M. Sombra, 2020

VI. SVD and principal components analysis

References:

[D] Demmel's book.

[S] Strang: Linear algebra and learning from data, Wellesley–Cambridge Press, 2019.

VI.1 Singular values and singular vectors [S, § I.8] [D, § 3.2.3]

Let S ~~n~~ $n \times n$ symmetric positive definite. Then

$$S = \underset{\text{orthogonal}}{Q} \; \overset{\text{diagonal with positive entries}}{\Lambda} \; Q^T$$

S has real and positive eigenvalues and orthogonal eigenvectors: the best we can hope!

The singular value decomposition (SVD) extends this factorization to any matrix, ~~even~~ outside the square case.

Let

$$A \quad m \times n\text{-matrix} \qquad (m \geq n)$$

There are two sets of singular vectors:

$u_1, \dots, u_m \in \mathbb{R}^m$ "left" singular vectors

$v_1, \dots, v_n \in \mathbb{R}^n$ "right" singular vectors

Both are orthogonal and are connected by

$$A v_i = \sigma_i u_i \qquad i = 1, \dots, n$$

for $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$

singular values

Set $r = \operatorname{rank}(A)$. Then

$u_1, \dots, u_r$ basis of $\operatorname{Im}(A)$

$v_{r+1}, \dots, v_n$ ——— $\operatorname{Ker}(A)$

Matrix form:

$U = (u_1 \cdots u_m)$ $m \times m$-orthogonal

$V = (v_1 \cdots v_n)$ $n \times n$-orthogonal

Both U and V are square nonsingular and

$$U^{-1} = U^T \qquad V^{-1} = V^T$$

$$\Sigma = \begin{array}{cc} \left(\begin{array}{c} \begin{matrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{matrix} \\ \hline 0 \end{array} \right) & \begin{array}{l} n \\ \\ m-n \end{array} \\ n & \end{array} \qquad m \times n$$

The (full) SVD is $A \cdot V = U \cdot \Sigma$ or equivalently

$$\boxed{A = U \cdot \Sigma \cdot V^T}$$

Example

$$\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 0.32 & -0.95 \\ 0.95 & 0.32 \end{pmatrix} \cdot \begin{pmatrix} 6.71 & 0 \\ 0 & 2.24 \end{pmatrix} \cdot \begin{pmatrix} 0.71 & 0.71 \\ -0.71 & 0.71 \end{pmatrix}$$

$$A \qquad\qquad U \qquad\qquad \Sigma \qquad\qquad V^T$$

The column-row multiplication of $U \cdot \Sigma$ and V^T separates A into r pieces of rank 1:

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

In the example:

$$\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = 6.71 \begin{pmatrix} 0.32 \\ 0.95 \end{pmatrix} \cdot \begin{pmatrix} 0.71 & 0.71 \end{pmatrix} + 2.24 \begin{pmatrix} -0.95 \\ 0.32 \end{pmatrix} \begin{pmatrix} -0.71 & 0.71 \end{pmatrix}$$

$$= \frac{3}{2} \begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 3 & -3 \\ -1 & 1 \end{pmatrix}$$

The first piece is "more important" (or representative of A) because $\sigma_1 = 6.71 > \sigma_2 = 2.24$: this will be given a precise sense later.

The thin SVD avoids the 0's in the lower part of Σ and uses a diagonal matrix for the singular values:

$$A = U_n \cdot \Sigma_n \cdot V^T$$

with $U_n = (u_1 \cdots u_n)$ $m \times n$-orthogonal

$\Sigma_n = \mathrm{diag}(\sigma_1, \dots, \sigma_n)$ $n \times n$-diagonal

The <u>reduced SVD</u> uses the nonzero singular values to remove the parts that are sure to produce zeros:

$$A = U_r \Sigma_r V_r^T$$

with $U_r = (u_1 \cdots u_r)$ $m \times r$ - orthogonal

$\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$ $r \times r$ - diagonal nonsingular

$V_r = (v_1 \cdots v_r)$ $n \times r$ - orthogonal

We still have

$$U_r^T U_r = \mathbb{1}_r \qquad V_r^T V_r = \mathbb{1}_r$$

but they are <u>not invertible</u> (since they are rectangular)

