MATH337 Reference

Krish A. Patel

December 15, 2025

1 Lesson 1

1.1 Content

1. Column vectors can hold complex numbers.

1.2 Memory Items

1. A ***scalar*** is a number, typically either real or complex, by which vector may be multiplied.

2. A ***linear combination*** of a set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is any vector that can be expressed as $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ where $\alpha_1, \ldots, \alpha_k$ are scalars.

3. When $\mathbf{x} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ the scalars are said to be the **coordinates**, **expansion coefficients**, or simply **coefficients** of $\mathbf{x}$ with respect to $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$.

4. A linear combination is an **affine combination** if and only if the sum of its coefficients is 1.

5. A set of vectors is **affine** if and only if it is closed under affine combination.

6. A nonempty affine set of vectors in $\mathbb{R}^3$ is a point, a line, a plane or a space. (geometric observation)

7. The dot product of two vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots + x_ny_n$, equal the product of the lengths of the vectors, $||\mathbf{x}||$ and $||\mathbf{y}||$, with the cosine of the angle between them. (geometric observation) ($\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}||||\mathbf{y}|| \cos\theta$)

8. Cauchy-Scwarz: $|\mathbf{x} \cdot \mathbf{y}| \leq ||\mathbf{x}|| \, ||\mathbf{y}||$

9. Two vectors are **orthogonal** if and only if their dot product is zero.

10. Perpendicular vectors are orthogonal. (geometrix observation)

11. The zero vector is orthogonal to all vectors having the same number of components, but perpendicular to none of them.

2 Lesson 2

1. The solution set of a system of linear equations always contains 0, 1 or infinitely many solutions.
2. The solution of a system of equations with three unknowns is the intersection of a set of planes. (geometric observation)

3 Lesson 3

1. The elementary row operations transform a matrix into a matrix that is row equivalent to the original.
2. Matrices that are row equivalent are also solution equivalent (but not necessarily vice versa).
3. An elimination step uses a nonzero entry in a column to produce a row equivalent.
4. Every elementary row operations can be reversed (undone) with an elementary row operation.

4 Lesson 4

1. The Gauss-Jordan algorithm uses elimination to find the reduced row-echelon form of a matrix.
2. The Gauss-Jordan algorithm has three major steps: the forward phase, normalization, and the backward phase.
3. Elimination in the forward phase of Gauss-Jordan is generally where most of the computational effort is required.
4. The reduced row-echelon form of a matrix is the unique reduced row-echelon matrix that is row equivalent to the matrix.

5 Lesson 5

1. The solution set of a linear system is empty if and only if the final column of the augmented matrix is a pivot column.
2. The particular solution in the parametric vector form is distinguished from all other solutions by having all free components equal to zero.
3. For linear systems with a nonempty solutions set, the number of basic solutions equals the number of columns of the coefficient matrix without pivots (nullity).

6 Lesson 6

1. Every column of a matrix product is a linear combination of the columns of the first factor with coefficients from a column of the second factor.
2. Every entry of a matrix product is the dot-like product of row of the first factor with a column of the second factor.
3. Every matrix product is the sum of the products of the columns of the first factor with the corresponding rows of the second factor (outer product expansion).
4. The Gram matrix contains all the relative geometric information regarding the columns of a matrix.
5. $(A + B)^2 = A^2 + AB + BA + B^2$
6. Multiplying a matrix on the left by a diagonal matrix scales the rows of the matrix.

7 Lesson 7

1. The inverse matrix of a square matrix, if it exists, is unique.
2. The inverse of a product is the product of the inverse taken in the opposite order.
3. The inverse of a transpose is the transpose of the inverse.
4. Linear systems with an invertible coefficient matrix have unique solutions.
5. Inverse linear transformations correspond with inverse matrices.

8 Lesson 8

1. Every invertible matrix is the product of elementary matrices.
2. Products of lower triangular matrices are lower triangular.
3. Products of lower unitriangular matrices are unitriangular.
4. LU is easily written down from the computation of the forward phase of Gauss-Jordan.
5. CR is easily written down from the matrix and its reduced row-echelon form.

9 Lesson 9

9.1 Memory Items

1. The Gram matrix of an $m \times n$ matrix is is an $n \times n$ (Hermitian) symmetric symmetric matrix that is always positive semi-definite (and often positive definite).

2. The columns of a matrix are orthonormal if and only if the Gram matrix is the identity.

3. The set of $n \times n$ orthogonal matrices $\mathrm{O}(n)$ form an algebraic group: i.e., (1) $\mathrm{O}(n)$ is closed under multiplication, (2) multiplication is associative, (3) $I \in \mathrm{O}(n)$, (4) each orthogonal matrix has a (multiplicative) inverse (easily computed as a transpose).

4. The set of 2×2 orthogonal matrices $\mathrm{O}(2)$ consist of the rotation matrices (which have determinant equal to 1) and the reflection matrices (which have determinant equal to -1).

5. The set of permutation matrices S_n forms an algebraic group (a subgroup of $\mathrm{O}(n)$) that contains $n!$ elements.

