Principal Component Analysis & SVD + COVAR

- Eigen analysis, Karhunen-Loeve transform
- **Eigenvectors:** derived from Eigen decomposition of the **scatter matrix**
- A projection set that best explains the distribution of the representative features of an object of interest.
- PCA techniques choose a dimensionality-reducing linear projection that maximizes the scatter of all projected samples.

Principal Component Analysis Contd.

- Let us consider a set of N sample images $\{x_1, x_2, \ldots\ldots, x_N\}$ taking values in n-dimensional image space.
- Each image belongs to one of c classes $\{X_1, X_2, \ldots\ldots, X_c\}$.
- Let us consider a linear transformation, mapping the original n-dimensional *image space* to m-dimensional *feature space*, where $m < n$.
- The new feature vectors $y_k \in R^m$ are defined by the linear transformation –

$$y_k = W^T x_k \qquad k = 1, 2, \ldots\ldots, N$$

where, $W \in R^{nxm}$ is a matrix with orthogonal columns representing the basis in feature space.

Principal Component Analysis Contd..

- **Total scatter matrix S_T is defined as**

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

where, *N* is the number of samples , and $\mu \in R^n$ is the mean image of all samples .

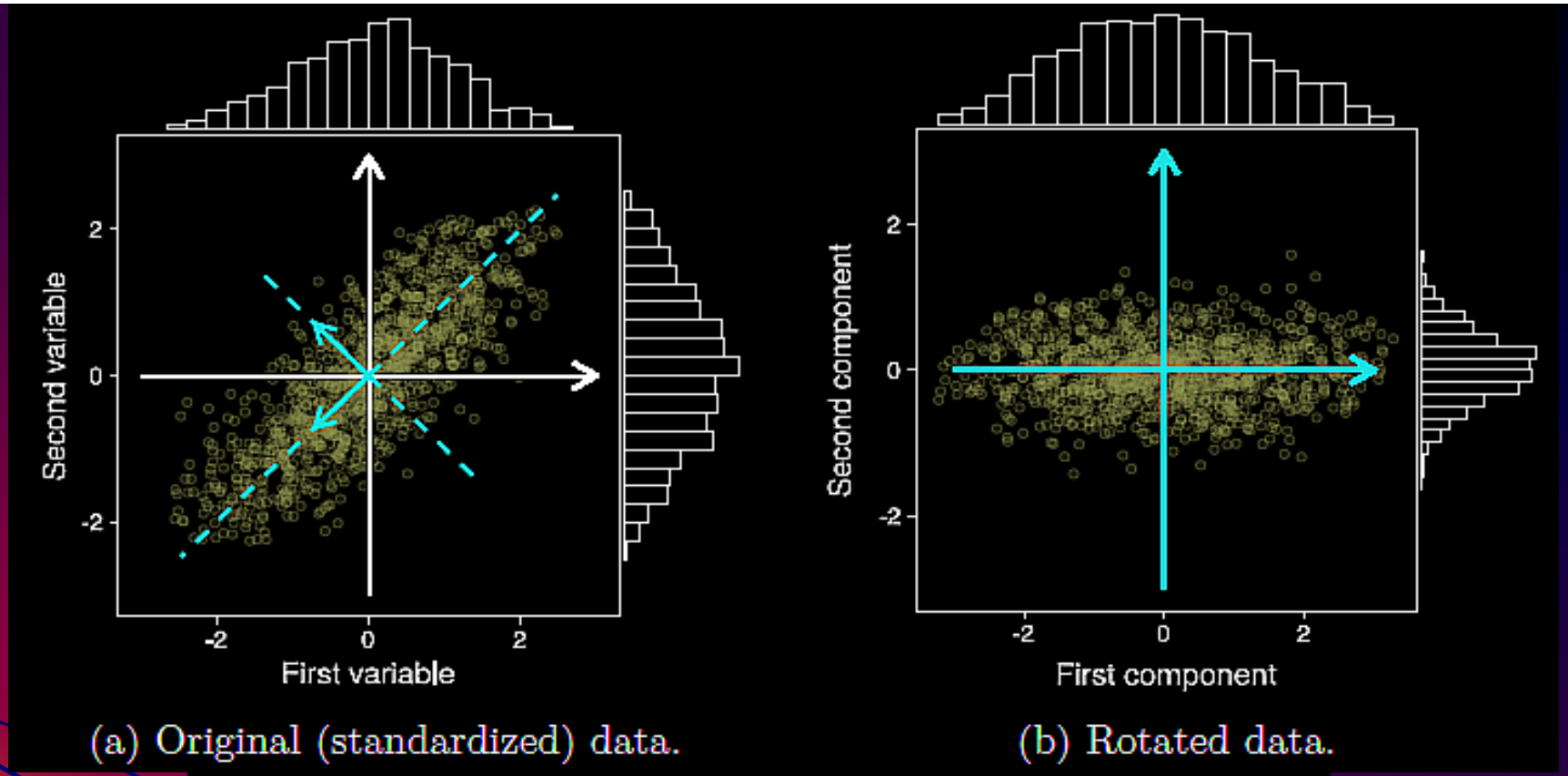
$$\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

- **The scatter of transformed feature vectors $\{y_1, y_2, \ldots, y_N\}$ is $W^T S_T W$.**

- **In PCA, W_{opt} is chosen to maximize the determinant of the total scatter matrix of projected samples, *i.e.*,**

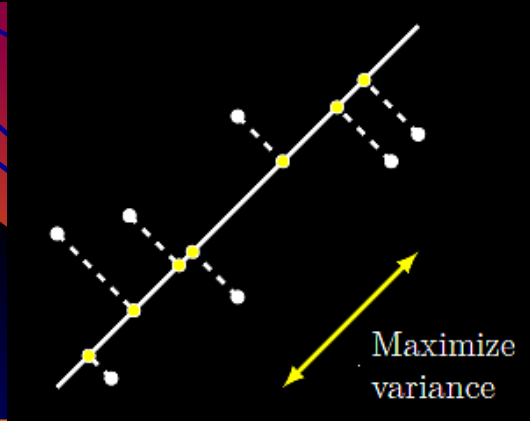
$$W_{opt} = \arg\max_{W} \left| W^T S_T W \right|$$

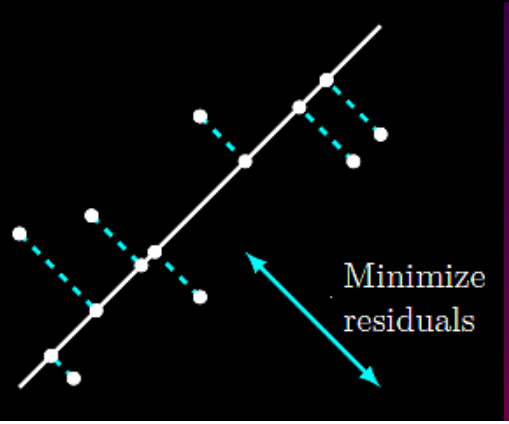
where $\{w_i \mid i = 1,2,\ldots,m\}$ is the set of *n* dimensional eigenvectors of S_T corresponding to *m* largest eigenvalues (check proof).



(a) Original (standardized) data.

(b) Rotated data.





For a data matrix, X^T, with zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), where each *column* is made up of results for a different subject, and each *row* the results from a different probe. This will mean that the PCA for our data matrix X will be given by:

$$Y = W^T X = \Sigma V^T,$$

where $W\Sigma V^T$ is the singular value decomposition (SVD) of X.

Goal of PCA:

Find some orthonormal matrix W^T, where $Y = W^T X$; such that

$$\mathrm{COV}(Y) \equiv (1/(n-1))YY^T \quad \text{is diagonalized.}$$

The rows of W are the principal components of X, which are also the eigenvectors of COV(X).

Unlike other linear transforms (DCT, DFT, DWT etc.), PCA does not have a fixed set of basis vectors. Its basis vectors depend on the data set.

Singular Value Decomposition

$A = U\Sigma V^T$ is known as the "**SVD**" or the ***singular value decomposition***.

The SVD is closely associated with the eigenvalue-eigenvector factorization $Q\Lambda Q^T$ of a positive definite matrix.

Any $m \times n$ matrix A can be factored into

$$A = U\Sigma V^{\mathrm{T}} = (\mathbf{orthogonal})(\mathbf{diagonal})(\mathbf{orthogonal})$$

The columns of U ($m \times m$) are ***eigenvectors of*** $\boldsymbol{AA^T}$, and the columns of V ($n \times n$) are ***eigenvectors of*** $\boldsymbol{A^TA}$.

The r singular values on the diagonal of Σ ($m \times n$) are the ***square roots of the nonzero eigenvalues*** of both AA^T and A^TA.

See next few slides for variants →

Strang - Sec. 6.3 – PP 367

Singular Value Decomposition: Any m by n matrix A can be factored into

$$A = U\Sigma V^{\mathrm{T}} = (\textbf{orthogonal})(\textbf{diagonal})(\textbf{orthogonal}).$$

The columns of U (m by m) are eigenvectors of AA^{T}, and the columns of V (n by n) are eigenvectors of $A^{\mathrm{T}}A$. The r singular values on the diagonal of Σ (m by n) are the square roots of the nonzero eigenvalues of both AA^{T} and $A^{\mathrm{T}}A$.

Given the $N \times p$ data matrix $\mathbf{X}$, let

Hastie - Sec. 18.3.5 – PP 659

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (18.12)$$

$$= \mathbf{R}\mathbf{V}^T \quad (18.13)$$

be the singular-value decomposition (SVD) of $\mathbf{X}$; that is, $\mathbf{V}$ is $p \times N$ with orthonormal columns, $\mathbf{U}$ is $N \times N$ orthogonal, and $\mathbf{D}$ a diagonal matrix with elements $d_1 \geq d_2 \geq d_N \geq 0$. The matrix $\mathbf{R}$ is $N \times N$, with rows r_i^T.

Stack the (centered) observations into the rows of an $N \times p$ matrix $\mathbf{X}$. We construct the *singular value decomposition* of $\mathbf{X}$:

Sec. 14.5 – PP 535

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{14.54}$$

This is a standard decomposition in numerical analysis, and many algorithms exist for its computation (Golub and Van Loan, 1983, for example). Here $\mathbf{U}$ is an $N \times p$ orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$) whose columns $\mathbf{u}_j$ are called the *left singular vectors*; $\mathbf{V}$ is a $p \times p$ orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}_p$) with columns v_j called the *right singular vectors*, and $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ known as the *sin-*

Here $\mathbf{U}$ and $\mathbf{V}$ are $N \times p$ and $p \times p$ orthogonal matrices, with the columns of $\mathbf{U}$ spanning the column space of $\mathbf{X}$, and the columns of $\mathbf{V}$ spanning the row space. $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ called the singular values of $\mathbf{X}$. If one or more values $d_j = 0$, $\mathbf{X}$ is singular.

Using the singular value decomposition we can write the least squares fitted vector as

$$\begin{aligned} \mathbf{X}\hat{\beta}^{\text{ls}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= \mathbf{U}\mathbf{U}^T\mathbf{y}, \end{aligned} \tag{3.46}$$

Singular Value Decomposition

Remark 1.

- For positive definite matrices, Σ is Λ and $U\Sigma V^T$ is identical to $Q\Lambda Q^T$.
- For other symmetric matrices, any negative eigenvalues in Λ become positive in Σ.
- For complex matrices, Σ remains real but U and V become *unitary* (the complex version of orthogonal).

