Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
Department of Mathematics

Robert Weismantel

Week 14:From symmetric matrices to the singular value theorem

The point of departure:

The spectral theorem: Let A be a real $n \times n$ symmetric matrix

Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A and $\lambda_1, \ldots, \lambda_n$ the associated eigenvalues. Then $A = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$

Proposition [Rayleigh Quotient]

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. The Rayleigh Quotient, defined for $x \in \mathbb{R}^n \setminus \{0\}$, as

$$\text{For } x \in \mathbb{R}^n \setminus \{0\}, \text{ let } R(x) = \frac{x^\top A x}{x^\top x}.$$

R attains its maximum at $R(v_{\max}) = \lambda_{\max}$ and its minimum at $R(v_{\min}) = \lambda_{\min}$ where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of A and $v_{\max}$, $v_{\min}$ their associated eigenvectors.

Proof.

Since $R(v_{\max}) = \lambda_{\max}$ and $R(v_{\min}) = \lambda_{\min}$ it is enough to show

$$\lambda_{\min} \leq R(x) \leq \lambda_{\max} \text{ for all } x \in \mathbb{R}^n \setminus \{0\}.$$

The proof continued

From the spectral theorem

- $$\text{For } x \in \mathbb{R}^n \setminus \{0\}, \quad R(x) = \frac{x^\top \left(\sum_{i=1}^n \lambda_i v_i v_i^\top\right) x}{\|x\|^2} = \frac{\sum_{i=1}^n \lambda_i \left(x^\top v_i\right)^2}{\|x\|^2},$$

  where $v_1, \ldots, v_n$ form an orthonormal basis of eigenvectors of A and $\lambda_1, \ldots, \lambda_n$ are the associated eigenvalues.
- $$\text{For all } 1 \leq i \leq n \quad \lambda_{\min} \left(x^\top v_i\right)^2 \leq \lambda_i \left(x^\top v_i\right)^2 \leq \lambda_{\max} \left(x^\top v_i\right)^2.$$
- Collecting all these inequalities we get

  $$\lambda_{\min} \frac{\sum_{i=1}^n \left(x^\top v_i\right)^2}{\|x\|^2} \leq \frac{\sum_{i=1}^n \lambda_i \left(x^\top v_i\right)^2}{\|x\|^2} \leq \lambda_{\max} \frac{\sum_{i=1}^n \left(x^\top v_i\right)^2}{\|x\|^2}.$$
- The v_i's are orthonormal, the matrix V with the v_i's as columns is orthogonal and $\sum_{i=1}^n \left(x^\top v_i\right)^2 = \|Vx\|^2 = \|x\|^2$ and so $\frac{\sum_{i=1}^n \left(x^\top v_i\right)^2}{\|x\|^2} = 1$.

Positive definite matrices

Definition (Positive Definite and Positive Semidefinite matrix)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is said to be Positive Semidefinite / Positive Definite (PSD / PD) if all its eigenvalues are non-negative / positive.

Proposition derived from the Rayleigh Quotient

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is PSD if and only if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$.
A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is PD if and only if $x^\top Ax > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Lemma

If $A, B \in \mathbb{R}^{n \times n}$ are symmetric and PSD, then A+B is PSD.

Proof.

If $x^T Ax \geq 0$ and $x^T Bx \geq 0$ for all $x \in \mathbb{R}^n$, then

$$x^T(A+B)x = x^T(Ax+Bx) = x^T Ax + x^T Bx \geq 0.$$

A key-observation: Gram matrices are PSD.

Definition (Gram Matrix)

Given n vectors, $v_1, \dots, v_n$ in $\mathbb{R}^m$, let $V \in \mathbb{R}^{m \times n}$ be the matrix with columns v_i. The Gram Matrix of V is defined to be the $n \times n$ matrix of inner products

$$G_{ij} = v_i^\top v_j.$$

In matrix notation, $G = V^\top V$.

Proposition

Let $A \in \mathbb{R}^{m \times n}$. The non-zero eigenvalues of $A^\top A \in \mathbb{R}^{n \times n}$ are the same as the ones of $AA^\top \in \mathbb{R}^{m \times m}$. Both matrices are also symmetric and PSD.

Proof.

$A^\top A$ and $AA^\top$ are symmetric. We have $x^\top A^\top A x = \|Ax\|^2 \geq 0$ for all x which implies $A^\top A$ is PSD. The same argument applies to $AA^\top$.