6. The action of an orthogonal matrix is an isometry.

9.2 Notes

1. Rotation and reflection matrices:

$$\mathrm{Rot}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \mathrm{Ref}(\phi) = \begin{bmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{bmatrix}$$

10 Lesson 10

10.1 Memory Items

1. A projection matrix is a matrix that is equal to its square ($P^2 = P$).

2. The null space of a projection matrix is the column space of it complementary projection.

3. The fixed points of a projection equal its column space.

4. A rank one matrix $A = \mathbf{u}\mathbf{v}^T$ is a projection iff (if and only if) $\mathbf{u} \cdot \mathbf{v} = 1$.

5. An orthogonal projection matrix with rank one has the form $\mathbf{u}\mathbf{u}^T$ where $||\mathbf{u}|| = 1$.

10.2 Other Remarks

1. (Follow Up Problem 10.3) The trace (sum of diagonal elements) of a projection matrix equals its rank.
2. Review 11/12/2025:
   (a) (Follow Up Problem 17.12) The determinant of a projection matrix can only be 0 or 1.
   (b) The eigenvalues of a projection matrix can only be 0 or 1.
   (c) The rank of P equals the trace of P, which both equal those of its diagonal matrix of eigenvalues D.

11 Lesson 11 (Informal)

Informal memory items since none were officially provided.

1. Determinants only exist for square matrices.
2. The determinant of a triangular (either upper or lower) is the product of the diagonal elements.
   (a) Thus, the determinant of a unitriangular (either upper or lower) is 1.
3. Elementary row operations:
   (a) Exchanging two rows of a matrix reverses the sign of the determinant.
   (b) Scaling a matrix by multiplication of one row by a factor α multiplies the determinant by α.
   (c) Subtracting a multiple of one row from another row leaves the determinant unchanged.
4. Shortcut to finding determinants of large matrices: perform Gauss-Jordan and apply the above sub-bullets.
5. A matrix is singular iff the determinant is 0.
6. $|AB| = |A|\,|B|$
7. $|A^T| = |A|$
8. The determinant of a 2x2 matrix $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$ is the area of the parallelogram generated by the columns of A (i.e. the set of all linear combinations $\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2$ where $0 \le \alpha_1 \le 1$ $0 \le \alpha_2 \le 1$).
9. The sign of the determinant is called the **orientation** of the ordered set of vectors.
10. $|U| = \pm 1$ (for any orthogonal matrix U)

(a) When $|U| = +1$, these matrices form the special orthogonal group $SO(n)$, which is closed under matrix multiplication.

11. The set of permutation matrices with determinant 1 are the **even permutations**, and the set with determinant -1 are the **odd permutations**.

12 Lesson 12

12.1 Memory Items

1. An eigenvector $\mathbf{v}$ of a square matrix A with its eigenvalue λ satisfies $A\mathbf{v} = \lambda\mathbf{v}$.
2. All the eigenvalues of a square matrix satisfy the characteristic equation $|A - \lambda I| = 0$.
3. The spectrum of an $n \times n$ matrix A has from 1 to n elements.
4. The sum of the multiplicities of the eigenvalues of an $n \times n$ matrix A is always n.
5. The trace equals the sum of the eigenvalues, and the determinant equals the product of the eigenvalues.
6. Distinct eigenvalues guarantee diagonalizability but not conversely.

12.2 Other Remarks

1. An eigenvector must not be the zero vector.
2. The characteristic polynomial for 2x2 matrices is $p_A(\lambda) = \lambda^2 - \text{tr}(A)\lambda + \det(A)$.
3. The set of all eigenvalues of a matrix A are the **spectrum**, denoted by $\sigma(A)$.
4. The multiplicity of an eigenvalue λ_ℓ is often denoted by μ_ℓ.
5. Eigenvalues with $\mu_\ell = 1$ are **distinct**, while those with $\mu_\ell > 1$ are **repeated**.
6. **Cayley-Hamilton Theorem**: Every square matrix satisfies its characteristic equation, and A^n is a linear combination of its lower powers $A^{n-1}, A^{n-2}, \ldots, A^2, A^1, A^0$.
7. An $n \times n$ matrix B is a **similarity transformation** of an $n \times n$ matrix M iff there exists a nonsingular $n \times n$ matrix S such that $B = SMS^{-1}$.
8. The **diagonalization** of a matrix A is a similary transformation of a diagonal matrix D: $A = SDS^{-1}$

(a) This is useful because computing powers of a matrix is given by $A^p = SD^pS^{-1}$, and computing the powers of a diagonal matrix is much easier.

9. The matrix D contains the eigenvalues of a matrix A in any order along the diagonal. The matrix S contains the eigenvectors for each corresponding eigenvalue in D in the same order.

10. Find the eigenvectors of each eigenvalue by finding the basic solutions of $(A - \lambda_\ell I)\mathbf{v} = 0$ for $\ell = 1, 2, \ldots, n$.

11. A basic solution is guaranteed to exist for the homogeneous system for each ℓ.

12. An eigenvector remains an eigenvector when multiplied by any non-zero constant.

13. A matrix A is **defective** if it is not diagonalizable. This is often when the number of basic solutions for an eigenvalue is less than the multiplicity (when $\mu_\ell > 1$).

14. For a 2×2 matrix with a repeated eigenvalue to be diagonalizable, it must be a multiple of I_2.

15. When $A - \lambda_1 I$ has only one free variable, it is obviously not the zero matrix. Nonetheless, $(A - \lambda_1 I)^2$ will be the zero matrix.

