M677_W04

September 26, 2021

1 W04: Eigenvalue decomposition

Last updated: 26 September 2021

1.1 Eigenvalues and eigenvectors

1.1.1 Theory

Definition: Let $L : V \to V$ be a linear map from a vector space V into itself. The scalar λ is called an *eigenvalue* for L if there is a nonzoero $v \in V$—an associated *eigenvector*—such that

$$L(v) = \lambda v.$$

The set of eigenvectors associated with λ (augmented with the zero vector), i.e.,

$$E_\lambda = \{v \in V : L(v) = \lambda v\}$$

is a linear subspace of V, called the *eigenspace* associated with λ.

Let $\underline{b}$ be a basis for the (finite-dimensional) vector space V and let A be the matrix of L in the basis $\underline{b}$. Recall that $\mathrm{coeff}_{\underline{b}}(L(v)) = A\,\mathrm{coeff}_{\underline{b}}(v)$, so that

$$[L(v) = \lambda v] \iff [Ax = \lambda x,\ x := \mathrm{coeff}_{\underline{b}}(v)].$$

Thus, an eigenvalue problem for a linear map translates right away into an eigenvalue problem for a matrix (and make sure you understand, at the end of the class, why the choice of $\underline{b}$ is inconsequential), so we shall only consider matrices below.

Q: How to find the eigenvalues of $A \in \mathbb{R}^{n\times n}$? A: Find the roots of the characteristic polynomial.

This is based on the following chain of equivalences:

$$
\begin{aligned}
[\lambda \text{ is an eigenvalue for } A] &\iff [\text{there exists } v \in \mathbb{R}^n \setminus \{0\} : Av = \lambda v] \\
&\iff [\text{there exists } v \in \mathbb{R}^n \setminus \{0\} : (A - \lambda I_n)v = 0] \\
&\iff [A - \lambda I_n \text{ is not injective}] \\
&\iff [A - \lambda I_n \text{ is not bijective}] \\
&\iff [\det(A - \lambda I_n) = 0] \\
&\iff [\lambda \text{ is a root of the characteristic polynomial } P_A].
\end{aligned}
$$

Here, the characteristic polynomial is defined by

$$P_A(x) = \det(A - xI_n).$$

Notice that it is indeed a polynomial. Its degree is n, its leading term is $(-1)^n$, and its constant term is $\det(A)$.

By properties of polynomials, one can deduce: - A has at most n eigenvalues; - $A \in \mathbb{C}^{n \times n}$ always has at least one eigenvalue—notice that the scalar field is $\mathbb{C}$; - $A \in \mathbb{R}^{n \times n}$ has at least one eigenvalue when n is odd—notice that the scalar field is $\mathbb{R}$.

Q: How to find the eigenvectors? A: Solve a homogeneous linear system.

Indeed, to find an eigenvector corresponding to an eigenvalue λ (which has been determined in the previous step), we have to find a nontrivial solution to

$$(A - \lambda I_n)v = 0.$$

Proposition: If $\lambda_1, \ldots, \lambda_k$ are *distinct* eigenvalues for A and if $v_1, \ldots, v_k$ are associated eigenvectors, then the system $(v_1, \ldots, v_k)$ is linearly independent.

Proof: By induction, see in class.

For scalars $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct) and vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, let $D \in \mathbb{R}^{n \times n}$ be the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal and let $V \in \mathbb{R}^{n \times n}$ be the matrix with columns $v_1, \ldots, v_n$. We notice that

$$
\begin{aligned}
[Av_j = \lambda_j v_j \text{ for all } j = 1, \ldots, n] &\iff [AVe_j = VDe_j \text{ for all } j = 1, \ldots, n] \\
&\iff [AV = VD].
\end{aligned}
$$

This observation is the core of the link between eigenvalues/eigenvectors and diagonalization.

Definition: A matrix $A \in \mathbb{R}^{n \times n}$ is called *diagonalizable* if there exist a diagonal matrix $D \in \mathbb{R}^{n \times n}$ and an invertible matrix $V \in \mathbb{R}^{n \times n}$ such that

$$A = VDV^{-1}.$$

According to the above observation, A is diagonalizable if and only if there exists a basis of eigenvectors for A. This happens for instance of A has n distinct eigenvalues.

