Lecture 4

Math Foundations Team

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Matrix decompositions

- We studied vectors and how to manipulate them in preceding lectures.
- Mappings and transformations of vectors can be conveniently described in terms of operations performed by matrices.
- In this lecture we shall study three aspects of matrices: how to summarize matrices, how matrices can be decomposed, and how the decompositions can be used for matrix approximations.

Characteristic polynomial

- For $\lambda \in \mathbb{R}$ and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ we can define $p_{\boldsymbol{A}}(\lambda) = \det(A - \lambda \boldsymbol{I})$ and show that it can be written as $c_0 + c_1\lambda + \ldots c_{n-1}\lambda^n + (-1)^n\lambda^n$ where $c_0, c_1 \ldots c_{n-1} \in R$.
- We can show that $c_0 = \det(\boldsymbol{A})$ and $c_{n-1} = (-1)^{n-1}tr(\boldsymbol{A})$
- To see that $c_0 = \det(\boldsymbol{A})$, set $\lambda = 0$ in $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ to get $p_{\boldsymbol{A}}(0) = \det(\boldsymbol{A}) = c_0$
- The formula for c_{n-1} takes a little bit of work - let us expand a 3×3 determinant $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix}$

Characteristic polynomial

- Expanding the determinant along the first row we see that the $(a_{11} - \lambda)C_{11}$ term contains the product $\prod_{i=1}^{i=3}(a_{ii} - \lambda)$ which contains the powers λ^3 and λ^2. The other contributors to the determinant i.e $a_{12}C_{12}$ and $a_{12}C_{13}$ expand into terms where the maximum power of $\lambda = 1$.
- Carrying this analogy over to the general case of $n > 3$ we see that expanding along the first row the first contributor to the determinant will have the term $\prod_{i=1}^{i=n}(a_{ii} - \lambda)$ and subsequent contributors will have a maximum power of λ^{n-2} as the minors for each such contributor will kill off a term containing λ in a given row and column.

Characteristic polynomial

- Thus in the determinant expansion to obtain the characteristic polynomial we see that coefficient to λ^{n-1} can only come from the expansion of $\prod_{i=1}^{i=n}(a_{ii} - \lambda)$ and can be seen to be seen to be $(-1)^{n-1}\sum_{i=1}^{i=n} a_{ii} = (-1)^{n-1}tr(\boldsymbol{A})$.
- As a corollary to this argument we can see that the coefficient to λ^n in the characteristic polynomial is $(-1)^n$
- We will use the characteristic polynomial to compute eigenvalues and eigenvectors.

Eigenvalues and eigenvectors

- Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{R}^n \setminus 0$ is the corresponding eigenvector of λ if $\boldsymbol{Ax} = \lambda \boldsymbol{x}$. This equation is called the eigenvalue equation.
- The following statements are equivalent:
  - λ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.
  - There exists an $\boldsymbol{x} \in \mathbb{R}^n \setminus 0$ with $\boldsymbol{Ax} = \lambda \boldsymbol{x}$, or equivalently, $(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{x} = \boldsymbol{0}$ can be solved non-trivially, i.e., $\boldsymbol{x} \neq \boldsymbol{0}$.
  - $\mathrm{rank}(\boldsymbol{A} - \lambda \boldsymbol{I}_n) < n$.
  - $\det(\boldsymbol{A} - \lambda \boldsymbol{I}_n) = 0$.
- If $\boldsymbol{x}$ is an eigenvector corresponding to a particular eigenvalue λ, $c\boldsymbol{x}, c \in \mathbb{R} \setminus 0$ is also an eigenvector.

