Recap: 1) Eigendecompositions

2) Classification of fixed pts. of linear systems

3) Linearization, nonlinear stability

Defective e-vals:

$$w_i(t) \propto \underset{\substack{\uparrow \\ \text{polynomial}}}{p_i(t)} e^{\lambda t}$$

Can represent any $\underline{x} = \sum_{i=1}^{N} w_i \underline{v}_i$

$$A\underline{x} = \sum_{i=1}^{N} w_i A\underline{v}_i = \sum_{i=1}^{N} w_i (\lambda \underline{v}_i)$$

$\underline{w}$ can be thought of as eigenvector representation of $\underline{x}$ (change of basis)

$$\underline{x} = V\underline{w}, \quad V = \begin{pmatrix} \underline{v}_1 & \cdots & \underline{v}_N \\ \downarrow & & \downarrow \end{pmatrix}$$

$$A\underline{x} = AV\underline{w} = V\underline{\Lambda}\underline{w} \qquad \underline{\Lambda} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix}$$

Holds for all $\underline{w}$: $AV = V\underline{\Lambda}$,

$$A = V\underline{\Lambda}V^{-1} \quad \text{"diagonalization"}$$

Application: Randomly connected rate network.

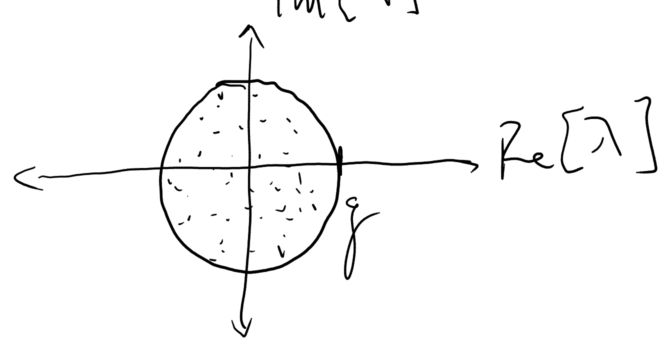
$$\dot{\underline{r}} = (-\underline{I} + W)\,\underline{r} \qquad W_{ij} \sim p(w),$$

$\underline{0}$ is a fixed pt.

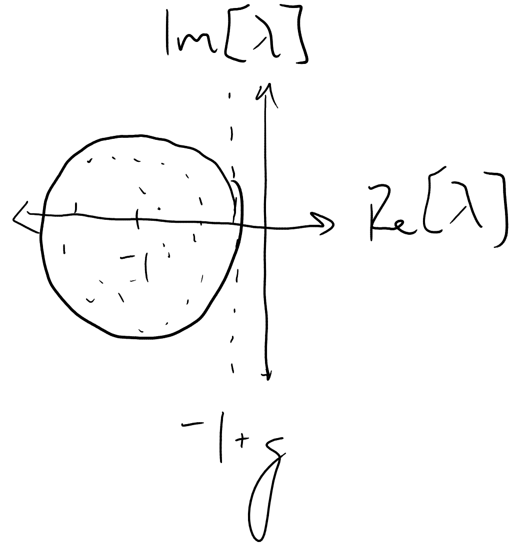
$$\mathrm{Var}(W_{ij}) = g^2/N$$
$$E[W_{ij}] = 0$$

Circular law (Ginibre, Girko):

E-vals of W uniformly distributed in circle of radius g (as $N \to \infty$).

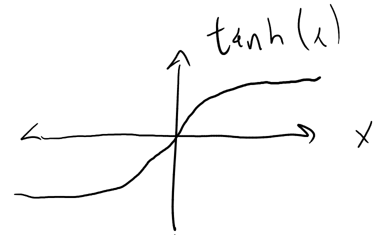


E-vals of $-\underline{I}+W$:



$g<1$: stable

$g>1$: unstable



Nonlinear network:

$$\dot{x}_i = -x_i + \sum_k W_{ik}\tanh(x_k) \equiv f_i(\underline{r})$$

$$\frac{\partial f_i}{\partial x_j} = \begin{cases} -1 + W_{ii}\tanh'(x_k) & i=j \\ W_{ij}\tanh'(x_j) & i\neq j \end{cases}$$

Note $\underline{0}$ is still a fixed pt.

