CS 132 – Geometric Algorithms
Summary Sheet

September 28, 2023

For full Lecture Notes, please refer to https://www.cs.bu.edu/fac/crovella/cs132-book/landing-page.html

Summary
L_0&L_1&L_2

1. **Linear Equation:** A linear equation in the variables $x_1, x_2, ..., x_n$ is an equiation that can be written in the form:

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

   where b and the coefficients $a_1, ...a_n$ are real or complex numbers.

2. **System of linear equations or linear system:** is a collection of one or more linear equations involving the same variables.

3. **Solution:** A solution of the system is a list of numbers $(s_1, ..., s_n)$ that makes each equation a true statement when substituted for $x_1, ...x_n$. Note that a system can have multiple solutions.

4. **Solution Set:** The set of all possible solution is the solution set of the linear system. Note that two linear systems are called *equivalent* if they have the same solution set.

5. A system of linear equations may have: no solutions (inconsistent); or exactly one solution (consistent); or infinitely many solutions (consistent).

6. **Vector Space:** $\mathbb{R}^n$ is a n-dimensional space that collects a set of list of numbers $(s_1, ..., s_n)$. If we consider linear equations with n unknowns, the solutions are points in $\mathbb{R}^n$.

7. **Elementary Row Operations:**

   Replacement - replacement one row by the sum of itself and a multiple of another row;

   Interchange - interchange two rows;

   Scaling - multiply all entries in a row by a non zero constant.

   Note that (a) two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other. (b) If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

8. **Consistency** of a linear system: A linear system is consistent if and only if it has one or infinitely many solutions.

9. **Uniqueness** of a linear system: A linear system has a unique solution if it has one and only one solution.

L_3&L_4

1. *Non zero row*: row with at least one nonzero entry (not including the last row in augmented matrix)

2. *Leading entry*: the first nonzero element in a row.

3. **Echelon Form** (or row echelon form) of a matrix:
   a. All non zero rows are above any rows of all zeros.
   b. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
   c. All entries in a column below a leading entry are zeros.
4. **Reduced Echelon Form** (or reduced row echelon form) of a matrix:
   a. The matrix is in echelon form.
   b. The leading entry in each nonzero row is 1.
   c. Each leading 1 is the only non zero entry in its column.
5. (Properties) Echelon forms are *not unique*; the reduced echelon form of a matrix *is unique*.
6. **Theorem:** Each matrix is equivalent to one and only one reduced echelon matrix.
7. **Pivot position:** A pivot position in a matrix A is the position of a leading 1 in the reduced echelon form of A.
8. **Gaussian Elimination:** Given an augmented matrix A representing a linear system
   a. (Elimination stage) Convert A to one of its echelon forms say U.
   b. (Backsubstitution stage) Convert U to A's reduced row echelon form.

   Each step iterates over the rows of A, starting with the first row.
9. The number of operations of Gaussian elimination is $\sim \frac{2}{3}n^3$ for a linear system with n equations and n unknowns. i.e, Gaussian elimination has arithmetic complexity of $\mathscr{O}(n^3)$. More specifically, we have:
   a. The elimination stage is $\mathscr{O}(n^3)$.
   b. The backsubstitution stage is $\mathscr{O}(n^2)$.
10. When a system is consistent, the solution set can be described explicitly by solving the reduced system of equations for the *basic variables* in terms of the *free variables*.
11. **Column vector** or a vector: a column vector is a matrix with only one column. e.g:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2 \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$$

    Two vectors are equal if and only if their corresponding entries are equal.
12. Let $\mathbf{a}, \mathbf{b}$ be two vectors in $\mathbb{R}^n$ and let c be a scalar. Then the sum and scalar multiplication can be defined as

$$\text{Sum: } \mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} \qquad \text{Scalar multiplication: } c\mathbf{a} = \begin{bmatrix} c \cdot a_1 \\ c \cdot a_2 \\ \vdots \\ c \cdot a_n \end{bmatrix}$$

    Note that if the vectors are not in the same size, the sum will be undefined.
13. **Properties**. Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors and let c, d be scalars.
    a. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
    b. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
    c. $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

d. $\mathbf{u}+(-\mathbf{u}) = (-\mathbf{u})+\mathbf{u} = \mathbf{0}$
e. $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}$
f. $(c+d)\mathbf{u} = c\mathbf{u}+d\mathbf{u}$
g. $c(d\mathbf{u}) = (cd)\mathbf{u}$
h. $1\cdot\mathbf{u} = \mathbf{u}$

14. **Linear Combinations:** a linear combinations of vectors $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}$ is: $\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_p\mathbf{v_p}$, where the $c_i \in \mathbb{R}$ are weights, and it can be zero.

15. A vector equation $x_1\mathbf{a_1} + x_2\mathbf{a_2} + ... + x_n\mathbf{a_n} = \mathbf{b}$ has the same solution set as the linear system whose augmented matrix is $\begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & ... & \mathbf{a_n} & \mathbf{b}\end{bmatrix}$

16. **Span**: Suppose that $\mathbf{b}$ can be expressed as a linear combination of $\mathbf{a_1}$ and $\mathbf{a_2}$. That is, there are some x_1, x_2 such that $x_1\mathbf{a_1} + x_2\mathbf{a_2} = \mathbf{b}$. Then $\mathbf{b}$ is in the span of the set of vectors $\{\mathbf{a_1}, \mathbf{a_2}\}$, i.e, $\mathbf{b} \in \text{Span}\{\mathbf{a_1}, \mathbf{a_2}\}$.