To <u>identify</u> the <u>singular values</u> and <u>vectors</u> we can consider the <u>symmetric semipositive definite</u> matrices

$$A^T A = (V \Sigma^T U^T) \cdot (U \Sigma V^T) = V \cdot \Sigma^T \Sigma \cdot V^T \qquad n \times n$$

$$A \cdot A^T = (U \Sigma V^T) \cdot (V \Sigma^T U^T) = U \cdot \Sigma \Sigma^T U^T \qquad m \times m$$

Then V contains orthonormal eigenvectors of $A^T A$

U ——————————————— $A A^T$

$\sigma_1, \dots, \sigma_r$ are the nonzero eigenvalues of both $A^T A$ and $A A^T$

To see the existence of the (full) SVD (and a way to compute it) we start with the diagonalization

$$A^T A = Q \Lambda Q^T$$

Let $v_1, \dots, v_n$ be the columns of Q, ordered so that $v_1, \dots, v_r$ correspond to the nonzero eigenvalues

$$\lambda_1 \geq \dots \geq \lambda_r \geq 0$$

Set $\sigma_k = \lambda_k^{1/2}$ $\quad k = 1, \dots, r$

Then $u_k = \sigma_k^{-1} A v_k$ $\quad k = 1, \dots, r$

gives

$$A = (u_1 \cdots u_r) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix}$$ reduced SVD

because $u_1, \dots, u_r$ are orthonormal: indeed

$$\langle u_j, u_k \rangle = u_j^T u_k = (\sigma_j^{-1} A v_j)^T (\sigma_k^{-1} A v_k) =$$

$$= \sigma_j^{-1} \sigma_k^{-1} v_j^T A^T A v_k = v_j^T v_k = \langle v_j, v_k \rangle \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

Exercice: show that $u_1, \dots, u_r$ are eigenvalues of AA^T with eigenvalues $\sigma_1^2, \dots, \sigma_r^2$.

To compute a full SVD, take

$$v_{r+1}, \dots, v_n \in \mathbb{R}^n \quad \text{and} \quad u_{r+1}, \dots, u_m \in \mathbb{R}^m$$

completing $v_1, \dots, v_r$ and $u_1, \dots, u_m$ to orthonormal basis. Then

$$A = U \cdot \Sigma \cdot V^T$$

with $U = (u_1, \dots, u_m)$

$V = (v_1 \dots v_n)$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \; 0 \\ & 0 & & \end{pmatrix}$$

<u>Example:</u> $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$

Then

$$A^T A = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix} \quad \text{and} \quad A A^T = \begin{pmatrix} 9 & 12 \\ 12 & 41 \end{pmatrix}$$

The <u>eigenvectors</u> and <u>eigenvalues</u> of $A^T A$ are

$$\begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 45 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 5 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

<u>Right singular vectors</u>

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix}$$

<u>Singular values</u>

$$\sigma_1 = \sqrt{45} = 6.71 \qquad \sigma_2 = \sqrt{5} = 2.24$$

<u>Left singular values</u>

$$u_1 = \sigma_1^{-1} A v_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.32 \\ 0.95 \end{pmatrix} \qquad u_2 = \sigma_2^{-1} A v_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.95 \\ 0.32 \end{pmatrix}$$

We conclude that $A = U \Sigma V^T$ with

$$U = \begin{pmatrix} 0.32 & -0.95 \\ 0.95 & 0.32 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 6.71 & \\ & 2.24 \end{pmatrix} \qquad V = \begin{pmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{pmatrix}$$

Some <u>particular cases</u>:

(1) Let S symmetric positive definite and
$$S = Q \Lambda Q^T$$
its diagonalization. Then $U = V = Q$ and $\Sigma = \Lambda$

(2) For an orthonormal $n \times n$-matrix Q, all its singular values are equal to 1

