Last time: Markov chains

Def: i) A vector $v$ is a state vector if $0 \leq v_j \leq 1$ for all $j$ and $\Sigma v_j = 1$

ii) A matrix $P = (P_{ij})_{ij} \in \mathbb{R}^{m \times n}$ is a stochastic matrix if

a) $0 \leq P_{ij} \leq 1$; 

b) Each column of $P$ adds up to 1: $\Sigma_{i=1}^n P_{ij} = 1$ for $j$.

iii) The entries $P_{ij}$ in a stochastic matrix $P$ are called transition probabilities. They indicate the probability to transition from state $j$ to state $i$.

Question: Given a state vector $x_n$ and transition probabilities $P_{ij}$, can we find $x_{n+1}$, i.e., the state of the system at time $n+1$?

Answer: Yes, $x_{n+1} = P \times x_n$

Note: If we know the initial state $x_0$ of the system, then the formula for the system at time $n$ is $x_n = P^n \times x_0$. This can be "analyzed" as an inhomogeneous recursion relation.

Question: What is the large $n$ behaviour of the system? Does there exist a steady state?

Def: Given a Markov chain with a transition matrix $P$, the steady state solution of the Markov chain is the eigenvector $\mathbf{x}$ corresponding to the eigenvalue $\lambda = 1$.

Facts: Let $P$ be a stochastic matrix.

1. If $\mathbf{x}$ is a state vector, then so is $P \mathbf{x}$.
2. $P$ has an eigenvalue $\lambda = 1$.
3. All (other) eigenvalues of $P$ satisfy $|\lambda| < 1$.
4. The eigenvector for $\lambda = 1$ has non-negative entries.
5. If $P (= P^2$ for some $n \in \mathbb{N})$ has all positive entries, then $\lambda = 1$, $\mathbf{x}$ has all positive entries and $|\lambda_j| < 1$ for $j \neq 1$. 

This is crucial!
Last time continued: Google Page rank

Idea: A web surfer is on some initial website of the "world-wide-web". On average, she/he clicks on every link that is available on that site with the same probability. If there is no link on a website, she/he randomly chooses one of all the websites in the "world-wide-web" to go to next.

The resulting (in real-life very big) matrix with all the transition probabilities is a stochastic matrix.

1. Case: \( P \) has all positive entries. Then we can directly use the Power Method to calculate the eigenvector \( \mathbf{v}_1 \) corresponding to the eigenvalue \( \lambda_1 = 1 \) of \( P \).

2. Case: There are zero entries in \( P \). Thus we need to introduce some randomness.
   1. Step: Create a matrix \( Q \) same size as \( P \), such that all entries of \( Q \) are identical and \( Q \) is stochastic.
   2. Step: Pick a "damping factor" \( \alpha \in [0, 1] \).
   3. Step: Set up the "Google matrix": \( G = \alpha \mathbf{P} + (1-\alpha) \mathbf{Q} \).
   Remark: \( G \) is a stochastic matrix by definition and as long as \( 0 < \alpha < 1 \) all entries in \( G \) are positive. \( \Rightarrow \) Power Method is applicable.
   4. Step: Use the Power method to calculate the eigenvector \( \mathbf{v}_1 \) corresponding to \( \lambda_1 = 1 \) of \( G \).

Def: The PageRank of a website is the corresponding entry of the eigenvector \( \mathbf{v}_1 \), normalized and that its entries add up to 1.
Singular value decomposition

Idea: Can we generalize "diagonalization" to arbitrary (non-square, non-square) matrices?

Why is this important?

1. If \( A = SDS^{-1} \) with invertible matrix \( S \) and diagonal matrix \( D \), i.e. diagonalizable, then
   \[
   A^2 = SD^2 S^{-1}, \quad h \geq 0
   \]
   and also for \( h < 0 \) if \( D_{jj} = \lambda_j \neq 0 \) for all \( j \).

2. If \( A \) is unitarily diagonalizable (e.g. \( A \) is Hermitian), then
   \[
   \|A\|_op = \max \{ |\lambda_j| \}
   \]
   where \( \lambda_j \) are the eigenvalues of \( A \).

These are very useful properties so how do we generalize "diagonalization"?