Remark 2.

U and V give orthonormal bases for all four fundamental subspaces:

first	r	columns of U:	**column space** of A
last	$m-r$	columns of U:	**left nullspace** of A
first	r	columns of V:	**row space** of A
last	$n-r$	columns of V:	**nullspace** of A

Singular Value Decomposition

Remark 3.

Eigenvectors of AA^T and A^TA must go into the columns of U and V:

$$AA^{\mathrm{T}} = (U\Sigma V^{\mathrm{T}})(V\Sigma^{\mathrm{T}}U^{\mathrm{T}}) = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}} \quad \text{and, similarly,} \quad A^{\mathrm{T}}A = V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}$$

- U must be the eigenvector matrix for AA^T .
- The eigenvalue matrix in the middle is $\Sigma\Sigma^T$ — which is $m \times m$ with $\sigma_1^2, \ldots, \sigma_r^2$ on the diagonal.

- From the $A^TA = V\Sigma^T\Sigma V^T$, the V matrix must be the eigenvector matrix for A^TA.

SVD – the theorem (*Src; WIKI ++*)

Suppose M is an m-by-n matrix whose entries come from the field K, which is either the field of real numbers or the field of complex numbers. Then there exists a factorization of the form

$$M = U\Sigma V^*$$

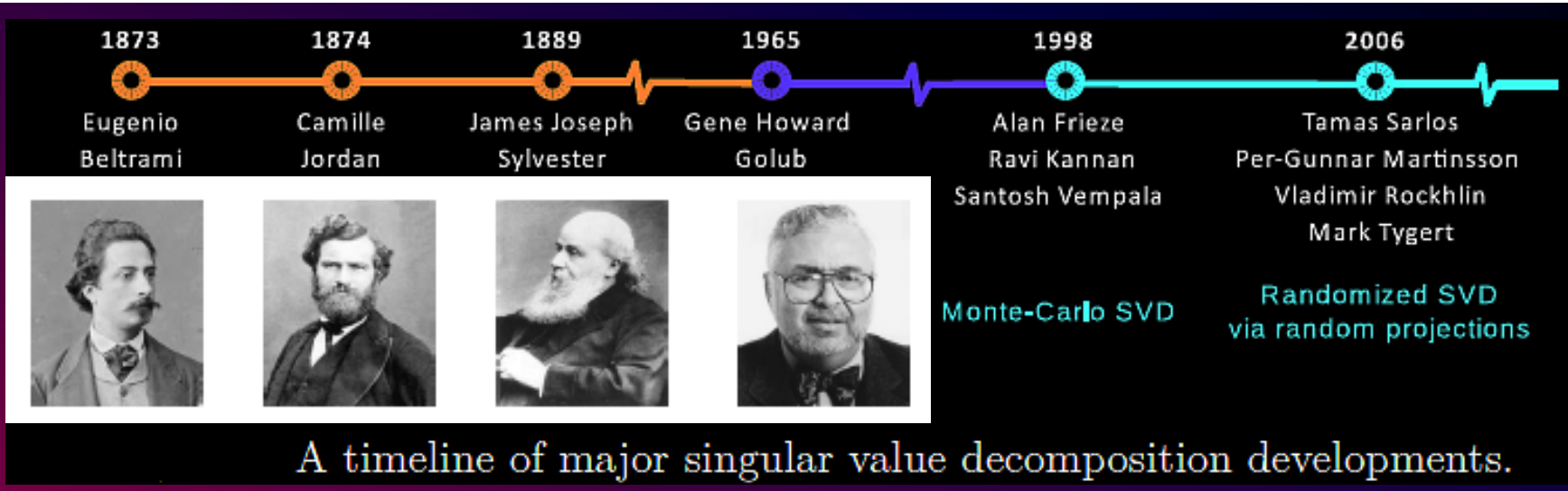
where, **U is an m-by-m** unitary matrix, the matrix **Σ is m-by-n** with nonnegative numbers on the diagonal and zeros off the diagonal, and V* denotes the conjugate transpose of **V, an n-by-n** unitary matrix over K. Such a factorization is called a (Full) *singular-value decomposition of M.*

The matrix V thus contains a set of orthonormal "input" or "analysing" basis vector directions for M.

The matrix U contains a set of orthonormal "output" basis vector directions for M. The matrix Σ contains the singular values, which can be thought of as scalar "gain controls" by which each corresponding input is multiplied to give a corresponding output.

A common convention is to order the values $\Sigma_{i,i}$ in non-increasing fashion. In this case, the diagonal matrix Σ is uniquely determined by M (though the matrices U and V are not).

For p = min(m,n) - ***U is m-by-p, Σ is p-by-p, and V is n-by-p.***



A timeline of major singular value decomposition developments.

Erichson, N. B., Voronin, S., **Brunton, S. L.**, & Kutz, J. N. (2019). Randomized Matrix

The columns of U $(m \times m)$ are ***eigenvectors of*** AA^T , and the columns of V $(n \times n)$ are ***eigenvectors of*** A^TA

Given a real matrix $\mathrm{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$, the singular value decomposition takes the form

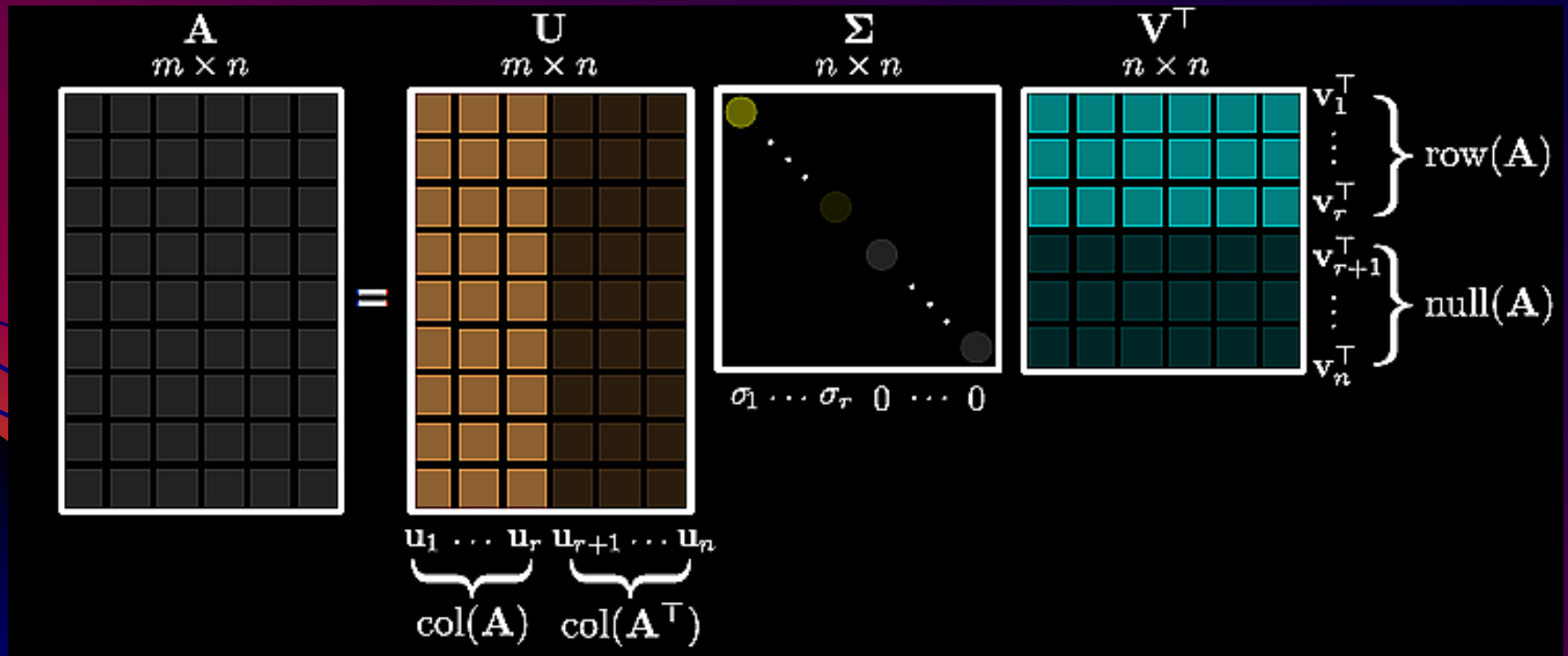
$$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\top}.$$

The matrices $\mathrm{U} = [\mathrm{u}_1, \ldots, \mathrm{u}_m] \in \mathbb{R}^{m \times m}$ and $\mathrm{V} = [\mathrm{v}_1, \ldots, \mathrm{v}_n] \in \mathbb{R}^{n \times n}$ are orthonormal so that $\mathrm{U}^{\top}\mathrm{U} = \mathrm{I}$ and $\mathrm{V}^{\top}\mathrm{V} = \mathrm{I}$. The left singular vectors in U provide a basis for the range (column space), and the right singular vectors in V provide a basis for the domain (row space) of the matrix A. The rectangular diagonal matrix $\Sigma \in \mathbb{R}^{m \times n}$ contains the corresponding non-negative singular values $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$, describing the spectrum of the data.

The so called "economy" or "thin" SVD computes only the left singular vectors and singular values corresponding to the number (i.e., n) of right singular vectors

$$A = U\Sigma V = [u_1, \ldots, u_n]\,\mathrm{diag}(\sigma_1, \ldots, \sigma_n)\,[v_1, \ldots, v_n]^\top.$$

If the number of right singular vectors is small (i.e. *n* << *m*), this is a more compact factorization than the full SVD.

