SINGULAR VALUE DECOMPOSITION (SVD)

THE SVD PROVIDES A NUMERICALLY STABLE MATRIX DECOMPOSITION AND IS GUARANTEE TO EXIST. SVD CAN BE USED FOR:

- OBTAINING LOW RANK APPROXIMATIONS OF MATRICES
- PERFORMING PSEUDO-INVERSES OF NON-SQUARE MATRICES
- DECOMPOSE HIGH DIMENSIONAL DATA INTO ITS MOST STATISTICALLY DESCRIPTIVE FACTORS.

COMPLEX SYSTEMS GENERATE A LOT OF DATA THAT CAN BE ARRANGED IN LARGE MATRICES. OTHERWISE, TYPICALLY, THESE MATRICES ARE LOW RANK, MEANING THAT JUST A SUBSPACE OF THEM IS REALLY IMPORTANT AND THERE ARE FEW DOMINANT PATTERNS THAT EXPLAIN THE HIGH DIMENSIONAL DATA. SVD IS A NUMERICALLY STABLE AND EFFICIENT METHOD TO EXTRACT THESE PATTERNS. WE CAN CALL THESE DOMINANT PATTERNS A MANIFOLDS OR LOW DIMENSIONAL ATTRACTOR

SVD ALLOWS TO DISCOVER THE LOW-DIMENSIONAL REPRESENTATION OF THE DATA IN A PURELY DATA-DRIVEN WAY WITHOUT THE ADDITION OF EXPERT KNOWLEDGE OR INTUITION.

DIFFERENTLY FROM THE EIGENDECOMPOSITION THE SVD IS GUARANTEE TO EXIST.

LET CONSIDER A LARGE DATA SET $\underline{\underline{X}} \in \mathbb{C}^{n \times m}$

$$\underline{\underline{X}} = \begin{bmatrix} | & | & & | \\ \underline{x}_1 & \underline{x}_2 & \dots & \underline{x}_m \\ | & | & & | \end{bmatrix}$$

WHERE THE COLUMNS $\underline{x}_k \in \mathbb{C}^n$ MAY BE MEASUREMENTS FROM SOME SIMULATIONS. IN THE CASE OF TIME-SERIES DATA $\underline{x}_k = \underline{x}(k\Delta t)$. THE COLUMNS ARE OFTEN CALLED SNAPSHOTS, AND m IS THE NUMBER OF THEM.

BEFORE INTRODUCING THE SVD LET'S GO BACK A LITTLE BIT. WHEN A VECTOR IS MULTIPLIED BY A MATRIX $\underline{\underline{A}}$ IT PRODUCES A NEW VECTOR THAT IS NOW ALIGNED IN A NEW DIRECTION WITH A NEW LENGTH.

THE ROTATION AND STRETCHING OF A TRANSFORMATION CAN BE PRECISELY CONTROLLED BY PROPER CONSTRUCTION OF THE MATRIX $\underline{\underline{A}}$.

FOR EXAMPLE,

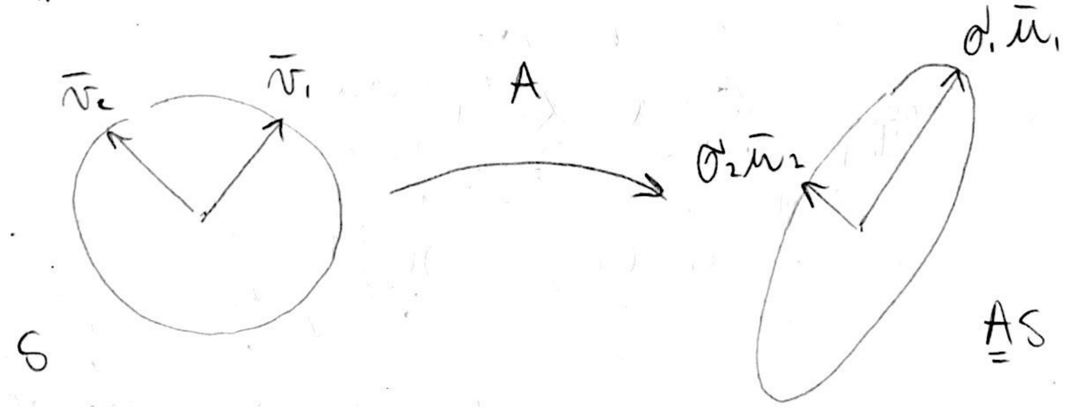
$$\underline{\underline{A}} = \begin{bmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{bmatrix}$$

