Linear Algebra for Structural Bioinformatics

Motivation

- Structural bioinformatics will require mathematical tools to handle:
  - Visualization of molecules
  - Geometry of macromolecules
    - In this course we concentrate on proteins.
  - Coordinate transformations
    - Eg.: rotation and translation of a molecule in $\mathbb{R}^3$.
  - Machine learning that uses vector data.

    - This is a partial list…

Introduction

- We will assume that you know Math136:
  - Systems of linear equations, matrix algebra, elementary matrices, computational issues, real and complex n-space, vector spaces and subspaces, basis and dimension, rank of a matrix, linear transformations and their matrix representations, inner products, angles and orthogonality.

- Beyond this we will need:
  - Orthogonal matrices and transformations, eigenvalues, eigenvectors, orthogonal diagonalization, and singular value decomposition.

Review of Vector Spaces

- By a vector space V we mean a nonempty set V with two mapping operations:
  - $(x, y) \mapsto x + y$ ***addition*** mapping $V \times V$ into V and
  - $(\lambda, x) \mapsto \lambda x$ ***multiplication by a scalar*** mapping $\mathbb{R} \times V$ into V with the following rules:
    - Commutativity: $x + y = y + x$
    - Associativity: $(x + y) + z = x + (y + z)$
    - Additive identity: $\exists\, 0 \in V \ni x + 0 = x$
    - Additive inverse: $\forall\ x \in V\ \exists\, \text{-}x \in V \ni x + (-x) = 0$
    - Scalar associativity: $\alpha(\beta x) = (\alpha\beta)x$
    - Distributivity across scalar sums: $(\alpha + \beta)x = \alpha x + \beta x$
    - Distributivity across vector sums: $\alpha(x + y) = \alpha x + \alpha y$
    - Scalar identity: $1x = x$
    - Scalar zero: $0x = 0$

Norms

- A norm is a function $\|\cdot\|: V \to \mathbb{R}_0^+$.
  - That is, it maps a vector $v \in V$ to a nonnegative real number.
  - For any $u, v \in V$ and $\alpha \in \mathbb{R}$, a norm must have the following properties:
    - $\|v\| > 0 \text{ if } v \neq 0$
    - $\|\alpha v\| = |\alpha| \|v\|$ and
    - $\|u + v\| \leq \|u\| + \|v\|$.
  - In general, once we have a norm, we can define a metric as $d(x, y) = \|x - y\|$.

Inner Product Review

- For a vector space V, an inner product satisfies the following rules:
  - $\langle u, v \rangle = \langle v, u \rangle$
  - $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
  - $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$
  - $\langle v, v \rangle \geq 0 \text{ and } \langle v, v \rangle = 0 \Leftrightarrow v = 0$

Note that the first two equations grant us:

$$\langle w, u + v \rangle = \langle w, u \rangle + \langle w, v \rangle$$

We can also derive:

$$\langle 0, v \rangle = \langle v, 0 \rangle = 0 \quad \text{and}$$

$$\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$$

Cauchy-Schwarz Inequality

- For all $u, v \in V$:

$$\langle u,v \rangle^2 \le \langle u,u \rangle \langle v,v \rangle$$

Minkowski's Inequality

- For all $u, v \in V$:

$$\sqrt{\langle u+v, u+v \rangle} \le \sqrt{\langle u,u \rangle} + \sqrt{\langle v,v \rangle}$$

Induced Norms

- If we have an inner product defined for a vector space then we can define a norm for the space by using: $\|v\| := \sqrt{\langle v,v \rangle}$
  - This is usually described by saying that the inner product *induces* the norm.
  - We can rewrite the Cauchy-Schwarz and Minkowski inequalities as:

$\langle u,v \rangle \le \|u\| \, \|v\|$ and $\|u+v\| \le \|u\| + \|v\|$.

Triangle inequality.

Metric Spaces

- A **metric space** $\boldsymbol{M}$ is a set of points with a function *d* that maps any pair of points in M to a nonnegative real value that can be regarded as the distance between the two points.
  - More formally, the metric space is a pair (M, d) where $d: M \times M \to [0, \infty)$ and the points in M can be almost any mathematical object as long as the mapping satisfies the metric properties listed as follows:

1. $0 \le d(x, y) < \infty \quad \text{for all pairs } x, y \in M,$
2. $d(x, y) = 0 \quad \text{if and only if } x = y,$
3. $d(x, y) = d(y, x) \quad \text{for all pairs } x, y \in M,$
4. $d(x, y) \le d(x, z) + d(z, y) \quad \text{for all } x, y, z \in M,$

Inner product, Norm and Metric for Column Vectors

- The above discussion deals with the abstract definitions of inner product, norm, and metric.
  - We now focus on an n-dimensional Euclidean space $\mathbb{R}^n$ as the vector space.
  - Recall that a typical vector is represented as the column vector $v = [v_1, v_2, \ldots, v_n]^{\mathrm{T}}$.
  - The inner product is:

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i = u^{\mathrm{T}} v.$$

Inner product, Norm and Metric for Column Vectors (cont.)