(3) Let $A = x \cdot y^T$ be an $m \times n$-matrix of rank 1 ($x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$). Its SVD (reduced) is

$$A = \underbrace{\frac{x}{\|x\|}}_{U_1} \cdot \underbrace{(\|x\| \cdot \|y\|)}_{\Sigma_1} \underbrace{\frac{y^T}{\|y\|}}_{V_1^T}$$

The geometry of the SVD:

The SVD separates the matrix into

orthogonal × diagonal × orthogonal

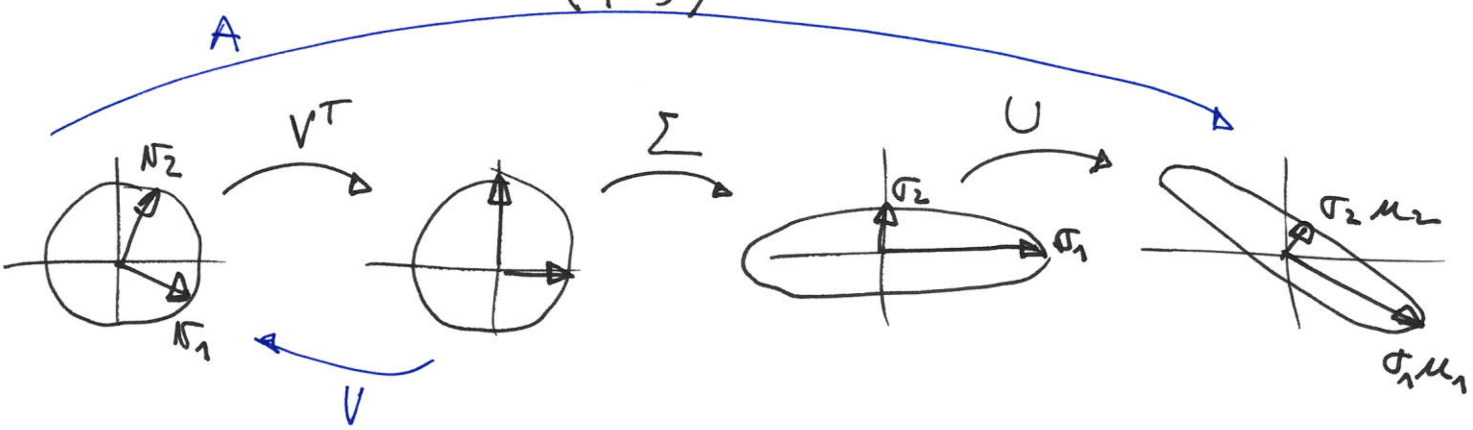
The unit sphere S_n of $\mathbb{R}^n$ is send to an ellipsoid $A S_n$ of $\mathbb{R}^n$ centered at the origin and with axes

$$\sigma_i u_i \qquad i = 1, \dots, r$$

In two dimensions, we can draw the process.

Example

$$A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$$



VI. 2 The best low rank approximation

[D, § 3.2.3], [S, § I.9]

The 2-norm can be computed in terms of the SVD

$$A = U \cdot \Sigma \cdot V^T$$

We have $\|A\|_2 = \sigma_1$

Indeed, the 2-norm is invariant under multiplication by orthogonal matrices and so

$$\|A\|_2 = \| U^T \cdot A \cdot V \|_2 = \|\Sigma\|_2 = \sigma_1$$

The Frobenius norm is also invariant under multiplication by orthogonal matrices and so

$$\underset{\left(\sum a_{ij}^2\right)^{1/2}}{\underset{\|}{\|A\|_F}} = \| U^T \cdot A \cdot V \|_F = \|\Sigma\|_F = (\sigma_1^2 + \dots + \sigma_r^2)^{1/2}$$

The Eckart–Young theorem says that for $k = 1, \dots, r$ the matrix

$$A_k = \sigma_1 u_1 v_1^T + \dots \sigma_k u_k v_k^T = U_k \cdot \Sigma_k \cdot V_k^T$$

is the best rank k approximation of A in ~~either the~~ both the 2-norm and the Frobenius norm.