Theorem (Cayley–Hamilton): A matrix $A \in \mathbb{R}^{n\times n}$ annihilates its characteristic polynomial, i.e.,

$$P_A(A) = 0.$$

Proof: The following argument is *incorrect*: since $P_A(x) = \det(A - Ax)$, then $P_A(A) = \det(A - AI_n) = \det(0) = 0$. Here is a correct argument in case A is diagonalizable: let $(v_1, \dots, v_n)$ be a basis of eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$; then $P_A(x) = (\lambda_1 - x)\cdots(\lambda_n - x)$, so that $P_A(A) = (\lambda_1 I_n - A)\cdots(\lambda_n I_n - A)$; next, since the factors commute, $P_A(A)(v_j) = \Big(\prod_{i\neq j}(\lambda_i I_n - A)\Big)(\lambda_j I_n - A)(v_j) = \Big(\cdot\Big)(0) = 0$; and this being true for each element of the basis $(v_1, \dots, v_n)$ implies $P_A(A) = 0$.

1.1.2 Some Python

[1]:
```python
from sympy import *
M = Matrix([[3, -2,  4, -2], [5,  3, -3, -2], [5, -2,  2, -2], [5, -2, -3,  3]])
lamda = symbols('lamda')
p = M.charpoly(lamda)
factor(p.as_expr())
```

[1]: $(\lambda - 5)^2 (\lambda - 3)(\lambda + 2)$

[2]:
```python
from scipy.linalg import funm
funm(M, lambda x: (x-5)**2 * (x-3) * (x+2) )
```

```
funm result may be inaccurate, approximate err = 0.721687836487032
```

[2]:
```
array([[-8.88178420e-15, -8.88178420e-15,  1.77635684e-14,
         -8.88178420e-15],
        [-2.26485497e-13, -8.88178420e-15,  2.35367281e-13,
         -8.88178420e-15],
        [-2.26485497e-13, -8.88178420e-15,  2.35367281e-13,
         -8.88178420e-15],
        [-2.26485497e-13, -8.88178420e-15,  2.35367281e-13,
         -8.88178420e-15]])
```

1.1.3 Exercises

1. Prove that A and $A^\top$ have the same eigenvalues.
2. Show that the eigenvalues of a diagonally dominant matrix are all positive.
3. If the matrices $A, B \in \mathbb{R}^{n\times n}$ are similar, i.e., $A = PBP^{-1}$ for some invertible matrix $P \in \mathbb{R}^{n\times n}$, prove that A and B have the same characteristic polynomial.
4. Prove that the matrix $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable.

5. Suppose that λ is an eignevalue of $A \in \mathbb{R}^{n\times n}$. Given a polynomial P, prove that $P(\lambda)$ is an eigenvalues for $P(A)$. Furthemore, if A is invertible, prove that $\lambda \neq 0$ and that λ^{-1} is an eigenvalue for A^{-1}.

1.2 Self-adjoint matrices

1.2.1 Theory

We recall that the *adjoint* of a matrix $A \in \mathbb{C}^{m\times n}$ is the matrix $A^* \in \mathbb{C}^{n\times m}$ with entries

$$A^*_{i,j} = \overline{A_{j,i}} \qquad \text{for all } i = 1, ..., n \text{ and } j = 1, ..., m.$$

It is characterized by the relation

$$\langle A^*x, y\rangle = \langle x, Ay\rangle \qquad \text{for all } x \in \mathbb{R}^m \text{ and all } y \in \mathbb{R}^n.$$

A square matrix $A \in \mathbb{C}^{n\times n}$ is called *self-adjoint*, or *hermitian*, if

$$A^* = A.$$

It is called *skew-hermitian* if

$$A^* = -A.$$

Hermitian and skew-hermitian matrices both belong to the set of *normal* matrices, i.e., matrices satisfying

$$AA^* = A^*A.$$

Unitary matrices are also normal.

An example of a self-adjoint (symmetric, in this case) matrix is given by the *Hessian* at some $x \in \mathbb{R}^n$ of a twice continuously differentiable function f defined on $\Omega \subseteq \mathbb{R}^n$. This is the matrix

$$\text{Hessian}(f, x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]_{\substack{i=1,...,n \\ j=1,...,n}}.$$

If $A \in \mathbb{C}^{n\times n}$ is self-adjoint, then $\langle Ax, x\rangle$ is a real number for any $x \in \mathbb{C}^n$. In the same line of ideas, we point out the following observation as a theorem.

