N obs. of M variables   Matrix

$$\underline{x}_1 = \begin{bmatrix} x_1[1] \\ x_1[2] \\ \vdots \\ x_1[N] \end{bmatrix} \quad \vdots \quad \underline{X} = \begin{bmatrix} x_1[1] & x_2[1] & \cdots & x_M[1] \\ x_1[2] & x_2[2] & & x_M[2] \\ \vdots & \vdots & & \vdots \\ x_1[N] & x_2[N] & & x_M[N] \end{bmatrix}$$

✓

Dependencies

1. Rank
2. Null space ✓
3. Singular values

$$\rightarrow \begin{bmatrix} x_1[1] & x_1[2] & \cdots & x_1[N] \\ x_2[1] & x_2[2] & \cdots & x_2[N] \\ \vdots & & & \\ x_M[1] & x_M[2] & \cdots & x_M[N] \end{bmatrix}$$

Matrices

Matrices

A **matrix** of size $N \times M$ is a (column wise) stacked arrangement of M N-dimensional vectors or a (row wise) stacked arrangement of N M-dimensional vectors.

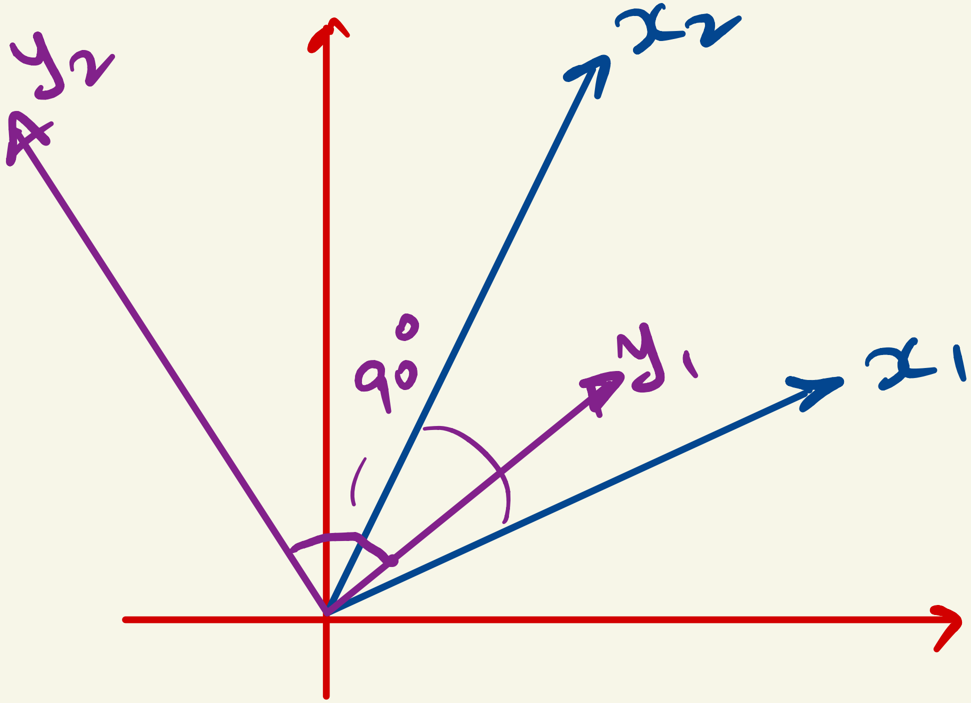
$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_M \end{bmatrix}, \ \mathbf{x}_i \in \mathbb{R}^N \qquad \text{OR} \qquad \mathbf{X} = \begin{bmatrix} \tilde{\mathbf{x}}_1^T \\ \tilde{\mathbf{x}}_2^T \\ \vdots \\ \tilde{\mathbf{x}}_N^T \end{bmatrix}, \ \tilde{\mathbf{x}}_j \in \mathbb{R}^M \tag{14}$$

They are useful in a wide variety of applications:

1. Transforming (rotating) a vector from one system of representation (basis) to another
2. Solving a system of linear equations
3. Analyzing a set of (vector) observations of different variables simultaneously

Matrices are always to be interpreted in the context, i.e., the purpose for which they have been constructed.