Eigenvalues and eigenvectors - example

- Consider the matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. The characteristic polynomial $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (1-\lambda)^2 - 1$ and setting it to zero gives us the roots of the characetristic polynomial: $(1-\lambda)^2 - 1 = 0$ has roots $\lambda = 2, 0$.
- What are the eigenvectors? For $\lambda = 0$ we solve for $\boldsymbol{Ax} = 0\boldsymbol{x}$, so we find the nullspace of the matrix $\boldsymbol{A}$. Using Gaussian elimination we convert $\boldsymbol{Ax} = \boldsymbol{0}$ to $\boldsymbol{Ux} = \boldsymbol{0}$ where $\boldsymbol{U} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.

  Thus we discover the eigenvector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for $\lambda = 0$.
- Similarly we discover the eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $\lambda = 2$.

Eigenvalues and eigenvectors - example

- The general procedure to find eigenvalues and eigenvectors is to first find the roots of the characteristic polynomials and then find the nullspaces of the matrices $\boldsymbol{A} - \lambda \boldsymbol{I}$ for the different roots λ.
- Does every $n \times n$ matrix have a full set of eigenvectors, i.e n eigenvectors?
- Look at $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. What are its eigenvalues and eigenvectors?
- **Point to ponder** Looking at the equation $\boldsymbol{Ax} = \lambda \boldsymbol{x}$ it seems that the action of $\boldsymbol{A}$ on $\boldsymbol{x}$ is to preserve the direction of $\boldsymbol{x}$ but scale it up or down according to λ. Does this mean that a rotation matrix has no eigenvalues and eigenvectors?

Some additional properties

- λ is an eigenvalue of $\boldsymbol{A}$ if and only if λ is a root of the characteristic polynomial $p_{\boldsymbol{A}}(\lambda)$ of $\boldsymbol{A}$. This can be easily seen as a consequence of the definition of $p_{\boldsymbol{A}}(\lambda)$.
- For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the set of eigenvectors corresponding to an eigenvalue λ spans a subspace of $\mathbb{R}^n$ called the Eigenspace of $\boldsymbol{A}$ with respect to λ and is denoted by E_λ.
- The set of all eigenvalues of $\boldsymbol{A}$ is called the spectrum of $\boldsymbol{A}$.
- Look at the eigenvalues and eigenspace of the $n \times n$ identity matrix $\boldsymbol{I}_n$. It has one eigenvalue $\lambda = 1$ and the eigenspace is $\mathbb{R}^n$. Every canonical vector is a basis vector for the eigenspace.

Some additional properties

- A matrix and its transpose have the same eigenvalues. To see this, first note that $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T)$. Then $\det(\boldsymbol{A} - \lambda I) = \det((\boldsymbol{A} - \lambda \boldsymbol{I})^T) = \det(\boldsymbol{A}^T - \lambda \boldsymbol{I}^T) = \det(\boldsymbol{A}^T - \lambda \boldsymbol{I})$. The last expression in the chain of equalities is the characteristic polynomial for $p_{\boldsymbol{A}^T}(\lambda)$. Thus we have $p_{\boldsymbol{A}}(\lambda) = p_{\boldsymbol{A}^T}(\lambda)$ which means the characteristic polynomials are equal and so the roots of the polynomials or the eigenvalues must be equal.
- The eigenspace E_λ is the nullspace of $\boldsymbol{A} - \lambda \boldsymbol{I}$.
- Symmetric, positive-definite matrices always have positive, real eigenvalues.

Some theorems

- The eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n$ of a $n \times n$ matrix $\boldsymbol{A}$ with n **distinct** eigenvalues are linearly independent $\rightarrow$ why?
- Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ we can show that $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a symmetric, positive-definite matrix when the rank of $\boldsymbol{A} = n$. Why is this true? Clearly $\boldsymbol{A}^T\boldsymbol{A}$ is a symmetric matrix and it is positive definite since $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{A}\boldsymbol{x}\|^2 > 0 \;\; \forall \boldsymbol{x} \in \mathbb{R}^n \setminus 0$ since the nullspaces of $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}$ are the same, and $\boldsymbol{A}$ is a full column rank matrix.
- The matrix $\boldsymbol{A}^T\boldsymbol{A}$ is important in machine learning since it figures in the least-squares solution to a data matrix represented as $\boldsymbol{A}$ where n represents the number of features and m is the number of data vectors.