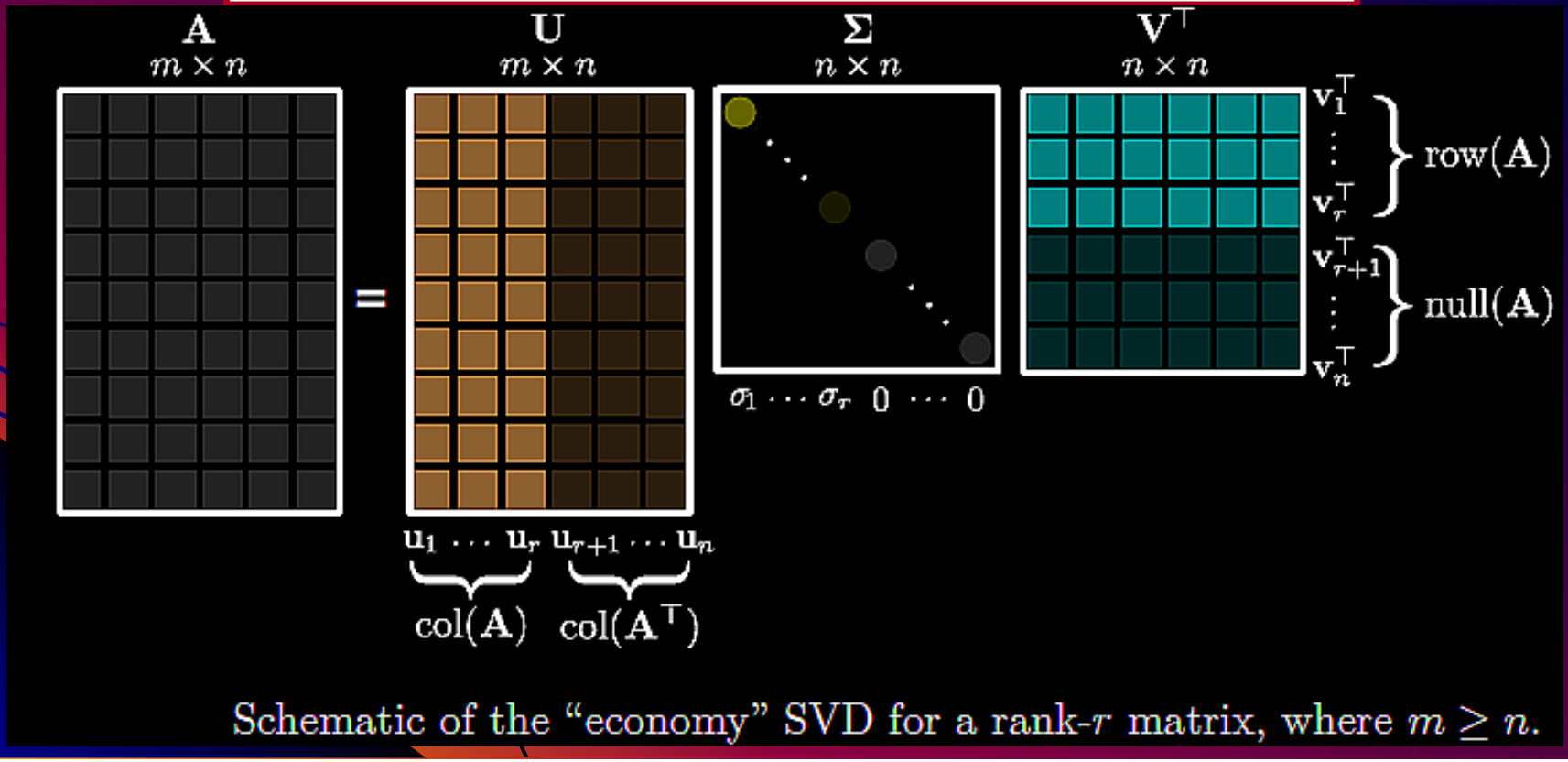


Schematic of the "economy" SVD for a rank-r matrix, where $m \geq n$.

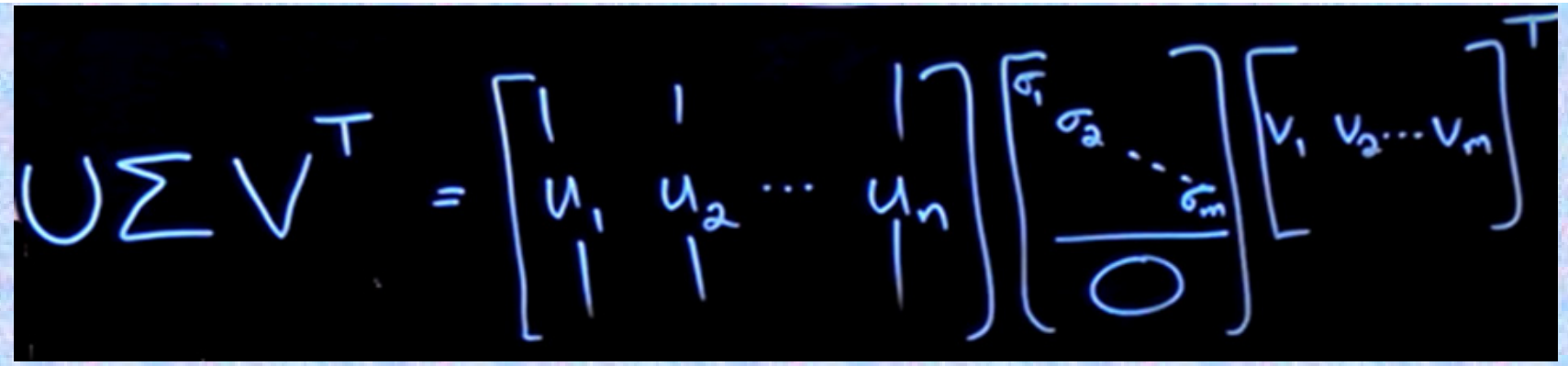
Low-rank matrices feature a rank (r) that is smaller than the dimensio and rows.

H o (0), and the nt null spaces.

first	r	columns of U:	**column space** of A
last	$m-r$	columns of U:	**left nullspace** of A
first	r	columns of V:	**row space** of A
last	$n-r$	columns of V:	**nullspace** of A



Schematic of the "economy" SVD for a rank-r matrix, where $m \geq n$.



$X_{n \times m} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix}$ (n × m)

$(X_{n \times p})^T$

Dim → n

samples → m

v_j^T → Mixture Coeffs of U_i's to create x_j

$= \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}_{n \times n} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \sigma_m \\ \hline 0 & & & \end{bmatrix}_{n \times m} \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{bmatrix}_{m \times m}$

$v_i \in R^m$; $u_i, x_i \in R^n$

$U^T U = U U^T = I_{n \times n}$ (L.S.V.)

$V^T V = V V^T = I_{m \times m}$ (R.S.V.)

In practical applications matrices are often contaminated by errors, and the effective rank of a matrix can be smaller than its exact rank *r*.

In this case, the matrix can be well approximated by including only those singular vectors which correspond to singular values of a *significant magnitude*. Hence, it is often desirable to compute a *reduced version of the SVD*, as:

$$\mathrm{A}_k := \mathrm{U}_k \Sigma_k \mathrm{V}_k = [\mathrm{u}_1, \ldots, \mathrm{u}_k] \mathrm{diag}(\sigma_1, \ldots, \sigma_k) [\mathrm{v}_1, \ldots, \mathrm{v}_k]^\top,$$

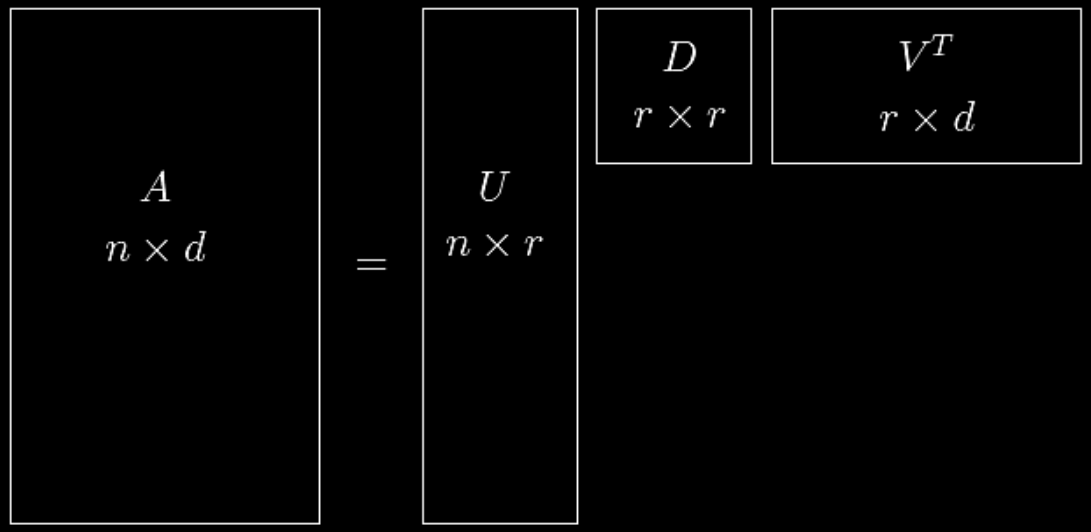
where k denotes the desired target rank of the approximation. In other words, this reduced form of the SVD allows one to express A approximately by the sum of k rank-one matrices

$$\mathrm{A}_k \approx \sum_{i=1}^{k} \sigma_i \mathrm{u}_i \mathrm{v}_i^\top.$$

Choosing an optimal target rank k is highly dependent on the task.

For massive datasets, however, the truncated/reduced SVD is costly to compute. The cost to compute the full SVD of an m × n matrix is of the order $O(mn^2)$, from which the first k components can then be extracted to form A_k.

k should be chosen close to the effective rank – data representation Applcn.; while, chosen much smaller (<< r) for dimension reduction (PCA).



The SVD decomposition of an $n \times d$ matrix.

Theorem 1.5 *Let A be an $n \times d$ matrix with right singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$, left singular vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}$, and corresponding singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$. Then*

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T.$$

$U^T U = \mathbf{I_{r*r}} <> UU^T$

$VV^T <> \mathbf{I_{r*r}} = V^T V$

B.Tech, CSE_ IIT Madras (1997);
Ph.D., MIT (2001);

Miller Research Fellow, UC Berkeley (2001-02);
CMU; Berkeley

$\mathbf{A}=\mathbf{U\Sigma V}^T$ and $\mathbf{A}^T=\mathbf{V\Sigma U}^T$

$\mathbf{A}^T\mathbf{A} = \mathbf{V\Sigma U}^T\mathbf{U\Sigma V}^T = \mathbf{V\Sigma}^2\mathbf{V}^T$
$\mathbf{A}^T\mathbf{AV} = \mathbf{V\Sigma}^2$; $\mathbf{V}^T\mathbf{A}^T\mathbf{AV} = \mathbf{\Sigma}^2$;

$\mathbf{A.A}^T = ??$

The diagonal entries $\sigma_i = \Sigma_{ii}$ of Σ are uniquely determined by $\mathbf{M}$ and are known as the singular values of $\mathbf{M}$. The number of non-zero singular values is equal to the rank of $\mathbf{M}$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called left-singular vectors and right-singular vectors of $\mathbf{M}$, respectively. They form two sets of orthonormal bases $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and if they are sorted so that the singular values σ_i with value zero are all in the highest-numbered columns (or rows), the singular value decomposition can be written as $\mathbf{M} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$, where $r \leq \min\{m, n\}$ is the rank of $\mathbf{M}$.