This also implies that their properties are to be analyzed contextually.



Scale vectors

$\alpha_1 \underline{x}_1, \quad \alpha_2 \underline{x}_2$

$\underline{A} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \quad X = \underset{N\times 2}{[\underline{x}_1 \quad \underline{x}_2]}$

$\underset{N\times 2}{\tilde{X}} = \underset{N\times 2}{X}\,\underset{2\times 2}{A} = [\alpha_1 x_1 \quad \alpha_2 \underline{x}_2]$

Rotate vectors

XA

A: full matrix

$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}; \quad Y = [\underbrace{\underline{x}_1 a_{11} + \underline{x}_2 a_{21}}_{\underline{y}_1} \quad \underbrace{\underline{x}_1 a_{12} + \underline{x}_2 a_{22}}_{\underline{y}_2}]$

Popular application in data analysis

100 x 5 $\longrightarrow$ 100 x 2

X   Y = XA   Y

5 x 2

Suppose we wish to test for linear independence among a set of vectors $\mathbf{v}_i, l = 1, \cdots, M$. This is also equivalent to

- Determining if a low-dimensional representation is possible (compression)
- Finding relations between variables / features (modelling)

Examples:

1. Examine if $\mathbf{a}_1 = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}^T, \mathbf{a}_2 = \begin{bmatrix} 4 & 7 & -9 \end{bmatrix}^T, \mathbf{a}_3 = \begin{bmatrix} -2 & -11 & 15 \end{bmatrix}^T$ are linearly independent. Matrix $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ should be of *full rank* or non-singular.
2. Linear regression: $y[k] = \beta_1 x_1[k] + \beta_2 x_2[k] + \beta_3 x_3[k] + e[k]$. Minimize $||\mathbf{y}_N - \mathbf{X}_N \mathbf{c}||_2^2$, where (vector) $\mathbf{y}_N$ and (matrix) $\mathbf{X}_N$ are constructed from data and $\mathbf{c}$ is the parameter vector.

Linear transformations

Linear transformations or rotations of vectors are very important mathematical operations in several applications.

The linear transformation (map) of a (finite-dimensional) vector $\mathbf{x} \in \mathcal{V}$ (on $\mathbb{R}^m$) to $\mathbf{y} \in \mathcal{W}$ (on $\mathbb{R}^n$) is achieved by pre-multiplying with a real-valued matrix $\mathbf{A}_{n\times m}$

$$f : \mathcal{V} \longrightarrow \mathcal{W}, \quad f(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{y} \tag{15}$$

wherein the matrix $\mathbf{A}$ determines the nature of transformation.

Examples: For vectors in 2-D space,

(i) **Rotation:** $\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ (ii) **Scaling:** $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$ (iii) **Projection:** $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

(iv) **Linear regression:** $\mathbf{A} = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$

Matrix norms

$A_{N \times M} \longrightarrow \text{vectorize}(A)$

Matrix norms are not defined in the usual vector sense, but in an **induced** sense.

1-norm

The $\boxed{\text{induced 1-norm}}$ of a matrix $\mathbf{A} \in \mathbb{C}^{N \times M}$ is the *maximum of column sums*.

$$||\mathbf{A}||_1 = \max_{j \in \mathbb{Z}_M} \sum_{i \in \mathbb{Z}_N} |a_{ij}| \tag{16}$$

2-norm

The induced 2-norm of a matrix $\mathbf{A} \in \mathbb{C}^{N \times M}$ is the $\boxed{\textit{largest singular value of } \mathbf{A}:}$

$$||\mathbf{A}||_2 = \max_{j=1,2,\cdots,M} \sqrt{\lambda_j(\mathbf{A}^H\mathbf{A})} \tag{17}$$

$$A_{2\times 3} = \begin{bmatrix} 1 & -2 & 3 \\ -7.1 & 4 & 0 \end{bmatrix} : \underbrace{\|A\|_1}_{\text{induced 1-norm}} = 8.1$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 7.1 & 4 & 0 \end{bmatrix} \downarrow \begin{bmatrix} 8.1 & 6 & 3 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & -7.1 \\ -2 & 4 \\ 3 & 0 \end{bmatrix}$$