Theorem: The eigenvalues of a self-adjoint matrix are real numbers.

Proof: In class.

Moreover, a self-adjoint matrix is always diagonalizable (and if it is real-valued, then it is diagonalizable in $\mathbb{R}$). The following result, known as the spectral theorem, says much more.

Theorem: A matrix $A \in \mathbb{C}^{n \times n}$ is normal if and only if it if unitarily diagonalizable, meaning that there exist a diagonal matrix $D \in \mathbb{C}^{n \times n}$ and a unitary matrix $V \in \mathbb{C}^{n \times n}$ such that

$$A = VDV^*.$$

Proof: The reverse implication is clear. We prove (in class) the direct implication only in the case where A is self-adjoint and we do so by induction.

The eigenvalues of a self-adjoint matrix $A \in \mathbb{C}^{n \times n}$ (which are real) are usually ordered in a nonincreasing fashion, i.e., as

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A).$$

Theorem: If $A \in \mathbb{C}^{n \times n}$ is self-adjoint, then

$$\lambda_1(A) = \max_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle};$$
$$\lambda_n(A) = \min_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

Proof: In class.

As a corollary, for any matrix $M \in \mathbb{C}^{m \times n}$, not necessarily self-adjoint nor even square, we derive (see class) that the matrix norm induced by the ℓ_2-norm on $\mathbb{C}^n$ is obtained as the largest eigenvalue of the self-adjoint matrix MM^*. In short,

$$|||M||| = \lambda_1(MM^*)^{1/2}.$$

Theorem (Courant–Fischer): If $A \in \mathbb{C}^{n \times n}$ is self-adjoint, then, for any $i = 1, \ldots, n$,

$$\lambda_i(A) = \min_{\dim(V)=n-i+1} \max_{x \in V} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \max_{\dim(V)=i} \min_{x \in V} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

Proof: In class, time allowing.

There are several methods to (numerically) find the eigenvalues of a self-adjoint matrix. They do not consist in finding the roots of the characteristic polynomial! One of them is the power method, discussed briefly in class.

1.2.2 Some Python

[3]:
```
import numpy as np
n = 7
A = np.random.normal(0,1,(n,n))
np.linalg.eigvals(A)
```

```
[3]: array([ 3.65175909+0.j        ,  1.57241483+2.83863579j,
         1.57241483-2.83863579j, -3.75179247+0.j        ,
        -1.48757667+1.16390223j, -1.48757667-1.16390223j,
         0.25618186+0.j        ])
```

```
[4]: np.linalg.eigvalsh(A+A.T)
```

```
[4]: array([-8.25152348, -5.02170853, -1.94258681,  0.46856535,  3.59732294,
         4.02542152,  7.77615861])
```

1.2.3 Exercises

1. Prove that any square matrix can be written uniquely as the sum of a hermitian matrix and a skew-hermitian matrix.
2. Prove that a matrix $A \in \mathbb{C}^{n\times n}$ is normal if and only if

$$\langle Ax, Ay\rangle = \langle A^*x, A^*y\rangle \qquad \text{for all } x, y \in \mathbb{R}^n.$$

3. For a self-adjoint matrix $A \in \mathbb{C}^{n\times n}$, prove that

$$|||A||| = \max\{|\lambda_i(A)|, i = 1, \ldots, n\}.$$

1.3 Positive semidefinite matrices

1.3.1 Theory

Definition: A square matrix $A \in \mathbb{C}^{n\times n}$ is called *positive semidefinite*, resp. *positive definite* if it is self-adjoint and if - $\langle Ax, x\rangle \geq 0$, resp. $\langle Ax, x\rangle > 0$, for all $x \in \mathbb{C}^n \setminus \{0\}$.

Equivalently, it is positive semidefinite, resp. positive definite, if it is self-adjoint and if all its eignevalues are nonnegative, resp. positive.

Proof of the equivalence: Using the spectral theorem, details in class.

An example of a positive semidefinite matrix is given by the Hessian of a twice continuously differentiable function f at one of its local minimizers.