The "formula"

\[
A = U \Sigma V^* \]

Note that "diagonal" for non-square matrices means

Example:

\[
\begin{bmatrix}
1 & 1 \\
1 & 1 \\
0 & 1
\end{bmatrix}
= A
\]

\[
\begin{bmatrix}
1/\sqrt{3} & -1/2 & -1/6 \\
-1/\sqrt{3} & -1/2 & 1/6 \\
1/\sqrt{3} & 0 & 2/\sqrt{6}
\end{bmatrix}
= U
\]

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & \sqrt{2} & 0 \\
0 & 0 & 0
\end{bmatrix}
= \Sigma
\]

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 0
\end{bmatrix}
= V^*
\]
How do we find $U$, $\Sigma$, $V$ given $A$?

Let $A$ be $m \times n$.

**Proposition 1:** All eigenvalues of $A^*A$ are non-negative (and strictly positive if $A$ is invertible).

**Proof:** Let $A^*A v = \lambda v$ for some $v \neq 0$. Then

$$v^* A^* A v = \lambda v^* v = \lambda v^* v$$

$$\implies \quad <Av, Av> = \lambda <v, v>$$

$$\implies \quad \|Av\|^2 = \lambda \|v\|^2 \implies \lambda \geq 0.$$

If $A$ is invertible, then $Av \neq 0$ since $v \neq 0$. Thus we obtain

$$\|Av\|^2 = \lambda \|v\|^2 \implies \lambda > 0.$$

**Proposition 2:** $A^*A$ and $AA^*$ have identical non-zero eigenvalues.

**Proof:** Suppose $\lambda \neq 0$ and $A^*A v = \lambda v$ for some $v \neq 0$. Then

$$A A^* (Av) = A(A^*A)v = \lambda Av$$

Thus $\lambda$, $Av$ are an eigenvalue/eigenvector pair for $AA^*$.

**Proposition 3:** $A^*A$ and $AA^*$ are hermitian and therefore unitarily diagonalizable.

**Proof:** i) $(A^*A)^* = A^*A^* = A^*A$; ii) $(AA^*)^* = A^{**}A^* = AA^*$. Thus they are unitarily diagonalizable and we have

$$A^*A = V \Sigma^2 V^*$$

$$AA^* = U \Sigma^2 U^*$$

Recall Proposition 2!
Recall from last page: \( A^*A = V \Sigma_1^2 V^* \) and \( AA^* = U \Sigma_2^2 U^* \) where the non-zero entries of \( \Sigma_1^2 \) and \( \Sigma_2^2 \) are the same.

General example for \( m < n \):

\[
\Sigma_1^2 = \begin{bmatrix}
\sigma_1^2 & & \\
& \sigma_2^2 & \\
& & \sigma_m^2
\end{bmatrix}_{n \times n}, \quad \Sigma_2^2 = \begin{bmatrix}
\sigma_1^2 & & \\
& \sigma_2^2 & \\
& & \sigma_m^2
\end{bmatrix}_{m \times m}
\]

We define \( \Phi_i = \Phi_j \) and set

\[
\Sigma = \begin{bmatrix}
\sigma_1 & & \\
& \sigma_2 & \\
& & \sigma_m
\end{bmatrix}_{m \times n}
\]

**Claim:** \( A = U \Sigma V^* \)

**Proof:** The general case is out of the scope for this course. However, we can prove the special case \( m = n \) and \( A \) is invertible.

Let \( m = n \) and \( A \) be invertible. Then \( \Sigma_1 = \Sigma_2 = \Sigma \) and \( \sigma_1, \sigma_2, \ldots, \sigma_m \) are all non-zero. Thus \( \Sigma \) is invertible. Moreover, since \( A^*A \) is Hermitian, we can find a unitary matrix \( V \) such that \( A^*A = V \Sigma_1^2 V^* = V \Sigma_2^2 V^* \).

We define \( U := AV \Sigma^{-1} \). Then \( U \) is unitary, since \( U^*U = \Sigma^{-1} V^* A^* A V \Sigma^{-1} = \Sigma^{-1} \Sigma^2 \Sigma^{-1} = I \).

Moreover, \( U \Sigma^2 U^* = AV \Sigma^{-1} \Sigma^2 V^* A^* = AV \Sigma^2 V^* A^* = AA^* \).