ROTATATES $\underline{x}$ BY AN ANGLE ϑ. THE TRANSFORMATION PRODUCED BY $\underline{\underline{A}}$ IS KNOWN AS A UNITARY TRANSFORMATION SINCE $A^{-1} = \bar{A}^{T}$. TO SCALE THE VECTOR LENGTH WE APPLY THE MATRIX

$$\underline{\underline{A}} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} = \alpha \underline{\underline{I}}$$

IN THIS WAY WE CAN SEPARATELY CONTROL THE ROTATION AND SCALING IN A TWO DIMENSIONAL VECTOR SPACE. THE SVD IS ESSENTIALLY A TRANSFORMATION THAT STRETCHES/COMPRESSES AND ROTATES A GIVEN SET OF VECTORS.

WHAT HAPPEN TO A HYPER-SPHERE UNDER A MATRIX MULTIPLICATION? IT BECOMES AN HYPER-ELLIPSE.



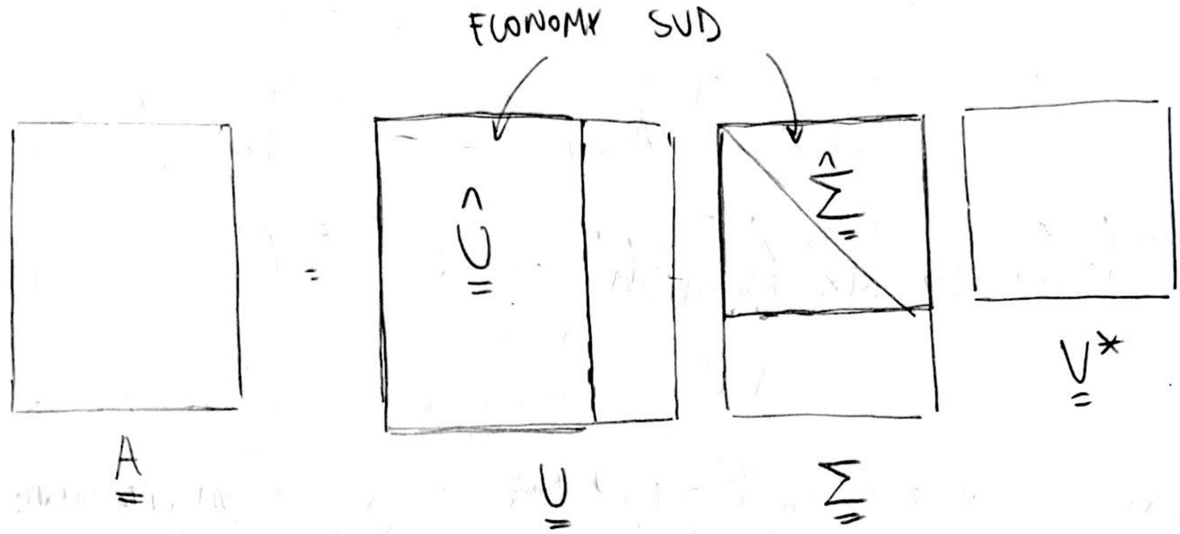
THE UNIT SPHERE IS DESCRIBED BY m UNITH VECTORS $\bar{v}_i$ IN THE $\mathbb{R}^m$ SPACE. IF WE APPLY $\underline{\underline{A}}$ TO THE SPHERE S WE WILL OBTAIN THE HYPER-ELLIPSE $\underline{\underline{A}}S$ WHERE THE VECTORS $\bar{v}_i$ ARE STRETCHED BY σ_i AND ARE ROTATED AS $\bar{u}_i$. THE QUANTITIES $\sigma_j \underline{u}_j$ ARE THE PRINCIPAL SEMI-AXES OF THE HYPER-ELLIPSE.