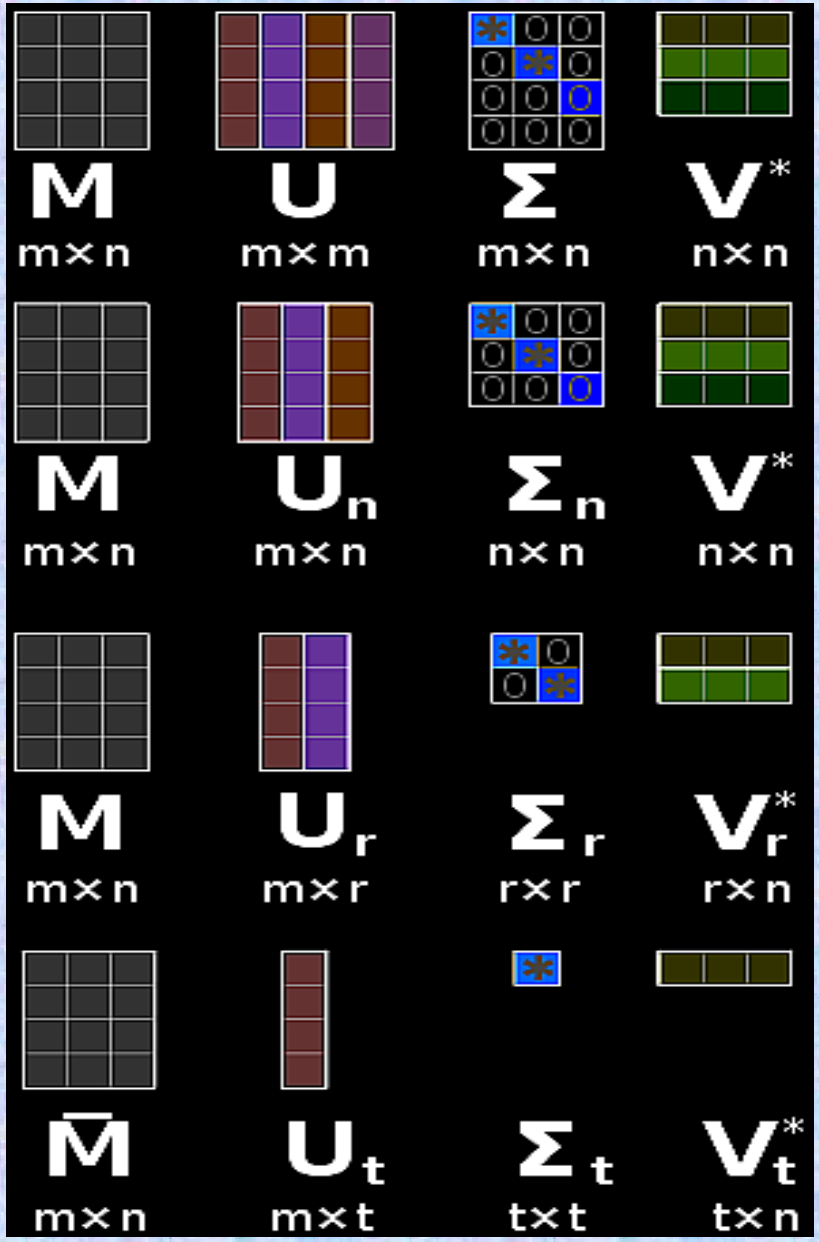
The term sometimes refers to the **compact SVD**, a similar decomposition $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^*$ in which Σ is square diagonal of size $r \times r$, where $r \leq \min\{m, n\}$ is the rank of $\mathbf{M}$, and has only the non-zero singular values. In this variant, $\mathbf{U}$ is an $m \times r$ semi-unitary matrix and $\mathbf{V}$ is an $n \times r$ semi-unitary matrix, such that $\mathbf{U}^*\mathbf{U} = \mathbf{V}^*\mathbf{V} = \mathbf{I}_r$.

1: Full SVD,

2: Thin SVD (remove columns of U not corresponding to rows of V*),

3: Compact SVD (remove vanishing singular values and corresponding columns/rows in U and V*),

4: Truncated SVD (keep only largest t singular values and corresponding columns/rows in U and V*)



A : m × n matrix of rank r

eigenvectors

$A A^T \longrightarrow \underbrace{u_1, u_2, u_3, \ldots, u_r}_{\text{span column space of } A}, \underbrace{\ldots, u_m}_{\text{span left nullspace of } A\text{: } N(A^T)}$

$A^T A \longrightarrow \underbrace{v_1, v_2, v_3, \ldots, v_r}_{\text{span row space of } A}, \underbrace{\ldots, v_n}_{\text{span nullspace of } A\text{: } N(A)}$

$$\underset{A}{\begin{pmatrix} x_{11} & x_{12} & & x_{1n} \\ & \ddots & & \\ x_{m1} & & & x_{mn} \end{pmatrix}} = \underset{U}{\begin{pmatrix} u_{11} & & u_{m1} \\ & \ddots & \\ u_{1m} & & u_{mm} \end{pmatrix}} \underset{S}{\begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & 0 \end{pmatrix}} \underset{V^T}{\begin{pmatrix} v_{11} & & v_{1n} \\ & \ddots & \\ v_{n1} & & v_{nn} \end{pmatrix}}$$

$$\underset{\substack{A \\ m \times n}}{\begin{pmatrix} x_{11} & x_{12} & & x_{1n} \\ & \ddots & & \\ x_{m1} & & & x_{mn} \end{pmatrix}} \underset{\substack{V \\ n \times r}}{\begin{pmatrix} v_{11} & & v_{r1} \\ & \ddots & \\ v_{1n} & & v_{rn} \end{pmatrix}} = \underset{\substack{U \\ m \times r}}{\begin{pmatrix} u_{11} & & u_{r1} \\ & \ddots & \\ u_{1m} & & u_{rm} \end{pmatrix}} \underset{\substack{S \\ r \times r}}{\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{pmatrix}}$$

Singular Value Decomposition

Example 1.

This A has only one column: rank r = 1. Then Σ has only $\sigma_1 = 3$:

$$\textbf{SVD} \qquad A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = U_{3\times 3}\Sigma_{3\times 1}V^{\mathrm{T}}_{1\times 1}$$

$A^{\mathrm{T}}A$ is 1 by 1, whereas AA^{T} is 3 by 3. They both have eigenvalue 9 (whose square root is the 3 in Σ). The two zero eigenvalues of AA^{T} leave some freedom for the eigenvectors in columns 2 and 3 of U. We kept that matrix orthogonal.

Singular Value Decomposition

Example 2.

Now A has rank 2, and $AA^T = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ with $\lambda = 3$ and 1:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = U\Sigma V^{\mathrm{T}} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{matrix} /\sqrt{6} \\ /\sqrt{2} \\ /\sqrt{3} \end{matrix}$$

Notice $\sqrt{3}$ and $\sqrt{1}$. The columns of U are *left singular vectors* (unit eigenvectors of $\boldsymbol{AA^T}$).

The columns of V are *right singular vectors* (unit eigenvectors of $\boldsymbol{A^T A}$).

In general, the best rank-r approximation to A is given by

$$\widehat{A} = \begin{bmatrix} \vdots & \vdots & \dots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_r \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}_{m\times r} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \sigma_r \end{bmatrix}_{r\times r} \begin{bmatrix} \vdots & \vdots & \dots & \vdots \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_r \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}^T_{r\times n}$$
$$= \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \ldots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T$$

and it can be shown that

$$\|A - \widehat{A}\|_F = \sqrt{\sigma_{r+1}^2 + \sigma_{r+2}^2 + \ldots + \sigma_n^2} \qquad (2)$$

The MATLAB command

```
[U, S, V] = svd(A)
```

returns the SVD decomposition of the matrix A, that is, it returns matrices U, S and V such that $A = USV^T$.

Example

The SVD of the following matrix A is:

$$A = \begin{bmatrix} -2 & 8 & 20 \\ 14 & 19 & 10 \\ 2 & -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} & 0 \\ \frac{4}{5} & \frac{3}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 30 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

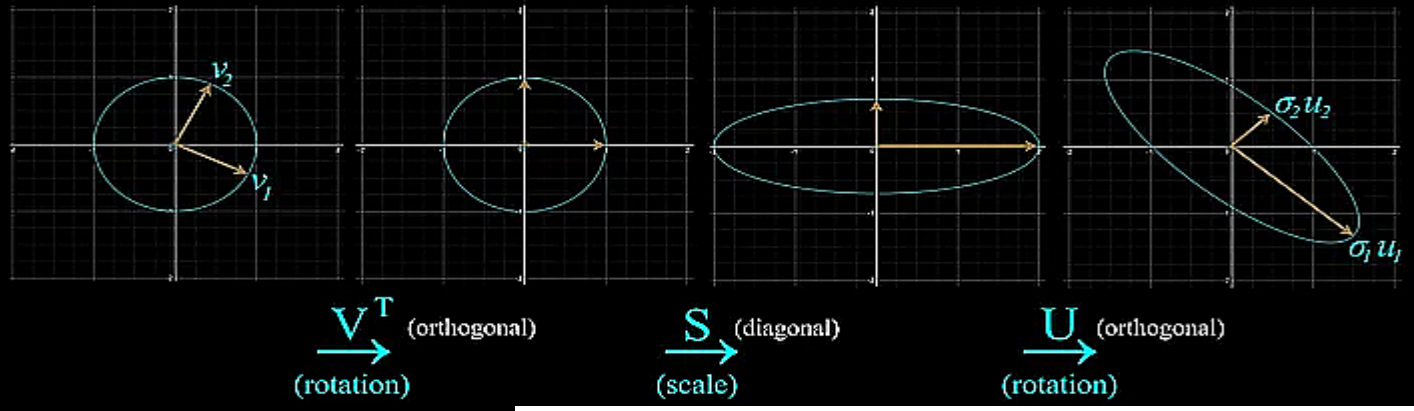
(a) Enter the matrix A and compute U, S and V using the svd comand. Verify that $A = USV^T$.

Answer:

```
>> A=[-2,8,20;14,19,10;2, -2, 1];
>> [U,S,V]=svd(A)
U =
    0.6000   -0.8000    0.0000
    0.8000    0.6000    0.0000
    0.0000   -0.0000    1.0000
S =
   30.0000         0         0
         0   15.0000         0
         0         0    3.0000
V =
    0.3333    0.6667    0.6667
    0.6667    0.3333   -0.6667
    0.6667   -0.6667    0.3333
```

We can easily verify that `U*S*V'` returns the matrix A.

$Ax = USV^T x$



Let $A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$. Then

$$AA^T = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix}.$$

$$A^TA = \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix}$$

$$\begin{aligned} &= \det(xI - AA^T) = \begin{vmatrix} x-11 & -5 \\ -5 & x-11 \end{vmatrix} \\ &= (x-11)^2 - 25 \\ &= x^2 - 22x + 121 - 25 \\ &= x^2 - 22x + 96 \\ &= (x-16)(x-6). \end{aligned}$$

$$\begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix} = \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\right) \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \left(\frac{1}{\sqrt{17}} \begin{bmatrix} 1 & -4 \\ 4 & 1 \end{bmatrix}\right)$$

$$V = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & -\sqrt{2} & -1 \\ 0 & -\sqrt{2} & 2 \\ \sqrt{3} & \sqrt{2} & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix},$$

$$\begin{aligned} A &= \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right) \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix} \left(\frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ -\sqrt{2} & -\sqrt{2} & \sqrt{2} \\ -1 & 2 & 1 \end{bmatrix}\right) \end{aligned}$$

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

$$A = USV^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{pmatrix}$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{(singular, inverse does not exist)}$$

$$A = U\Sigma V^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$A^+ = V\Sigma^+U^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$n = 5$:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}$$

As seen here, the SVD is given by $A = U\tilde{S}V^T$, with

$$U = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tilde{S} = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad V^T = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{pmatrix}.$$