- The induced norm is: $$\|v\| := \sqrt{\langle v, v \rangle} = \sqrt{\sum_{i=1}^{n} v_i^2}.$$
- The metric for $\mathbb{R}^n$ is: $$d(u,v) := \|u - v\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}.$$
- Cauchy-Schwarz and Minkowski inequalities:

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \le \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right)$$

$$\sqrt{\sum_{i=1}^{n} (u_i + v_i)^2} \le \sqrt{\sum_{i=1}^{n} u_i^2} + \sqrt{\sum_{i=1}^{n} v_i^2}.$$

Eigenvectors and Eigenvalues

- Suppose A is a square $n \times n$ matrix.
- The eigenvector equation is:

$$Av = \mu v$$

where vector v is called the eigenvector and the constant μ is called the eigenvalue.

  - The eigenvectors of A define those directions in which the effect of A on v is simply to change the length of v while leaving its direction unchanged: Multiplying an eigenvector v of A by A does not rotate v.
    - If v is an eigenvector of A then any multiple of that vector will also satisfy the equation. This implies, we can normalize v so that $\|v\| = 1$.

An Equation for Eigenvalues

- The eigenvector equation can be rewritten as:

$$(A-\mu I)v=0.$$

  - Matrix theory tells us that this equation has nonzero solution v if and only if $A-\mu I$ is a singular matrix, that is, when $\det(A-\mu I)=0.$
    - For an n by n matrix A, the left side of the last equation represents a polynomial of degree n in μ.
    - The roots of this polynomial are the eigenvalues of $Av=\mu v.$
      - When n is large, other techniques are used to derive eigenvalues.

The Eigenvector Matrix

- If we consider all n eigenvector equations:

$$Av^{(i)}=v^{(i)}\mu_i \qquad i=1,2,\ldots,n$$

we can combine all of them by using the single matrix equation

$$AV=V\,diag[\mu_1,\mu_2,\ldots,\mu_n]$$

where V and the diagonal matrix are:

$$V=\begin{bmatrix} v^{(1)} & v^{(2)} & \cdots & v^{(n)} \end{bmatrix} = \begin{pmatrix} v_1^{(1)} & \cdots & v_1^{(n)} \\ \vdots & \ddots & \vdots \\ v_n^{(1)} & \cdots & v_n^{(n)} \end{pmatrix}.$$

$$diag[\mu_1,\mu_2,\ldots,\mu_n] = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{pmatrix}$$

Motivation for Eigenvectors

- What is so special about eigenvectors?
  - Eventually we want to write:

$$A = V\, diag\left[\mu_1, \mu_2, \ldots, \mu_n\right] V^{-1}$$

These values make up the **eigenvalue spectrum**.

  - This can be done if V is invertible.
    - Later we will provide the conditions that guarantee this invertability.
  - This eigendecomposition of A is very important because it helps us characterize many of the behaviours of the matrix A in various applications.

Similar Matrices

- Suppose A and B are both $n \times n$ matrices, then we say that B is **similar** to A if there exists an invertible matrix P such that

$$B = P^{-1}AP.$$

  - Note: $$A = PBP^{-1}.$$
  - So, the equation $A = V\, diag\left[\mu_1, \mu_2, \ldots, \mu_n\right] V^{-1}$ really states that A is similar to a diagonal matrix.

Similarity is an equivalence relation: reflexive, symmetric, and transitive.

Similar Matrices and Eigenvalues

- Theorem: Similar matrices have the same eigenvalue spectrum.
  - In fact, if we let $B = P^{-1}AP$ then if $\{\lambda, x\}$ is an eigenpair of B then $\{\lambda, Px\}$ is an eigenpair of A.
    - SO: both A and B have the same eigenvalues (but the eigenvectors corresponding to these eigenvalues are not necessarily the same).

"Eigenpair" is an eigenvalue and eigenvector pair that satisfy the eigenvector equation.

Diagonalizable Matrices

- A matrix A is diagonalizable if it is similar to a diagonal matrix.
  - That is, A is diagonalizable if there exists invertible P such that

$$D = P^{-1}AP$$

  where D is a diagonal matrix.
- Theorem: An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors.

Distinct Eigenvalues

- Theorem: Suppose A has k eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ that are all **distinct**.
  Then the corresponding eigenvectors $v^{(1)}, v^{(2)}, \ldots, v^{(k)}$ are **linearly independent**.

