

Note 3: Properties of Linear Transformations

Math 198: Math for Machine Learning

1 Determinant & Trace

1.1 Determinant

There are multiple ways to define the *determinant* of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Geometrically, the determinant of \mathbf{A} , $\det(\mathbf{A})$, conveys how \mathbf{A} changes the volume of the unit cube. If $C = \{(c_1, \dots, c_n)^\top \in \mathbb{R}^n : c_i \in [0, 1]\}$ is the unit n -cube in \mathbb{R}^n , then $|\det(\mathbf{A})|$ is the volume of $\mathbf{A}(C) = \{\mathbf{A}x : x \in C\}$, and the sign of $\det(\mathbf{A})$ is positive if and only if \mathbf{A} preserves the orientation of C . You can think of $\det(\mathbf{A})$ as the factor by which the action of \mathbf{A} scales volume.

For computational purposes, we will formally define determinant in terms of minors. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. We define the determinant recursively as $\det(\mathbf{A}) = \sum_{j=1}^n (-1)^j a_{1j} \det(\mathbf{A}_{1j})$, where $\mathbf{A}_{ij} \in \mathbb{R}^{(n-1) \times (n-1)}$ represents \mathbf{A} with the i 'th row and j 'th column removed.

For example, let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Then

$$\begin{aligned} \det(\mathbf{A}) &= \det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \\ &= 1 \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2 \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3 \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1(5 \cdot 9 - 6 \cdot 8) - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) \\ &= 0. \end{aligned}$$

The determinant of this matrix is 0. If we view this from our geometric standpoint, then we conclude that \mathbf{A} sends the unit n -cube to a set with volume 0 in \mathbb{R}^3 , i.e. a flat parallelogram, a line segment, or 0. This picture tells us that \mathbf{A} flattens space, removing one or more dimensions. Thus, \mathbf{A} has a nontrivial kernel and is not injective. We conclude that \mathbf{A} is not an isomorphism. With this picture in hand, it is easy to see that the converse is true as well. **Claim:** \mathbf{A} is invertible if and only if $\det(\mathbf{A}) \neq 0$. This result is an extremely useful tool that we'll use when studying eigenvalues.

Now, let us list some properties of the determinant.

- (a) $\det(\mathbf{I}_n) = 1$.
- (b) $\det(\mathbf{A}^\top) = \det(\mathbf{A})$.
- (c) $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$.
- (d) $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ for $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$.¹

¹The set of $n \times n$ matrices with nonzero determinant form a group under matrix multiplication known as the *general linear group* of degree n , denoted $GL(n)$. The set of $n \times n$ matrices with determinant 1 form a normal subgroup of $GL(n)$ called the *special linear group* of degree n , denoted $SL(n)$. These groups are typical examples of *Lie groups*, smooth manifolds equipped with a group structure. They are important in physics.

- (e) $\det(\mathbf{a}_1, \dots, c\mathbf{x} + \mathbf{y}, \dots, \mathbf{a}_n) = c \det(\mathbf{a}_1, \dots, \mathbf{x}, \dots, \mathbf{a}_n) + \det(\mathbf{a}_1, \dots, \mathbf{y}, \dots, \mathbf{a}_n)$ for column vectors \mathbf{x}, \mathbf{y} .
- (f) If any row or column in \mathbf{A} is the zero vector $\mathbf{0} \in \mathbb{R}^n$, then $\det(\mathbf{A}) = 0$. Why is this obvious? Because it would imply that \mathbf{A} has nontrivial kernel, and thus is not invertible.

1.2 Trace

The *trace* of an $n \times n$ matrix \mathbf{A} , denoted $\text{tr}(\mathbf{A})$, is given by $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$, the sum of the entries on the diagonal of \mathbf{A} . Here are some properties of the trace that are easy to prove:

- (a) Trace is a linear map $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}$: $\text{tr}(c\mathbf{A} + \mathbf{B}) = c\text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.
- (b) Trace is preserved under taking the transpose: $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$.
- (c) Trace is preserved under *cyclical permutations* – that is, for square matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, $\text{tr}(\mathbf{ABCD}) = \text{tr}(\mathbf{BCDA}) = \text{tr}(\mathbf{CDAB}) = \text{tr}(\mathbf{DABC})$. However, trace is not preserved under arbitrary permutations.

Additionally, trace is **invariant under change of coordinates**. Mathematically, this means that for any invertible matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$, $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{BAB}^{-1})$.