At $\underline{0}$, $\tanh'(x_k)=0$

$\Longrightarrow J(\underline{0}) = -\underline{I}+W$

$g<1$: stable

$g>1$: chaotic (Sompolinsky, Crisanti, Sommers)

Singular value decomposition:

Any matrix $A \in \mathbb{R}^{M \times N}$ can be written as

$A = U \Sigma V^T$, where

Σ is diagonal, $\Sigma_{ii} = \sigma_i$ is singular value

U, V are <u>unitary</u> matrices, $U^T U = V^T V = 1$

Columns of U, V are orthonormal basis.

$$\underset{A}{\overset{N}{M\begin{bmatrix} \quad & \quad \\ & \\ & \end{bmatrix}}} = \underset{U}{\overset{M}{M\begin{bmatrix} \underline{u}_1 & \cdots & \cdots & \underline{u}_M \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix}}} \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{bmatrix} \overset{N}{\begin{bmatrix} \underline{v}_1 \longrightarrow \\ \vdots \longrightarrow \\ \underline{v}_N \longrightarrow \end{bmatrix}} N$$

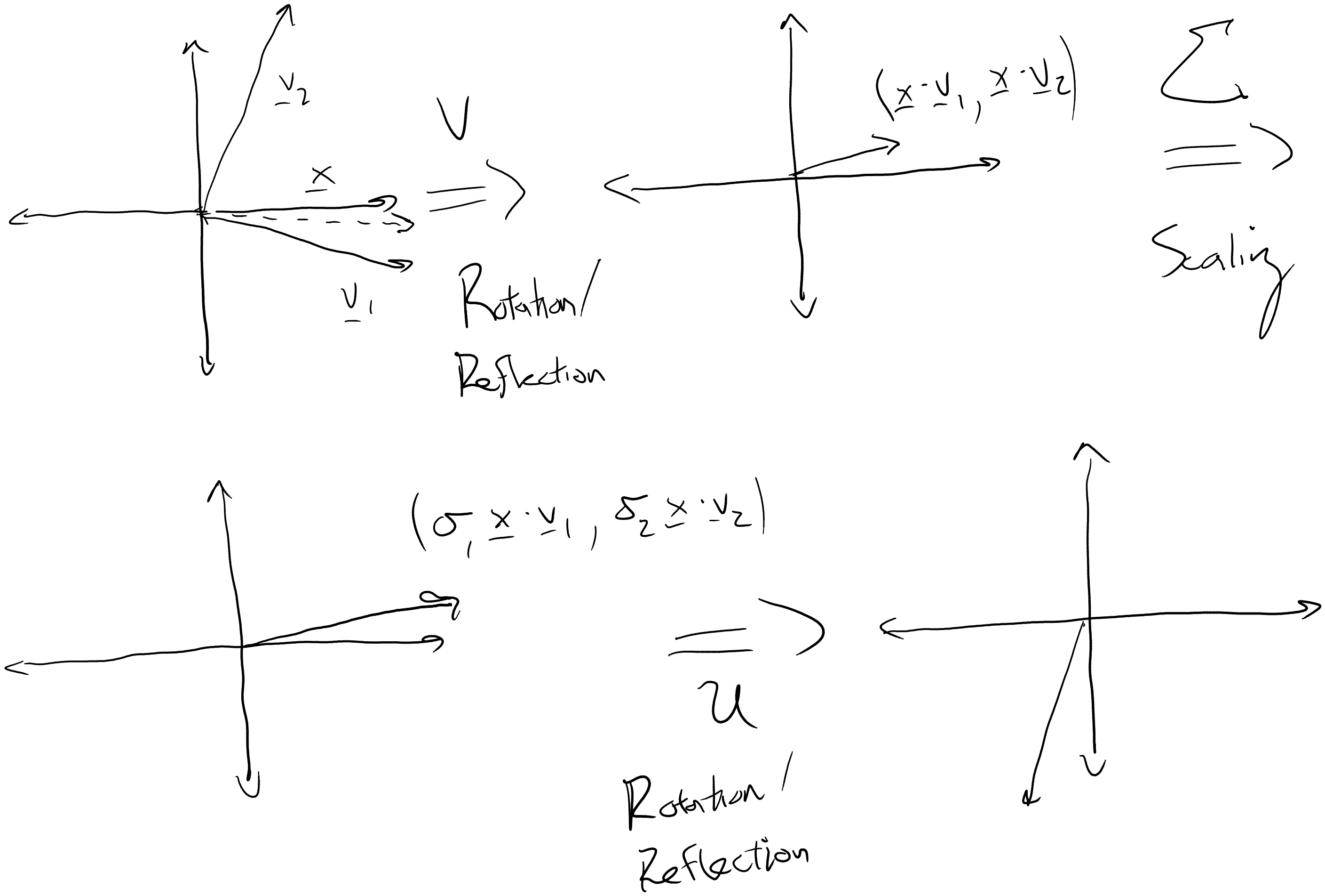
$\underline{v}_i$: right singular vectors