- Theorem: If an $n \times n$ matrix A has n distinct eigenvalues, then it is diagonalizable.

Complete Set of Eigenvectors

- A **complete** set of eigenvectors for an $n \times n$ matrix A is a set of n **linearly independent** set of eigenvectors for A.
  - A matrix that does not have a complete set of eigenvectors is said to be **deficient**.

- Summary:
  n distinct eigenvalues ➔ complete set ➔ diagonalizability.

Diagonalizability & Distinct Eigenvalues

- What if eigenvalues are not distinct?
  - If there are repeated eigenvalues then there is still the possibility that a complete set of eigenvectors can exist.
  - The analysis of this situation is a bit complicated (covered in the text: Theorems 6,7, and 8 in Appendix C).
    - We will omit these topics (eigenspaces, Thm. 6, algebraic & geometric multiplicity, Thm. 7, and Thm. 8).
    - Instead we will simply assume that the matrices that we encounter in our applications have distinct eigenvalues.

Orthogonal Matrix

- A square matrix Q is said to be an orthogonal matrix if

$$Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I.$$

  - If Q is nonsingular, then we have a wonderful simplicity for the inverse of the matrix Q, in fact, $Q^{\mathrm{T}}Q = I$ implies $Q^{-1} = Q^{\mathrm{T}}$.
  - Note: $Q^{\mathrm{T}}Q = I$ if and only if the columns of Q are such that:

$$q^{(i)\mathrm{T}}q^{(j)} = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Columns of Q

Orthogonal Diagonalization

- Consider an $n \times n$ matrix A such that there exists an **orthogonal** nonsingular matrix Q and

$$Q^{\mathrm{T}} A Q = D$$

  for some diagonal matrix D.
- Then we say that A is **orthogonally diagonalizable**.

Symmetric Matrices

- An $n \times n$ matrix A is **symmetric** if it has the property: $A = A^{\mathrm{T}}$.
  - This additional property of A allows us to say more about eigenvalues, eigenvectors, and diagonalizability.
  - There are two significant benefits that are gained when A is symmetric:
    1. A is always diagonalizable.
    2. A is not just diagonalizable, it is **orthogonally diagonalizable**.

This means that the eigenvectors are mutually orthogonal and hence provide us with a very useful orthonormal basis.

Use of an Orthogonal Basis

- If we use the orthogonal basis vectors to form a change-of-basis matrix then we have the following benefit: Orthogonal operations on a vector (multiplying a vector by an orthogonal matrix) do not change the length of a vector.

$$\begin{aligned}\|x\|^2 &= x^{\mathrm{T}} x \\ &= x^{\mathrm{T}} Q^{\mathrm{T}} Q x \\ &= (Qx)^{\mathrm{T}} Qx \\ &= \|Qx\|^2\end{aligned}$$

Angle Between Two Vectors

- The angle between two vectors is given by

$$\cos\theta = \langle u, v\rangle / \left(\|u\|\|v\|\right).$$

- Note that this remains invariant after an orthogonal operation on both vectors since the vector lengths in the denominator are unchanged and

$$\langle u, v\rangle = u^{\mathrm{T}} v = u^{\mathrm{T}} Q^{\mathrm{T}} Q v = (Qu)^{\mathrm{T}} Qv = \langle Qu, Qv\rangle.$$

Symmetry iff Orthogonal Diagonalizability

- Theorem: Matrix A is symmetric if and only if A is orthogonally diagonalizable.

- Corollary: Suppose $n \times n$ symmetric matrix A has n (not necessarily distinct) eigenvalues written as $d_1, d_2, \ldots, d_n$.
Then there exists $n \times n$ orthogonal Q such that

$$Q^{\mathrm{T}} A Q = diag\left(d_1, d_2, \ldots, d_n\right).$$

Spectral Decomposition of a Symmetric Matrix

- Consider an $n \times n$ symmetric matrix A.
  - The foregoing discussion asserts the existence of Q and D such that

$$A = QDQ^{\mathrm{T}}.$$

  - We can rewrite this last equation as:

$$A = \sum_{i=1}^{n} d_i q^{(i)} q^{(i)\mathrm{T}}$$

where $q^{(i)}$ is the i^{th} column of Q.

Symmetric Matrices in Practice

- Symmetric matrices commonly occur in the following situations:
  - The n^2 entries of A may be derived from all possible distances between n points in some metric space:
    $$a_{ij} = dist\left(x^{(i)}, x^{(j)}\right)$$
    then the symmetry of a distance calculation
    $dist\left(x^{(i)}, x^{(j)}\right) = dist\left(x^{(j)}, x^{(i)}\right)$
    leads to symmetry of A, that is, $a_{ij} = a_{ji}$.
  - Given *any* arbitrary $m \times n$ matrix M, note that the $m \times m$ matrix MM^{T} and the $n \times n$ matrix $M^{\mathrm{T}}M$ are both symmetric.