To see what this means from an intuitive perspective, let \mathbf{A} be the matrix representation of a linear map T with respect to the standard basis. Let \mathcal{S} be any other basis for \mathbb{R}^n , and let \mathbf{A}' be the matrix representation of T with respect to \mathcal{S} . Then there exists an invertible matrix \mathbf{B} such that $\mathbf{A}' = \mathbf{BAB}^{-1}$. The claim here is that $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}')$. It doesn't matter with respect to which basis you represent T ; all the corresponding matrices will have the same trace. Trace is an intrinsic property of a linear map that is the same under every coordinate system. Thus, we can define the trace of a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ to be the trace of a corresponding matrix \mathbf{A} with respect to any basis.

Some notation: If $\mathbf{A}' = \mathbf{BAB}^{-1}$ for some invertible \mathbf{B} , then we say that \mathbf{A} and \mathbf{A}' are *similar*. Similarity induces an equivalence class on $\mathbb{R}^{n \times n}$.

Trace and rank. If \mathbf{I}_n is the identity matrix, then clearly $\text{tr}(\mathbf{I}_n) = n$. This hints at a relationship between $\text{tr}(\mathbf{A})$ and $\text{rank}(\mathbf{A})$. This relationship becomes explicit when \mathbf{A} is a (not necessarily orthogonal) projection matrix, i.e. when $\mathbf{A}^2 = \mathbf{A}$. Such a matrix is called *idempotent*. Claim: If $\mathbf{A}^2 = \mathbf{A}$, then $\text{tr}(\mathbf{A}) = \text{rank}(\mathbf{A})$. The proof of this fact is easy and relies on eigenvalues, which we define below.

A quick aside about the geometric meaning of trace. Recall that the $\det(\mathbf{A})$ encodes how \mathbf{A} scales the volume of the unit n -cube. We can also view trace from this perspective. Consider this excerpt from V. I. Arnold's "Ordinary Differential Equations":

Suppose small changes are made in the edges of a parallelepiped. Then the main contribution to the change in volume of the parallelepiped is due to the change of each edge in its own direction, changes in the direction of the other edges making only a second-order contribution to the change in volume.

The "change of each edge in its own direction" is determined by the values on the main diagonal of \mathbf{A} and is thus encoded in the trace of \mathbf{A} .

Formally, this fact is best expressed as a vector ordinary differential equation: $\mathbf{y}'(t) = \mathbf{A}\mathbf{y}(t)$, where $\mathbf{y} : \mathbb{R} \rightarrow \mathbb{R}^n$ is a vector-valued function. The situation encoded here is as follows: Suppose at time $t = 0$, we begin with a unit n -cube. In accordance with the differential equation, observe how the unit cube changes in time as \mathbf{A} acts on it. The solution to this ODE is $\mathbf{y}(t) = \exp(t\mathbf{A})\mathbf{y}(0)$. Thus, the volume of the transformed unit n -cube at time t is

$$\det(\exp(t\mathbf{A})) = 1 + t \text{tr}(\mathbf{A}) + o(t^2).$$

As we can see, the change in volume from time $t = 0$ to time $t = t_0$ depends linearly on $\text{tr}(\mathbf{A})$. Everything else is second order and can be ignored.

2 Eigenvalues & Eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\lambda \in \mathbb{R}$ is an *eigenvalue* of \mathbf{A} if $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ has a nontrivial solution $\mathbf{v} \in \mathbb{R}^n$. A nonzero vector \mathbf{v} satisfying $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ is called an *eigenvector* of A corresponding to λ .² A matrix acts on its eigenvectors simply by scaling them (potentially by a negative eigenvalue, causing a flip).

Examples.

- (a) If $\mathbf{A} = \mathbf{I}_n$, then its only eigenvalue is 1, and every nontrivial vector is a corresponding eigenvector.
- (b) If \mathbf{A} is the zero matrix, then its only eigenvalue is 0, and every nontrivial vector is a corresponding eigenvector.
- (c) If \mathbf{A} is diagonal with diagonal elements (a_i) , then for each $i = 1, \dots, n$, we have that \mathbf{e}_i is an eigenvector corresponding to eigenvalue a_i .
- (d) If $V = C^1(\mathbb{R})$, the vector space of continuously differentiable real-valued functions on \mathbb{R} , and $T : V \rightarrow V$ is the differential operator given by $f \mapsto f'$, then the exponential function $t \mapsto e^{\lambda t}$ is an eigenvector of T corresponding to eigenvalue λ . This is a big reason that the exponential function is so important. Note that C^1 is infinite dimensional and not isomorphic to \mathbb{R}^n , so we can't represent T with a finite matrix.

