Session 22
April 29, 2020

Last time we introduced normal matrices. These are the $n\times n$ matrices that satisfy $A^\dagger A = AA^\dagger$. Normal matrices are so nice because these are exactly the matrices that can be diagonalized with orthonormal eigenvectors. Whether or not a matrix is normal is very easy to find out: We simply have to check if $A^\dagger A = AA^\dagger$ (just two matrix multiplications).

Today, we will discuss a few more "subtypes" of normal matrices, and what we can say about their eigenvalues and eigenvectors.

A) Hermitian or Self-adjoint matrices

Again, we allow for $n\times n$ matrices A to have complex entries.

In many applications, matrices have some symmetry. For example we defined the Hessian matrix with entries $\frac{\partial^2 f}{\partial x_i \partial x_j}$. This does not change if we interchange i and j. In the general case, we define the following:

Definition: An $n\times n$ matrix A is called Hermitian or self-adjoint if $A^\dagger = A$. In components, this means $\bar{A}_{ji} = A_{ij}$ $\forall\, i,j = 1,\dots,n$.

Similarly, one calls matrices with $A^\dagger = -A$ anti-Hermitian or skew-Hermitian.

Note: • We can write any matrix as $A = \underbrace{\tfrac{1}{2}(A+A^\dagger)}_{\substack{\text{Hermitian, since} \\ (A+A^\dagger)^\dagger = A+A^\dagger}} + \underbrace{\tfrac{1}{2}(A-A^\dagger)}_{\substack{\text{anti-Hermitian, since} \\ (A-A^\dagger)^\dagger = -(A-A^\dagger)}}$, so these two types of matrices are very elementary.

• If $i=j$, the equation $A^\dagger = A$ tells us that $\bar{A}_{jj} = A_{jj}$. Thus the diagonal entries of a Hermitian matrix are real.

And note: If $A^\dagger = A$, then clearly $A^\dagger A = AA = AA^\dagger$, i.e., every Hermitian matrix is normal (but not necessarily the other way around)

↳ also every anti-Hermitian matrix is normal

But we already showed that if a normal matrix A has eigenvalue λ, then $A^\dagger$ has eigenvalue $\overline{\lambda}$ (the complex conjugate). If now $A = A^\dagger$ it follows that $\lambda = \overline{\lambda}$, i.e., λ was real! We found:

<u>Theorem:</u> All eigenvalues of a Hermitian matrix are real.

Note: • Exactly this fact makes it possible to describe the statistics of experiments in quantum mechanics with Hermitian matrices.

• For anti-Hermitian matrices one can show similarly that all eigenvalues are purely imaginary or zero.

• Since Hermitian matrices are in particular normal, they can be diagonalized with orthonormal eigenvectors. This important result is called $\underbrace{\text{spectral theorem}}_{\text{there are many interesting versions and generalizations of this}}$.

As simple example, consider $A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} = A^\dagger$. We find

$\det\begin{pmatrix} 1-\lambda & i \\ -i & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - i(-i) = 1 - 2\lambda + \lambda^2 - 1 = \lambda(\lambda - 2)$, so the two eigenvalues are

$\lambda_1 = 0$ and $\lambda_2 = 2$, both real. Eigenvectors:

↳ $\lambda_1 = 0$: $\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x + iy = 0 \Rightarrow \begin{pmatrix} -iy \\ y \end{pmatrix}$ are the eigenvectors, i.e.,

for any $y \in \mathbb{C}, y \neq 0$

$E_{\lambda_1}(A) = \operatorname{span}\left\{\begin{pmatrix} -i \\ 1 \end{pmatrix}\right\}$.

When is $\begin{pmatrix} -iy \\ y \end{pmatrix}$ normalized? We want $1 = \left\langle \begin{pmatrix} -iy \\ y \end{pmatrix}, \begin{pmatrix} -iy \\ y \end{pmatrix} \right\rangle = \overline{\begin{pmatrix} -iy \\ y \end{pmatrix}} \cdot \begin{pmatrix} -iy \\ y \end{pmatrix} = \begin{pmatrix} i\overline{y} \\ \overline{y} \end{pmatrix} \cdot \begin{pmatrix} -iy \\ y \end{pmatrix} = 2|y|^2$.