$$\text{vec}(A) = \begin{bmatrix} 1 \\ -7.1 \\ -2 \\ 4 \\ 3 \\ 0 \end{bmatrix}$$

Ax - 3×1

$$\|A\|_\infty : \begin{bmatrix} 6 \\ 11.1 \end{bmatrix} = 11.1$$

$$\|A\|_1 = \|\text{vec}(A)\|_1$$

$$\|A\|_2 = \|\text{vec}(A)\|_2 \longrightarrow \text{Frobenius norm of } A$$

∞-norm

The induced ∞-norm of a matrix $\mathbf{A} \in \mathbb{C}^{N \times M}$ is the *maximum row sum* of $\mathbf{A}$:

$$||\mathbf{A}||_\infty = \max_{i \in \mathbb{Z}_N} \sum_{j \in \mathbb{Z}_M} |a_{ij}| \tag{18}$$

The property of a matrix, in general, is measured by the transformation it induces in a vector $\mathbf{x}$.

Definition

The induced $||.||_p$ norm of a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ is given by

$$\boxed{||\mathbf{A}||_p = \sup \frac{||\mathbf{Ax}||_p}{||\mathbf{x}||_p} = \sup_{\mathbf{x}, ||\mathbf{x}||_p = 1} ||\mathbf{Ax}||_p} \tag{19}$$

Frobenius norm

A^T

The Frobenius norm is the matrix equivalent of the 2-norm in the vector world.

Frobenius norm

The Perron-Frobenius norm is defined as

trace

$A^T A$

$$||\mathbf{A}||_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{A}^H \mathbf{A})} \tag{20}$$

Examples:

1. Suppose a measurement $\mathbf{y}_N = \mathbf{x}_N + \mathbf{v}_N$, where $\mathbf{v}_N$ is the noise s.t. $||\mathbf{v}_N||_2 = 1$. Then the linearly transformed measurement $\mathbf{Ay}$, $\mathbf{A} \in \mathbb{R}^{N \times N}$ has the 2-norm of $||\mathbf{A}||_2$.
2. Suppose the noise $\mathbf{v}_N$ is such that it affects only one observation. Then the *average* (across all possibilities) of the squared scaled error norm is $\mathbf{Av}$ is $(1/N) \times ||\mathbf{A}||_F^2$.

Some Matrix Facts

Trace (A) = sum of diagonal elements

1. A real-valued matrix satsifying $\mathbf{A} = \mathbf{A}^T$ is said to be **symmetric.** (Hermitian for complex-valued matrices).
2. A matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$ is said to be **orthogonal** (**orthonormal**) if $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$.
3. A matrix $\mathbf{B} \in \mathbb{C}^{M \times M}$ is said to be **unitary** if $\mathbf{B}^H\mathbf{B} = \mathbf{B}\mathbf{B}^H = \mathbf{I}$.
4. For any matrix $\mathbf{A}$, the eigenvalues of matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are always real and non-negative.
5. For any matrix $\mathbf{A}$, the matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ can always be diagonalized by unitary matrices.

Eigenvalues of a matrix : A (square)

Given $A \underline{v}$? $= \lambda \underline{v}$?

eigenvector

eigenvalue

Linear transform of $\underline{v}$

Scaling of $\underline{v}$

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} \underbrace{a_{11} v_{11} + a_{21} v_{21}} \\ \underbrace{a_{12} v_{11} + a_{22} v_{21}} \end{bmatrix}$$

$A : N \times N$

'N' eig. values

eig. vectors

$\lambda_1, \lambda_2, \ldots \lambda_N$

$\underline{v}_1, v_2, \ldots v_N$

$(\lambda_1, \underline{v}_1) \cdot (\lambda_2, \underline{v}_2)$

$\det(A) = $ product of eig.(A)

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

$$\begin{bmatrix} 1 & 2 \\ 5 & -6 \end{bmatrix}$$

$\det(A)$
$= 1 \times -6$
$- 5 \times 2$
$= \underline{-16}$

If A is rank deficient then necessarily at least one eigenvalue is zero!

$$A = [\underline{a}_1 \quad \underline{a}_2 \quad \cdots \quad \underline{a}_M]$$

$$\alpha_1 \underline{a}_1 + \alpha_2 \underline{a}_2 + \cdots + \alpha_M \underline{a}_M = 0$$

$$A \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix}}_{\underline{x}} = 0 \, \underline{x}$$

Eigenvalues, Eigenvectors and Singular Value Decomposition

Eigenvalues and Eigenvectors

For a given *dimension-preserving* linear transformation, there exist (almost always) a special set of vectors that at most change by a scalar (stretch or shrink).

