CSci 4270 and 6270
Computational Vision, Spring 2025
Lecture 3: Linear Algebra
January 13, 2025

Overview

- Points and vectors
- Matrices and matrix operations
- Eigenvalues and eigenvectors
- Decompositions
- NumPy implementations throughout
  - We'll look at both NumPy itself and its linear algebra module through examples written in class.

These notes are a severely condensed summary / introduction. We will not go into them in great detail, but will refer back to them frequently.

Points

- A point is a location in n-dimensional space.
- A point can be described by a tuple of scalars $(x_1, \ldots, x_n)$ relative to a given coordinate system. This tuple is a **vector**.
- Euclidean n space $\mathbb{R}^n$ is the set all of possible point location vectors together with a distance metric on the points.
- The distance between two points $\mathbf{x} = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ is

$$\sqrt{(x_1 - y_1)^2 + \ldots (x_n - y_n)^2} = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{1/2} \tag{1}$$

Vectors

- Vectors are used to describe point locations and directions. The geometric intention of what is meant will be determined by the context.

- Vectors have an algebra, including a series of important properties. What follows is a summary.

- The length, magnitude or norm of a vector $\mathbf{x} = (x_1, \ldots, x_n)$ is

$$\|\mathbf{x}\| = (x_1^2 + \ldots + x_n^2)^{1/2} = \left[\sum_{i=1}^{n} x_i^2\right]^{1/2}. \tag{2}$$

- When a vector $\mathbf{x}$ is multiplied by a scalar, c, the result is a vector with each of its components scaled individually.

$$c\mathbf{x} = (cx_1, \ldots, cx_n). \tag{3}$$

- The *sum* of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is another vector formed by adding each of the respective components.

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{4}$$

  We will look at the geometric interpretation of this.

- The *dot* or *scalar* product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is a scalar.

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i. \tag{5}$$

- In $\mathbb{R}^3$, the cross-product of two vectors is another vector:

$$\mathbf{x} \times \mathbf{y} = (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1). \tag{6}$$

- Note that $\|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$

- A vector $\mathbf{x}$ is a unit vector if $\|\mathbf{x}\| = 1$

- We use a special notation for a unit vector: $\hat{\mathbf{x}}$.

- A vector can be converted to a unit vector — just divide by the magnitude:

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{7}$$

- The angle between two vectors is found from the dot product:

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}. \tag{8}$$

- We can decide if two vectors are perpendicular, parallel, or antiparallel by looking at the dot product between the vectors. These follow from the definition of the angle between two vectors.

- The direction between two points is the vector joining them, i.e. $\mathbf{x} - \mathbf{y}$. It is unique only up to a scale multiple.
- Vectors can be projected onto each other. In particular, if $\hat{\mathbf{y}}$ is a unit vector then
$$(\mathbf{x} \cdot \hat{\mathbf{y}})\hat{\mathbf{y}} \tag{9}$$
is a vector in direction $\hat{\mathbf{y}}$.
- Given a vector $\mathbf{x}$ and a unit vector $\hat{\mathbf{y}}$, $\mathbf{x}$ can be decomposed into two vectors, one parallel to $\hat{\mathbf{y}}$ and one perpendicular to $\hat{\mathbf{y}}$

Matrices

- Matrices are rectangular arrays of numbers, with each number subscripted by two indices:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{10}$$

- A short-hand notation for this is

$$\mathbf{A} = (a_{ij}). \tag{11}$$

- The dimensions of this matrix are written $m \times n$ for m rows and n columns.

- Matrices will be denoted with bold capital letters, generally taken from the beginning and middle of the alphabet, e.g.

$$\mathbf{A}, \mathbf{K}, \mathbf{M}. \tag{12}$$

- The $n \times n$ identity matrix, denoted by $\mathbf{I}_n$ or $\mathbf{I}_{n\times n}$ when the the value of n must be made clear, is a square matrix such that:

$$\mathbf{I}_n = (\delta_{ij}) \tag{13}$$

  where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{14}$$

  is the Kronecker delta.

- Vectors will generally be treated as column vectors ($m \times 1$ matrices) instead of row vectors ($1 \times n$ matrices).

Matrix Operations — Transpose and Symmetry

- The *transpose* of a matrix is obtained by interchanging its rows and columns. It is denoted with the superscript T, as in

$$\mathbf{A}^\top. \tag{15}$$

- Note that $(\mathbf{A}^\top)^\top = \mathbf{A}$.
- When $\mathbf{A}^\top = \mathbf{A}$, then $\mathbf{A}$ is said to be *symmetric*.
- If

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

  is a column vector, then we can write

$$\mathbf{x}^\top = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}$$

  to describe the row vector formed from the components of $\mathbf{x}$.