Computing eigenvalues and eigenvectors. How do we compute eigenvalues and eigenvectors? We start with eigenvalues. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Note that

$$\begin{aligned}\lambda \text{ is an eigenvalue of } \mathbf{A} &\iff \text{there exists nonzero } \mathbf{v} \text{ such that } \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \\ &\iff \mathbf{A} - \lambda\mathbf{I} \text{ has a nontrivial kernel} \\ &\iff \mathbf{A} - \lambda\mathbf{I} \text{ is not invertible} \\ &\iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0.\end{aligned}$$

The function $p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ is known as the *characteristic polynomial* of \mathbf{A} . Note that it is a polynomial in λ and that its roots are the eigenvalues of \mathbf{A} . To compute the eigenvalues of \mathbf{A} , simply find the roots of its characteristic equation.

To find eigenvectors corresponding to the eigenvalues we've found, note that

$$\begin{aligned}\mathbf{v} \text{ is an eigenvector of } \mathbf{A} \text{ corresponding to } \lambda &\iff \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \\ &\iff (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0 \\ &\iff \mathbf{v} \in \ker(\mathbf{A} - \lambda\mathbf{I}).\end{aligned}$$

Thus the set of eigenvectors corresponding to λ is $\ker(\mathbf{A} - \lambda\mathbf{I})$. It follows that the set of eigenvectors is a subspace of \mathbb{R}^n , which we will denote E_{λ} and call the *eigenspace* corresponding to λ .

To sum up, the algorithm for finding eigenvalues and corresponding eigenvectors is as follows:

1. Find $p_{\mathbf{A}}(\lambda)$, find its roots.
2. For each root λ , compute (a basis for) the eigenspace $\ker(\mathbf{A} - \lambda\mathbf{I})$.

Facts about eigenvalues. Let $A \in \mathbb{R}^{n \times n}$.

- (a) Trace is the sum of the eigenvalues (with algebraic multiplicity)
- (b) Determinant is the product of the eigenvalues (with algebraic multiplicity)
- (c) The eigenvalues of \mathbf{A}^k are λ_i^k , where the λ_i are eigenvalues of \mathbf{A} .
- (d) If \mathbf{A}^{-1} exists, then its eigenvalues are $\frac{1}{\lambda_i}$.
- (e) If $\mathbf{A} = \mathbf{A}^{\top}$, then all its eigenvalues are real.

²In functional analysis/operator theory, the set of eigenvalues of a linear operator a is called the *spectrum* of a , hence the prevalence of the word "spectral" in the study of eigenvalues.

3 Diagonalization & Spectral Theorem for Real Symmetric Matrices

Why do we care about eigenvectors and eigenvalues of \mathbf{A} ? Because they reveal key information about the action of \mathbf{A} . Eigen-stuff allows you to understand a linear transformation in terms of the simplest type of linear action: scaling. Eigenvalues are sometimes called *characteristic values* because they represent the fundamental, 1-dimensional action of \mathbf{A} .

If \mathbf{A} is a square matrix and there exists a basis for \mathbb{R}^n consisting of eigenvectors (a.k.a. an eigenbasis), then we can *diagonalize* \mathbf{A} like so: Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be n linearly independent eigenvectors of \mathbf{A} corresponding to $\lambda_1, \dots, \lambda_n$. Note that the λ_i could be equal to one another. Let

$$\mathbf{Q} = (\mathbf{x}_1 \quad \dots \quad \mathbf{x}_n) \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ & & \ddots & & \\ 0 & \dots & 0 & 0 & \lambda_n \end{pmatrix}.$$

Then $\mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{\Lambda}$, so we can write $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$. This decomposition of \mathbf{A} is called an *eigendecomposition*. We have decomposed \mathbf{A} into three simpler maps: First, we change coordinates via the isomorphism \mathbf{Q}^{-1} . Then, we act via the diagonal matrix $\mathbf{\Lambda}$, which simply scales each coordinate by its eigenvalue. Finally, we change back to our original basis via \mathbf{Q} .

It turns out that if $\mathbf{A}^\top = \mathbf{A}$, i.e. \mathbf{A} is *symmetric*, then there are always n linearly independent eigenvectors with which we can diagonalize \mathbf{A} . In fact, the claim is even stronger.

Spectral Theorem for Real Symmetric Matrices. If $\mathbf{A} \in \mathbb{R}^{n \times n}$ and \mathbf{A} is symmetric, then there exists an orthonormal basis for \mathbb{R}^n consisting of eigenvectors of \mathbf{A} . In this case, \mathbf{A} is said to be *unitarily diagonalizable*.

We will explore this theorem in the next lecture, but if you want to read ahead, check out this resource: <http://www.math.lsa.umich.edu/~speyer/417/SpectralTheorem.pdf>.