These special set of vectors are known as the **eigenvectors**, while the factors by which they stretch or shrink are termed as the **eigenvalues**.

Given a dimension-preserving linear transformation of a (finite-dimensional) vector $\mathbf{x} \in \mathbb{R}^n$ through a real-valued matrix $\mathbf{A}_{n\times n}$, the eigenvalues λ and eigenvectors $\mathbf{v}$ satisfy

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{21}$$

The non-trivial solution to (21) is determined by first solving (21) for λ

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{22}$$

which is also known as the **characteristic equation** of $\mathbf{A}$.

Subsequently, for each eigenvalue λ_i, $i = 1, \cdots, n$, the eigenvectors are found by solving (21). The collection of all eigenvectors is usually denoted by $\mathbf{V}$ and (21) is written as

$$\mathbf{AV} = \mathbf{V}\Lambda \tag{23}$$

Subsequently, for each eigenvalue $\lambda_i,\ i = 1, \cdots, n$, the eigenvectors are found by solving (21). The collection of all eigenvectors is usually denoted by $\mathbf{V}$ and (21) is written as

$$\mathbf{AV} = \mathbf{V}\Lambda \tag{23}$$

1. For every matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, there exist n eigenvalues.
2. If the $\lambda_i,\ i = 1, \cdots, n$s are distinct, the eigenvectors form a basis for $\mathbb{R}^n$.
3. A square matrix is diagonalizable if and only if the eigenvectors form a basis.
4. Symmetric matrices have real-valued λs, orthonormal eigenvectors and are diagnolizable.
5. The trace of a matrix is tr$(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$ and $\det(\mathbf{A}) = \Pi_{i=1}^{n}\lambda_i$
6. Rank of a square matrix is the number of non-zero eigenvalues

$$\underset{3\times 3}{X} = \begin{array}{c} \begin{array}{ccc} -1 & 2x & \end{array} \\ \begin{bmatrix} 1 & 5 & 9 \\ -3 & 4 & 11 \\ 9 & 11 & 13 \end{bmatrix} \end{array} ;$$

Max. possible rank$(X) = 3$

Rank$(X) = 2$

Rank deficiency $= 1$

Three eig. values $\lambda_1, \lambda_2, \lambda_3$

$\lambda_1 = 23.1067$; $\lambda_2 = -5.1067$; $\lambda_3 = 0$

$$X \underline{v}_3 = \lambda_3 \underline{v}_3 = 0 ;$$

Constraint $\underline{v}_3 = \begin{bmatrix} 0.4082 \\ -0.8164 \\ 0.4082 \end{bmatrix}$;

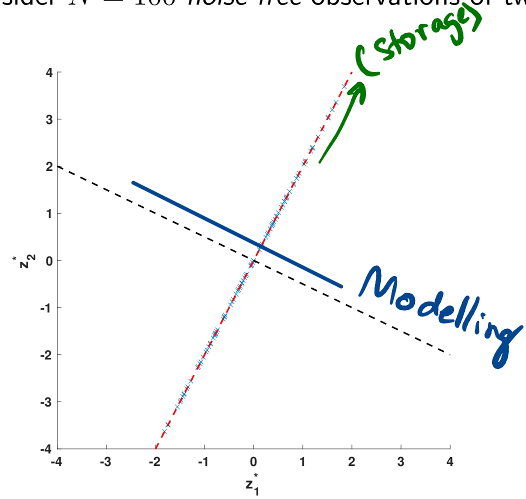
$$[\underline{x}_1 \;\; \underline{x}_2 \;\; \underline{x}_3] \begin{bmatrix} 0.4082 \\ -0.8164 \\ 0.4082 \end{bmatrix} = 0 \implies \boxed{\underline{x}_1 - 2\underline{x}_2 + \underline{x}_3 = 0}$$

