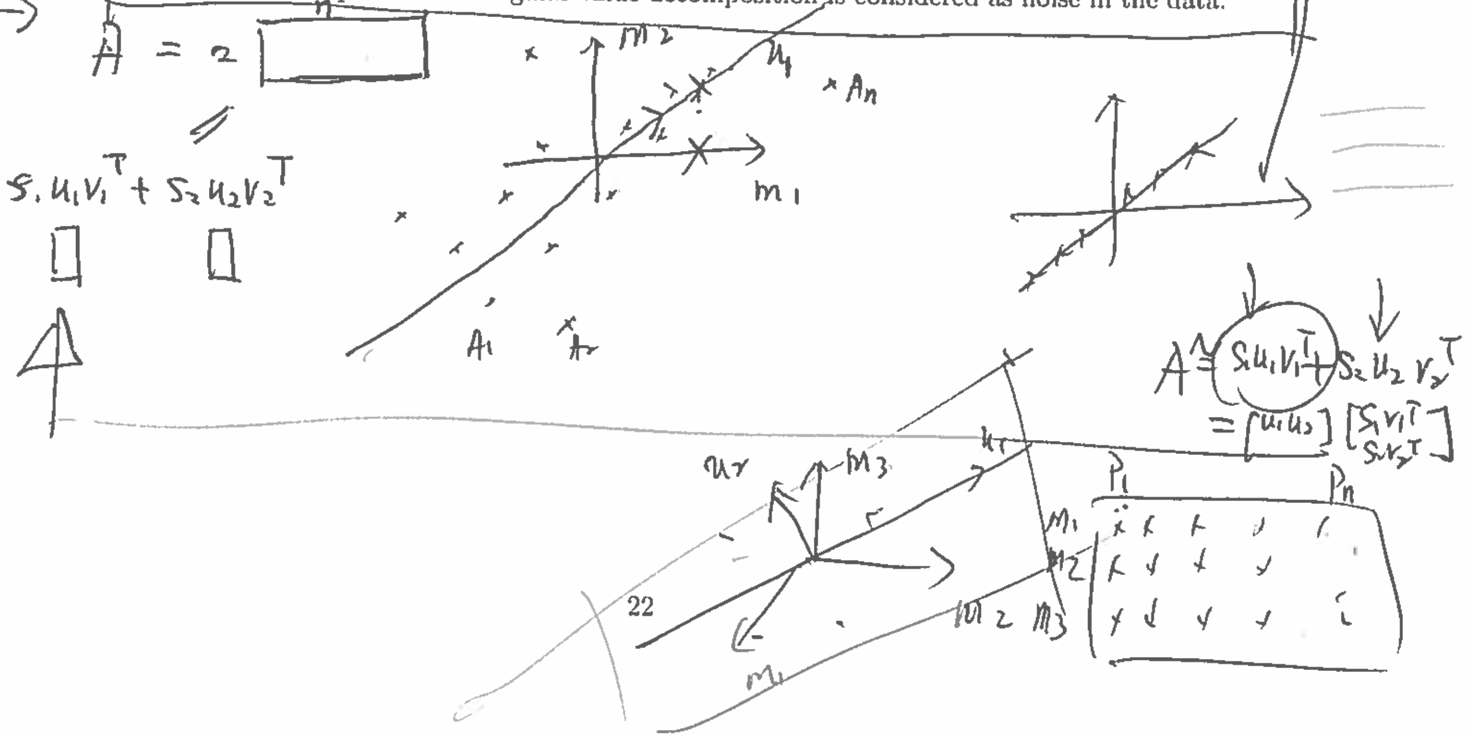
$$= \operatorname{tr} AA^T/(n-1) =$$

$$T = \operatorname{tr} A^T A/(n-1) = \operatorname{tr} S/(n-1) = (\sum_{j=1}^{r} s_j^2)/(n-1),$$

  where $s_1 \geq \cdots \geq s_r > 0$ are the singular values of A.

$$A = s_1 u_1 v_1^T + \cdots + s_k u_k v_k^T$$

- The first k singular vectors capture more information than other vectors; u_j is referred to as the jth principal component of the data that accounts for the fraction s_j^2/T of the variance.
- The effective rank k of A or S is the number of singular values larger than certain threshold so that the other part in the singular value decomposition is considered as noise in the data.