Indeed for both norms

$$\|A - A_k\| \underset{\text{orthogonal invariance}}{=} \|\Sigma - \Sigma_k\| = \left\| \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \sigma_{k+1} & \\ & & & & \ddots \\ & & & & & \sigma_r \end{pmatrix} \right\|$$

and so

$$\|A - A_k\|_2 = \sigma_{k+1} \qquad \|A - A_k\|_F = (\sigma_{k+1}^2 + \cdots + \sigma_r^2)^{1/2}$$

The theorem says that for both norms

$$\|A - B\| \geq \|A - A_k\|$$

for any other k-rank $m \times n$-matrix B.

An application: <u>image compression</u>

A B/W image of $m \times n$ pixels can be coded by an $m \times n$-matrix A with entries $0 \leq a_{ij} \leq 1$ (the brightness of the pixel (i,j))

black		white
0	.... gray ----	1

The <u>size</u> of the image is mn

Instead of storing/transmitting it, we can <u>replace it</u> by the <u>k-rank</u> approximation

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

of size $k \cdot (1 + m + n)$ (or $k(m+n)$ storing $\sigma_i u_i$)

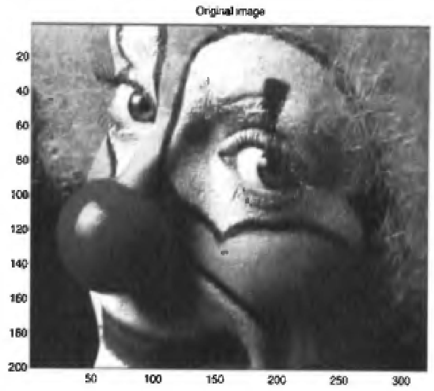
The relative error of the approximation is

$$\frac{\|A-A_k\|_2}{\|A\|_2} = \frac{\sigma_{k+1}}{\sigma_1}$$

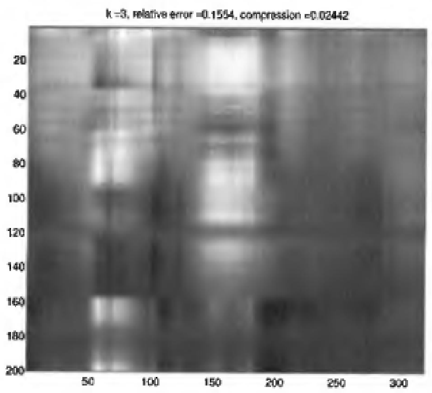
and the compression ratio is

$$\frac{k\,(m+n)}{m\cdot n}$$

A 320×200 photo of a clown and its approximations

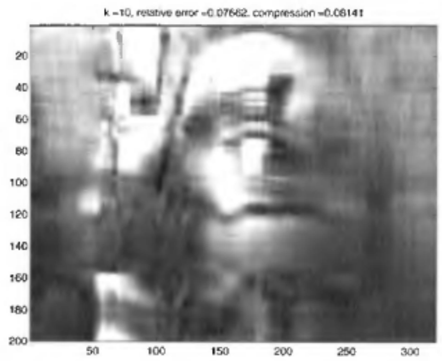


(a)



(b)

Fig. 3.3. *Image compression using the SVD.* (a) *Original image.* (b) *Rank $k = 3$ approximation.*



(c)

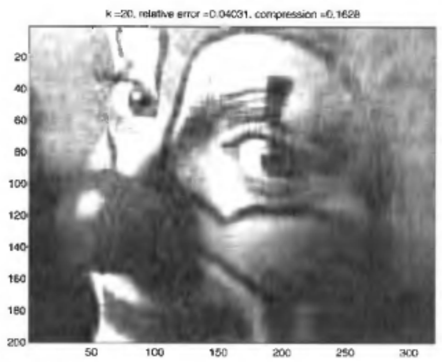


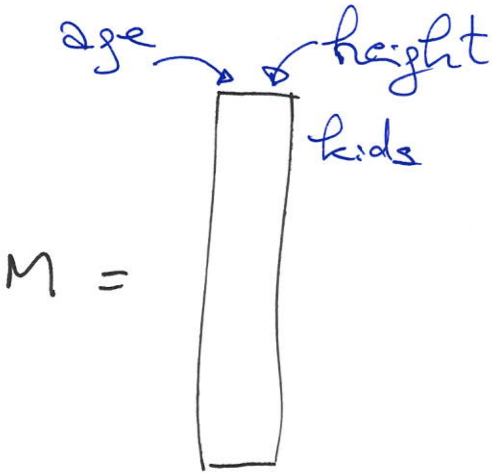
Fig. 3.3. *Continued.* (c) *Rank $k = 10$ approximation.* (d) *Rank $k = 20$ approximation.*