$$A \bar{v}_i = \sigma_i \bar{u}_i$$

$U \in \mathbb{C}^{m \times m}$ IS UNITARY

$V \in \mathbb{C}^{n \times n}$ IS UNITARY

$\Sigma \in \mathbb{R}^{m \times n}$ IS DIAGONAL



IT IS ASSUMED THAT THE DIAGONAL ENTRIES OF Σ ARE NONNEGATIVE AND ORDERED FROM LARGEST TO SMALLEST

Theorem: EVERY MATRIX $A \in \mathbb{C}^{m \times n}$ HAS A SINGULAR VALUE DECOMPOSITION. FURTHERMORE, THE SINGULAR VALUES σ_j ARE UNIQUELY DETERMINED, AND, IF A IS SQUARE AND THE σ_j DISTINCT, THE SINGULAR VECTORS $\{u_j\}$ AND $\{v_j\}$ ARE UNIQUELY DETERMINED UP TO A COMPLEX SIGNS.

HOW WE COMPUTE THE SVD ?

$$A^T A = (U \Sigma V^*)^T (U \Sigma V^*) = V \Sigma U^* U \Sigma V^* = V \Sigma^2 V^*$$

$$A A^T = (U \Sigma V^*)(U \Sigma V^*)^T = U \Sigma V^* V \Sigma U^* = U \Sigma^2 U^*$$

$$A^T A V = V \Sigma^2 \qquad (\cdot)$$

$$A A^T U = U \Sigma^2$$

Scanned with CamScanner

THE TWO RELATIONS IN (·) ARE NOTHING BUT TWO CONSISTENT EIGENVALUES PROBLEMS. WE CAN SEE THAT $\underline{A}^T\underline{A}$ AND $\underline{A}\,\underline{A}^T$ SHARE THE SAME EIGENVALUES.

▷ DIAGONALIZATION

CONSIDER THE SYSTEM OF DIFFERENTIAL EQUATIONS:

$$\frac{d}{dt}\underline{y} = \underline{\underline{A}}\,\underline{y}$$

ASSUMING A SOLUTION OF THE FORM

$$\underline{y} = \underline{x}\exp(\lambda t)$$

WE OBTAIN AN EGV PROBLEM

$$\underline{\underline{A}}\,\underline{x} = \lambda\underline{x}$$

$$\underline{\underline{A}}\underline{x} - \lambda\underline{\underline{I}}\,\underline{x} = (\underline{\underline{A}} - \lambda\underline{\underline{I}})\underline{x} = \underline{0}$$

1. FIRST OPTION: THE DETERMINANT OF $(\underline{\underline{A}} - \lambda\underline{\underline{I}})$ IS NOT ZERO. THE MATRIX IS NONSINGULAR AND ITS INVERSE $(\underline{\underline{A}} - \lambda\underline{\underline{I}})^{-1}$ CAN BE FOUND. THE SOLUTION IN THIS CASE IS

$$\underline{x} = (\underline{\underline{A}} - \lambda\underline{\underline{I}})^{-1}\underline{0} = \underline{0}$$

THIS IS THE TRIVIAL SOLUTION.

2. SECOND OPTION: THE DETERMINANT IS ZERO, THE MATRIX IS SINGULAR AND ITS INVERSE CANNOT BE FOUND. THERE IS NO GUARANTEE THAT A SOLUTION EXISTS BUT THIS IS THE ONLY CONDITION TO OBTAIN A NON TRIVIAL SOLUTION.

$$\left.\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ &\vdots \\ Ax_n &= \lambda_n x_n \end{aligned}\right\} \quad \underline{\underline{A}}\,\underline{\underline{X}} = \underline{\underline{X}}\,\underline{\underline{\Lambda}}$$

WE CAN COMPARE THE FACTORIZATION OF THE MATRIX $\underline{\underline{A}}$ THROUGH EIGENVALUE DECOMPOSITION AND THROUGH SVD:

$$\text{EVD} \rightarrow \quad \underline{\underline{A}} = \underline{\underline{X}}\,\underline{\underline{\Lambda}}\,\underline{\underline{X}}^{-1} \qquad \text{ONLY ONE BASIS } \underline{\underline{X}}$$

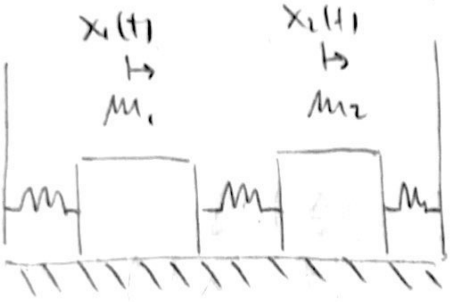
$$\text{SVD} \rightarrow \quad \underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^{*} \qquad \text{TWO BASIS } \underline{\underline{U}},\ \underline{\underline{V}}$$