The matrix is rank $r = 3$. A rank-two approximation is given by zeroing out the smallest singular value, which produces

$$\hat{A}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

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

$$
\begin{array}{cccc}
R & & U & \Sigma & V^T
\end{array}
$$

$$
\begin{bmatrix}
5 & 5 & 0 & 0 & 1 \\
4 & 5 & 1 & 1 & 0 \\
5 & 4 & 1 & 1 & 0 \\
0 & 0 & 4 & 4 & 4 \\
0 & 0 & 5 & 5 & 5 \\
1 & 1 & 4 & 4 & 4
\end{bmatrix}
\approx
\begin{bmatrix}
-0{,}27 & 0{,}55 & -0{,}78 & 0 \\
-0{,}29 & 0{,}47 & 0{,}44 & -0{,}71 \\
-0{,}29 & 0{,}47 & 0{,}44 & 0{,}71 \\
-0{,}45 & -0{,}29 & -0{,}01 & 0 \\
-0{,}56 & -0{,}36 & -0{,}02 & 0 \\
-0{,}50 & -0{,}18 & -0{,}05 & 0
\end{bmatrix}
\begin{bmatrix}
13{,}74 & 0 & 0 & 0 \\
0 & 10.88 & 0 & 0 \\
0 & 0 & 1.36 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
-0{,}32 & -0{,}32 & -0{,}52 & -0{,}52 & -0{,}5 \\
0{,}63 & 0{,}63 & -0{,}25 & -0{,}25 & -0{,}29 \\
-0{,}02 & -0{,}02 & 0{,}41 & 0{,}41 & -0{,}82 \\
0{,}71 & -0{,}71 & 0 & 0 & 0
\end{bmatrix}
$$

$$
\begin{bmatrix}
5 & 5 & 0 & 0 & 1 \\
4 & 5 & 1 & 1 & 0 \\
5 & 4 & 1 & 1 & 0 \\
0 & 0 & 4 & 4 & 4 \\
0 & 0 & 5 & 5 & 5 \\
1 & 1 & 4 & 4 & 4
\end{bmatrix}
\approx
\begin{bmatrix}
-0{,}27 & 0{,}55 \\
-0{,}29 & 0{,}47 \\
-0{,}29 & 0{,}47 \\
-0{,}45 & -0{,}29 \\
-0{,}56 & -0{,}36 \\
-0{,}50 & -0{,}18
\end{bmatrix}
\begin{bmatrix}
13{,}74 & 0 \\
0 & 10.88
\end{bmatrix}
\begin{bmatrix}
-0{,}32 & -0{,}32 & -0{,}52 & -0{,}52 & -0{,}5 \\
0{,}63 & 0{,}63 & -0{,}25 & -0{,}25 & -0{,}29
\end{bmatrix}
$$

$$
\begin{bmatrix}
4.95 & 4.95 & 0.43 & 0.43 & 0.11 \\
4.49 & 4.49 & 0.79 & 0.79 & 0.50 \\
4.49 & 4.49 & 0.79 & 0.79 & 0.50 \\
-0.00 & -0.00 & 4.00 & 4.00 & 4.00 \\
-0.00 & -0.00 & 4.98 & 4.98 & 4.98 \\
0.96 & 0.96 & 4.06 & 4.06 & 4.00
\end{bmatrix}
=
\begin{bmatrix}
-0{,}27 & 0{,}55 \\
-0{,}29 & 0{,}47 \\
-0{,}29 & 0{,}47 \\
-0{,}45 & -0{,}29 \\
-0{,}56 & -0{,}36 \\
-0{,}50 & -0{,}18
\end{bmatrix}
\begin{bmatrix}
13{,}74 & 0 \\
0 & 10.88
\end{bmatrix}
\begin{bmatrix}
-0{,}32 & -0{,}32 & -0{,}52 & -0{,}52 & -0{,}5 \\
0{,}63 & 0{,}63 & -0{,}25 & -0{,}25 & -0{,}29
\end{bmatrix}
$$

X:	U	Σ	V^T
[0 1 2] [-1 0 3] [-2 -3 0]	[-0.5976 0. 0.8018] [-0.7171 0.4472 -0.5345] [0.3586 0.8944 0.2673]	[3.7417 0. 0.] [0. 3.7417 0.] [0. 0. 0.]	[[-0. -0.4472 -0.8944] [-0.5976 -0.7171 0.3586] [0.8018 -0.5345 0.2673]]
[0 1 2] [-1 1 3] [-2 -3 0]	[0.5167 0.1746 0.8382] [0.6713 0.5251 -0.5232] [-0.5314 0.833 0.1541]	[4.1449 0. 0.] [0. 3.4265 0.] [0. 0. 0.2816]	[0.0945 0.6713 0.7352] [-0.6394 -0.5251 0.5616] [0.763 -0.5232 0.3796]]
[0 1.01 2] [-1 0 3] [-2 -3 0]	[-0.5238 -0.2893 0.8012] [-0.4091 -0.7395 -0.5345] [0.7472 -0.6078 0.269]	[3.7455 0. 0.] [0. 3.7405 0.] [0. 0. 0.0043]	[-0.2897 -0.7397 -0.6074] [0.5227 0.4093 -0.7478] [-0.8018 0.5341 -0.268]]
[0 1 2] [0 1.5 3] [0 0 0]	[-0.5547 0.8321 0.] [-0.8321 -0.5547 0.] [0. 0. 1.]	[4.0311 0. 0.] [0. 0. 0.] [0. 0. 0.]	[0. -0.4472 -0.8944] [-0. 0.8944 -0.4472] [1. 0. 0.]

X:	U	Σ	V^T
[40 -10 110] 80 -20 220] 120 -30 330]	[-0.2673 0.9561 0.1204] [-0.5345 -0.0431 -0.8441] [-0.8018 -0.29 0.5226]	[439.5452 0. 0.] [0. 0. 0.] [0. 0. 0.]	[-0.3405 0.0851 -0.9364] [-0.9379 0.0396 0.3447] [-0.0664 -0.9956 -0.0664]
[0 0 0] [-1 -2 -3] [-2 -4 -6]	[0. 0.8944 0.4472] [0.4472 -0.4 0.8] [0.8944 0.2 -0.4]	[8.3666 0. 0.] [0. 0. 0.] [0. 0. 0.]	[-0.2673 -0.5345 -0.8018] [0. -0.8321 0.5547] [0.9636 -0.1482 -0.2224]

X:	U	Σ	V^T

$$\begin{bmatrix} 0 & 0 & -3 & 0 \\ -\sqrt{2} & \frac{4\sqrt{3}}{3} & 0 & 0 \\ -\sqrt{2} & -\frac{4\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{4\sqrt{3}}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & \frac{-1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{2}{\sqrt{6}} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Applications of Singular Value Decomposition

Image Processing.

- Suppose a satellite takes a picture, and wants to send it to Earth.
- The picture may contain 1000×1000 "pixels"—a million little squares, each with a definite color.
- We can code the colors, and send back $1{,}000{,}000$ numbers.
- It is ***better to find the essential information inside the*** $\mathbf{1000} \times \mathbf{1000}$ ***matrix***, and send only that.

In SVD some σ's are significant and others are extremely small.

If we keep 20 and throw away 980, then we send only the corresponding 20 columns of U and V.

The other 980 columns are multiplied in $U\Sigma V^T$ by the small σ's that are being ignored. ***If only 20 terms are kept, we send 20 times 2000 numbers instead of a million (25 to 1 compression).***

Definition [edit]

Throughout this article, boldfaced unsubscripted $\mathbf{X}$ and $\mathbf{Y}$ are used to refer to random vectors, and unboldfaced subscripted X_i and Y_i are used to refer to scalar random variables.

If the entries in the column vector

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$$

are random variables, each with finite variance and expected value, then the covariance matrix $\mathrm{K}_{\mathbf{XX}}$ is the matrix whose (i, j) entry is the covariance[1]: p. 177

$$\mathrm{K}_{X_i X_j} = \operatorname{cov}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])]$$

where the operator E denotes the expected value (mean) of its argument.

Conflicting nomenclatures and notations [edit]

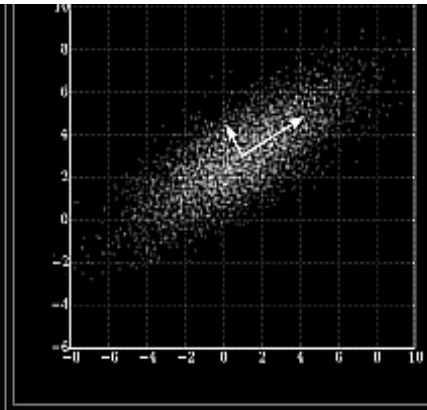
Nomenclatures differ. Some statisticians, following the probabilist William Feller in his two-volume book *An Introduction to Probability Theory and Its Applications*,[2] call the matrix $\mathrm{K}_{\mathbf{XX}}$ the **variance** of the random vector $\mathbf{X}$, because it is the natural generalization to higher dimensions of the 1-dimensional variance. Others call it the **covariance matrix**, because it is the matrix of covariances between the scalar components of the vector $\mathbf{X}$.

$$\operatorname{var}(\mathbf{X}) = \operatorname{cov}(\mathbf{X}, \mathbf{X}) = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\mathrm{T}}\right].$$

Both forms are quite standard, and there is no ambiguity between them. The matrix $\mathrm{K}_{\mathbf{XX}}$ is also often called the *variance-covariance matrix*, since the diagonal terms are in fact variances.

By comparison, the notation for the cross-covariance matrix *between* two vectors is

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathrm{K}_{\mathbf{XY}} = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^{\mathrm{T}}\right].$$



Sample points from a bivariate Gaussian distribution with a standard deviation of 3 in roughly the lower left-upper right direction and of 1 in the orthogonal direction. Because the x and y components co-vary, the variances of x and y do not fully describe the distribution. A 2×2 covariance matrix is needed; the directions of the arrows correspond to the eigenvectors of this covariance matrix and their lengths to the square roots of the eigenvalues.

Basic properties

For $\operatorname{K}_{\mathbf{XX}} = \operatorname{var}(\mathbf{X}) = \operatorname{E}\left[(\mathbf{X} - \operatorname{E}[\mathbf{X}])\,(\mathbf{X} - \operatorname{E}[\mathbf{X}])^{\mathrm{T}}\right]$ and $\mu_{\mathbf{X}} = \operatorname{E}[\mathbf{X}]$, where $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ is a n-dimensional random variable, the following basic properties apply:[4]

1. $\operatorname{K}_{\mathbf{XX}} = \operatorname{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}}) - \mu_{\mathbf{X}}{\mu_{\mathbf{X}}}^{\mathrm{T}}$
2. $\operatorname{K}_{\mathbf{XX}}$ is positive-semidefinite, i.e. $\mathbf{a}^T \operatorname{K}_{\mathbf{XX}}\, \mathbf{a} \geq 0 \quad \text{for all } \mathbf{a} \in \mathbb{R}^n$
3. $\operatorname{K}_{\mathbf{XX}}$ is symmetric, i.e. $\operatorname{K}_{\mathbf{XX}}^{\mathrm{T}} = \operatorname{K}_{\mathbf{XX}}$
4. For any constant (i.e. non-random) $m \times n$ matrix $\mathbf{A}$ and constant $m \times 1$ vector $\mathbf{a}$, one has $\operatorname{var}(\mathbf{A}\mathbf{X} + \mathbf{a}) = \mathbf{A}\, \operatorname{var}(\mathbf{X})\, \mathbf{A}^{\mathrm{T}}$
5. If $\mathbf{Y}$ is another random vector with the same dimension as $\mathbf{X}$, then $\operatorname{var}(\mathbf{X} + \mathbf{Y}) = \operatorname{var}(\mathbf{X}) + \operatorname{cov}(\mathbf{X}, \mathbf{Y}) + \operatorname{cov}(\mathbf{Y}, \mathbf{X}) + \operatorname{var}(\mathbf{Y})$ where $\operatorname{cov}(\mathbf{X}, \mathbf{Y})$ is the cross-covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$.