Spectral theorem

Theorem: If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric there exists an orthonormal basis of the corresponding vector space V consisting of the eigenvectors of $\boldsymbol{A}$, and each eigenvalue is real.

Proof: We will not attempt a full proof of this theorem but provide some intuitions about why it is true. The theorem relies on the following three statements, shown in the next slide.

Spectral theorem

- All roots of the characteristic polynomial $p_{\boldsymbol{A}}(\lambda)$ are real.
- For each eigenvalue λ we can compute an orthonormal basis for its eigenspace. We can string together the orthonormal bases for the different eigenvalues of $\boldsymbol{A}$ to come up with the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots$
- The dimension of the eigenspace E_λ, called its geometric multiplicity, is the same as the algebraic multiplicity of λ which is the number of times λ appears as a root of the characteristic polynomial.
- All the basis vectors from the different Eigenspaces combine to provide an orthonormal basis for $\mathbb{R}^n$.

Complex vectors

- In the old formulation with real vectors, length-squared according to the Euclidean norm was $x_1^2 + x_2^2 + \ldots x_n^2$. If the x_i are complex we should take length-squared to be $|x_1|^2 + |x_2|^2 + \ldots |x_n|^2$ where $|.|$ denotes modulus. For the complex number $a + bi$, the modulus is $\sqrt{(a+bi)(a-bi)} = \sqrt{a^2 + b^2}$
- For complex vectors we would like to preserve the idea as possible that $\|x\|^2 = \boldsymbol{x}^T\boldsymbol{x}$. If we keep the old definition of inner product for complex vectors we will not get a real number as length as shown in the next bullet.
- Let $\boldsymbol{x} = \begin{bmatrix} 1+i \\ 2+i \end{bmatrix}$. We have $\boldsymbol{x}^T\boldsymbol{x} = (1+i)^2 + (2+i)^2 = 1 + 2i + i^2 + 4 + 4i + i^2 = 6i + 3.$

Hermitian matrices

- We modify the inner product between two complex vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ to $\boldsymbol{x}^H\boldsymbol{y}$, where $\boldsymbol{x}^H = \overline{\boldsymbol{x}}^T$.
- Now $\boldsymbol{x}^H\boldsymbol{x} = \overline{x_1}x_1 + \ldots \overline{x_n}x_n = \|x\|^2$ according to the new definition of length.
- A Hermitian matrix is a generalization of a symmetric matrix.
- Instead of requiring $\boldsymbol{A}^T = \boldsymbol{A}$, we say a matrix is Hermitian if it is equal to its conjugate-transpose, ie $\boldsymbol{A}$ is a Hermitian matrix if $\boldsymbol{A}^H = \boldsymbol{A}$ or $\overline{\boldsymbol{A}}^T = \boldsymbol{A}$
- As an example consider the matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 3-i \\ 3+i & 4 \end{bmatrix}$. It is a Hermitian matrix since $\boldsymbol{A}^H = \begin{bmatrix} 1 & 3-i \\ 3+i & 4 \end{bmatrix} = \boldsymbol{A}$.

Spectral theorem

We shall now show that all eigenvalues for a symmetric matrix are real. Let $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. Then premultiplying with $\boldsymbol{x}^H$ on both sides we have $\boldsymbol{x}^H\boldsymbol{Ax} = \lambda\boldsymbol{x}^H\boldsymbol{x}$

Now $\boldsymbol{x}^H\boldsymbol{Ax}$ is a 1×1 matrix. Taking the Hermitian of this matrix we have $(\boldsymbol{x}^H\boldsymbol{Ax})^H = \boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{x} = \boldsymbol{x}^H\boldsymbol{Ax}$, so the Hermitian of the matrix is itself which means that the matrix is real.