16. Shortcut to finding eigenvectors for 2×2 diagonalization:
$A - \lambda I = \begin{bmatrix} u & v \\ w & y \end{bmatrix} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}$ when $\mathbf{x} = \begin{bmatrix} -v \\ u \end{bmatrix}$ or $\mathbf{x} = \begin{bmatrix} -y \\ w \end{bmatrix}$

    (a) If one of the vectors is zero, we must choose the non-zero vector.

    (b) If both are zero, then the matrix is the zero matrix, and so, every non-zero vector is an eigenvector.

    (c) Remember that complex solutions come in conjugates, so the other eigenvector is simply the complex conjugate of the first one.

17. (Follow Up Problem 12.6) For a projection matrix P,

    (a) the eigenvector corresponding to an eigenvalue of 0 is a null vector of P.

    (b) the eigenvector corresponding to the eigenvalue of 1 is a fixed point of P.

    (c) diagonalization yields a diagonal projection matrix as D.

18. (Follow Up Problem 12.11) The similarity transformation of a projection matrix is a projection matrix:
$H = SPS^{-1}$
$H^2 = (SPS^{-1})^2 = SP(S^{-1}S)PS^{-1} = SP^2S^{-1} = SPS^{-1} = H$

19. (Follow Up Problem 12.12) If A is a similarity transform of B, then B is a similarity transform of A. For $S, \tilde{S} \in GL_n$:

$$\begin{aligned} A &= SBS^{-1} \\ S^{-1}AS &= B \\ B &= \tilde{S}A\tilde{S}^{-1} \\ \tilde{S} &= S^{-1} \end{aligned}$$

13 Lesson 13

13.1 Memory Items

1. Vector spaces are nonempty sets of vectors closed under formation of linear combinations.
2. Vector spaces are affine sets containing the zero vector.
3. Every vector space is a span, and every span is a vector space.
4. Every vector space is a subspace, and every subspace is a vector space.
5. The only linear combination of a linearly independent set of vectors equal to the zero vector is trivial.
6. Representation in terms of a given basis is unique.
7. The dimension of a vector space equals the number of vectors in a basis.

13.2 Other Remarks

1. Vector spaces must contain the zero vector as a consequence of their definition.
2. The **span** is all linear combinations of elements of the set. A special case is $\mathrm{Span}(\emptyset) = \{\mathbf{0}\}$ to ensure that all spans are vector spaces.
3. Given a subspace $V \subset \mathbb{R}^2$, **a spanning set** S for V is a set with the property $\mathrm{Span}(S) = V$.
4. The **trivial subspace** is simply $\{\mathbf{0}\}$.
5. Subspaces are closed under intersection.
6. The only linear combination of indepedent vectors that leads to the zero vector is one in which all the coefficients are 0 (trivial).
7. A set of vectors is linearly dependent if:

   (a) it contains $\mathbf{0}$.

(b) it contains more vectors than the number of components in each vector.

(c) lacks at least one pivot when put into a matrix and converted to rref.

8. Orthonormal sets are linearly independent.

9. An $n \times n$ matrix is diagonalizable iff there is a collection of n linearly independent eigenvectors of the matrix.

10. The eigenvectors for unique eigenvalues are linearly independent.

11. A basis for a subspace V is a linearly indepedent spanning set.

12. There are numerous bases for every subspace, but all of them must contain the same number of vectors (have the same **dimension**).

14 Lesson 14

14.1 Memory Items

1. The pivot columns are a basis for the column space.

2. Inclusion relations for subspaces can be identified through Gauss-Jordan.

3. Bases for an intersection can be identified through Gauss-Jordan.

4. Bases for a sum can be identified through Gauss-Jordan.

14.2 Other Remarks

1. **Theorem**: If $F = \begin{bmatrix}\mathbf{f}_1 & \cdots & \mathbf{f}_n\end{bmatrix}$ and $G = \begin{bmatrix}\mathbf{g}_1 & \cdots & \mathbf{g}_n\end{bmatrix}$ are two row equivalent $m \times n$ matrices and suppose $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$, then

   $\{\mathbf{f}_{i_1}, \cdots, \mathbf{f}_{i_k}\}$ is linearly dependent $\Leftrightarrow$

   $\{\mathbf{g}_{i_1}, \cdots, \mathbf{g}_{i_k}\}$ is linearly dependent

   Essentially, corresponding subsets of columns of two matrices are either both linearly independent or both linearly dependent.

2. A basis for a vector space $U = \text{span}(\mathcal{A})$ is the set of pivot columns of a matrix A formed by the vectors in $\mathcal{A}$, found through (the forward phase of) Gauss-Jordan elimination.

3. **Inclusion**: If $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$, $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$, $A = \begin{bmatrix}\mathbf{a}_1 & \cdots \mathbf{a}_k\end{bmatrix}$, and $B = \begin{bmatrix}\mathbf{b}_1 & \cdots \mathbf{b}_k\end{bmatrix}$,

   (a) $\operatorname{col} B \subset \operatorname{col} A$ (equivalently $\text{span}(\mathcal{B}) \subset \text{span}(\mathcal{A})$) iff $\text{pref}[A \mid B]$ has no pivots in the columns at the positions occupied by B.