VI.6 Principal components analysis [S, I.9], [Schlens]

[S] Strang's book

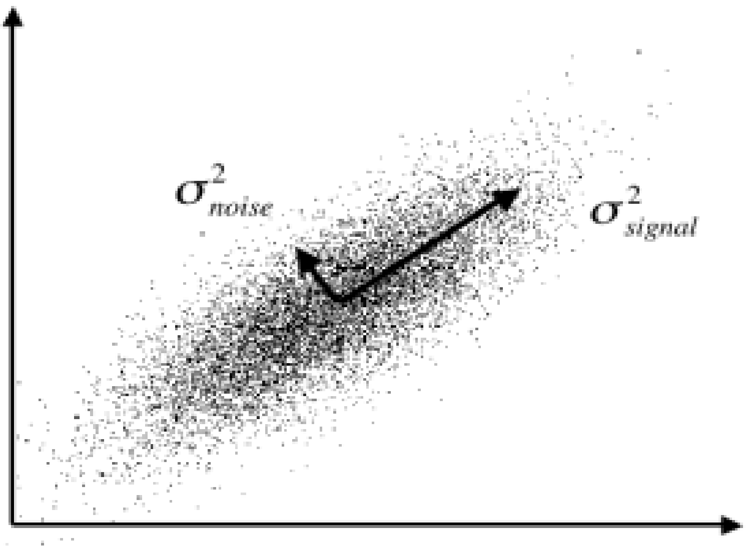
[Schlens] Schlens, A tutorial on PCA, 2003.

Let M be an $m \times n$ data matrix: the rows are samples and the columns are variables, like in



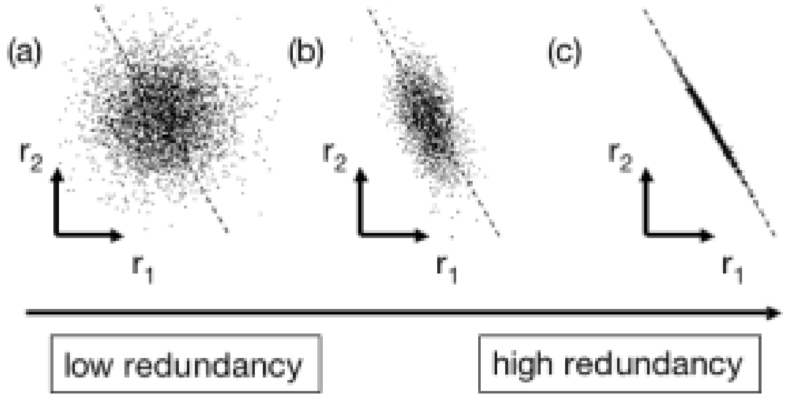
$m \times 2$ matrix

Can we simplify the description of this data?
Plotting the samples in $\mathbb{R}^n$ and centering them, they might correlated:



This correlation might be higher or lower

more significative   less significative



The key parameters in probability and statistics:

$\boxed{\text{mean}}$ $\boxed{\text{variance}}$

The mean is

$$\mu = \frac{1}{m} \sum_{i=1}^{m} M_i \quad \in \mathbb{R}^n$$

The centered data is

$$A = M - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu^T = \begin{pmatrix} M_1 - \mu^T \\ \vdots \\ M_m - \mu^T \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}$$

The mean of each variable in A is 0.

column

$\cancel{\langle a_i, a_i \rangle}$ $\langle a_i, a_i \rangle$ $\quad$ $\langle \cancel{a_i}, a_j \rangle$ $i \neq j$