Matrix Operations — Addition and Multiplication

- The *sum* of two $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ is simply

$$(a_{ij} + b_{ij}). \tag{16}$$

- The product of a constant c and a matrix $\mathbf{A}$ is

$$c\mathbf{A} = \mathbf{A}c = (ca_{ij}). \tag{17}$$

- The *product* of an $m \times n$ matrix $\mathbf{A}$ and an $n \times p$ matrix $\mathbf{B}$ is an $m \times p$ matrix:

$$\mathbf{AB} = \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right) \tag{18}$$

- Matrix multiplication is associative but not commutative. It distributes over matrix addition (for appropriately sized matrices):

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \tag{19}$$

- Note in general that

$$(\mathbf{AB})^{\top} = \mathbf{B}^{\top}\mathbf{A}^{\top}. \tag{20}$$

- **Hint:** Whenever doing matrix multiplication (as part of a larger set of matrix manipulations), use the fact that the number of columns in the first matrix must equal the number of rows of the second matrix as a sanity check.

More on Matrix Multiplication

- In matrix notation, the dot product of two vectors may be written as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \mathbf{v}^\top \mathbf{u}. \tag{21}$$

We will use this fact repeatedly.

- The *outer product* of an $m \times 1$ vector $\mathbf{u}$ and a $n \times 1$ vector $\mathbf{v}$ is a $m \times n$ matrix:

$$\mathbf{u}\mathbf{v}^\top. \tag{22}$$

Note that the transpose operation has higher precedence than the matrix multiplication.

- The squared-magnitude of a vector $\mathbf{x}$ is

$$\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x} \tag{23}$$

- The multiplication of a $m \times n$ matrix $\mathbf{A}$ times a $n \times 1$ vector $\mathbf{b}$ may be expressed concisely when $\mathbf{A}$ is written in terms of a stack of row vectors:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \cdots \\ \mathbf{a}_m^\top \end{pmatrix} \tag{24}$$

where each $\mathbf{a}_i^\top$ contains n entries. Then

$$\mathbf{A}\mathbf{b} = \begin{pmatrix} \mathbf{a}_1^\top \mathbf{b} \\ \mathbf{a}_2^\top \mathbf{b} \\ \cdots \\ \mathbf{a}_m^\top \mathbf{b} \end{pmatrix} \tag{25}$$

- The multiplication of two matrices written in block form:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \tag{26}$$

may be written succinctly as

$$\mathbf{A}\mathbf{B} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \tag{27}$$

provided the necessary dimensions work out.

Matrix Operations — Trace and Determinants

These operations are all applied to *square* matrices, i.e. matrices of dimension $n \times n$ for integer $n \geq 1$:

- The *trace* of a matrix $\mathbf{A}$ is the sum of the elements on its main diagonal.
- The *determinant* of a matrix, $\mathbf{A}$, is written

$$\det(\mathbf{A}) \qquad \text{or} \qquad |\,\mathbf{A}\,|\,. \tag{28}$$

  Although this is not the way it is commonly defined in linear algebra texts, the definition I prefer to give is recursive:

  - As a preliminary, let $\mathbf{M}_{ij}$ be the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ by deleting row i and column j.
  - When $n = 1$,
$$\det(\mathbf{A}) = a_{11}. \tag{29}$$
  - When $n = 2$,
$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}. \tag{30}$$
  - For $n \geq 2$, choose any row i of $\mathbf{A}$. Then,
$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{M}_{ij}). \tag{31}$$
    Alternatively, we can choose any column j of $\mathbf{A}$:
$$\det(\mathbf{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{M}_{ij}). \tag{32}$$

- Properties of the determinant:
  - The determinant changes sign when two rows (or columns) are interchanged; it is 0 when two rows (or two columns) are repeated.
  - $\det(\mathbf{I}) = 1$.
  - $\det(\mathbf{AB}) = \det(\mathbf{A})\ \det(\mathbf{B})$.
  - $\det(\mathbf{A}^\top) = \det(\mathbf{A})$.

Matrix Operations — Inverse

- For matrix $\mathbf{A}$, a matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{I} \tag{33}$$

is said to be the *right inverse* of $\mathbf{A}$. A matrix $\mathbf{C}$ such that

$$\mathbf{CA} = \mathbf{I} \tag{34}$$

is said to be the *left inverse* of $\mathbf{A}$.