   (b) $\operatorname{col} A = \operatorname{col} B$ (equivalently span(A) = span(B)) iff $\operatorname{col} A \subset \operatorname{col} B$ and $\operatorname{col} B \subset \operatorname{col} A$.

4. **Basis for intersection**: If $\mathbf{v} \in U \cap V$ then there are vectors $\alpha \in \mathbb{F}^k$ and $\beta \in \mathbb{F}^\ell$ such that

$$\mathbf{v} = A\alpha = B\beta,$$

and each vector $\mathbf{v}$ corresponds to the solution of

$$[A \quad B] \begin{bmatrix} \alpha \\ -\beta \end{bmatrix} = \mathbf{0}.$$

So, every basic solution yields a basis for the intersection.

5. **Basis for sum**: A spanning set for the sum of two vector spaces $U + V$ is the union of their corresponding spanning sets $\mathcal{A} \cup \mathcal{B}$, and a basis is given from the set of pivot columns of the matrix whose columns are the vectors in $\mathcal{A} \cup \mathcal{B}$.

6. $\dim(U + V) \;=\; \dim(U) \;+\; \dim(V) \;-\; \dim(U \cap V)$

7. (Follow Up Problem 14.8) For two subspaces U and V,

$$U \cap V = U + V \qquad \Leftrightarrow \qquad U = V$$

15 Lesson 15

15.1 Memory Items

1. The column space of A is the set of all vectors of the form $A\mathbf{x}$.
2. The null space of A is the set of all vectors satisfying $A\mathbf{x} = \mathbf{0}$.
3. The row space and the left null space are the orthogonal complements of the null space and the column space, respectively.
4. The CR factorization displays bases for the column space and row space.
5. Eigenvectors are either in the column space or the null space.

15.2 Table of Fundamental Subspaces

Space	Notation	Standard Basis	Dimension	$\subseteq$
null	$\text{nul}(A)$	basic solutions of $A\mathbf{x} = \mathbf{0}$	$n - r$	$\mathbb{R}^n$
column	$\text{col}(A)$	pivot columns of A	r	$\mathbb{R}^m$
row	$\text{row}(A)$	nonzero rows of $\text{rref}(A)$	r	$\mathbb{R}^n$
left null	$\text{nul}(A^T)$	last $m - r$ rows of W $\text{rref}\big([A\ I]\big) = [\text{rref}(A)\ W]$	$m - r$	$\mathbb{R}^m$

15.3 Other Remarks

- Assume an $m \times n$ matrix.
- ν represents nullity.
- r represents rank.

1. A linear system like $A\mathbf{x} = \mathbf{0}$ is an *implicit* representation of a vector space, whereas a spanning set is an *explicit* representation.
2. **Null space**:
   (a) $\text{nul}(A) \in \mathbb{R}^n$ (domain)
   (b) $\dim(\text{nul}(A)) = \nu$
   (c) A basis for $\text{nul}(A)$ is the set of all basic solutions of $A\mathbf{x} = \mathbf{0}$.
3. **Column space**:
   (a) $\text{col}(A) \in \mathbb{R}^m$ (codomain)
   (b) $\dim(\text{col}(A)) = r$
   (c) A basis for $\text{col}(A)$ is the set of all pivot columns.
4. **Row space**:
   (a) $\text{row}(A) = \text{col}(A^T)$
   (b) $\text{row}(A) \in \mathbb{R}^n$ (domain)
   (c) $\dim(\text{row}(A)) = r$
   (d) A basis for $\text{row}(A)$ is the set of all non-zero rows in $\text{rref}(A)$.
5. **Left null space**:

(a) $\text{nul}(A^T)$

(b) $\text{nul}(A^T) \in \mathbb{R}^m$ (codomain)

(c) $\dim(\text{nul}(A^T)) = m - r$

(d) After performing Gauss-Jordan elimination,

$$\text{rref}[A \mid I] = [\text{rref}(A) \mid W],$$

a basis for $\text{nul}(A^T)$ is the set of all the rows in W next to rows of zeros in $\text{rref}(A)$.

6. Linear mapping: $T_A : \mathbb{R}^n \to \mathbb{R}^m$

7. $A\mathbf{x} = \mathbf{b}$ has at least one solution $\quad \Leftrightarrow \quad \mathbf{b} \in \text{col}(A)$

8. $A\mathbf{x}_1 = \mathbf{b} \quad \text{and} \quad A\mathbf{x}_2 = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x}_1 - \mathbf{x}_2 \in \text{nul}(A)$

9. The smaller the null space, the less ambiguity there is, with $\{\mathbf{0}\}$ representing no ambiguity (unique solutions).

10. The action of an $m \times n$ matrix defines a one-to-one correspondence between its row space and its column space.

11. **Orthogonal Complements**:

(a) The set of vectors orthogonal to a vector space U is called the **orthogonal complement** $U^\perp$.

$$U^\perp = \{\mathbf{x} \ : \ \mathbf{x} \cdot \mathbf{u} = 0 \text{ for all } \mathbf{u} \in U\}$$

(b) $U^\perp$ is a subspace.