Model:

$$\boxed{\begin{array}{l} \underline{x}_3 = 2\underline{x}_2 - \underline{x}_1 \\ \text{OR} \\ \underline{x}_2 = \frac{1}{2}(\underline{x}_1 + \underline{x}_2) \end{array}}$$

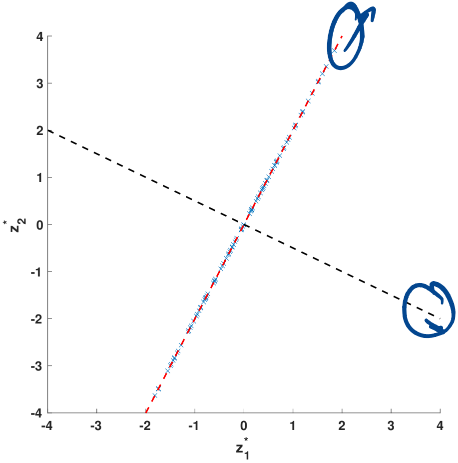
SVD: A motivational example

Consider $N = 100$ *noise-free* observations of two variables $z_1^\star$ and $z_2^\star$.



SVD: A motivational example

Consider $N = 100$ *noise-free* observations of two variables $z_1^\star$ and $z_2^\star$.

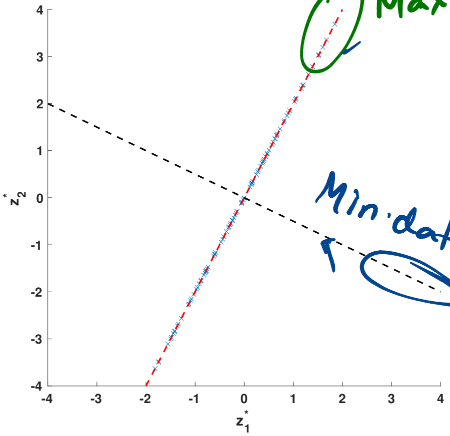


Q: *Can we express the data in a lower dimensional space?*

1. Detect and discover the direction of **maximal data spread**

SVD: A motivational example

Consider $N = 100$ *noise-free* observations of two variables $z_1^\star$ and $z_2^\star$.



Q: *Can we express the data in a lower dimensional space?*

1. Detect and discover the direction of **maximal data spread**
2. **Complementary** approach: find directions of **no (or minimal) data spread.**
3. This amounts to **identifying the best fit line** (constraint tying $z_1^\star$ and $z_2^\star$).

First set up

$$a_{11} z_1^* + a_{12} z_2^* = 0 \; ?$$

100×2

$$\boxed{Z^* \underline{a} = 0}$$

$$\mathbf{Z}^\star = \begin{bmatrix} \mathbf{z}_1^\star & \mathbf{z}_2^\star \end{bmatrix} = \begin{array}{cc} z_1^* & z_2^* \\ \begin{bmatrix} z_1^\star[0] & z_2^\star[0] \\ z_1^\star[1] & z_2^\star[1] \\ \cdots & \cdots \\ z_1^\star[99] & z_2^\star[99] \end{bmatrix} \end{array} \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \tag{24}$$

$$\underline{a} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}$$

1. **Approach I:** Detect redundancies (dependencies) in the data. If yes, re-express data as linear independent basis vectors and work with minimal dimensions.
   - Dependencies (redundancies) can be **linear** / non-linear. We shall confine to the first type.
   - Linear dependence is best detected by computing **rank** (through SVD or eigenvalue analysis)
2. **Approach II:** Detect directions of maximal variations and project data along those directions (will NOT be pursued here).

For any X: (linear dependencies)

if there exists $\underline{a}$ such that $X\underline{a} = 0$?