- For a square matrix $\mathbf{A}$, the left inverse exists if and only if the right inverse exists. In this case, the two inverses are equal and $\mathbf{A}$ is said to be *invertible*. The inverse is generally denoted

$$\mathbf{A}^{-1}. \tag{35}$$

- A matrix that is not invertible is called *singular*.
- Important properties:
  - $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$.
  - If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$ and both invertible, then

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{36}$$

  - $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.
- The inverse of a 2x2 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

if it exists, is

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \tag{37}$$

Special Forms of Square Matrices

- A *diagonal matrix* $\mathbf{A}$ is a square matrix where $a_{ij} = 0$ if $i \neq j$. Diagonal matrices are often written using the notation:

$$\mathrm{diag}(a_{11}, \ldots, a_{nn}) \qquad \text{or} \qquad \mathrm{diag}(a_1, \ldots, a_n) \tag{38}$$

- A *lower triangular matrix* $\mathbf{A}$ has $a_{ij} = 0$ whenever $i < j$.
- An *upper triangular matrix* $\mathbf{A}$ has $a_{ij} = 0$ whenever $i > j$.
- Products of diagonal matrices are diagonal. Products of upper triangular matrices are upper triangular. Products of lower triangular matrices are lower triangular.
- Determinant calculations are trivial for these special forms:

$$\det \mathbf{A} = \prod_{i=1}^{n} a_{ii}. \tag{39}$$

- Square matrix $\mathbf{A}$ is orthogonal if its inverse equals its transpose:

$$\mathbf{A}^{-1} = \mathbf{A}^{\top}$$

  Orthogonal matrices have determinants that are either 1 or -1.

  - When the determinant is 1 and all the entries in the matrix are real numbers, it is called a "rotation matrix".

Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are only defined for square matrices.

- The vector $\mathbf{x}$ is an *eigenvector* of matrix $\mathbf{A}$ and the scalar λ is an *eigenvalue* of $\mathbf{A}$ if
$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$
- We can show that equivalently, the scalar λ is an eigenvalue of $\mathbf{A}$ if
$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$
(This is the *characteristic equation* for $\mathbf{A}$.) and $\mathbf{x}$ is an eigenvector of $\mathbf{x}$ if $\mathbf{x}$ is in the nullspace of $\mathbf{A} - \lambda\mathbf{I}$.
- The rank of a $n \times n$ square matrix $\mathbf{A}$ is the number of non-zero eigenvalues.
    - If the rank of $\mathbf{A}$ is less than n, we say that $\mathbf{A}$ is *rank deficient*.

  Note (as a technicality) that an eigenvalue may be appear as multiple roots of the characteristic equation — in this case it is counted multiple times as a non-zero eigenvalue.

Example: Rotation

Consider the following matrix — a rotation matrix in three dimensions:

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- You can easily check that its determinant is 1 and that its transpose is its inverse.
- This matrix has 1 real and 2 complex eigenvalues:
  - The real eigenvalue is 1
  - The complex eigenvalues, which always come in "conjugate pairs," are
    $$e^{i\theta} = \cos\theta \pm i \sin\theta$$
- The eigenvector corresponding to the real eigenvalue is $(0, 0, 1)^T$, which is the z-axis
- Points along this axis are unchanged by the rotation, and therefore these are "fixed points" of the rotation.
- All rotations in $\mathbb{R}^3$ have this same structure, i.e.
  - Two complex eigenvalues that give the angle of rotation
  - A real eigenvalue, always 1, and its associated eigenvector is the axis of rotation

  We will reconsider this in Lecture 4.

Properties of Eigenvalues and Eigenvectors

- Suppose $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\mathbf{A}$, then

$$\begin{aligned} \sum \lambda_i &= \text{trace}(\mathbf{A}), \qquad \text{and} \\ \prod \lambda_i &= \det(\mathbf{A}). \end{aligned}$$

- For a triangular matrix, the eigenvalues are the diagonal elements.
- $\mathbf{A}$ and $\mathbf{A}^2$ have the same eigenvectors and if λ_i is an eigenvalue of $\mathbf{A}$, then λ_i^2 is an eigenvalue of $\mathbf{A}^2$.
- If $\mathbf{A} = \mathbf{M}^{-1}\mathbf{B}\mathbf{M}$ for invertible matrix $\mathbf{M}$, then $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues, but not the same eigenvectors.

Computing Eigenvalues and Eigenvectors

- There is no simple form for computing eigenvalues and eigenvectors because there is no expression for the roots of a quintic or higher polynomial.

- Numerical methods are generally used, and we will just make use of eigenvalue / eigenvector solver functions in NumPy without studying them.

Symmetric Matrices and Spectral (Eigenvalue) Decompositions

- The eigenvalues of symmetric matrices are all real, though not necessarily positive.
- The eigenvectors corresponding to distinct eigenvalues of a symmetric matrix are orthogonal.
- Any symmetric $n \times n$ matrix $\mathbf{A}$ can be written as

  $$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T,$$

  where $\mathbf{V}$ is an orthogonal matrix whose columns are normalized eigenvectors $\mathbf{v}_i$ of $\mathbf{A}$, and $\mathbf{D}$ is a diagonal matrix containing the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$.
- It follows that the eigenvectors of $\mathbf{A}$ are orthogonal to each other ("orthonormal")

Singular Value Decomposition — SVD

- Any $m \times n$ matrix $\mathbf{A}$ — notice not necessarily square — of real values can be decomposed