$A = U \Sigma V^T$

$U^T A V = \Sigma$

- Note that the line is different from finding the best fit $y = ax + b$. In that case, ~~we want to~~ find best a, b such that $ax_i + b = y_i$ for $i = 1, \ldots, n$ without centering the data. We consider $\tilde{A}(a,b)^T = (y_1, \ldots, y_n)^T$ and find the least square solution:

$$\tilde{A}^T \tilde{A}(a,b)^T = \tilde{A}^T (y_1, \ldots, y_n)^T.$$

$\left\| \begin{bmatrix} A_1^T \\ A_i^T \\ A_n^T \end{bmatrix} [u_1\ u_2] \right\|_F$

This is known as standard least square.

- In our case, we consider the centered data, and

$$\|A^T\|_F^2 = \|A^T u_1\|_F^2 + \cdots + \|A^T u_m\|_F^2$$

$(\Sigma |a_{ij}|^2)$

$(\Sigma |a_{ij}|^2) = \| [\ ] [u_1\ u_2] \|$

$= \|A^T u_1\|_F^2 + \|A^T u_2\|_F^2$

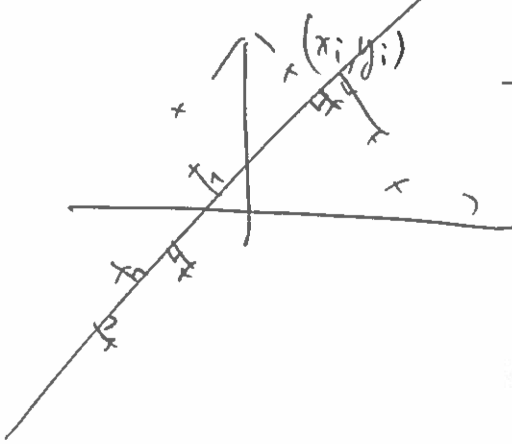
so that

the sum of squared distances from the data points to $u_1, \ldots, u_k$ is a minimum.

There are interesting discussion of the Hilbert matrix

$$H = [a_{ij}] = [1/(i+j-1)]$$

and the zero-one matrix representing the picture of square, triangle, circe, etc. See pp. 78-79.



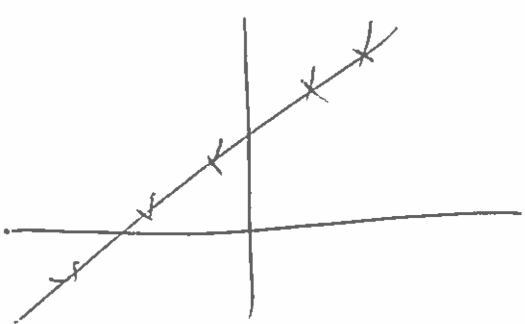
$y = ax + b$

$ax_1 + b = y_1$

$ax_2 + b = y_2$

$\vdots$

$ax_n + b = y_n$



$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$\hat{A} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_n \end{bmatrix}$

$\hat{A}^T \hat{A} \begin{bmatrix} a \\ b \end{bmatrix} = \hat{A}^T \begin{bmatrix} y_1 \\ y_n \end{bmatrix}$

$= \begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

Linear regression

$$H_1 = [\,1\,]$$

$$H_2 = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}$$

$$H_3 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} = \sum_{i=1}^{3} s_i u_i u_i^T$$

$$H_n = \sum s_i u_i u_i^T$$

If $A = A^*$ is psd. then $A = \lambda_1 u_1 u_1^* + \cdots + \lambda_r u_r u_r^*$.

$\lambda_1, \cdots, \lambda_r$ are the positive eigenvalues.

$$U^* A U = \left[\begin{array}{c|c} \begin{matrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right] = \lambda_1 u_1 u_1^* + \quad + \lambda_r u_r u_r^*$$

If $A = A^*$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$

& $\lambda_n \leq \lambda_{n-1} \cdots \leq \lambda_{n-p+1} < 0$

$\lambda_{k+1} = \cdots = \lambda_{n-p} = 0$.

---

and $A = \lambda_1 u_1 u_1^* + \cdots + \lambda_k u_k u_k^* + \lambda_{n-p+1} u_{n-p+1} u_{n-p+1}^*$

$+ \cdots + \lambda_n u_n u_n^*$.

Then the SVD is:

$$A = \lambda_1 u_1 u_1^* + \cdots + \lambda_k u_k u_k^*$$
$$+ |\lambda_{n-p+1}| u_{n-p+1} (-u_{n-p+1})^* + \cdots + |\lambda_n| u_n (-u_n)^*$$

$\{u_1 \cdots u_k \; -u_{n-p+1}, \cdots -u_n\}$

$\Uparrow$

$\{v_1 \cdots v_k \; v_{n-p+1} \cdots v_n\}$

$A^* A$ has e.v. $|\lambda_1|^2, \cdots, |\lambda_n|^2$