$\equiv$ Finding null space of X

Null space and SVD

Identifying linear constraints on variables amounts to finding the $\texttt{null}(\mathbf{Z}^\star)$. A numerically robust solution (to finding the null space) is provided by the *singular value decomposition* (SVD)

SVD

Given any matrix $\mathbf{X} \in \mathbb{R}^{N\times M}$, it can be written as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{25}$$

where $\mathbf{U}$ and $\mathbf{V}$ are said to be the **left** and **right** singular vectors, respectively, and $\mathbf{S}$ is the matrix of singular values.

Null space and SVD

Suppose X was a square matrix

$Xv = \lambda v$ OR $XV = V\Lambda$ $\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{bmatrix}$

Identifying linear constraints on variables amounts to finding the $\texttt{null}(\mathbf{Z}^\star)$. A numerically robust solution (to finding the null space) is provided by the *singular value decomposition* (SVD)

$\boxed{X = V \Lambda V^T}$ $V^T V = I$

SVD

Given any matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, it can be written as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{25}$$

where $\mathbf{U}$ and $\mathbf{V}$ are said to be the **left** and **right** singular vectors, respectively, and $\mathbf{S}$ is the matrix of singular values.

The right singular vectors corresponding to **zero singular values** of a matrix $\mathbf{X}$ provide a **basis** for the null space of $\mathbf{X}$.

$V^T V = I$ (true of SVD) $U^T U = I$

rect.

$$X = USV^T$$

$$\Rightarrow \boxed{XV = US}$$

$$U^T X = SV^T$$

Both scaling & rotation

$$X = \sum_{i=1}^{M} \sigma_i \underline{u}_i v_i^T$$

Square

$$X\underline{V} = V\Lambda$$

$$X\underline{v} = \lambda \underline{v}$$

Only scaling
No rotation

$$X = USV^T \Rightarrow \underset{N\times M}{X} = \sum_{i=1}^{M} \sigma_i \, \underset{N\times 1}{\underline{u}_i} \, \underset{M\times 1}{v_i^T}$$

scalar

$$V^T V = I$$

$$\Rightarrow \underline{v}_i^T \underline{v}_j = 0, \; i \neq j$$
$$= 1, \; i = j$$

$M < N$

decomposition

rank-1

$N\times M$

$$= \sigma_1 \underbrace{\underline{u}_1 \underline{v}_1^T}_{N\times M} + \sigma_2 \underline{u}_2 \underline{v}_2^T + \cdots + \sigma_M \underbrace{\underline{u}_M \underline{v}_M^T}_{N\times M}$$

rank-1

$$X \underline{v}_1 = \boxed{\sigma_1 \underline{u}_1 \underbrace{\underline{v}_1^T \underline{v}_1}_{=1}} + \cancel{\sigma_2 \underline{u}_2 \underline{v}_2^T \underline{v}_1} \to 0 + \cdots + \cancel{\sigma_M \underline{u}_M \underline{v}_M^T \underline{v}_1} \to 0$$

$$\boxed{X \underline{v}_1 = \sigma_1 \underline{u}_1}$$

If X was a sq.

$$X \underline{v}_1 = \lambda_1 v_1$$

v_1 is the direction of max. spread.

SVD: Properties

$X : N \times M$

V_1 V_2

$S_1 = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_M \end{bmatrix}$

1. $\mathbf{U} \in \mathbb{R}^{N \times N}$ and $\mathbf{V} \in \mathbb{R}^{M \times M}$ are such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.
2. The matrix $\mathbf{S} \in \mathbb{R}^{N \times M}$ has a special structure

$$\boxed{\mathbf{S} : \begin{bmatrix} \underbrace{\mathbf{S}_1}_{M \times M} & \mathbf{0}_{M \times (N-M)} \\ \mathbf{0}_{(N-M) \times M} & \mathbf{0}_{(N-M) \times (N-M)} \end{bmatrix}} \qquad (M - \text{diagonal}) \qquad (26)$$

where $\mathbf{S}_1 = \text{diag}(\sigma_1, \cdots, \sigma_M)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M$ is a diagonal matrix of *singular values* of $\mathbf{X}$.

3. $\mathbf{X} = \sum_{i=1}^{M} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ (additive decomposition of $\mathbf{X}$ into M rank-one matrices).