(c) $(U^\perp)^\perp = U$

(d) $U \subset V \quad \Leftrightarrow \quad V^\perp \subset U^\perp$

(e) $\text{nul}(A)$ and $\text{row}(A)$ (in the domain) are orthogonal complements.

(f) $\text{nul}(A^T)$ and $\text{col}(A)$ (in the codomain) are orthogonal complements.

12. The fundamental subspaces of an invertible $m \times n$ matrix A:

$$\begin{aligned} \text{nul}(A) &= \{\mathbf{0}\}, \\ \text{col}(A) &= \mathbb{R}^n, \\ \text{row}(A) &= \mathbb{R}^n, \\ \text{nul}(A^T) &= \{\mathbf{0}\} \end{aligned}$$

13. Given two matrices A and B, it is always possible to find a matrix M, where $\text{col}(M) = \text{col}(A)$ and $\text{row}(M) = \text{col}(B) = \text{row}(B^T)$:

$$M = AXB^T, \qquad X \in \text{GL}_r$$

14. When $\text{col}(A) \subset \text{nul}(A)$

$$A^2 = A \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{a}_1 & \cdots & A\mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

    That is, A is the square root of the zero matrix.

15. In CR factorization $A = CR$,

$$\text{col}(A) = \text{col}(C) \qquad \text{and} \qquad \text{row}(A) = \text{row}(R),$$

    and thus,

$$\text{nul}(A^T) = \text{nul}(C^T) \qquad \text{and} \qquad \text{nul}(A) = \text{nul}(R).$$

16. For the multiplication of matrices B and C,

$$\text{rank}(BC) \leq \min\bigl(\text{rank}(B), \text{rank}(C)\bigr).$$

17. Given an eigenvector $\mathbf{v}$ of A with eigenvalue λ, $A\mathbf{x} = \lambda\mathbf{x}$,

$$\begin{aligned} \lambda = 0 &\Rightarrow \mathbf{x} \in \text{nul}(A) \\ \lambda \neq 0 &\Rightarrow \mathbf{x} \in \text{col}(A) \end{aligned}$$

    Thus, every eigenvector of A is in the column space of A, in the null space of A, or possibly both.

16 Lesson 16

16.1 Notes

1. If a matrix is triangular, all its eigenvalues are on the main diagonal.

2. A matrix is defective if it doesn't have as many free variables as the multiplicity of the eigenvalue. (Refer to **Example 16.6**)

3. The characteristic polynomial of a 3×3 matrix is given by

$$\begin{aligned} p_A(\lambda) = &-\lambda^3 \\ &+ \text{tr}(A)\lambda^2 \\ &- \left(\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} + \begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} + \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} \right) \lambda \\ &+ \det(A) \end{aligned}$$

    Notice that the signs alternate, beginning with minus for odd degrees like 3. The cofficient of the λ term is the negative of the trace of the minor matrix.

4. To find the roots of a characteristic equation:

(a) Use the rational roots theorem to generate candidate roots:

$$\frac{\pm\,\text{factors of the constant}}{\pm\,\text{factors of the leading term}}$$

(b) Try each candidate until one root is found.

(c) Divide the polynomial by this root (preferably using synthetic division).

(d) For a cubic polynomial, this yields a linear factor and a quadratic that can be easily solved. Otherwise, repeat until there are various linear roots and a quadratic.

5. Shortcuts when computing eigenvectors of a 3×3 matrix A:

   (a) When a row that is a linear combination of the others is replaced by a row of zeros, the rref is unchanged and hence the null space is unchanged.

   (b) The eigenvector can be found via the cross product method: the i^{th} element of the eigenvector is found from the determinant of the minor created by crossing the 3rd row (of all zeros) and i^{th} column of A; the result is multiplied by the cofactor, which starts as positive 1 and alternates its sign for each element.

6. Recall that $A = SDS^{-1}$.

17 Lesson 17

17.1 Notes

1. An orthogonal projection matrix P projects a vector $\mathbf{x}$ from a vector space to a subspace V, resulting in $\mathbf{p}$.

$$\mathbf{p} = P\mathbf{x}$$

   It essentially finds a point in V closest to the point in the original vector space.

2. P is both a projection matrix and symmetric.

$$P^2 = P = P^T$$

3. If $\mathbf{x} \in V$ already, then the closest point it $\mathbf{x}$ itself, so

$$\begin{aligned} \mathbf{x} &= P\mathbf{x}, \\ \operatorname{col}(P) &= \operatorname{col}(A) = V, \\ \operatorname{col}(P) &= \operatorname{row}(P). \end{aligned}$$

   The last point is because $P^T = P$.

4. The orthogonal projection matrix P is given by forming a matrix A whose columns are the basis of a vector space and using the equation

$$P \ = \ A(A^T A)^{-1} A^T.$$

5. For orthonormal bases, $A^T A$ reduces to I, so the original equation simplifies to

$$P = AA^T,$$

which is also the sum of the outer products of the columns

$$P = \mathbf{a}_1 \otimes \mathbf{a}_1 + \cdots + \mathbf{a}_k \otimes \mathbf{a}_k.$$

6. The complementary matrix $Q = I - P$ is also an orthogonal projection matrix:

$$Q^2 = Q = Q^T$$

7. Relation of subspaces of P and Q:

$$V = \text{col}(P) = \text{nul}(Q)$$
$$V^{\perp} = \text{col}(P)^{\perp} = \text{col}(Q)$$

Notice that $\text{nul}(Q)^{\perp} = \text{col}(Q) = \text{row}(Q)$ because Q is symmetric.