On the right hand side we note that $\boldsymbol{x}^H\boldsymbol{x}$ is real, so this means that λ must be real.

Spectral theorem

Let us show that eigenvectors belonging to different eigenvalues are orthogonal. Let $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ and $\boldsymbol{A}\boldsymbol{y} = \mu \boldsymbol{y}$. Then we have

$$\begin{aligned}
\boldsymbol{y}^H \boldsymbol{A}\boldsymbol{x} &= \lambda \boldsymbol{y}^H \boldsymbol{x} \\
\boldsymbol{x}^H \boldsymbol{A}\boldsymbol{y} &= \mu \boldsymbol{x}^H \boldsymbol{y}
\end{aligned}$$

But $\boldsymbol{x}^H \boldsymbol{A}\boldsymbol{y} = (\boldsymbol{y}^H \boldsymbol{A}^H \boldsymbol{x})^H = (\boldsymbol{y}^H \boldsymbol{A}\boldsymbol{x})^H = \lambda \boldsymbol{x}^H \boldsymbol{y}$. We already know that $\boldsymbol{x}^H \boldsymbol{A}\boldsymbol{y} = \mu \boldsymbol{x}^H \boldsymbol{y}$. This means $\lambda \boldsymbol{x}^H \boldsymbol{y} = \mu \boldsymbol{x}^H \boldsymbol{y}$. Since $\lambda \neq \mu$, this must mean $\boldsymbol{x}^H \boldsymbol{y} = 0$.

This shows that eigenvectors corresponding to different eigenvalues are orthogonal.

Spectral theorem

- So we see that the eigenvalues of a symmetric matrix are real and eigenvectors belonging to different eigenvalues are orthogonal.
- This suggests that one can string together all the orthonormal bases for the different eigenvalues and get an orthonormal basis for $\mathbb{R}^n$.
- But who is to say that when we string together the basis vectors for all the eigenvalues, we will have enough vectors to describe $\mathbb{R}^n$? We need n basis vectors and might end up having fewer than n vectors.
- If the eigenvalues are all different, we can see that we will indeed have enough basis vectors. But what about when there are repeating eignevalues?

Spectral theorem

- We need one more piece to complete the puzzle and show that we will have enough eigenvectors to complete the orthonormal basis - this part we shall not prove!
- As a consequence of the spectral theorem we can write a real symmetric matrix $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^T$ where $\boldsymbol{Q}$ is an orthonormal basis (think orthonormal basis vectors for example), and Λ is a diagonal matrix consisting of non-zero entries only along the diagonal.
- The spectral theorem can be used in a machine learning context since we can take the data matrix $\boldsymbol{A}$ and create a symmetric matrix out of it - $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^T$ which are both used in Singular-Value Decomposition and PCA.

Trace and eigenvalues

- We can show that the sum of the eigenvalues of a matrix is equal to the trace of the matrix, i.e $\sum_{i=1}^{i=n} \lambda_i = \sum_{i=1}^{i=n} a_{ii}$. To see why this is true, note that the characteristic polynomial $p_{\boldsymbol{A}}(\lambda)$ can be written as $\prod_{i=1}^{i=n} (\lambda_i - \lambda)$. The coefficient to λ^{n-1} in this expansion is $(-1)^{n-1} \sum_{i=1}^{i=n} \lambda_i$. Early on in this lecture we showed from a direct expansion of the determinant that the coefficient of λ^{n-1} is $(-1)^{n-1} \sum_{i=1}^{i=n} a_{ii}$. Thus we have our result.
- The product of all eigenvalues is the determinant of the matrix, i.e $\det(\boldsymbol{A}) = \prod_{i=1}^{i=n} \lambda_i$. To see why this is true, let us once again look at the factorisation of $p_{\boldsymbol{A}}(\lambda)$ as $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = p_{\boldsymbol{A}}(\lambda) = \prod_{i=1}^{i=n} (\lambda_i - \lambda)$. Setting $\lambda = 0$ in this equation gives the result.