The <u>variances</u> and <u>covariances</u> of $A = (a_1 \cdots a_n)$ are the <u>diagonal</u> and <u>off-diagonal</u> entries of

$$A^T \cdot A$$

The <u>covariance matrix</u> is

$$S = \frac{1}{m-1} A^T \cdot A$$

$n \times n$ symmetric semipositive definite

$m-1$ degrees of freedom

<u>Example</u>

$$A = \begin{pmatrix} 3 & 7 \\ -4 & -6 \\ 7 & 8 \\ 1 & -1 \\ -4 & -1 \\ -3 & -7 \end{pmatrix}$$

centered data matrix of ages and heights

Then

$$S = \frac{1}{6-1} A^T A = \begin{pmatrix} 20 & 25 \\ 25 & 40 \end{pmatrix}$$

covariance matrix

Let

$$A = U \cdot \Sigma \cdot V^T \qquad \text{thin SVD}$$

and set

$$B = A \cdot V = U \Sigma \qquad m \times n$$

Then

$$\frac{1}{m-1} B^T B = \frac{1}{m-1} \Sigma^2 \qquad \text{diagonal}$$

the variables in B are not correlated and ordered by their variance $\langle b_i, b_i \rangle = \sigma_i^2$ $\langle b_i, b_j \rangle = 0$ $i \neq j$

The matrix

$$V = (v_1 \cdots v_n)$$

gives an <u>orthonormal basis</u> of $\mathbb{R}^n$.

The <u>representation</u> of a vector $x \in \mathbb{R}^n$ w.r. to this basis is

$$x = \sum_{j=1}^{n} \langle x_i, v_j \rangle v_j$$

Hence

components of A_i w.r. to the basis $v_1, \dots, v_n$

$$B = A \cdot V = \left(\langle A_i, v_j \rangle \right)_{\substack{1 \le i \le m \\ 1 \le j \le n}}$$

For $1 \le k \le n$ the first k columns of B

$$b_1, \dots, b_k \in \mathbb{R}^m$$

are <u>uncorrelated</u> and <u>have total variance</u>

$$\frac{1}{m-1}\left(\sigma_1^2 + \cdots + \sigma_k^2\right)$$

~~this ...~~ These columns give the <u>projection</u> of the samples in A to the linear subspace

$$\mathrm{Vect}(v_1, \dots, v_k)$$

The projected data has <u>maximal total variance</u> (among all possible orthogonal projections of the data to a k-linear subspace of $\mathbb{R}^n$)

Also the <u>sum of the squares of the distances between the samples and its projections</u> is <u>minimal</u> for this k-linear subspace:

$$\sum_{i=1}^{m} \| A_i - \sum_{j=1}^{k} \langle A_i, v_j \rangle v_j \|_2^2 = \sum_{i=1}^{m} \sum_{j=k+1}^{n} \langle A_i, v_j \rangle^2 = \|A - A_k\|_F^2$$