8. P and Q decompose a vector $\mathbf{x}$ into orthogonal components $\mathbf{p} \in P$ and $\mathbf{q} \in Q$:

$$\mathbf{x} = \mathbf{p} + \mathbf{q}$$
$$\mathbf{p} \cdot \mathbf{q} = 0$$

9. A linear system $A\mathbf{x} = \mathbf{b}$ may not have a solution if $\mathbf{b} \notin \text{col}(A)$, which is more often the case for linear data collected in real life.

10. Thus, least-squares solutions aim to minimize $\|A\mathbf{x}-\mathbf{b}\|$, satisfying $A\mathbf{x} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is an orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$.

11. The **normal solutions** for least-squares solutions:

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

    (a) If the original system had solutions, the the normal equations have the same solutions.
    (b) If the original system had no solutions, this is guaranteed to yield a solution, specifically minimizing $\|A\mathbf{x} - \mathbf{b}\|$.

12. When given a set of points $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, the linear system fitting the model $y = \alpha x + \beta$ is

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},$$

from which the normal solutions can be found and solved for. If fitting a quadratic model $y = \alpha x^2 + \beta x + \gamma$, simply add a column to the left matrix, making sure to square the x values.

13. The gram matrix of a matrix with independent columns is invertible.

18 Lesson 18

18.1 Notes

1. Recall that the eigenvectors for unique eigenvalues are linearly independent.

2. The rational canonical form of diagonalization avoids using complex numbers. Recall that eigenvalues and their eigenvectors come in complex conjugate pairs:

$$\lambda_\pm = \alpha \pm \beta i$$
$$\mathbf{v}_\pm = \mathbf{x} + i\mathbf{y}$$

The rational canonical form of the diagonalization of A can be written as

$$A = \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix}^{-1}.$$

3. **Scalar multiples**:

$$\alpha A = \alpha\, SDS^{-1} = S(\alpha D)S^{-1}$$

That is, αA is also diagonalizable with the same eigenvectors but each eigenvalue scaled by α.

4. **Shifts**:

$$A + \alpha I \;=\; SDS^{-1} + \alpha SIS^{-1} \;=\; S(D + \alpha I)S^{-1}$$

That is, $A + \alpha I$ is also diagonalizable with the same eigenvectors but α added to each eigenvalue.

5. **Powers**:

$$A^p \;=\; (SDS^{-1})^p \;=\; SD^pS^{-1}$$

That is, A^p (when it is defined) is also diagonalizable with the same eigenvectors but each eigenvalue raised to p. This is valid for negative powers as well.

6. **Polynomial functions**:

$$p(A) = S\,p(D)\,S^{-1}$$

That is, $p(A)$ is also diagonalizable with the same eigenvectors but p applied to each eigenvalue.

7. If $\mathrm{col}(A) \cap \mathrm{nul}(A)$ is nontrivial, then A is defective.

8. Diagonalization $A = SDS^-1$ can be thought of geometrically.

$$
\begin{array}{ccccc}
 & & 4 & & \\
 & \mathbf{x} & \longrightarrow & A\mathbf{x} = S\beta & \\
1 & \downarrow & & \uparrow & 3 \\
 & \alpha = S^{-1}\mathbf{x} & \longrightarrow & \beta = D\alpha & \\
 & & 2 & &
\end{array}
$$

1 : change of coordinates to the eigenbasis
2 : action of A with respect to the eigenbasis
3 : hange of coordinates to the standard basis
4 : action of A with respect to the standard basis

9. Two matrices A_1 and A_2 are **simultaneously diagonalizable** if

$$A_1 = SD_1S^{-1},$$
$$A_2 = SD_2S^{-1}.$$

That is, both matrices have the same eigenvectors but not necessarily the same eigenvalues.

10. Simultaneously diagonalizable matrices commute.

11. An **eigenspace** associated with the matrix A is denoted by

$$E_\lambda(A) = \mathrm{nul}(A - \lambda I).$$

(a) Every eigenvector of A with eigenvalue λ is in $E_\lambda(A)$.

(b) Every nonzero vector in $E_\lambda(A)$ is an eigenvector of A with eigenvalue λ.

(c) The intersection of eigenspaces corresponding to different eigenvalues is the trivial subspace.

(d) If A is a diagonalizable matrix, then the direct sum of its eigenspaces is $\mathbb{R}^n$.

12. (Follow Up Problem 18.10) Two diagonalizations of the same matrix must share the same eigenvalues and eigenspaces, but there is substantial freedom in choosing the eigenbasis. That is, for an $n \times n$ matrix A, there are n matrices possible for D (changing the order of the eigenvalues) but many more possible matrices S (as long as $\mathrm{col}(S) = E_\lambda(A)$), such that $A = SDS^{-1}$.