$\underline{u}_i$: left singular vectors

Choose $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_N$

$$A = \sum_i \sigma_i \underline{u}_i \underline{v}_i^T = \sigma_1 \begin{pmatrix} \underline{u}_1 \\ \downarrow \end{pmatrix} \begin{pmatrix} \underline{v}_1 \longrightarrow \end{pmatrix} + \sigma_2 \begin{pmatrix} \underline{u}_2 \\ \downarrow \end{pmatrix} \begin{pmatrix} \underline{v}_2 \longrightarrow \end{pmatrix} + \ldots$$

Interpretation as geometrical map:



Construction of SVD:

1) Find direction $\underline{v}_1$ that maximizes $\|A\underline{v}_1\|$, $\|\underline{v}_1\| = 1$

2) $\sigma_1 = \|A\underline{v}_1\|$

3) $\underline{u}_1 = \frac{1}{\sigma_1} A \underline{v}_1$

4) Repeat, for each $\underline{v}_i$ in orthogonal space.

Exercise: Why are $\underline{u}_i$ orthogonal?

Properties:

1) # of $\sigma_i \neq 0$ is <u>rank</u> of matrix

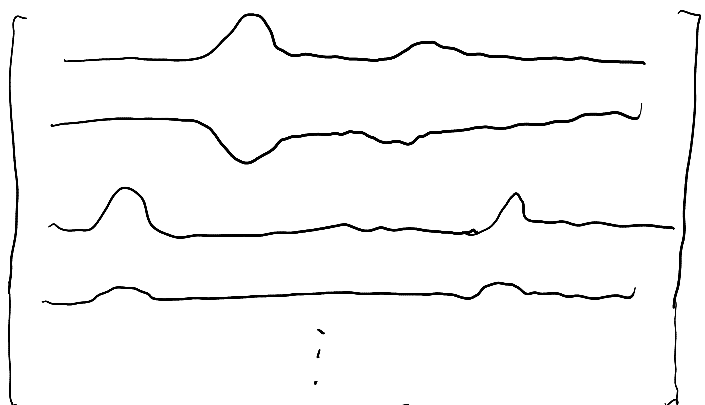
2) $\underline{u}_i$ for which $\sigma_i \neq 0$ span <u>range</u> of A

3) $\underline{v}_i$ for which $\sigma_i = 0$ span <u>null space</u> of A

Interpretation as dimensionality reduction:

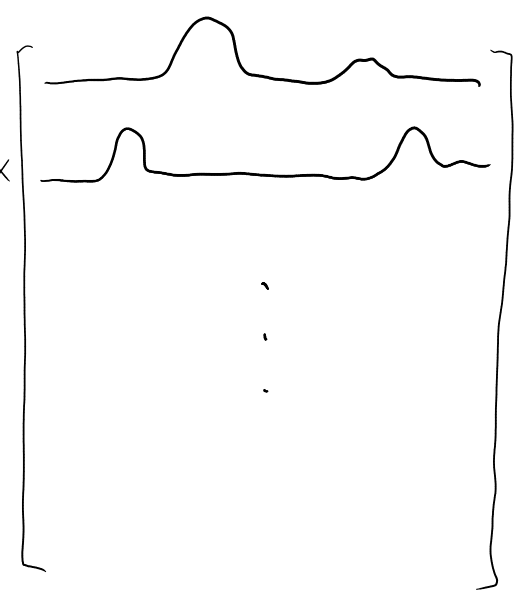
T

$A = N$



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

$$= \begin{bmatrix} 1/\sqrt{2} & 0 & \\ -1/\sqrt{2} & 0 & \cdots \\ 0 & \sqrt{3}/2 & \\ 0 & 1/2 & \end{bmatrix} \times \begin{bmatrix} 1 & & & \\ & 1/2 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} \times$$



N samples of T-dim. space,

$\underline{v}_1$: Direction that explains most variance

Relationship to PCA & eigendecomposition:

$C = A^T A$ "covariance matrix"

Symmetric $\Longrightarrow$ $\lambda_i > 0$, $\underline{v}_i$ orthogonal

$$A^T A = V \Sigma \underbrace{U^T U}_{I} \Sigma V^T = V \Sigma^2 V^T$$

V diagonalizes C.