So for any $y \in \mathbb{C}$ such that $|y| = \frac{1}{\sqrt{2}}$ the vector has length 1. E.g., we could choose $y = \frac{1}{\sqrt{2}}$.

↳ it does not matter which choice we make here

Then the normalized eigenvector is $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix}$

$\hookrightarrow \lambda_2 = 2: \begin{pmatrix} 1-2 & i \\ -i & 1-2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & i \\ -i & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow -x+iy=0$

$\Rightarrow \begin{pmatrix} iy \\ y \end{pmatrix}$ are the eigenvectors, i.e., $E_{\lambda_2}(A) = \mathrm{span}\left\{\begin{pmatrix} i \\ 1 \end{pmatrix}\right\}$.

A normalized eigenvector is $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix}$ (i.e., when $y = \frac{1}{\sqrt{2}}$).

We find that $\langle \vec{v}_1, \vec{v}_2 \rangle = \overline{\frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix}} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix} = \frac{1}{2}\left(\underbrace{i \cdot i}_{i^2=-1} + 1 \cdot 1\right) = 0,$

so $\vec{v}_1$ and $\vec{v}_2$ are indeed orthonormal.

B) Real Symmetric matrices

If we assume that all matrix entries are real, then the condition $\underbrace{A^{+}}_{\text{A dagger}} = A$ from before just becomes $\underbrace{A^{T}}_{\text{A transpose}} = A$. ← i.e., the matrix does not change if we interchange rows and columns.

This is exactly what we had when we discussed the Hessian.

$\boxed{\text{Definition: An } n\times n \text{ matrix } A \text{ is called symmetric if } A^T = A.}$

Note: • One could also consider complex symmetric matrices, but usually one is interested into real symmetric matrices (all entries are real).

As in A), we see that for real symmetric matrices, all eigenvalues are real, and they can be diagonalized with orthonormal eigenvectors.

Let us now combine our knowledge of eigenvalues and eigenvectors, in particular for symmetric matrices, with our knowledge of the Hessian of a function $f: \mathbb{R}^n \to \mathbb{R}$.

In Chapter 2.3, we answered the question of whether a critical point of $f:\mathbb{R}^n \to \mathbb{R}$ is a local maximum or minimum or saddle point only for $n=2$. But now we can answer this question for any n.

Recall the setting: For a function $f:\mathbb{R}^n \to \mathbb{R}$ a Taylor expansion gives

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + \overbrace{\underbrace{(\vec{\nabla} f)(\vec{a})}_{=\vec{0} \text{ at stationary points}} \cdot \vec{h}}^{=\langle \vec{\nabla} f(\vec{a}), \vec{h} \rangle} + \frac{1}{2} \underbrace{\vec{h} \cdot \left(H_f(\vec{a}) \vec{h}\right)}_{=\langle \vec{h}, H_f(\vec{a}) \vec{h} \rangle \text{ in the notation we used before}} + \text{Rest} \quad , \text{ with } \underbrace{\left(H_f(\vec{a})\right)_{ij}}_{\text{Hessian of } f \text{ at } \vec{a}} = \frac{\partial^2 f(\vec{a})}{\partial x_i \partial x_j}$$

We consider stationary points $\vec{a}$, i.e., $(\vec{\nabla} f)(\vec{a}) = \vec{0}$.

We conclude that:
- if $\langle \vec{h}, H_f(\vec{a}) \vec{h} \rangle > 0$ for all $\vec{h}$, then $\vec{a}$ is a local min.
- if $\langle \vec{h}, H_f(\vec{a}) \vec{h} \rangle < 0$ for all $\vec{h}$, then $\vec{a}$ is a local max.
- if $\langle \vec{h}, H_f(\vec{a}) \vec{h} \rangle > 0$ for some $\vec{h}$ and < 0 for others, then $\vec{a}$ is a saddle point
- if $\langle \vec{h}, H_f(\vec{a}) \vec{h} \rangle = 0$ for some $\vec{h}$, we need to look at higher order derivatives

Since $H_f(\vec{a})$ is real and symmetric, we know it has n real eigenvalues $\lambda_1, \dots, \lambda_n$ (in this notation some λ_i might be the same), and n $\underbrace{\text{orthonormal eigenvectors } \vec{v}_1, \dots, \vec{v}_n}_{\text{i.e., they form a basis}}$.