13. (Follow Up Problem 18.14) When the multiplicity of an eigenvalue is 2 or more, the square root of the diagonal matrix need not be diagonal, e.g.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

14. (Follow Up Problems 19.1-19.3) **Spectral decomposition**:

$$\begin{aligned} A = SDS^{-1} &= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{v}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{u}_n^T, \end{aligned}$$

where each $\mathbf{v}_\ell$ is a column from S, and each $\mathbf{u}_\ell^T$ is a row from S^{-1}.

19 Lesson 19

19.1 Notes

1. **Discrete linear dynamics** refers to the sequences of vectors resulting from the repeated action of a square matrix.

2. Starting at $\mathbf{x}_0$, the general form is

$$\mathbf{x}_{\ell+1} = A\mathbf{x}_\ell,$$

which can be written directly in terms of $\mathbf{x}_0$ as

$$\mathbf{x}_\ell = A^\ell \mathbf{x}_0.$$

With diagonalization,

$$\mathbf{x}_\ell = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix} \begin{bmatrix}\lambda_1^\ell & & & \\ & \lambda_2^\ell & & \\ & & \ddots & \\ & & & \lambda_n^\ell\end{bmatrix} \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix}^{-1} \mathbf{x}_0.$$

Thus, the solution is a linear combination of eigenvectors with the eigenvalues raised to ℓ.

$$\mathbf{x}_\ell = \lambda_1^\ell \mathbf{w}_1 + \lambda_2^\ell \mathbf{w}_2 + \cdots + \lambda_n^\ell \mathbf{w}_n$$
$$,$$

where

$$\mathbf{w}_j = s_j \mathbf{v}_j \quad \text{and} \quad \begin{bmatrix}s_1\\ s_2\\ \vdots\\ s_n\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix}^{-1} \mathbf{x}_0.$$

3. The spectrum of a matrix can give considerable insight into the behavior of the iterates:

   (a) **Decay to 0**: If all eigenvalues $|\lambda_j| < 1$, then as $\ell \to \infty$,

   $$\lambda_j^\ell \to 0 \quad \text{and so} \quad \mathbf{x}_\ell \to \mathbf{0}.$$

   (b) **Convergence to a general limit**: Suppose $\lambda_1 = 1$ and the other eigenvalues have magnitudes less than 1, then as $\ell \to \infty$,

   $$\mathbf{x}_\ell \to \mathbf{w}_1.$$

   This can be generalized to when one or more eigenvectors are 1 while others have magnitudes less than 1.

   (c) **Divergence in an eigendirection**: Suppose $\lambda_1 > 1$ and the remaining eigenvalues have magnitudes less than 1, then as $\ell \to \infty$,

   $$\mathbf{x}_\ell \approx \lambda_1^\ell \mathbf{w}_1$$

   because the solution is dominated by the term with the largest eigenvalue.

   (d) **Orbit about the origin**: Suppose $\lambda_\pm = e^{\pm i\theta}$, then the solution will obrit about the origin.

4. **General second-order difference equation**:

$$y_{n+2} = a y_{n+1} + b y_n$$

$$\begin{cases} y_{n+2} &= ay_{n+1} + by_n \\ y_{n+1} &= y_{n+1} \end{cases}$$

$$\underbrace{\begin{bmatrix} y_{n+2} \\ y_{n+1} \end{bmatrix}}_{\mathbf{x}_{n+1}} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} y_{n+1} \\ y_n \end{bmatrix}}_{\mathbf{x}_n}$$

$$\underbrace{\begin{bmatrix} y_{n+2} \\ y_{n+1} \end{bmatrix}}_{\mathbf{x}_{n+1}} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}^{n+1}}_{A^{n+1}} \mathbf{x}_0$$

This can easily be generalized to higher order equations. The first row is always the the equation itself, and the subsequent rows are lower order terms equal to themselves to fulfill the square matrix requirement.

5. **Phase curves**: Curves traced by

$$\mathbf{x}_t = \lambda_1^t s_1 \mathbf{v}_1 + \lambda_2^t s_2 \mathbf{v}_2 + \cdots + \lambda_n^t s_n \mathbf{v}_n,$$

where the discrete parameter ℓ has been replaced by the continuous parameter t.

6. Matrix for Fib numbers. Diagonalization makes it easy to compute A^n.

7. Diagonalizing this 2×2: one e-val will always be 1, the other will always be less than 1.

8. Both numbers in [an+1, an] will go to the same number in the limit of n to infinity since they're very close.

20 Lesson 20

20.1 Notes

1. Convert a second-order linear differential equation into a system of two first-order linear differential equations by creating a new variable. For example, the initial value problem

$$\begin{cases} \ddot{y} - y = 0 \\ y(0) = 1 \\ \dot{y}(0) = 0 \end{cases}$$

can be converted into the system using a new variable $v = \dot{y}$:

$$\begin{cases} \dot{v} &= y \\ \dot{y} &= v \\ y(0) &= 1 \\ v(0) &= 0 \end{cases}$$

This can be written in matrix form as

$$\begin{bmatrix} \dot{v} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v \\ y \end{bmatrix}$$
$$\begin{bmatrix} v(0) \\ y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

where A is the square matrix above.

2. When solving second-order linear differential equations, assume the form

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0).$$

3. Raising e to a matrix is simply raising e to the power of the diagonal elements after diagonalizing the matrix:

$$A = SDS^{-1}$$
$$e^A = S \begin{bmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_n} \end{bmatrix} S^{-1}$$

4. $e^O = I$, where O is the zero matrix.
5. $e^{A+B} = e^A e^B$ if matrices A and B commute.

