1212: Linear Algebra II

Dr. Vladimir Dotsenko (Vlad)

Lecture 10

Orthogonal 3×3-matrices

We begin, as promised last time, with discussing a geometric way to view orthogonal 3×3-matrices.

Let us first consider an orthogonal 3×3-matrix A with $\det(A) = 1$. We begin with showing that A has an eigenvalue 1. We have

$$\chi_A(t) = \det(A - tI) = \det(A) + a_1 t + a_2 t^2 - t^3 = 1 + a_1 t + a_2 t^2 - t^3.$$

Since this is a polynomial of degree 3, it assumes values of opposite signs as $t \to \pm\infty$, so by the intermediate value theorem, it has a real root, so A has a real eigenvalue λ. Let v be such that $Av = \lambda v$. Then $(v, v) = (Av, Av) = (\lambda v, \lambda v) = \lambda^2(v, v)$, so $\lambda^2 = 1$. If $\lambda = 1$, we are done. Otherwise, A has an eigenvalue -1, and so $\chi_A(t) = (1+t)(1+at-t^2)$, where the polynomial $1+at-t^2$ must have real roots because it assumes a positive value at 0 and negative values as $t \to \pm\infty$. Therefore, all eigenvalues of A are real. If one of them is equal to 1, we are done; otherwise they are all equal to -1, and $\chi_A(t)$ is proportional to $(1+t)^3$, which is impossible (compare the leading coefficients and the constant terms).

If $Av = v$, then $U = \text{span}(v)^\perp$ is an invariant subspace. Indeed, $u \in U$ if and only if $(u, v) = 0$. If $u \in U$, we have

$$(Au, v) = (u, A^T v) = (u, A^{-1} v) = (u, v) = 0$$

since A is orthogonal and $Av = v$. On U, the transformation A induces an orthogonal linear transformation. Moreover, it is easy to see that the determinant of that linear transformation is equal to 1, so it is a 2D rotation. Consequently, in 3D the original matrix A represents a rotation about the line containing v.

If A is an orthogonal 3×3-matrix with determinant -1, then $-A$ is an orthogonal 3×3-matrix with determinant $(-1)^3 \det(A) = 1$, so every orthogonal matrix in 3D with determinant -1 is a rotation about some axis followed by a central symmetry about the origin.

General bilinear and quadratic forms: motivation

Right now, we are going to change our outlook temporarily, and examine symmetric matrices in a different context. While until now matrices represented linear transformations, we shall take the outlook which views symmetric matrices in the spirit of one of the proofs from last week, and interpret them via the associated *quadratic forms*. Let us start with a motivating example.

Example 1. Consider a function $f(x_1, \ldots, x_n)$ of n scalar real arguments. Let $x^0 = x_1^0 e_1 + \cdots + x_n^0 e_n$ be a point in $\mathbb{R}^n$, and assume that the function f is smooth enough to consider its Taylor series to order two near the point x^0:

$$f(x) = f(x^0) + \sum_{i=1}^{k} (x_i - x_i^0) \frac{\partial f}{\partial x_i}(x^0) + \frac{1}{2} \sum_{i,j=1}^{n} (x_i - x_i^0)(x_j - x_j^0) \frac{\partial^2 f}{\partial x_i x_j}(x^0) + o(|x - x^0|^2).$$

Suppose that we would like to know whether f attains its locally minimal/maximal value at x^0. Then, since in the first order of magnitude we have

$$f(x) = f(x^0) + \sum_{i=1}^{k} (x_i - x_i^0) \frac{\partial f}{\partial x_i}(x^0) + o(|x - x^0|),$$

we conclude that a necessary condition is $\frac{\partial f}{\partial x_i}(x^0) = 0$ for all i, that is the *gradient* of f vanishes at x^0. In this case, we have

$$f(x) = f(x^0) + \frac{1}{2} \sum_{i,j=1}^{n} (x_i - x_i^0)(x_j - x_j^0) \frac{\partial^2 f}{\partial x_i x_j}(x^0) + o(|x - x^0|^2),$$

so the difference between $f(x)$ and $f(x^0)$, when x is close to x^0, is approximately equal to $\frac{1}{2}q(x - x^0)$, where

$$q(y) = a_{11}y_1^2 + 2a_{12}y_1y_2 + \cdots + 2a_{1n}y_1y_n + y_2^2 + 2a_{23}y_2y_3 + \cdots + a_{nn}y_n^2,$$

where for brevity we denote $a_{ij} = \frac{\partial^2 f}{\partial x_i x_j}(x^0)$; we have $a_{ij} = a_{ji}$ whenever the function f is smooth enough. The function $q(y)$ is a very typical example of a *quadratic form*.

Definition 1. Let V be a vector space. A function $q\colon V \to \mathbb{R}$ is said to be a *quadratic form* if for some basis $e_1, \ldots, e_n$ of V we have

$$q(x_1e_1 + \cdots + x_ne_n) = \sum_{1 \leqslant i \leqslant j \leqslant n} a_{ij}x_ix_j,$$

that is values of q are quadratic polynomials in coordinates of a vector.

Remark 1. It is easy to see that if the condition from the definition holds for some basis, then it holds for any basis, since coordinates relative to different bases are related by transition matrices in a linear way.

One simple example of a quadratic form is

$$q(x) = x_1^2 + \cdots + x_n^2.$$

In general, if V is a Euclidean vector space then $q(x) = (x, x)$ is certainly a quadratic form. This can be generalised; every *bilinear form* gives rise to a quadratic form.

Definition 2. Let V be a vector space. A function $V \times V \to \mathbb{R}$, $v_1, v_2 \mapsto b(v_1, v_2)$ is called a bilinear form if for all vectors v, v_1, v_2 the following conditions are satisfied:

$$b(c_1v_1 + c_2v_2, v) = c_1b(v_1, v) + c_2b(v_2, v) \quad \text{and} \quad b(v, c_1v_1 + c_2v_2) = c_1b(v, v_1) + c_2b(v, v_2).$$