best k-rank approximation

Frobenius norm

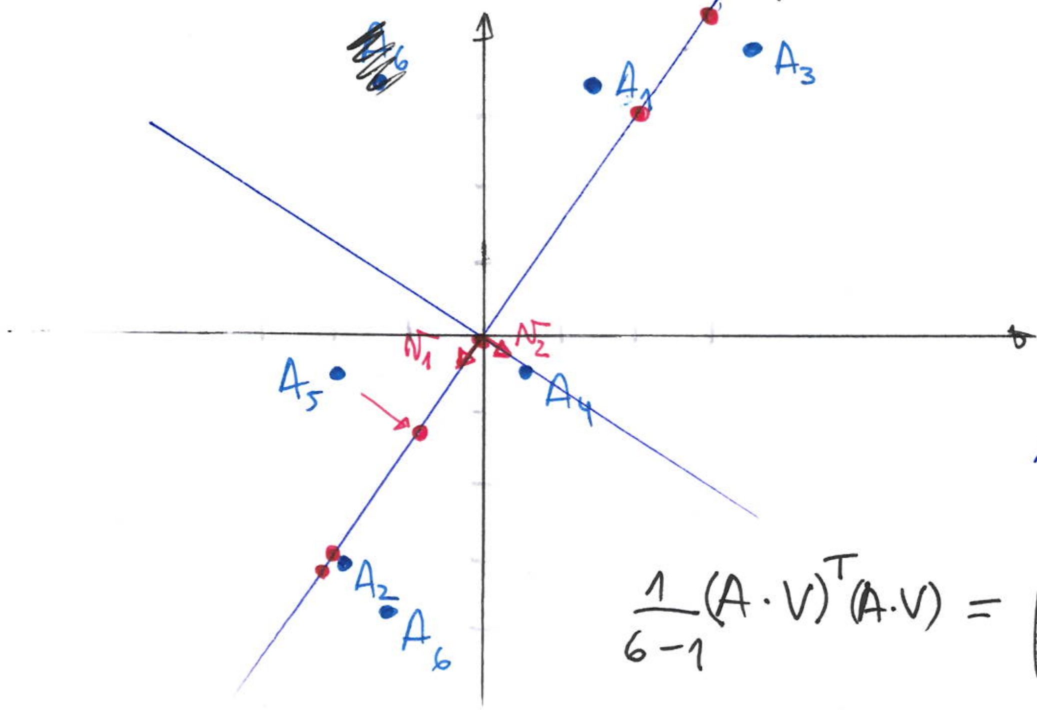
This is a consequence of the Eckart–Young theorem.

<u>Example</u> (cont)

1st ppal component

2nd ppal component

$$A = \begin{pmatrix} 3 & 7 \\ -4 & -6 \\ 7 & 8 \\ 1 & -1 \\ -4 & -1 \\ -3 & -7 \end{pmatrix} = U \cdot \Sigma \cdot V^T = \begin{pmatrix} -0.44 & -0.36 \\ 0.43 & 0.01 \\ -0.63 & 0.33 \\ 0.02 & 0.35 \\ 0.18 & -0.70 \\ 0.44 & 0.37 \end{pmatrix} \cdot \begin{pmatrix} 16.87 & 0 \\ 0 & 3.92 \end{pmatrix} \begin{pmatrix} -0.56 & -0.83 \\ +0.83 & -0.56 \end{pmatrix}$$



1st variance

$$\frac{1}{6-1} (A \cdot V)^T (A \cdot V) = \begin{pmatrix} 56.92 & 0 \\ 0 & 3.07 \end{pmatrix}$$

2nd variance

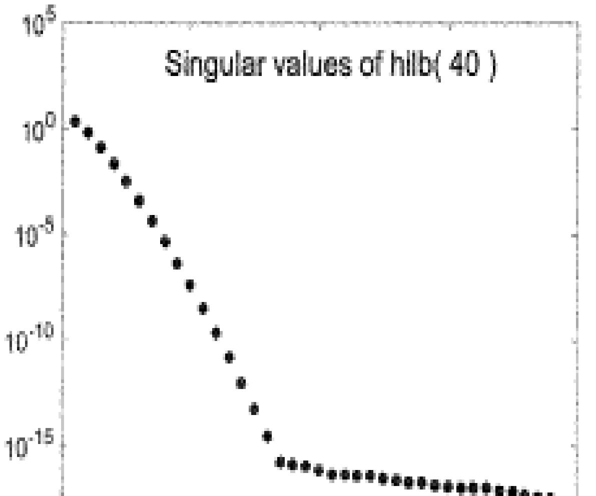
The principal components

$$b_i = \sigma_i u_i \qquad i=1,\dots,k$$

account for

$$\rho_k = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{n} \sigma_i^2}$$

of the total variance of the data. The choice of k should keep the true signal and discard the noise



$$H_{ij} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdot & \cdot \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} = \frac{1}{(i+j-1)}$$

Figure I.12: Scree plot of $\sigma_1, \dots, \sigma_{39}$ ($\sigma_{40} = 0$) for the evil Hilbert matrix, with elbow at the effective rank: $r \approx 17$ and $\sigma_r \approx 10^{-16}$.