21 Lesson 21

21.1 Notes

1. QR factorization of a matrix A involves finding a matrix Q and R such that $A = QR$.
   (a) Any $m \times n$ matrix A can be factored as $A = QR$ where Q is $m \times r$ with orthonormal columns and R is $r \times n$ and upper trapazoidal, where r is the rank of A.
   (b) If A is $n \times n$ and the columns of A are independent, then Q is $n \times n$ and orthogonal and R is $n \times n$ and upper triangular.
   (c) Upper triangular and upper trapazoidal are representations of proto-row echelon form.
2. If $A = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ and $Q = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$, to find orthogonal bases, start with $\mathbf{v}_1 = \mathbf{u}_1$ and apply the equation

$$\mathbf{v}_{\ell+1} = \mathbf{u}_{\ell+1} - \sum_{j=1}^{\ell} \left(\frac{\mathbf{v}_j \cdot \mathbf{u}_{\ell+1}}{\mathbf{v}_j \cdot \mathbf{v}_j} \right) \mathbf{v}_j$$

for $\ell = 1, 2, \ldots, n-1$.

3. If the columns of A are dependent, then you can optionally extend the orthonormal basis to $\mathbb{R}^m$ by augmenting the original matrix with the identity matrix and continuing Gram-Schmidt until the missing bases are found. Start with $\mathbf{e}_1$ as the vector to orthonormalize, if that does not work, then try $\mathbf{e}_2$, and so on. Now, $A = QR$ where
   (a) Q is $m \times m$ orthogonal matrix where the first r columns form an orthonormal basis for $\text{col}(A)$ and the remaining $m - r$ columns form an orthonormal basis for $\text{nul}(A^T) = \text{col}(A)^\perp$ and
   (b) R is a proto-row-echelon matrix ($R \sim \text{rref}(A)$).
4. **Cholesky factorization** Every square matrix shares its Gram matrix with an upper triangular matrix:
$$G = A^T A = (QR)^T (QR) = R^T Q^T QR = R^T R$$

22 Lesson 22

22.1 Notes

1. Real symmetric matrices ($A^T = A$) are Hermitian symmetric ($A^H = A$).
2. All Hermitian symmetric matrices are diagonalizable.
3. All eigenvalues of Hermitian symmetric matrices are real.
4. All Hermitian symmetric matrices have an orthonormal eigenbasis.
5. Eigenvectors of Hermitian symmetric matrices corresponding to distinct eigenvalues are necessarily orthogonal.
   (a) Thus, Gram-Schmidt only needs to be applied once to the basis of eigenvectors corresponding to a repeated eigenvalue.
6. For Hermitian symmetric matrices, the process is called **unitary diagonalization**, while for real symmetric matrices it is called **orthogonal diagonalization**.
7. Orthogonal diagonalization is represented by
$$A = UDU^T,$$
   where U is an orthogonal matrix and D is a diagonal matrix.
   (a) Diagonalize A as SDS^{-1}.
   (b) D remains the same.
   (c) The vectors in S corresponding to repeated eigenvalues need to be orthogonalized by Gram-Schmidt and then normalized. The rest of the vectors corresponding to unique eigenvalues simply need to be normalized. The resulting matrix U is then orthogonal.

8. Real symmetric matrices can be written as a linear combination of rank one orthogonal projections. Geometrically, this means that the unit n-sphere defined by u_ℓ is streched in n directions corresponding to each eigenvector by factors defined by the corresponding eigenvalue, creating an ellipsoid.

9. All normal matrices (defined as $AA^H = A^H A$) are unitary diagonalizable, not just Hermitian symmetric matrices.

23 Lesson 23

23.1 Notes

1. Single value decomposition (SVD) is a factorization of a real $m \times n$ matrix A into

$$A = U\Sigma V^T,$$

   where

   - U is an $m \times m$ orthogonal matrix,
   - Σ is an $m \times n$ "diagonal" matrix with positive real entries on the main diagonal in decreasing order (these non-zero diagonal entries are called the **singular values**), and
   - V is an $n \times n$ orthogonal matrix.

2. For a symmetric matrix with positive eigenvalues, the SVD is the same as the orthogonal diagonalization with the eigenvalues arranged in decreasing order.

3. To find the SVD of a real $m \times n$ matrix A:

   (a) Compute the orthogonal diagonalization of $A^T A = VDV^T$ (V is now found).

   (b) Square root the r non-zero values of D to find the $m \times n$ matrix Σ. If any eigenvalue is negative, simply apply the negative to the respective eigenvector, but only in either U or V but not both.

   (c) Letting $r = \text{rank}(A)$ (or equivalently the number of singular values), use

$$\mathbf{u}_\ell = \frac{1}{\sigma_\ell}\mathbf{A}\mathbf{v}_\ell$$

   for $\ell = 1, \ldots, r$ to find the first r columns of U.

   (d) Find the remaining columns of U by finding the basis for $\text{nul}(A^T)$. Use Gram-Schmidt to orthonormalize the basis.

4. There is typically more than one SVD for a given matrix, but the singular values are shared by all of them.