ANOTHER USEFUL APPLICATION OF EGVD AND EGVTs IS THE POWER OF A MATRIX. COMPUTING $\underline{\underline{A}}^M$ DIRECTLY IS EXPENSIVE BUT:

$$\underline{\underline{A}}^M = (X\Lambda X^{-1})^M = \underset{1}{(X\Lambda X^{-1})}\underset{2}{(X\Lambda X^{-1})}\dots\underset{M}{(X\Lambda X^{-1})}$$

$$= X\Lambda^M X^{-1}$$

THE EXPONENTIAL OF A DIAGONAL MATRIX IS CHEAP TO COMPUTE.



DIAGONALIZATION IS ALSO USEFUL IN REAL SYSTEM ANALYSIS. INSTEAD OF WORKING IN THE ORIGINAL COORDINATE SYSTEM WE CAN DIAGONALIZE AND WORK IN A NEW SPACE WHERE THE TWO DIRECTION OF MOTION ARE INDEPENDENT.

LETS CONSIDER NOW THE DIAGONALIZATION IN THE CONTEXT OF THE SVD. SINCE $\underline{\underline{U}}$ AND $\underline{\underline{V}}$ ARE ORTHONORMAL BASES IN $\mathbb{C}^{m\times m}$ AND $\mathbb{C}^{n\times n}$ RESPECTIVELY, THEN ANY VECTOR IN THESE SPACES CAN BE EXPANDED IN THEIR BASIS. CONSIDER $\underline{b} \in \mathbb{C}^m$ AND $\underline{x} \in \mathbb{C}^n$:

$$\underline{b} = \underline{\underline{U}}\,\hat{\underline{b}}$$

$$\underline{x} = \underline{\underline{V}}\,\hat{\underline{x}}$$

NOW CONSIDER:

$$\underline{\underline{A}}\,\underline{x} = \underline{b} \rightarrow \underline{\underline{U}}^*\underline{b} = \underline{\underline{U}}^*\underline{\underline{A}}\,\underline{x} = \underline{\underline{U}}^*\underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^*\underline{x} \rightarrow \hat{\underline{b}} = \underline{\underline{\Sigma}}\,\hat{\underline{x}}$$

WE SAW THAT MATRICES CAN BE DIAGONALIZED VIA EITHER SVD DECOMPOSITION OR EIGENVALUE DECOMPOSITION BUT, WHAT ARE THE DIFFERENCES?

	SVD	EGVD
1.	2 basis $\underline{\underline{U}}, \underline{\underline{V}}$ ↓ more flexibility since orthonormal	1 basis $\underline{\underline{X}}$ ↓ orthonormality is not guarantee
2.	orthonormal	not always orthonormal ↓ orthonormality just with hermitian matrix
3.	exists always	not alway exist

▷ SVD THEOREMS

Theorem: IF MATRIX $\underline{\underline{A}}$ IS RANK r, THEN THERE EXIST r NON ZERO SVD.

Theorem: THE RANGE of $\underline{\underline{A}}$ is $\text{range}(\underline{\underline{A}}) = \langle \underline{u}_1, \dots, \underline{u}_r \rangle$
THE NULL OF $\underline{\underline{A}}$ IS $\text{null}(\underline{\underline{A}}) = \langle \underline{v}_{r+1}, \underline{v}_{r+2}, \dots, \underline{v}_n \rangle$

Theorem: THE NORM $\|\underline{\underline{A}}\|_2 = \sigma_1$ AND $\|\underline{\underline{A}}\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$ WHERE $\|\cdot\|_F$ STANDS FOR THE FROBENIUS NORM. THEY ESSENTIALLY MEASURE THE ENERGY OF A MATRIX

THE RATIO $\|A\|_2 / \|A\|_F$ MEASURES THE PORTION OF THE ENERGY IN THE SEMI-AXIS $\underline{u}_1$

Theorem: THE NON ZERO SINGULAR VALUES OF $\underline{\underline{A}}$ ARE THE SQUARE ROOTS OF THE NONZERO EIGENVALUES OF $\underline{\underline{A}}^*\underline{\underline{A}}$ OR $\underline{\underline{A}}\underline{\underline{A}}^*$