For random vectors $\mathbf{X}$ and $\mathbf{Y}$, each containing random elements whose expected value and variance exist, the **cross-covariance matrix** of $\mathbf{X}$ and $\mathbf{Y}$ is defined by[1]:p.336

$$\operatorname{K}_{\mathbf{XY}} = \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \stackrel{\text{def}}{=} \operatorname{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})^{\mathrm{T}}] \quad \textbf{(Eq.1)}$$

where $\mu_{\mathbf{X}} = \operatorname{E}[\mathbf{X}]$ and $\mu_{\mathbf{Y}} = \operatorname{E}[\mathbf{Y}]$ are vectors containing the expected values of $\mathbf{X}$ and $\mathbf{Y}$. The vectors $\mathbf{X}$ and $\mathbf{Y}$ need not have the same dimension, and either might be a scalar value.

The cross-covariance matrix is the matrix whose (i, j) entry is the covariance

$$\operatorname{K}_{X_i Y_j} = \operatorname{cov}[X_i, Y_j] = \operatorname{E}[(X_i - \operatorname{E}[X_i])(Y_j - \operatorname{E}[Y_j])]$$

For the cross-covariance matrix, the following basic properties apply:[2]

1. $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \operatorname{E}[\mathbf{X}\mathbf{Y}^{\mathrm{T}}] - \mu_{\mathbf{X}}{\mu_{\mathbf{Y}}}^{\mathrm{T}}$
2. $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \operatorname{cov}(\mathbf{Y}, \mathbf{X})^{\mathrm{T}}$
3. $\operatorname{cov}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{Y}) = \operatorname{cov}(\mathbf{X}_1, \mathbf{Y}) + \operatorname{cov}(\mathbf{X}_2, \mathbf{Y})$
4. $\operatorname{cov}(A\mathbf{X} + \mathbf{a}, B^{\mathrm{T}}\mathbf{Y} + \mathbf{b}) = A\operatorname{cov}(\mathbf{X}, \mathbf{Y})B$
5. If $\mathbf{X}$ and $\mathbf{Y}$ are independent (or somewhat less restrictedly, if every random variable in $\mathbf{X}$ is uncorrelated with every random variable in $\mathbf{Y}$), then $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = 0_{p\times q}$

where $\mathbf{X}$, $\mathbf{X_1}$ and $\mathbf{X_2}$ are random $p \times 1$ vectors, $\mathbf{Y}$ is a random $q \times 1$ vector, $\mathbf{a}$ is a $q \times 1$ vector, $\mathbf{b}$ is a $p \times 1$ vector, A and B are $q \times p$ matrices of constants, and $0_{p\times q}$ is a $p \times q$ matrix of zeroes.

Given a sample consisting of n independent observations $x_1,\ldots, x_n$ of a p-dimensional random vector $X \in \mathbf{R}^{p\times 1}$ (a $p\times 1$ column-vector), an unbiased estimator of the ($p\times p$) covariance matrix

$$\Sigma = \operatorname{E}\left[(X - \operatorname{E}[X])\,(X - \operatorname{E}[X])^{\mathrm{T}}\right]$$

is the sample covariance matrix

$$\mathbf{Q} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathrm{T}},$$

where x_i is the i-th observation of the p-dimensional random vector, and the vector

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the sample mean. This is true regardless of the distribution of the random variable X, provided of course that the theoretical means and covariances exist. The reason

$$\begin{aligned}\operatorname{Var}[bX] &= \mathrm{E}[(bX - \mathrm{E}[bX])(bX - \mathrm{E}[bX])^\top] \\ &= \mathrm{E}[(bX - b\mathrm{E}[X])(bX - b\mathrm{E}[X])^\top]\end{aligned}$$

Which matrices are covariance matrices?

let $\mathbf{b}$ be a $(p \times 1)$ real-valued vector, then

$$\operatorname{var}(\mathbf{b}^{\mathrm{T}}\mathbf{X}) = \mathbf{b}^{\mathrm{T}} \operatorname{var}(\mathbf{X})\mathbf{b},$$

which must always be nonnegative, since it is the variance of a real-valued random variable, so a covariance matrix is always a positive-semidefinite matrix.

The above argument can be expanded as follows:

$$\begin{aligned} & w^{\mathrm{T}} \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\mathrm{T}}\right] w = \mathrm{E}\left[w^{\mathrm{T}}(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\mathrm{T}} w\right] \\ & = \mathrm{E}\left[\left(w^{\mathrm{T}}(\mathbf{X} - \mathrm{E}[\mathbf{X}])\right)^2\right] \geq 0, \end{aligned}$$

where the last inequality follows from the observation that $w^{\mathrm{T}}(\mathbf{X} - \mathrm{E}[\mathbf{X}])$ is a scalar.

Conversely, every symmetric positive semi-definite matrix is a covariance matrix. To see this, suppose M is a $p \times p$ symmetric positive-semidefinite matrix. From the finite-dimensional case of the spectral theorem, it follows that M has a nonnegative symmetric square root, which can be denoted by $\mathbf{M}^{1/2}$. Let $\mathbf{X}$ be any $p \times 1$ column vector-valued random variable whose covariance matrix is the $p \times p$ identity matrix. Then

$$\operatorname{var}(\mathbf{M}^{1/2}\mathbf{X}) = \mathbf{M}^{1/2} \operatorname{var}(\mathbf{X}) \mathbf{M}^{1/2} = \mathbf{M}.$$

The **conjugate transpose**, also known as the **Hermitian transpose**, of an m × n complex matrix A is an n × m matrix obtained by transposing A and applying complex conjugate on each entry (the complex conjugate of a + ib being a − i b, for real numbers a and b)

$$\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}} = \left(\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}}.$$

$$\boldsymbol{A}^{\mathrm{H}} = \left(\overline{\boldsymbol{A}}\right)^{\mathsf{T}} = \overline{\boldsymbol{A}^{\mathsf{T}}}$$

A ***matrix is full row rank*** when each of the rows of the matrix are linearly independent and full column rank when each of the columns of the matrix are linearly independent.

For a ***square matrix*** these two concepts are equivalent and we say the matrix is ***full rank if all rows and columns are linearly independen***t. A square matrix is full rank if and only if its determinant is nonzero.

For a ***non-square matrix*** with m rows and n columns, it will always be the case that either the rows or columns (whichever is larger in number) are linearly dependent. Hence when we say that a non-square matrix is full rank, we mean that the row and column rank are as high as possible, given the shape of the matrix. So, if there are ***more rows than columns (m > n), then the matrix is full rank if the matrix is full column rank.***

The ***rank of A equals the number of non-zero singular values***, which is the same as the number of non-zero diagonal elements ***in Σ*** in the singular value decomposition A = U Σ V*

If A is a matrix over the real numbers then the rank of A and the rank of its corresponding Gram matrix are equal. Thus, for real matrices

$$\operatorname{rank}(A^{\mathrm{T}}A) = \operatorname{rank}(AA^{\mathrm{T}}) = \operatorname{rank}(A) = \operatorname{rank}(A^{\mathrm{T}}).$$

suppose A is an $n \times m$ matrix and $n \neq m$. It must be that
$rank(A^t) = rank(A) \leq \min(n, m) < \max(n, m)$.

Using the fact that $rank(AB) \leq rank(A)$ for any A, B for which the product is defined, we have that:

$$rank(AA^t) \leq rank(A) < \max(n, m)$$

$$rank(A^tA) \leq rank(A^t) < \max(n, m).$$

But it must be the case that the dimensions of AA^t or A^tA is $\max(n, m)$. Therefore at least one of them does not have full rank. For square matrices, not having full rank is equivalent to being singular.

Example 1

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\mathrm{A}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

$$C = AA^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix}$$

$$D = A^T A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix}$$

$$\mathrm{rank(A)} = \mathrm{rank}(A^T) = \mathrm{rank(C)} = \mathrm{rank(D)} = 2$$

Example 2

$$A = \begin{bmatrix} 3 & 6 & 1 & 1 & 7 \\ 1 & 2 & 2 & 3 & 1 \\ 2 & 4 & 5 & 8 & 4 \\ 0 & 0 & 1 & 2 & 2 \end{bmatrix}$$

$$\mathrm{A}^T = \begin{bmatrix} 3 & 1 & 2 & 0 \\ 6 & 2 & 4 & 0 \\ 1 & 2 & 5 & 1 \\ 1 & 3 & 8 & 2 \\ 7 & 1 & 4 & 2 \end{bmatrix}$$

$$C = AA^T = \begin{bmatrix} 96 & 27 & 71 & 17 \\ 27 & 19 & 48 & 10 \\ 71 & 48 & 125 & 29 \\ 17 & 10 & 29 & 9 \end{bmatrix}$$

$$D = A^T A = \begin{bmatrix} \mathbf{14} & \mathbf{28} & \mathbf{15} & \mathbf{22} & \mathbf{30} \\ \mathbf{28} & \mathbf{56} & \mathbf{30} & \mathbf{44} & \mathbf{60} \\ \mathbf{15} & \mathbf{30} & \mathbf{31} & \mathbf{49} & \mathbf{31} \\ \mathbf{22} & \mathbf{44} & \mathbf{49} & \mathbf{78} & \mathbf{46} \\ \mathbf{30} & \mathbf{60} & \mathbf{31} & \mathbf{46} & \mathbf{70} \end{bmatrix}$$

$$\mathrm{rank(A)} = \mathrm{rank}(A^T) = \mathrm{rank(C)} = \mathrm{rank(D)} = 3$$

Principal component analysis

Herve´ Abdi and Lynne J. Williams;
UT Dalas, Ontario
(Behavioural Brain Sc, Psychology)

WIREs Computational Statistics
Volume 2; Issue 4; ACM-DL
July 2010

John Wiley & Sons, Inc.
DOI: 10.1002/wics.101

The data table to be analyzed by PCA comprises I observations described by J variables and it is represented by the $I \times J$ matrix X, whose generic element is $x_{i,j}$. The matrix X has rank L where $L \leq \min\{I, J\}$.

In general, the data table will be preprocessed before the analysis. Almost always, the columns of X will be centered so that the mean of each column is equal to 0 (i.e., $\mathbf{X}^{\mathsf{T}}\mathbf{1} = 0$, where 0 is a J by 1 vector of zeros and 1 is an I by 1 vector of ones). If in addition, each element of X is divided by $\sqrt{I}$ (or $\sqrt{I-1}$), the analysis is referred to as a *covariance* PCA because, in this case, the matrix $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ is a covariance matrix. In addition to centering, when the variables are measured with different units, it is customary to standardize each variable to unit norm. This is obtained by dividing each variable by its norm (i.e., the square root of the sum of all the squared elements of this variable). In this case, the analysis is referred to as a *correlation* PCA because, then, the matrix $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ is a correlation matrix (most statistical packages use correlation preprocessing as a default).