Proof continued

It remains to show that the non-zero eigenvalues of $A^\top A \in \mathbb{R}^{n\times n}$ are the same as the ones of $AA^\top \in \mathbb{R}^{m\times m}$.

- Let r be the rank of A. We know

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

- $AA^\top$ and $A^\top A$ have a complete set of real eigenvalues and orthogonal eigenvectors.
- Let $\lambda_1, \ldots, \lambda_r$ be the r non-zero eigenvalues of $A^\top A$ and $v_1 \ldots, v_r$ the corresponding eigenvectors. Let $\mu_1, \ldots, \mu_r$ be the r non-zero eigenvalues of $AA^\top$ and $w_1 \ldots, w_r$ be the corresponding eigenvectors.
- $A^\top A v_k = \lambda_k v_k$. Hence, $AA^\top A v_k = \lambda_k A v_k$ and so λ_k is a nonzero eigenvalue of $AA^\top$ with eigenvector Av_k.
- $(A^\top A)A^\top w_i = A^\top(AA^\top w_i) = \mu_i A^\top w_i$ for all i. This shows that $\mu_1, \ldots, \mu_r$ are non-zero eigenvalues of $A^\top A$ with corresponding eigenvectors $A^\top w_1 \ldots, A^\top w_r$.
- Hence, $\{\mu_1, \ldots, \mu_r\} = \{\lambda_1, \ldots, \lambda_r\}$.

What else do we get for PSD matrices?

Proposition [Cholesky decomposition]

Every symmetric positive semidefinite matrix M is a Gram matrix of an upper triangular matrix C. $M = C^\top C$ is known as the Cholesky Decomposition.

Proof.

- There is a decomposition $M = V\Lambda V^\top$ with Λ a diagonal matrix with the eigenvalues of M in the diagonal. Since M is PSD, $\Lambda_{ii} \geq 0$.
- Define $\Lambda^{1/2}$ by taking the square root of each diagonal entry of Λ. Then $M = \left(V\Lambda^{1/2}\right)\left(V\Lambda^{1/2}\right)^\top$.
- To make the matrices upper triangular use the QR decomposition: $\left(V\Lambda^{1/2}\right)^\top = QR$ with Q such that $Q^\top Q = I$ and R upper triangular.
- $$M = \left(V\Lambda^{1/2}\right)\left(V\Lambda^{1/2}\right)^\top = (QR)^\top (QR) = R^\top Q^\top QR = R^\top R.$$

  Taking $C = R$ establishes the result.

How to establish a decomposition of the flavour of the spectral theorem for general matrices?

Definition (SVD — Singular Value Decomposition)

- Let $A \in \mathbb{R}^{m \times n}$. A singular value decomposition of A consists of orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U \Sigma V^\top, \tag{1}$$

  where $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix, $U^\top U = I$ and $V^\top V = I$.
  The columns of U (V) are the left (right) singular vectors of A. The diagonal elements of Σ, $\sigma_i = \Sigma_{ii}$ are called the singular values of A and are ordered as

$$\sigma_1 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0.$$

- If A has rank r we can write the SVD in compact form $A = U_r \Sigma_r V_r^\top$, where $U_r \in \mathbb{R}^{m \times r}$ contains the first r left singular vectors, $V_r \in \mathbb{R}^{n \times r}$ contains the first r right singular vectors and $\Sigma_r \in \mathbb{R}^{r \times r}$ is a diagonal matrix with the first r singular values.

What does an SVD give us?

Suppose $A \in \mathbb{R}^{m \times n}$ and $A = U\Sigma V^\top$ is its SVD.

- $$AA^\top = U\Sigma V^\top V\Sigma^\top U^\top = U\left(\Sigma\Sigma^\top\right)U^\top.$$

  Hence, the left singular vectors of A are the eigenvectors of $AA^\top$. The singular values of A are the square-root of the eigenvalues of $AA^\top$ (note that $\Sigma\Sigma^\top \in \mathbb{R}^{m \times m}$ is diagonal). If $m > n$, A has n singular values and $AA^\top$ has m eigenvalues (which is larger than n), but the "missing" ones are 0.

- $$A^\top A = V\left(\Sigma^\top\Sigma\right)V^\top.$$

  Hence, the right singular vectors of A are the eigenvectors of $A^\top A$ and the singular values of A are the square-root of the eigenvalues of $A^\top A$ (note that $\Sigma^\top\Sigma$ is $n \times n$ diagonal). If $n > m$, A has m singular values and $A^\top A$ has n eigenvalues (which is larger than m), but the "missing" ones are 0.