The notation used to write that a matrix $A \in \mathbb{C}^{n\times n}$ is positive semidefinite, resp. positive definite, is $A \succeq 0$, resp. $A \succ 0$. Obviously, writing $B \succeq C$, resp. $B \succ C$, means that $B - C$ is positive semidefinite, resp. positive definite. The following result follows from Courant–Fischer theorem.

Proposition: If $B \succeq C$, then $\lambda_i(B) \geq \lambda_i(C)$ for all $i = 1, \ldots, n$.

Proof: In class.

Proposition: For $A \in \mathbb{C}^{n\times n}$,

$$[A \succeq 0] \iff [\text{there exists } B \text{ such that } A = B^*B].$$

Proof: In class—the reverse implication follows from the definition; for the direct implication, use the spectral theorem.

Note that there was no resriction on the size of B. If one imposes B to be square, also self-adjoint, and positive semidefinite, then B is unique. It is called the square root of A.

Theorem (Schur): If $A, B \in \mathbb{C}^{n\times n}$ are positive semidefinite, then so is their pointwise product. In other words:

$$[A, B \succeq 0] \Rightarrow [A \odot B \succeq 0].$$

Proof: See class.

1.3.2 Some Python

[5]:
```
# the sorted eigenvalues of a self-adjoint matrix
n = 10;
k = 5;
aux = np.random.normal(0,1,(n,k))
A = aux@aux.T
-np.sort(-np.linalg.eigvalsh(A))
```

[5]:
```
array([ 2.45280659e+01,  1.14655625e+01,  9.53130646e+00,  5.55219537e+00,
        3.15442485e+00,  2.10656099e-15,  1.08248987e-15,  5.84844058e-16,
       -1.89519324e-15, -2.06130381e-15])
```

[6]:
```
# the matrix square-root and its sorted eigenvalues
from scipy.linalg import sqrtm
B = sqrtm(A)
-np.sort(-np.linalg.eigvalsh(B))
```

[6]:
```
array([ 4.95258174e+00,  3.38608365e+00,  3.08728140e+00,  2.35630969e+00,
        1.77607006e+00,  4.39139488e-08,  3.49257854e-08,  2.43591365e-08,
       -4.43349761e-09, -1.56441463e-08])
```

[7]:
```
# illustration of Schur theorem
aux = np.random.normal(0,1,(n,1))
C = aux@aux.T
-np.sort(-np.linalg.eigvalsh(B*C))
```

[7]:
```
array([ 1.51857388e+01,  6.07024648e+00,  2.47349485e+00,  1.56833377e+00,
        1.23332657e+00,  1.05736120e-07,  1.76529052e-08,  1.40502536e-08,
       -2.56067065e-09, -1.68311786e-08])
```

1.3.3 Exercises

1. Prove that the self-adjointness condition in the definition of positive semidefiniteness is redundant in the complex setting (it is not on the real setting). More precisely, given $A \in \mathbb{C}^{n\times n}$, prove that, if $\langle Ax, x\rangle \in \mathbb{R}$ for all $x \in \mathbb{C}^n$, then A is automatically self-adjoint.
2. If $A \in \mathbb{C}^{n\times n}$ is positive semidefinite, verify that $\langle Ax, y\rangle$, $x, y \in \mathbb{C}^n$, defines an inner product on $\mathbb{C}^n$.

3. Let A and B be positive semifefinite matrices. Show that $\mathrm{tr}(AB) \geq 0$. Is the product AB necessarily positive semidefinite?
4. Given two square matrices $B, C \in \mathbb{C}^{n\times n}$, suppose that $\lambda_i(B) \geq \lambda_i(C)$ for all $i = 1, \ldots, n$. Prove that there exists a unitary matrix $V \in \mathbb{C}^{n\times n}$ such that $B \succeq VCV^*$.

1.4 More Exercises

1. For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$, prove that AB and BA have the same nonzero eigenvalues.
2. If $A \in \mathbb{R}^{n\times n}$ is invertible, prove that A^{-1} can be expressed as a polynomial in A.
3. Diagonalize the matrix

$$\begin{bmatrix} 1 & t & t & \cdots & t \\ t & 1 & t & \cdots & t \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t & \cdots & t & 1 & t \\ t & \cdots & t & t & 1 \end{bmatrix}.$$

4. What is the characteristic polynomial of the so-called *companion matrix*

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}.$$