VI.4 The rank deficient LSP revisited [D, §3.5]

The SVD applies also to the LSP, and it is particularly appropriate for the rank deficient case.

Let

$$A \in \mathbb{R}^{m \times n} \qquad r = \operatorname{rank}(A)$$

If $r < n$ then the solution of the LSP

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

is not unique since

$$A\hat{x} = A(\hat{x} + y) \qquad \forall y \in \operatorname{Ker}(A) \simeq \mathbb{R}^{n-r}$$

A reasonable choice is the minimizer with the smallest norm

Let

$$\underset{m \times n}{A} = \underset{m \times m}{U} \; \underset{m \times n}{\Sigma} \; \underset{n \times n}{V^T} = \underset{m \times r}{U_r} \; \underset{r \times r}{\Sigma_r} \; \underset{r \times n}{V_r^T}$$

full SVD   reduced SVD

$$U = (u_1 \cdots u_m) \implies U_r = (u_1 \cdots u_r)$$

$$V = (v_1 \cdots v_n) \implies V_r = (v_1 \cdots v_r)$$

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{pmatrix} \implies \Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$$

The Moore-Penrose inverse of A is

$$A^+ = V_r \Sigma_r^{-1} U_r^T \qquad n \times m\text{-matrix}$$

The solution of the LSP is

$$\boxed{\hat{x} = A^+ b}$$

When A is rank deficient (i.e. $r<n$) it is the solution with the smallest norm

It is well conditioned if the smallest nonzero singular value of A is not too small:

changing b to $b+\delta b$ changes x to $x+\delta x$ with

$$\boxed{\|\delta x\|_2 \leq \frac{\|\delta b\|_2}{\sigma_r}}$$

because

$$\delta x = V_r \Sigma_r^{-1} U_r^T \delta b$$

and so $\quad \|\delta x\|_2 \leq \|\Sigma_r\|_2^{-1} \|\delta b\|_2 = \dfrac{\|\delta b\|_2}{\sigma_r}$

In practice, the difficulty is that the rank is not continuous, and so affected by small perturbations

Example $\quad A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Then

$$A^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot (1) \cdot (1\ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\hat{x} = A^+ b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$ with condition number $1/\sigma_1 = 1$

Let $\varepsilon > 0$ and

$$A_\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}$$

gives

$$\hat{x}_\varepsilon = \begin{pmatrix} 1 \\ 1/\varepsilon \end{pmatrix}$$

Hence round off will make perturbations of size $\mathcal{O}(\varepsilon) \|A\|_2$, that might increase the condition number from $1/\sigma_r$ to $1/\varepsilon$.

The SVD is backward stable : round-off with machine epsilon ε gives

$$(U + \delta U)(\Sigma + \delta\Sigma)(V + \delta V)^T = A + \delta A$$

with

$$\|\delta A\|_2 \leq \mathcal{O}(\varepsilon) \|A\|_2$$

Hence the computed singular values $\sigma_i + \delta\sigma_i$ verify

$$|\delta \sigma_i| \leq \mathcal{O}(\varepsilon) \|A\|_2$$

Let $\boxed{tol > 0}$ be a user supplied measure of uncertainty in A.

Round off implies

σ_1

$$tol \geq \varepsilon \|A\|_2$$

but it might be larger (eg. errors in measurements)

Given

$\tilde{U}, \tilde{\Sigma}, \tilde{V}$ computed SVD of A

set $$\hat{\sigma}_i = \begin{cases} \tilde{\sigma}_i & \text{if } \tilde{\sigma}_i \geq tol \\ 0 & \text{else} \end{cases}$$

and replace $\tilde{\Sigma}$ by

$$\hat{\Sigma} = \begin{pmatrix} \tilde{\sigma}_1 \cdots \tilde{\sigma}_r \, 0 \cdots 0 \\ 0 \end{pmatrix}$$ truncated SVD

and

$$\hat{x} = \tilde{V}_r \hat{\Sigma}_r^{-1} \tilde{U}_r^T$$

The error is bounded by $O(tol)$ and the condition number by $1/\sigma_r$.