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

- Where
  - $\mathbf{U}$ is $m \times m$ and contains orthonormal columns,
  - $\mathbf{S}$ is $m \times n$, diagonal, and contains decreasing, non-negative entries along the diagonal, and
  - $\mathbf{V}$ is $n \times n$ and contains orthonormal columns.
- In class we will look at the relationship between the SVD and the eigenvalue decomposition.
- The diagonal entries of $\mathbf{S}$ are referred to as *singular values*. Note that there are $\min(m, n)$ singular values.
- In the frequent case where $m > n$, some definitions of the SVD truncate $\mathbf{U}$ and $\mathbf{S}$
  - $\mathbf{U}$ is $m \times n$ and contains orthonormal columns, and
  - $\mathbf{S}$ is $n \times n$.

  everything else is the same.

Projections and Principle Component Analysis

- Consider a set of n m-dimensional data point vectors, $\{\mathbf{x}_i\}, i = 1, \dots, n$.
  - We'll draw pictures in class.
- To summarize this data, especially when each vector is high dimensional (m is large), we might like to know the directions along which the data varies most and only consider the variation along this direction.
- Let $\hat{\mathbf{v}}$ be a potential direction, and consider the projection of $\mathbf{x}_i$ onto $\hat{\mathbf{v}}$. The square magnitude of this projection is

$$\|\mathbf{x}_i^\top \hat{\mathbf{v}}\|^2 = \hat{\mathbf{v}}^\top \mathbf{x}_i \mathbf{x}_i^\top \hat{\mathbf{v}}.$$

- Summing across all n points gives the summed square magnitude of this projection onto (candidate) direction $\hat{\mathbf{v}}$:

$$\begin{aligned}\sum_{i=1}^{n} \hat{\mathbf{v}}^\top \mathbf{x}_i \mathbf{x}_i^\top \hat{\mathbf{v}} &= \hat{\mathbf{v}}^\top \big(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top\big) \hat{\mathbf{v}} \\ &= \hat{\mathbf{v}}^\top \mathbf{M} \hat{\mathbf{v}}\end{aligned}$$

- It can be shown that all eigenvalues of $\mathbf{M}$ are real and non-negative.
- The directions $\hat{\mathbf{v}}$ for which the data vary the most turn about to be the eigenvectors of $\mathbf{M}$ corresponding to the greatest eigenvalues of $\mathbf{M}$.
  - We will prove a simple version of this claim in Lecture 5.
- When we order the eigenvectors by *decreasing eigenvalue* $(\lambda_1, \hat{\mathbf{v}}_1), (\lambda_2, \hat{\mathbf{v}}_2), \dots, (\lambda_m, \hat{\mathbf{v}}_m)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ we have the *principle components* of $\mathbf{M}$. In other words,
  - $\hat{\mathbf{v}}_1$ is the first principle component,
  - $\hat{\mathbf{v}}_2$ is the second principle component,
  - Etc.
- For high-dimensional point vectors we usually only consider the first k vectors as the k principle components, for relatively small value of $k << n$,
- If we project each $\mathbf{x}_i$ value onto these k principle component vectors, we get a much shorter vector:

$$\mathbf{x}_i' = \begin{pmatrix} \hat{\mathbf{v}}_1^\top \mathbf{x}_i \\ \hat{\mathbf{v}}_2^\top \mathbf{x}_i \\ \dots \\ \hat{\mathbf{v}}_k^\top \mathbf{x}_i \end{pmatrix},$$

or

$$\mathbf{x}_i' = \mathbf{V}^\top \mathbf{x}_i,$$

where

$$\mathbf{V} = \big(\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_k\big).$$

- If the values of λ_j for $j > k$ are relatively small, then we have a much shorter description of the data without losing much information.

Avoiding Forming the Huge Matrix, $\mathbf{M}$

- In high-dimensional space (e.g. 10K values in each vector), where $m >> n$, we don't want the expense of forming $m \times m$ matrix $\mathbf{M}$.

- To avoid this, we form an $m \times n$ matrix from the $\mathbf{x}_i$ values

  $$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{pmatrix}.$$

  Then, playing around a bit with matrix multiplication (it is good practice), you can see that

  $$\mathbf{M} = \mathbf{X}\mathbf{X}^\top,$$

  but we never outright calculate $\mathbf{M}$.

- Instead we can calculate the first k principle components of the data directly from $\mathbf{X}$ using either the SVD, defined above, or the Karhunen-Loeve transform (which we will not define in these notes).

Summary

- There is a lot of information here, but I'm hoping much of it is review.
- In class and in the homework, we will practice with the calculations and computations to get a feel for the ideas.
- Throughout the semester, when you have questions about linear algebra, please refer back to this printed handout and and the notes you have taken. You will find most of the information is here.