The theorem

To wrap up the previous slide

It gives us an idea how to construct a SVD. We will use the spectral theorem applied to the symmetric matrices $A^\top A$ and $AA^\top$. The singular values and vectors of A are in relation with eigenvalues and eigenvectors of these matrices!

Theorem (The SVD Theorem)

Every matrix $A \in \mathbb{R}^{m\times n}$ has an SVD decomposition of the form (1)*.*
In other words:

Every linear transformation is diagonal when viewed in the bases of the singular vectors.

Notes on the proof

Let $A \in \mathbb{R}^{m\times n}$ of rank r. We build a compact SVD $A = U_r \Sigma_r V_r^\top$. From this one gets an SVD as in (1) by adding singular values that are zero and extending singular vectors in both U_r and V_r to orthonormal bases.

The proof I

The first steps

- From the spectral theorem $AA^\top$ has a complete set of orthonormal eigenvectors and can be written as

$$AA^\top = U \Lambda U^\top, \qquad (2)$$

  where $U \in \mathbb{R}^{m \times m}$ is orthogonal and Λ is diagonal.
- Let us write (2) by ordering the diagonal entries of Λ in decreasing order. (2) can be written in compact form, by keeping only the r non-zero eigenvalues and eigenvectors,

$$AA^\top = U_r \Lambda_r U_r^\top$$

  for $U_r \in \mathbb{R}^{m \times r}$ such that $U_r^\top U_r = I$ and Λ_r is $r \times r$ diagonal with the non-zero eigenvalues of $AA^\top$.
- The eigenvalues of $AA^\top$ are non-negative and so Λ_r has positive entries on the diagonal. Let $\Sigma_r \in \mathbb{R}^{r \times r}$ be the diagonal matrix with entries $\sigma_i := (\Sigma_r)_{ii} = \sqrt{\Lambda_{ii}}$.

Show that with $V_r := A^\top U_r \Sigma_r^{-1}$ we obtain a compact SVD.

1. $V_r^\top V_r = I$. Recall that $AA^\top = U_r \Lambda_r U_r^\top$:

$$\begin{aligned} V_r^\top V_r &= \left(A^\top U_r \Sigma_r^{-1}\right)^\top A^\top U_r \Sigma_r^{-1} = \Sigma_r^{-1} U_r^\top A A^\top U_r \Sigma_r^{-1} \\ &= \Sigma_r^{-1} U_r^\top U_r \Lambda_r U_r^\top U_r \Sigma_r^{-1} = \Sigma_r^{-1} \Lambda_r \Sigma_r^{-1} = I \end{aligned}$$

2. $A = U_r \Sigma_r V_r^\top$. Note that

$$U_r \Sigma_r V_r^\top = U_r \Sigma_r \left(A^\top U_r \Sigma_r^{-1}\right)^\top = U_r U_r^\top A.$$

Let us verify that $A = U_r U_r^T A$ by showing $Ax = U_r U_r^T Ax$ for all $x \in \mathbb{R}^n$.

- Let $x \in N(A)$. Then $Ax = 0 = U_r U_r^T x$
- Let $x \in C(A^T)$. It follows that $x = A^T y$ for $y \in \mathbb{R}^m$ and hence,

$$\begin{aligned} Ax &= AA^T y = U_r \Lambda_r U_r^T y = U_r I \Lambda_r U_r^T y \\ &= U_r U_r^T U_r \Lambda_r U_r^T y = U_r U_r^T A A^T y = U_r U_r^T A x. \end{aligned}$$

Consequence of the SVD

Theorem.

A rank-r matrix is a sum of r rank-1 matrices. Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank r. Let $\sigma_1, \ldots, \sigma_r$ be the non-zero singular values of A with left and right vectors $u_1, \ldots, u_r$, $v_1, \ldots, v_r$, respectively. Then

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^\top. \tag{3}$$

Final remark

- The SVD is a powerful tool. Many results presented in this course become significantly simpler with the SVD.
- For instance, if A is invertible and A has SVD $A = U\Sigma V^\top$, then A^{-1} has SVD $A^{-1} = V\Sigma^{-1}U^\top$.
- Similarly, one can define the Moore-Penrose Pseudoinverse by using the SVD.