Eigenvalues & Eigenvectors

- Given a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, λ is said to be an eigenvalue of $\mathbf{A}$ and vector $\vec{x}$ the corresponding eigenvector if

$$A\vec{x} = \lambda\vec{x}$$

- **Geometrical interpretation**
  We can think of the eigenvectors of a matrix A as those vectors which upon being operated by A are only scaled but not rotated.
- **Example**

$$A = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}, \vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} 35 \\ 7 \end{bmatrix} = 7\vec{x}$$

Characteristic Equation

- Trivially, the $\vec{0}$ vector would always be an eigenvector of any matrix. Hence, we only refer only to non-zero vectors as eigenvectors.
- Given a matrix A, how do we find all eigenvalue-eigenvector pairs?

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

The above will hold iff

$$|(A - \lambda I)| = 0$$

This equation is also referred to as the characteristic equation of A. Solving the equation gives us all the eigenvalues λ of A. Note that these eigenvalues can be **complex**.

Properties

1. The trace tr(A) of a matrix A also equals the sum of its n eigenvalues.

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

2. The determinant $|A|$ is equal to the product of the eigenvalues.

$$|A| = \prod_{i=1}^{n} \lambda_i$$

3. The rank of a matrix is equal to the number of non zero eigenvalues of A.
4. If A is invertible, then the eigenvalues of A^{-1} are of form $\frac{1}{\lambda_i}$, where λ_i are the eigenvalues of A.

Proof

$$\sum_{i=1}^{i=k} a_i \vec{v_i} = \vec{0}$$

$$(A - \lambda_k I) \sum_{i=1}^{i=k} a_i \vec{v_i} = \vec{0}$$

$$\sum_{i=1}^{i=k} (A - \lambda_k I) a_i \vec{v_i} = \vec{0}$$

$$\sum_{i=1}^{i=k} a_i (\lambda_i - \lambda_k) \vec{v_i} = \vec{0}$$

Since the eigenvalues are distinct, $\lambda_i \neq \lambda_k \forall i \neq k$. Thus the set of $(k-1)$ eigenvectors is also linearly dependent, violating our assumption of it being the smallest such set. This is a result of our incorrect starting assumption. Hence proved by contradiction.

Diagonalization

Given a matrix A, we consider the matrix S with each column being an eigenvector of A

$$S = \begin{bmatrix} \vdots & \vdots & \dots & \vdots \\ \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}$$

$$AS = \begin{bmatrix} \vdots & \vdots & \dots & \vdots \\ \lambda_1\vec{v_1} & \lambda_2\vec{v_2} & \dots & \lambda_n\vec{v_n} \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}$$

$$AS = \begin{bmatrix} \vdots & \vdots & \dots & \vdots \\ \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \\ \vdots & \vdots & \dots & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots \\ \vdots & \ddots & \dots \\ 0 & \dots & \lambda_n \end{bmatrix}$$

NPTEL

Diagonalization

$$AS = S\Lambda$$
$$A = S\Lambda S^{-1}$$

- $S^{-1}AS$ is diagonal
- Note that the above result is dependent on S being invertible. In the case where the eigenvalues are distinct, this will be true since the eigenvectors will be linearly independent

Properties of Diagonalization

1. A square matrix A is said to be **diagonalizable** if $\exists S$ such that $A = S\Lambda S^{-1}$.
2. Diagonalization can be used to simplify computation of the higher powers of a matrix A, if the diagonalized form is available

$$A^n = (S\Lambda S^{-1})(S\Lambda S^{-1})\ldots(S\Lambda S^{-1})$$
$$A^n = S\Lambda^n S^{-1}$$

Λ^n is simple to compute since it is a diagonal matrix.

Eigenvalues & Eigenvectors of Symmetric Matrices

- Two important properties for a symmetric matrix A:
  1. All the eigenvalues of A are real
  2. The eigenvectors of A are orthonormal, i.e., matrix S is orthogonal. Thus, $A = S\Lambda S^T$.
- Definiteness of a symmetric matrix depends entirely on the sign of its eigenvalues. Suppose $A = S\Lambda S^T$, then

$$x^T A x = x^T S\Lambda S^T x = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

- Since $y_i^2 \geq 0$, sign of expression depends entirely on the λ_i's. For example, if all $\lambda_i > 0$, then matrix A is positive definite.