21.1 The covariance matrix and principal component analysis

Suppose $\mathbf{S}$ is an $m \times n$ data matrix, in which the first dimension is the space-like dimension and the second is the time-like dimension. At each location i, we assume that we have subtracted off the time mean, so that $S_i(t)$ has a time-mean of zero. Then the covariance C_{ip} between measurement $S_{ij} = S_i(t_j)$ at location i and $S_{pj} = S_p(t_j)$ at location p is

$$C_{ip} = \frac{1}{n-1} \sum_{j=1}^{n} S_{ij} S_{pj}$$

Lecture 21: Principal Component Analysis

Winter 2014

The sum can be regarded as an inner product of the i'th row of $\mathbf{S}$ and the j'th column of $\mathbf{S}^T$. Hence we can assemble a covariance matrix between the m different space locations, whose i, p element is C_{ip}:

$$\mathbf{C}_S = \frac{1}{n-1} \mathbf{S}\mathbf{S}^T$$

The covariance matrix $\mathbf{C}_S$ is real and symmetric. Hence it can be diagonalized, which simplifies the covariance structure. We could do this in the normal way by finding its eigenvalues and eigenvectors. However, there is also a close relationship between $\mathbf{C}_S$ and the SVD of the data matrix which is advantageous to exploit.

$$COV(X) = XX^T = W\Sigma\Sigma^T W^T = WDW^T$$
COV (Y) = D

Let the SVD of the data matrix be:

$$\mathbf{S} = \mathbf{U}\Sigma\mathbf{V}^T$$

Basic properties of the SVD give the diagonalization of $\mathbf{C}_S$:

$$\mathbf{C}_S = \mathbf{U}\Lambda\mathbf{U}^T$$

where $\Lambda = \mathrm{diag}(\sigma_k^2)/(n-1)$. The left singular vectors $\mathbf{u}_k$ of the data matrix are the eigenvectors of the covariance matrix. If we rotate the data matrix into this basis by setting $\hat{\mathbf{S}} = \mathbf{U}^T\mathbf{S}$, then the rotated data has covariance matrix

$$\mathbf{C}_{\hat{S}} = \frac{1}{n-1}\hat{\mathbf{S}}\hat{\mathbf{S}}^T = \mathbf{U}^T\mathbf{C}_S\mathbf{U} = \Lambda$$

$$W_L^T X = \Sigma_L V_L^T$$

In this rotated coordinate system, the variance decouples (is mutually uncorrelated) between coordinate directions, with direction k contributing a variance $\sigma_k^2/(n-1)$. The time-dependent amplitude of the k'th rotated data component is $\mathbf{u}_k^T\mathbf{S} = \sigma_k\mathbf{v}_k$.

This decomposition of the covariance matrix is called principal component analysis (PCA). The vectors $\mathbf{u}_k$ are called the loading vectors or patterns. The vectors $\mathbf{v}_k$ are called the principal components or PCs. Various application-dependent normalizations are applied to the PCs

Steps for principal component analysis

- Principal component analysis using the covariance function should only be considered if all of the variables have the same units of measurement.

If the variables have different units of measurement, (i.e., pounds, feet, gallons, etc), or if we wish each variable to receive equal weight in the analysis, then the variables should be standardized before conducting a principal components analysis. To standardize a variable, subtract the mean and divide by the standard deviation:

$$Z_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j}$$

where

- X_{ij} = Data for variable *j* in sample unit *i*
- $\bar{x}_j$ = Sample mean for variable *j*
- s_j = Sample standard deviation for variable *j*

Note! The variance-covariance matrix of the standardized data is equal to the correlation matrix for the unstandardized data. Therefore, principal component analysis using the standardized data is equivalent to principal component analysis using the correlation matrix.

A Summary of the PCA Approach

- Standardize the data.
- Obtain the Eigenvectors and Eigenvalues from the covariance matrix or correlation matrix, or perform Singular Value Decomposition.
- Eigenvalues from SVD are sorted in descending order; so choose the **k** eigenvectors that correspond to the **k** largest eigenvalues where **k** is the number of dimensions of the new feature subspace (**k**≤d).
- Construct the projection matrix W from the selected **k** eigenvectors.

Using SVD on the data matrix has two advantages over just calling the Matlab function **eig** on the covariance matrix $\mathbf{C}_S$, which would give the $\sigma_k^2)/(n-1)$s as the eigenvalues, and the patterns $\mathbf{u}_k$ as the eigenvectors. First, if $m > n$ (more variables than samples), $\mathbf{C}_S$ is $m \times m$, which can become very large (m is over 2000 in our Pacific SST example, while n is only 396). Only n or less of these eigenvalues will be nonzero, but ihis can choke Matlab. Note there is a short version **svds** that, like **eigs** will just return a small number of leading singular modes, which is all we usually care about in PCA. That can minimize computation and memory requirements if the dataset is large.

Second, the right singular vectors automatically give the principal component time series for the patterns. To get these by eigendecomposition of the covariance matrix requires an extra step of projecting the data at each time onto the eigenvectors $\mathbf{u}_k$.

The Karhunen-Loève transform is therefore equivalent to finding the singular value decomposition of the data matrix *X*, and then obtaining the reduced-space data matrix Y by projecting X down into the reduced space defined by only the first *L* singular vectors, W_L:

$$X = W\Sigma V^T; \quad Y = W_L^T X = \Sigma_L V_L^T$$

The matrix W of singular vectors of X is equivalently the matrix W of eigenvectors of the matrix of observed covariances $C = X X^T$ *(find out?) =:*

$$COV(X) = XX^T = W\Sigma\Sigma^T W^T = WDW^T$$

The eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the data set. PCA is equivalent to empirical orthogonal functions (EOF).

PCA is a popular technique in pattern recognition. But it is not optimized for class separability. An alternative is the linear discriminant analysis, which does take this into account. PCA optimally minimizes reconstruction error under the L_2 norm.

PCA by COVARIANCE Method

We need to find a dxd orthonormal transformation matrix W^T, such that:

$$Y = W^T X$$

with the constraint that:
Cov(Y) is a diagonal matrix, and $W^{-1} = W^T$.

$$COV(Y) = E[YY^T] = E[(W^T X)(W^T X)^T]$$
$$= E[(W^T X)(X^T W)] = W^T E[XX^T]W$$
$$= W^T COV(X)W = W^T (WDW^T)W = D$$

$$WCOV(Y) = WW^T COV(X)W = COV(X)W$$

Can you derive from the above, that:

$$[\lambda_1 W_1, \lambda_2 W_2, ..., \lambda_d W_d] =$$
$$[COV(X)W_1, COV(X)W_2, ..., COV(X)W_d]$$

Maximise $\mathbf{u}^T\mathbf{x}\mathbf{x}^T\mathbf{u}$ s.t $\mathbf{u}^T\mathbf{u} = 1$

Construct Langrangian $\mathbf{u}^T\mathbf{x}\mathbf{x}^T\mathbf{u} - \lambda\mathbf{u}^T\mathbf{u}$

Vector of partial derivatives set to zero

$$\mathbf{x}\mathbf{x}^T\mathbf{u} - \lambda\mathbf{u} = (\mathbf{x}\mathbf{x}^T - \lambda\mathbf{I})\,\mathbf{u} = 0$$

As $\mathbf{u} \neq \mathbf{0}$ then $\mathbf{u}$ must be an eigenvector of $\mathbf{x}\mathbf{x}^T$ with eigenvalue λ

let $\mathbf{u}$ be an arbitrary vector of length 1, so that $\mathrm{u}^T S$ (a column vector of length n) is its projection on the data matrix. Let $\hat{\mathrm{u}} = \mathbf{U}^T\mathbf{u}$ be this unit vector expressed into the rotated basis (in which it will also have length 1). Then

$$\begin{aligned}
\mathrm{var}[\mathrm{u}^T S] &= \frac{1}{n-1}\mathrm{u}^T S S^T \mathrm{u} \\
&= \mathrm{u}^T \mathbf{U}\Lambda\mathbf{U}^T \mathrm{u} \\
&= \hat{\mathrm{u}}^T \Lambda \hat{\mathrm{u}} \\
&= \frac{1}{n-1}\sum_k \sigma_k^2 \hat{u}_k^2 \\
&\leq \frac{\sigma_1^2}{n-1}\sum_k \hat{u}_k^2 = \frac{\sigma_1^2}{n-1}.
\end{aligned}$$

Example of PCA

Samples: $x_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}; x_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}; x_3 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix};$ $\quad X = \begin{bmatrix} -1 & -2 & 4 \\ 1 & 3 & 0 \\ 2 & 1 & 3 \end{bmatrix}$

3-D problem, with N = 3.

Each column is an observation (sample) and each row a variable (dimension),

Mean of the samples: $\mu_x = \begin{bmatrix} 1/3 \\ 4/3 \\ 2 \end{bmatrix};$ $\quad \tilde{x}_1 = \begin{bmatrix} -4/3 \\ -1/3 \\ 0 \end{bmatrix}; \tilde{x}_2 = \begin{bmatrix} -7/3 \\ 5/3 \\ -1 \end{bmatrix}; \tilde{x}_3 = \begin{bmatrix} 11/3 \\ -4/3 \\ 1 \end{bmatrix};$

Method – 1 (easiest)

$\tilde{X} = \begin{bmatrix} -4/3 & -7/3 & 11/3 \\ -1/3 & 5/3 & -4/3 \\ 0 & -1 & 1 \end{bmatrix};$ $\quad$ COVAR = $(\tilde{X}\,\tilde{X}^T)/2 = (1/2)\begin{bmatrix} 62/3 & -25/3 & 6 \\ -25/3 & 14/3 & -3 \\ 6 & -3 & 2 \end{bmatrix}$

Method – 2 (PCA defn.)

$$S_T = \left(\frac{1}{N-1}\right)\sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T$$

$$\tilde{x}_1 = \begin{bmatrix} -4/3 \\ -1/3 \\ 0 \end{bmatrix}; \tilde{x}_2 = \begin{bmatrix} -7/3 \\ 5/3 \\ -1 \end{bmatrix}; \tilde{x}_3 = \begin{bmatrix} 11/3 \\ -4/3 \\ 1 \end{bmatrix};$$

C1 =

1.7778	0.4444	0
0.4444	0.1111	0
0	0	0

C2 =

5.4444	-3.8889	2.3333
-3.8889	2.7778	-1.6667
2.3333	-1.6667	1.0000

C3 =

13.4444	-4.8889	3.6667
-4.8889	1.7778	-1.3333
3.6667	-1.3333	1.0000

SigmaC =

20.6667	-8.3333	6.0000
-8.3333	4.6667	-3.0000
6.0000	-3.0000	2.0000

COVAR =
SigmaC/2 =

10.3333	-4.1667	3.0000
-4.1667	2.3333	-1.5000
3.0000	-1.5000	1.0000

Next do SVD, to get vectors.

For a face image with N samples and dimension d (=w*h, very large), we have:

The array X or X_{avg} of size d*N (N vertical samples stacked horizontally)

Thus XX^T will be of d*d, which will be very large. To perform eigen-analysis on such large dimension is time consuming and may be erroneous.

