1212: Linear Algebra II

Dr. Vladimir Dotsenko (Vlad)

Lecture 8

An application of Bessel's inequality

Recall that last time we proved Bessel's inequality: For any vector $v \in V$ and any orthonormal system e_1, ..., e_k (not necessarily a basis) we have

$$(v, v) \geqslant (v, e_1)^2 + \ldots + (v, e_k)^2.$$

Let us consider the Euclidean space of all continuous functions on $[-1, 1]$ with the scalar product

$$(f(t), g(t)) = \int_{-1}^{1} f(t)g(t)\, dt.$$

It is easy to see that the functions

$$e_0 = \frac{1}{\sqrt{2}}, e_1 = \cos \pi t, f_1 = \sin \pi t, \ldots, e_n = \cos \pi n t, f_n = \sin \pi n t$$

form an orthonormal system there. Consider the function $h(t) = t$. We have

$$(h(t), h(t)) = \frac{2}{3},$$
$$(h(t), e_0) = 0,$$
$$(h(t), e_k) = 0,$$
$$(h(t), f_k) = \frac{2(-1)^{k+1}}{k\pi},$$

(the latter integral requires integration by parts to compute it), so Bessel's inequality implies that

$$\frac{2}{3} \geqslant \frac{4}{\pi^2} + \frac{4}{4\pi^2} + \frac{4}{9\pi^2} + \ldots + \frac{4}{n^2\pi^2},$$

which can be rewritten as

$$\frac{\pi^2}{6} \geqslant 1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2}.$$

Actually $\sum\limits_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, which was first proved by Euler. We are not able to establish it here, but it is worth mentioning that Bessel's inequality gives a sharp bound for this sum.

Extremal properties of eigenvectors of symmetric matrices

Our next goal is to progress with linear transformations. Earlier, we mentioned that we have to use complex numbers as scalars to ensure that linear transformations have eigenvectors. Today, we shall make an exception, and explore the case when we can ensure existence of eigenvectors even for real scalars.

Definition 1. A matrix A is said to be symmetric if it is equal to its transpose: $A^T = A$.

Our goal today is to prove the following theorem:

Theorem 1. *Every symmetric matrix A with real coefficients has a real eigenvalue, and an a eigenvector with real coordinates.*

Proof. As usual in mathematics, we shall find that it is easier to prove a bit more than the theorem actually states. Let us view the given $n \times n$-matrix A as a linear transformation of $\mathbb{R}^n$, equip $\mathbb{R}^n$ with the standard scalar product, and consider the function g on $\mathbb{R}^n$ defined as

$$g(x) = (Ax, x).$$

If x has coordinates $x_1, \dots, x_n$, then

$$\begin{aligned}(Ax, x) = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)x_1 + \dots + (a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)x_n = \\ = a_{11}x_1^2 + 2a_{12}x_1x_2 + \dots + 2a_{1n}x_1x_n + a_{22}x_2^2 + 2a_{23}x_2x_3 + \dots + a_{nn}x_n^2.\end{aligned}$$

Let us consider the function $g(x)$ on the unit sphere $|x| = 1$. This function is continuous, and a continuous function reaches its maximal and minimal value on any compact (closed and bounded) set in $\mathbb{R}^n$. Therefore, for all x with $|x| = 1$ we have $m \leqslant g(x) \leqslant M$ for some m and M, and these inequalities become equalities for some x. Now, note that for $x \neq 0$ we have

$$g(x) = (Ax, x) = \left(A|x|\frac{1}{|x|}x, |x|\frac{1}{|x|}x\right) = |x|^2 \left(A\frac{1}{|x|}x, \frac{1}{|x|}x\right),$$

and the vector $\frac{1}{|x|}x$ is of length 1 for each $x \neq 0$, because $\left(\frac{1}{|x|}x, \frac{1}{|x|}x\right) = \frac{1}{|x|^2}(x, x) = 1$. Therefore, for each $x \neq 0$ we have

$$m(x, x) = m|x|^2 \leqslant g(x) \leqslant M|x|^2 = M(x, x),$$

and by inspection this holds for $x = 0$ also. In particular, this implies that

$$g(x) - M(x, x) \leqslant 0$$

for all x, so the values of x where $g(x) = M(x, x)$ are solutions to the local maximum problem for the function $f(x) = g(x) - M(x, x)$. This means that the gradient of $f(x)$ must be equal to zero. By examining the formula

$$\begin{aligned}f(x) &= (Ax, x) - M(x, x) = \\ &= (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)x_1 + \dots + (a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)x_n - M(x_1^2 + \dots + x_n^2) = \\ &= (a_{11} - M)x_1^2 + 2a_{12}x_1x_2 + \dots + 2a_{1n}x_1x_n + (a_{22} - M)x_2^2 + 2a_{23}x_2x_3 + \dots + (a_{nn} - M)x_n^2,\end{aligned}$$

we note that (recall that $a_{ij} = a_{ji}$ for all i, j)

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 2(a_{11} - M)x_1 + 2a_{12}x_2 + \dots + 2a_{1n}x_n, \\ \frac{\partial f}{\partial x_2} &= 2a_{12}x_1 + 2(a_{22} - M)x_2 + \dots + 2a_{2n}x_n, \\ &\dots \\ \frac{\partial f}{\partial x_n} &= 2a_{1n}x_1 + 2a_{2n}x_2 + \dots + 2(a_{nn} - M)x_n.\end{aligned}$$

Therefore, the gradient of f vanishes at x if and only if $(A - M \cdot I)x = 0$, that is x is an eigenvector of A with the eigenvalue M. □