NPTEL

Eigenvalues of a PSD Matrix

Consider a positive semi definite matrix A. Then, $\forall \vec{x}$ which are eigenvectors of A.

$$\vec{x}^T A \vec{x} \geq 0$$
$$\lambda \vec{x}^T \vec{x} \geq 0$$
$$\lambda ||\vec{x}||^2 \geq 0$$

Hence, all eigenvalues of a PSD matrix are non-negative.

Singular Value Decomposition

1. We saw that diagonalization is applicable only to square matrices. We need some analogue for rectangular matrices too, since we often encounter them, e.g the Document-Term matrix. For a rectangular matrix, we consider left singular and right singular vectors as two bases instead of a single base of eigenvectors for square matrices.
2. The Singular Value Decomposition is given by $A = U\Sigma V^T$ where $U \in R^{m\times m}$, $\Sigma \in R^{m\times n}$ and $V \in R^{n\times n}$.

NPTEL

Singular Value Decomposition

1. U is such that the m columns of U are the eigenvectors of AA^T, also known as the left singular vectors of A.
2. V is such that the n columns of V are the eigenvectors of A^TA, also known as the right singular vectors of A.
3. Σ is a rectangular diagonal matrix with each element being the square root of an eigenvalue of AA^T or A^TA

Significance: SVD allows us to construct a lower rank approximation of a rectangular matrix. We choose only the top r singular values in Σ, and the corresponding columns in U and rows in V^T

Matrix Calculus

1. **The Gradient**
   Consider a function $f : \Re^{m \times n} \rightarrow \Re$. The gradient $\nabla_A f(A)$ denotes the matrix of partial derivatives with respect to every element of the matrix A. Each element is given by $(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$
2. **The Hessian**
   Suppose a function $f : \Re^n \rightarrow \Re$ takes in vectors and returns real numbers. The Hessian, denoted as $\nabla_x^2 f(x)$ or H is the $n \times n$ matrix of partial derivatives. $(\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$. Note that the Hessian is always symmetric.
3. Note that the Hessian is not the gradient of the gradient, since the gradient is a vector, and we cannot take the gradient of the vector. However, if we do take elementwise gradients of every element of the gradient, then we can construct the Hessian.

NPTEL

Differentiating Linear and Quadratic Functions

If $f(x) = b^T x$, for some constant $b \in \Re^n$. Let us find the gradient of f.

$$f(x) = \sum_{i=1}^{i=n} b_i x_i$$

$$\frac{\partial f(x)}{x_k} = b_k$$

We can see that $\frac{\partial b^T x}{\partial x} = b$. We can intuitively see how this relates to differentiating $f(x) = ax$ with respect to x when a and x are real scalars.

Differentiating Linear and Quadratic Functions

Consider the function $f(x) = x^T Ax$ where $x \in R^n$ and $A \in \Re^{n\times n}$ is a known symmetric matrix.

$$f(x) = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} A_{ij}x_i x_j$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\sum_{i\neq k}\sum_{j\neq k} A_{ij}x_i x_j + \sum_{i\neq k} A_{ik}x_i x_k + \sum_{j\neq k} A_{kj}x_k x_j + A_{kk}{x_k}^2\right]$$

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i\neq k} A_{ik}x_i + \sum_{j\neq k} A_{kj}x_j + 2A_{kk}x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i=1}^{n} A_{ik}x_i + \sum_{j=1}^{n} A_{kj}x_j$$

$$\frac{\partial f(x)}{\partial x_k} = 2\sum_{i=1}^{n} A_{ki}x_i$$

Differentiating Linear and Quadratic Functions

Thus $\nabla_x(x^T Ax) = 2Ax$. Now, let us find the Hessian H.

$$\frac{\partial}{\partial x_k}\frac{\partial f(x)}{\partial x_l} = \frac{\partial}{\partial x_k}(2\sum_{i=1}^{i=n} A_{li}x_i) \qquad = 2A_{kl}$$

Hence, $\nabla_x^2(x^T Ax) = 2A$.