But then we could just write $\vec{h}$ in terms of the basis vectors $\vec{v}_1, \dots, \vec{v}_n$:

$$\vec{h} = \sum_{i=1}^{n} c_i \vec{v}_i \quad \text{for some coefficients } c_i .$$

$$H_f(\vec a)\vec v_i = \lambda_i \vec v_i$$

Then we find $\langle \vec h, H_f(\vec a)\vec h\rangle = \langle \vec h, H_f(\vec a)\sum_{i=1}^n c_i \vec v_i\rangle = \langle \vec h, \sum_{i=1}^n c_i \lambda_i \vec v_i\rangle$

$$= \langle \sum_{j=1}^n c_j \vec v_j, \sum_{i=1}^n c_i \lambda_i \vec v_i\rangle = \sum_{j=1}^n \sum_{i=1}^n \overline{c_j} c_i \lambda_i \underbrace{\langle \vec v_j, \vec v_i\rangle}_{=\begin{cases}0 & \text{if } i\neq j\\ 1 & \text{if } i=j\end{cases}}$$

this is the definition of orthonormal

$$= \sum_{i=1}^n |c_i|^2 \lambda_i$$

But $|c_i|^2$ is always ≥ 0. So the eigenvalues tell us whether the expression is always positive, always negative, sometimes positive sometimes negative, or sometimes zero!

We found:

$\boxed{\text{Theorem}}$

Theorem: Let $f: \mathbb{R}^n \to \mathbb{R}$ be of class C^2, and let $(\vec\nabla f)(\vec a) = \vec 0$ for some $\vec a \in \mathbb{R}^n$.
Let $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ denote the eigenvalues of the Hessian of f at $\vec a$. Then:

- if all $\lambda_i > 0$, then $\vec a$ is a local min,
- if all $\lambda_i < 0$, then $\vec a$ is a local max,
- if some $\lambda_i > 0$ and some $\lambda_i < 0$ (but none $= 0$), then $\vec a$ is a saddle point,
- if at least one of the $\lambda_i = 0$, then this test is inconclusive (can be either, we would need to consider higher derivatives).

Note: • In general, matrices A such that $\langle \vec x, A\vec x\rangle > 0 \ \forall \vec x \neq \vec 0$ are called positive definite (or negative definite if $\langle \vec x, A\vec x\rangle < 0 \ \forall \vec x \neq \vec 0$). As we saw above, for Hermitian or symmetric matrices this is the same as all eigenvalues being positive.

We could now also recover our result for $n=2$. In this case, let us write the Hessian as $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$, as we did in Chapter 2.3. In Homework 5, Problem 4, we computed that the two eigenvalues are $\lambda_\pm = \frac{\alpha+\gamma}{2} \pm \sqrt{\left(\frac{\alpha+\gamma}{2}\right)^2 - \alpha\gamma + \beta^2}$

$$\left(= \frac{\alpha+\gamma}{2} \pm \sqrt{\left(\frac{\alpha-\gamma}{2}\right)^2 + \beta^2} \right).$$

Thus:
- if $\beta^2 - \alpha\gamma < 0$ and $\alpha > 0$, then both $\lambda_+ > 0$ and $\lambda_- > 0$, $\Rightarrow$ local min.
  then also $\gamma > 0$
- if $\beta^2 - \alpha\gamma < 0$ and $\alpha < 0$, then both $\lambda_+ < 0$ and $\lambda_- < 0$, $\Rightarrow$ local max.
  then also $\gamma < 0$
- if $\beta^2 - \alpha\gamma > 0$, then $\lambda_+ > 0$, $\lambda_- < 0$, $\Rightarrow$ saddle point
- if $\beta^2 = \alpha\gamma$, then λ_+ or λ_- are zero. $\Rightarrow$ inconclusive

This is exactly what we found in Chapter 2.3.

More examples in the homework exercises.