Rank of $M^{\mathrm{T}}M$ and MM^{T}

- Theorem: For any $m \times n$ matrix M the symmetric matrices $M^{\mathrm{T}}M$ and MM^{T} have ranks given by:

$$rank(M) = rank\left(M^{\mathrm{T}}M\right) = rank\left(MM^{\mathrm{T}}\right).$$

Definiteness

- For $n \times n$ symmetric matrix A:
  - A is positive semi-definite iff $x^{\mathrm{T}} A x \geq 0$ for all $x \in \mathbb{R}^n$.
  - A is positive definite iff $x^{\mathrm{T}} A x > 0$ for all $x \in \mathbb{R}^n$.
  - A is negative semi-definite iff $x^{\mathrm{T}} A x \leq 0$ for all $x \in \mathbb{R}^n$.
  - A is negative definite iff $x^{\mathrm{T}} A x < 0$ for all $x \in \mathbb{R}^n$.
  - A is indefinite if $x^{\mathrm{T}} A x$ can have both positive and negative values for all $x \in \mathbb{R}^n$.
- For the symmetric matrix $M^{\mathrm{T}} M$ we can easily demonstrate that the first case holds:

$$x^{\mathrm{T}} M^{\mathrm{T}} M x = (Mx)^{\mathrm{T}} Mx = \|Mx\|^2 \geq 0.$$

PD and the Similarity Transformation

- Positive definiteness is invariant under a similarity transformation.
  - This is expressed in the following:
- Theorem: Suppose A is positive definite. Let

$$B = P^{\mathrm{T}} A P$$

with P invertible. Then B is also positive definite.
- Theorem: For symmetric matrix A, the following statements are equivalent:

  (a) A is positive definite.

  (b) All eigenvalues of A are positive.

“Square Root” of a PD Matrix

- Theorem: An $n \times n$ symmetric matrix A is positive definite if and only if there exists an $n \times n$ invertible matrix B such that

$$A = B^{\mathrm{T}} B$$

- Theorem: Suppose A is an $m \times n$ matrix. Then the positive eigenvalues of $A^{\mathrm{T}} A$ are equal to the positive eigenvalues of AA^{T}.

Singular Value Decomposition

- We now explore a very useful factorization that can be applied to *any* real matrix.
  - Suppose we assume (without proof) that an $m \times n$ real matrix A can be written as:

$$A = USV^{\mathrm{T}}.$$

  - Then: $AA^{\mathrm{T}} = US^2U^{\mathrm{T}}$ and $A^{\mathrm{T}}A = VS^2V^{\mathrm{T}}$.
    - We know that both $A^{\mathrm{T}}A$ and AA^{T} are symmetric positive definite matrices and have the same eigenvalues.
    - Both are diagonalizable (they can be written as portrayed in the last two equations).
      - With these observations, we can use the equations to compute U, V, and S.

Eigenvectors of AA^{T} and $A^{\mathrm{T}}A$

- Can we say anything about the eigenvectors of $A^{\mathrm{T}}A$ and AA^{T} ?
  - There is an elegant relationship among them: If $v^{(i)}$ is a column vector of V then it is an eigenvector. In fact:

    $$A^{\mathrm{T}}Av^{(i)} = s_i^2 v^{(i)}$$

    where s_i^2 is the i^{th} entry of the diagonal matrix S^2. Pre-multiplying this last equation by A, we get: $\left(AA^{\mathrm{T}}\right)\left(Av^{(i)}\right) = s_i^2\left(Av^{(i)}\right)$ which says that we can get an eigenvector of AA^{T} by taking an eigenvector of $A^{\mathrm{T}}A$ and pre-multiplying it by A.

The Singular Value Decomposition Theorem

- Theorem: If A is an $m \times n$ matrix with $rank(A) = r$, then there exists orthogonal $m \times m$ matrix U and orthogonal $n \times n$ matrix V, such that $A = USV^{\mathrm{T}}$

  where S is a diagonal matrix with nonnegative entries on its diagonal and:

  i) S is a full rank $r \times r$ diagonal matrix if $r = m = n$.
  ii) S is $r \times n$ and has extra 0 filled columns if $r = m < n$.
  iii) S is $m \times r$ and has extra 0 filled rows if $r = n < m$.
  iv) S is $m \times n$ and has extra 0 filled rows and columns if $r < m$ and $r < n$.