Thus often X^TX of dimension N*N is considered for eigen-analysis. Will it result in the same, after SVD? Lets check:

$$S = \tilde{X}\,\tilde{X}^{T} = (1/2)\begin{bmatrix} 62/3 & -25/3 & 6 \\ -25/3 & 14/3 & -3 \\ 6 & -3 & 2 \end{bmatrix} = \begin{matrix} 10.3333 & -4.1667 & 3.0000 \\ -4.1667 & 2.3333 & -1.5000 \\ 3.0000 & -1.5000 & 1.0000 \end{matrix}$$

$$S^{m} = \tilde{X}^{T}\,\tilde{X} = \begin{matrix} 0.9444 & 1.2778 & -2.2222 \\ 1.2778 & 4.6111 & -5.8889 \\ -2.2222 & -5.8889 & 8.1111 \end{matrix}$$

Lets do SVD of both:

$S = X \tilde{X}^T =$

10.3333	-4.1667	3.0000
-4.1667	2.3333	-1.5000
3.0000	-1.5000	1.0000

U =

-0.8846	-0.4554	-0.1010
0.3818	-0.8313	0.4041
-0.2680	0.3189	0.9091

S =

13.0404	0	0
0	0.6263	0
0	0	0.0000

V =

-0.8846	-0.4554	0.1010
0.3818	-0.8313	-0.4041
-0.2680	0.3189	-0.9091

$S^m = \tilde{X}^T \tilde{X} =$

0.9444	1.2778	-2.2222
1.2778	4.6111	-5.8889
-2.2222	-5.8889	8.1111

U =

-0.2060	0.7901	0.5774
-0.5812	-0.5735	0.5774
0.7872	-0.2166	0.5774

S =

13.0404	0	0
0	0.6263	0
0	0	0.0000

V =

-0.2060	0.7901	0.5774
-0.5812	-0.5735	0.5774
0.7872	-0.2166	0.5774

Samples:

Example, where d <> N:

$$x_1 = \begin{bmatrix} -3 \\ -3 \end{bmatrix}; x_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}; x_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}; x_4 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}; x_5 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}; x_6 = \begin{bmatrix} 6 \\ 7 \end{bmatrix};$$

2-D problem (d=2), with N = 6.

Each column is an observation (sample) and each row a variable (dimension),

X =

-3	**-2**	**-1**	**4**	**5**	**6**
-3	**-2**	**-1**	**4**	**5**	**7**

Mean of the samples:

$$\mu_x = \begin{bmatrix} 3 / 2 \\ 5 / 3 \end{bmatrix};$$

XM=

-4.5000	-3.5000	-2.5000	2.5000	3.5000	4.5000
-4.6667	-3.6667	-2.6667	2.3333	3.3333	5.3333

$XM^T * XM =$

42.0278	32.8611	23.6944	-22.1389	-31.3056	-45.1389
32.8611	25.6944	18.5278	-17.3056	-24.4722	-35.3056
23.6944	18.5278	13.3611	-12.4722	-17.6389	-25.4722
-22.1389	-17.3056	-12.4722	11.6944	16.5278	23.6944
-31.3056	-24.4722	-17.6389	16.5278	23.3611	33.5278
-45.1389	-35.3056	-25.4722	23.6944	33.5278	48.6944

$COVAR(X) = XM * XM^T$

=

77.5000	82.0000
82.0000	87.3333

COVAR(X) = XM * XM^T

$$= \begin{matrix} 77.5000 & 82.0000 \\ 82.0000 & 87.3333 \end{matrix}$$

U =

$$\begin{matrix} -0.6856 & -0.7280 \\ -0.7280 & 0.6856 \end{matrix}$$

S =

$$\begin{matrix} 164.5639 & 0 \\ 0 & 0.2694 \end{matrix}$$

V =

$$\begin{matrix} -0.6856 & -0.7280 \\ -0.7280 & 0.6856 \end{matrix}$$

$XM^T * XM =$

$$\begin{matrix} 42.0278 & 32.8611 & 23.6944 & -22.1389 & -31.3056 & -45.1389 \\ 32.8611 & 25.6944 & 18.5278 & -17.3056 & -24.4722 & -35.3056 \\ 23.6944 & 18.5278 & 13.3611 & -12.4722 & -17.6389 & -25.4722 \\ -22.1389 & -17.3056 & -12.4722 & 11.6944 & 16.5278 & 23.6944 \\ -31.3056 & -24.4722 & -17.6389 & 16.5278 & 23.3611 & 33.5278 \\ -45.1389 & -35.3056 & -25.4722 & 23.6944 & 33.5278 & 48.6944 \end{matrix}$$

U =

$$\begin{matrix} -0.5053 & -0.1469 & -0.7547 & 0.3882 & 0.0214 & 0.0486 \\ -0.3951 & -0.0654 & 0.3632 & 0.0984 & -0.4091 & 0.7284 \\ -0.2849 & 0.0162 & -0.0433 & -0.3456 & -0.7396 & -0.5002 \\ 0.2660 & 0.4241 & -0.5083 & -0.5306 & -0.1150 & 0.4429 \\ 0.3762 & 0.5057 & -0.0258 & 0.6601 & -0.4043 & -0.0539 \\ 0.5432 & -0.7337 & -0.1938 & 0.0541 & -0.3293 & 0.1332 \end{matrix}$$

S =

$$\begin{matrix} 164.5639 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2694 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0 \end{matrix}$$

V = U ??

X:

```
[5 5 0 0 1]
[4 5 1 1 0]
[5 4 1 1 0]
[0 0 4 4 4]
[0 0 5 5 5]
[1 1 4 4 4]
```

Covariance Matrix for X:

```
[ 4.917  4.75  -4.083 -4.083 -4.333]
[ 4.75   4.917 -4.083 -4.083 -4.333]
[-4.083 -4.083  3.583  3.583  3.667]
[-4.083 -4.083  3.583  3.583  3.667]
[-4.333 -4.333  3.667  3.667  4.222]
```

SVD applied on Covariance Matrix:

U:

```
[-0.482  0.076 -0.707 -0.511   0.   ]
[-0.482  0.076  0.707 -0.511   0.   ]
[ 0.413 -0.365 -0.    -0.443   0.707]
[ 0.413 -0.365 -0.    -0.443  -0.707]
[ 0.44    0.85   -0.   -0.289   0.   ]
```

D:

```
[20.611  0.308  0.167  0.137  0.   ]
```

V^T:

```
[-0.482 -0.482  0.413  0.413   0.44 ]
[ 0.076  0.076 -0.365 -0.365   0.85 ]
[-0.707  0.707  -0.     -0.     -0.   ]
[-0.511 -0.511 -0.443 -0.443  -0.289]
[  0.     -0.    0.707 -0.707   -0.   ]
```

X:

```
[5 5 0 0 1]
[4 5 1 1 0]
[5 4 1 1 0]
[0 0 4 4 4]
[0 0 5 5 5]
[1 1 4 4 4 ]
```

Covariance Matrix for X^T:

```
[ 5.36  4.16  4.16 -4.48 -5.6  -3.36]
[ 4.16  3.76  3.56 -3.68 -4.6  -2.76]
[ 4.16  3.56  3.76 -3.68 -4.6  -2.76]
[-4.48 -3.68 -3.68  3.84  4.8   2.88]
[-5.60  -4.60  -4.6   4.8   6.0   3.6 ]
[-3.36 -2.76 -2.76  2.88  3.6   2.16]
```

SVD applied on Covariance Matrix of X^T:

U:

```
[-0.462  0.669    -0.    -0.486  0.31    0.087]
[-0.383 -0.518  0.707 -0.243  0.155   0.043]
[-0.383 -0.518 -0.707 -0.243  0.155   0.043]
[ 0.397 -0.071     0.    -0.289  0.492  -0.715]
[ 0.497 -0.088    -0.      -0.72   -0.3    0.37 ]
[ 0.298 -0.053     0.       0.21   0.723  0.584]
```

D:

```
[24.292  0.388  0.2    0.     0.     0.   ]
```

V^T:

```
[-0.462 -0.383 -0.383  0.397  0.497  0.298]
[ 0.669 -0.518 -0.518 -0.071 -0.088 -0.053]
[-0.        0.707 -0.707     0.        -0.      0.   ]
[-0.462 -0.231 -0.231 -0.253 -0.74   0.261]
[-0.328 -0.164 -0.164 -0.608  0.298 -0.616]
[-0.135 -0.067 -0.067  0.635 -0.33  -0.679]
```

Q1

X:
[-2 8 20]
[14 19 10]
[2 -2 1]

Covariance Matrix of X:

[46.222 43.111 -14.222]
[43.111 73.556 29.889]
[-14.222 29.889 60.222]

SVD applied on Covariance Matrix of X:

U:
[-0.465 0.568 0.68]
[-0.814 0.028 -0.581]
[-0.349 -0.823 0.449]

D:
[111. 69.   0.]

V transposed:
[-0.465 -0.814 -0.349]
[0.568 0.028 -0.823]
[-0.68   0.581 -0.449]

Q3

X:

[0. 0. -3. 0.]
[-1.414 2.309 0. 0.]
[-1.414 -2.309 0. 0.]
[0. 2.309 0. 0.]
[0. 0. 0. 1.]

Cov(X):

[0.48 0.261 -0.339 0.113]
[0.261 2.987 0.277 -0.092]
[-0.339 0.277 1.44 0.12]
[0.113 -0.092 0.12 0.16]

SVD applied on Cov(X):

U:

[-0.079 0.324 -0.826 -0.455]
[-0.985 0.116 0.092 0.087]
[-0.151 -0.937 -0.241 -0.205]
[0.022 -0.063 -0.502 0.862]

D:

[3.052 1.531 0.421 0.063]

V_transpose:

[-0.079 -0.985 -0.151 0.022]
[0.324 0.116 -0.937 -0.063]
[-0.826 0.092 -0.241 -0.502]
[-0.455 0.087 -0.205 0.862]

Q4.1

X:

[1 0 0 0 2]
[0 0 3 0 0]
[0 0 0 0 0]
[0 4 0 0 0]

Covariance Matrix for X:

[0.188 -0.25 -0.188 0.   0.375]
[-0.25   3.   -0.75   0.   -0.5]
[-0.188 -0.75   1.688 0.   -0.375]
[0.   0.   0.   0.   0.   ]
[0.375 -0.5   -0.375 0.   0.75]

SVD applied on Covariance Matrix:

U:

[-0.065 -0.25 -0.365 -0.894 0.   ]
[0.918 0.23 -0.323 0.   0.   ]
[-0.368 0.796 -0.48   0.   0.   ]
[0.   0.   0.   0.   1.   ]
[-0.131 -0.5   -0.73   0.447 0.   ]

D:

[3.39 1.765 0.47 0.   0.   ]

V_transpose:

[-0.065 0.918 -0.368 0.   -0.131]
[-0.25   0.23   0.796 0.   -0.5]
[-0.365 -0.323 -0.48   0.   -0.73]
[0.894 0.   -0.   0.   -0.447]
[0.   0.   0.   1.   0.   ]

Q4.2

X:

[1 0 0 0 2]
[0 0 3 0 0]
[0 0 0 0 0]
[0 4 0 0 0]

Covariance Matrix for X transposed:

[0.64 -0.36 0.   -0.48]
[-0.36 1.44 0.   -0.48]
[0.   0.   0.   0.]
[-0.48 -0.48 0.   2.56]

SVD applied on Covariance Matrix of X transposed:

U:

[0.162 -0.457 -0.875 0.   ]
[0.292 0.869 -0.4   0.   ]
[0.   0.   0.   -1.   ]
[-0.943 0.191 -0.274 0.   ]

D:

[2.791 1.524 0.325 0.   ]

V_transpose:

[0.162 0.292 0.   -0.943]
[-0.457 0.869 0.   0.191]
[-0.875 -0.4   0.   -0.274]
[0.   0.   -1.   0.   ]