Right singular vectors = e-vecs of $A^T A$

Left singular vectors = e-vecs of $A A^T$

σ_i^2 = e-vals. of C.

Applications:

1) Low-rank approximation:

$$A \approx \sum_{i=1}^{K} \sigma_i \underline{u}_i \underline{v}_i^T$$ is rank-K approximation of A.

2) Pseudoinverse:

Define $A^\dagger = V \Sigma^\dagger U^T$, where

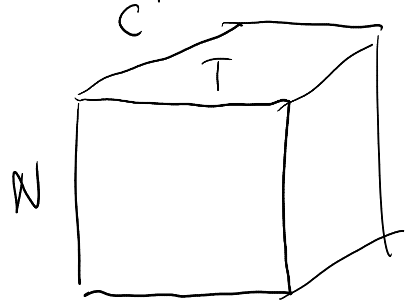
$\Sigma^\dagger_{ii} = 1/\Sigma_{ii}$.

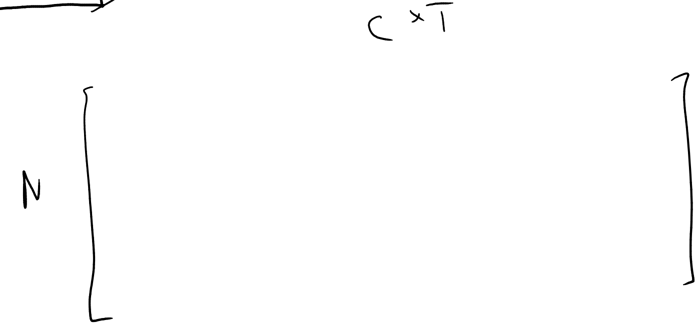
Then $$A^\dagger A = V \Sigma^\dagger \underbrace{U^T U}_{I} \Sigma V^T$$
$$= V \Sigma^\dagger \Sigma V^T$$

If full-rank, $\Sigma^\dagger \Sigma = I$.

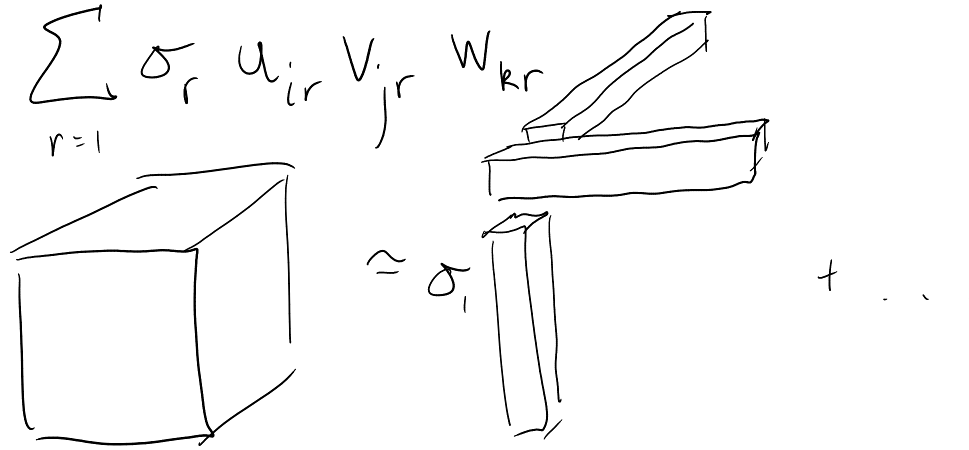
$$A A^\dagger = U \Sigma \Sigma^\dagger U^T$$

Tensor decomposition (e.g. Williams...Ganguli Neuron '18
Seely...Abbott P. Comp. Biol. '17)

A =


Flatten:


$$A_{ijk} \approx \sum_{r=1}^{R} \sigma_r \, u_{ir} \, v_{jr} \, w_{kr}$$



Properties: 1) Non-orthogonal

2) Maximal rank of arbitrary tensor unknown

3) Non-convex

4) Cannot be done iteratively