$\phi_j(x) = (x - x_0)^j$ TAYLOR EXPANSION

$\phi_j(x) = \cos(jx)$ DISCRETE COSINE TRANSFORM

$\phi_j(x) = \sin(jx)$ DISCRETE SINE TRANSFORM

$\phi_j(x) = \exp(jx)$ FOURIER TRANSFORM

$\phi_j(x) = \psi_{a,b}(x)$ WAVELET TRANSFORM

$\phi_j(x) = \phi_{\lambda_j}(x)$ EIGENFUNCTION EXPANSION

$\psi_{a,b}(x)$ IS THE BEST REPRESENTATION TO SHRINK DATA WHILE $\phi_{\lambda_j}(x)$ IS THE MOST INTERPRETABLE ONE. THE WEIGHTING COEFFICIENTS $a_j(t)$ CAN BE OBTAINED EASILY. SINCE THE BASIS FUNCTIONS ARE ORTHONORMAL:

$$\int \phi_j(x)\, \phi_m(x)\, dx = \begin{cases} 1 & j = m \\ 0 & j \neq m \end{cases}$$

SO:

$$a_j(t) = \int f(x,t)\, \phi_j(x)\, dx$$

SINCE ANY COMPLETE BASIS EXPANSION CAN REPRESENT THE FUNCTION $f(x,t)$ TO ANY DESIRED ORDER OF ACCURACY GIVEN N LARGE, WHICH IS THE BEST ONE? WE WANT A BASIS THAT ALLOWS TO USE THE SMALLEST N POSSIBLE WHILE ACHIEVING THE DESIRED LEVEL OF ACCURACY. THESE SPECIAL, ORDERED, ORTHONORMAL FUNCTIONS ARE CALLED THE PROPER ORTHOGONAL MODES (POD) FOR THE FUNCTION $f(x,t)$. THE BASIS FUNCTIONS ARE COMPUTED WITH THE SVD. THE POD MODES ARE THE COLUMNS OF $\underline{U}$ WHICH ARE SCALED BY THE SINGULAR VALUES AND THEIR EVOLUTION IN TIME IS GIVEN BY THE COLUMNS OF $\underline{\underline{V}}$.

Theorem: IF $\underline{\underline{A}} = \underline{\underline{A}}^*$ (self-adjoint) THEN THE SINGULAR VALUES OF $\underline{\underline{A}}$ ARE THE ABSOLUTE VALUES OF THE EIGENVALUES OF $\underline{\underline{A}}$

Theorem: FOR $\underline{\underline{A}} \in \mathbb{C}^{m \times m}$, THE DETERMINANT IS GIVEN BY

$$|\det(A)| = \prod_{j=1}^{m} \sigma_j$$

Theorem: $\underline{\underline{A}}$ IS THE SUM OF r RANK-ONE MATRICES

$$\underline{\underline{A}} = \sum_{j=1}^{r} \sigma_j \underline{u}_j \underline{v}_j^*$$

THIS LAST EXPRESSION REPRESENTS THE MATRIX $\underline{\underline{A}}$ AS A LINEAR SUPERPOSITION OF SVD MODES.

Theorem: FOR ANY N SO THAT $0 \leq N \leq r$ WE CAN DEFINE A PARTIAL SUM

$$\underline{\underline{A}}_N = \sum_{j=1}^{N} \sigma_j \underline{u}_j \underline{v}_j^*$$

AND IF $N = \min\{m, n\}$, DEFINE $\sigma_{N+1} = 0$. THEN

$$\| \underline{\underline{A}} - \underline{\underline{A}}_N \|_2 = \sigma_{N+1}$$

THE SVD GIVES A TYPE OF LEAST-SQUARE FITTING ALGORITHM ALLOWING US TO PROJECT THE MATRIX ONTO LOW DIMENSIONAL REPRESENTATIONS.

▷ MODAL EXPANSIONS

THE PROBLEM HERE IS TO FIND A SUITABLE BASE IN WHICH REPRESENT THE PROBLEM

$$f(x,t) \approx \sum_{j=1}^{N} a_j(t) \phi_j(x)$$

THE MOST COMMONLY USED EXPANSION BASIS ARE.