Now to the claim:

\[
U \Sigma V^* = AV \Sigma^{-1} \Sigma V^* = AVV^* = A
\]
**Defn.** \( A = U \Sigma V^* \) as above is called the Singular-Value-Decomposition (SVD) of \( A \). The numbers \( \sigma_1, \sigma_2, \ldots, \sigma_m \) are the singular values of \( A \).

**Example:** Let \( A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \). Find the SVD of \( A \).

**Solution:**

\[
A^*A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \Rightarrow \lambda_1 = 2, \quad \lambda_2 = 3
\]

The corresponding eigenvectors are \( v_1 = [0, 0] \), \( v_2 = [1, 1] \) and thus \( V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \).

Moreover we have found:

\[
\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix}
\]

It remains to find \( U \). For that we need to calculate the eigenvalue/eigenvector pairs of \( AA^* \).

\[
AA^* = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & -1 \\ 0 & 1 & 1 \end{bmatrix} \Rightarrow \lambda_1 = 2, \quad u_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}^T
\]

\[
\lambda_2 = 3, \quad u_2 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}^T
\]

\[
\lambda_3 = 0, \quad u_3 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}^T
\]

We obtain

\[
U = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{6}} & 2\frac{\sqrt{6}}{6} \end{bmatrix} = [u_1, u_2, u_3].
\]

Direct computation shows

\[
A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{6}} & 2\frac{\sqrt{6}}{6} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = U \Sigma V^*.
\]
The shapes of SVD: Let \( A \) be \( m \times n \)

**Case I:** \( m \geq n \)

\[
\begin{bmatrix}
A \\
U \\
V^*
\end{bmatrix}
\]

**Case II:** \( m < n \)

\[
\begin{bmatrix}
A \\
U \\
V^*
\end{bmatrix}
\]

**Example:** Let \( A \) be \( 4 \times 4 \).

\[
A = I_4 \begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \frac{1}{2}
\]

Find an orthogonal basis for \( W(A) \).

**Solution:** \( A \) is in SVD form with \( U = I_4 \), \( \Sigma = \begin{bmatrix} 3 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \) and \( V = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \)

Let \( \mathbf{x} \in W(A) \). Then \( \mathbf{Ax} = \mathbf{0} \)

\[
\Rightarrow \mathbf{Ax} = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \langle v_1, x \rangle \\ \langle v_2, x \rangle \\ \langle v_3, x \rangle \\ \langle v_4, x \rangle \end{bmatrix} = \begin{bmatrix} 3 \langle v_1, x \rangle \\ 2 \langle v_2, x \rangle \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\]

Thus \( \mathbf{x} \in W(A) \) if and only if \( \langle v_1, x \rangle = 0 \) and \( \langle v_2, x \rangle = 0 \). Hence \( \mathbf{x} \in \text{span}\{v_3, v_4\} \)

Since \( \{v_1, v_2, v_3, v_4\} \) is an orthonormal basis we have \( \mathbf{x} \in W(A) \iff \mathbf{x} \in \text{span}\{v_3, v_4\} \)

Note that \( v_3, v_4 \) are the columns of \( V^* \) that correspond to the zero singular values of \( A \).
In general, let $A = U \Sigma V^*$ be the SVD of $A$.

$$
A = \begin{bmatrix}
  u_1 & u_2 & \ldots & u_m
\end{bmatrix}
\begin{bmatrix}
  \sigma_1 & 0 & \ldots & 0 \\
  0 & \sigma_2 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & \sigma_r
\end{bmatrix}
\begin{bmatrix}
  v_1^* \\
  v_2^* \\
  \vdots \\
  v_r^*
\end{bmatrix}
\Sigma
\begin{bmatrix}
  v_1^* \\
  v_2^* \\
  \vdots \\
  v_n^*
\end{bmatrix}
$$

where $u_1, \ldots, u_m$ are the columns of unitary $U$, $v_1, \ldots, v_n$ are the columns of unitary $V$, and $\sigma_1, \ldots, \sigma_r$ are the non-zero singular values of $A$ ($r \leq \min\{m, n\}$).

1. Nullspace of $A$

As in the example we obtain:

$$
x \in \mathcal{N}(A) \iff x \perp v_1^*, \ldots, v_r^* \iff x \perp \text{span}\{v_1^*, \ldots, v_r^*\}
$$

Since $\{v_1, \ldots, v_n\}$ is ONB:

$$
x \in \mathcal{N}(A) \iff x \in \text{span}\{v_{r+1}^*, \ldots, v_n^*\}
$$

Example:

\[ A = \begin{bmatrix}
  1 & 1 \\
  1 & -1 \\
  0 & 1
\end{bmatrix} = U \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  0 & 0 \\
  0 & 0 \\
  0 & 0
\end{bmatrix}, \quad \mathcal{N}(A) = \{0\}
\]

\[ A^* = \begin{bmatrix}
  1 & 0 \\
  0 & 1
\end{bmatrix} \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
  \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\
  -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 2/\sqrt{6}
\end{bmatrix}, \quad \mathcal{N}(A^*) = \text{span} \begin{bmatrix}
  -\frac{1}{\sqrt{6}} \\
  \frac{1}{\sqrt{6}} \\
  2/\sqrt{6}
\end{bmatrix}
\]
2. Range of \( A \)

\[
R(A) = \{ Ax : x \in \mathbb{R}^n \} = \{ \sum_{i=1}^{r} \sigma_i u_i^T y_i : y_i \in \mathbb{R}^n \}
\]

\[
= \{ U \Sigma V^* x : x \in \mathbb{R}^n \}
\]

\[
= \{ U \Sigma y : y \in \mathbb{R}^n \}
\]

\[
= \left\{ \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & \sigma_2 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} : y \in \mathbb{R}^n \right\}
\]

\[
= \left\{ \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 y_1 \\ \sigma_2 y_2 \\ \vdots \\ \sigma_r y_r \end{bmatrix} : y \in \mathbb{R}^n \right\}
\]

\[
= \left\{ \sum_{j=1}^{r} \sigma_j y_j u_j : y \in \mathbb{R}^n \right\}
\]

\[
\Rightarrow R(A) = \text{span}\{ u_1, \ldots, u_r \}
\]
Implications from the results on $R(A)$

1. $r = \# \text{ of non-zero singular values of } A$
   - dimension of $R(A)$
   - rank $(A)$

2. Orthogonality relation: $R(A) = N(A^*)^\perp$
   - We know: $R(A) = \text{span} \{ u_1, \ldots, u_r \} \rightarrow N(A^*) = \text{span} \{ u_{r+1}, \ldots, u_m \}$
   - Since $\{v_j\}_{j=1}^m$ is OMB

3. By "rank-nullity" theorem: $\dim(R(A)) = r \Rightarrow \dim(N(A)) = m - r$

Matrix norm & SVD

Recall that for diagonal matrices the operator norm $\|\cdot\|_{op}$ is the absolute value of the (in absolute value) largest eigenvalue.

What about general matrix $A$?

Let $A = U \Sigma V^*$ (SVD of $A$) and suppose the diagonal entries of $\Sigma$ are ordered s.t. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, $\sigma_j = 0$ for $j > r$. (Note: no assumptions on $m, n$)

Then:
1. $\|A\|_{op} = \sigma_1$

   Proof: $\|A\|_{op} = \max_{\|x\|_2 = 1} \|Ax\|_2 = \max_{\|u\|_2 = 1} \|U^*x\|_2 = \max_{\|y\|_2 = 1} \|\Sigma y\|_2 = \sigma_1$

2. "min. stretch": $\min \|Ax\| = \begin{cases} \sigma_1, & \text{if } r = n \\ \sigma_j, & \text{if } j < r \end{cases}$

3. $\|A\|_{HS}^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2$
Other applications of the SVD

1) Pseudo inverse: The pseudo inverse of a matrix $A = U\Sigma V^*$ is

$$A^+ = V\Sigma^+ U^*$$

where $\Sigma^+$ is the pseudo inverse of $\Sigma$, which is formed by replacing all non-zero diagonal entries by their corresponding reciprocal and throwing the resulting matrix.

$pseudoinverses$ can be used to solve linear least square problems.

2) Principal component analysis: see Section IV.8. in the script.

3) Signal processing

4) Pattern recognition

5) Quantum information theory (Schmidt decomposition)

6) Numerical weather prediction

7) Gravitational wave form modeling

and many more ... (see for example the Wikipedia page about Singular value decomposition)