    Generally, let $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}$, for each $\mathbf{v} \in\in \mathbb{R}^n$ be vectors, then the set of all linear combinations of $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}$ is denoted by $\text{Span}\{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}\}$, and it is called the subset of $\mathbb{R}^n$ spaned by $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}$.

$\mathbf{L_5}$&$\mathbf{L_6}$

1. If A is an $m \times n$ matrix, with columns $\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_n}$, and if $\mathbf{x} \in \mathbb{R}^n$, then the product of A and $\mathbf{x}$, denoted $A\mathbf{x}$, is the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights; that is,

$$A\mathbf{x} = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & ... & \mathbf{a_n}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = x_1\mathbf{a_1} + x_2\mathbf{a_2} + ... + x_n\mathbf{a_n}$$

   Note that the number of columns of A must match the number of rows of $\mathbf{x}$.

2. If $A = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & ... & \mathbf{a_n}\end{bmatrix}$ and $\mathbf{x} \in \mathbb{R}^n$, then $A\mathbf{x} \in \text{Span}\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$.

3. **Matrix Equation**. If A is an $m \times n$ matrix, with columns $\mathbf{a_1}, \mathbf{a_2}, ...\mathbf{a_n}$, and if $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, then the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the vector equation $x_1\mathbf{a_1} + x_2\mathbf{a_2} + ... + x_n\mathbf{a_n} = \mathbf{b}$, and in turn, has the same solution set as the system of linear equations whose augmented matrix is $\begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & ... & \mathbf{a_n} & \mathbf{b}\end{bmatrix}$. (**$\mathbf{L_3}$&$\mathbf{L_4}$** Point 14, Page 3)

4. Theorem. Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true or they are all false.

   i. For each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.
   ii. Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.
   iii. The columns of A span $\mathbb{R}^m$.
   iv. A has a pivot position in every row.

5. Properties. Let A be an $m \times n$ matrix, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ are vectors, and c is a scalar. Then

   i. $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
   ii. $A(c\mathbf{u}) = c(A\mathbf{u})$

6. **Inner Product**. $\begin{bmatrix}x_1 & x_2 & ... & x_n\end{bmatrix}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} = \sum_{i=1}^{n} x_i y_i$

7. **Linear dependence**. A set of vectors $\mathbf{v_1}, \mathbf{v_2}, ...\mathbf{v_p}$, all of which are in $\mathbb{R}^n$ is said to be linearly dependent if there exist weights $\{c_1, ..., c_p\}$, not all zero, such that $c_1\mathbf{v_1} + ...c_p\mathbf{v_p} = \mathbf{0}$

8. **Linearly Independent**. A set of vectors $\mathbf{v_1}, \mathbf{v_2}, ...\mathbf{v_p}$, all of which are in $\mathbb{R}^n$ is said to be linearly independent if the vector equation has only the trivial solution $c_1 = ... = c_p = 0$

9. Testing if a set of vectors is linearly independent is to determine if there is a nontrivial solution of the vector equation: $x_1\mathbf{a_1} + x_2\mathbf{a_2} + ... + x_n\mathbf{a_n} = \mathbf{0}$

10. If a set contains more vectors than there are entries in each vector, then the set is linearly dependent.

L7

1. **Transformation**: A transformation (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T and $\mathbb{R}^m$ is the **codomain** of T. Denoted as: $T : \mathbb{R}^n \to \mathbb{R}^m$.

2. $\mathbf{x} \in \mathbb{R}^n$ is called the **image** of $\mathbf{x}$ under T. The set of all images $T(\mathbf{x})$ is called the **range** of T.

3. **(*Important)** A transformation T is linear if

   i. $T(\mathbf{u}+\mathbf{v}) = T\mathbf{u} + T\mathbf{v}$, $\forall \mathbf{u}, \mathbf{v} \in \text{Domain}(T)$
   ii. $T(c\mathbf{u}) = c(T\mathbf{u})$ for all scalars c and $\forall \mathbf{u} \in \text{Domain}(T)$

   Facts: $T(\mathbf{0}) = \mathbf{0}$ and $T(c\mathbf{u}+d\mathbf{v}) = cT\mathbf{u} + dT\mathbf{v}$

L8&L9

1. **Theorem:** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, then there is always a unique matrix A such that: $T(\mathbf{x}) = A\mathbf{x} \quad \forall \mathbf{x} \in \mathbb{R}^n$.

2. **Standard matrix**. Full table please see textbook Sec 1.9 Table 1-4.

3. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at least one $\mathbf{x}$ in $\mathbb{R}^n$. Alternatively, T is onto if there is a solution $\mathbf{x}$ of $T(\mathbf{x}) = \mathbf{b}$ for all possible $\mathbf{b} \in \mathbb{R}^m$.

4. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at most one $\mathbf{x}$ in $\mathbb{R}^n$.

5. If $T\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b}$, then T is onto but may not be one-to-one. If $T\mathbf{x} = \mathbf{b}$ is consistent and has a unique solution for all $\mathbf{b}$, then T is both one-to-one and onto.

6. **Matrix Multiplication**: If A is an $m \times n$ matrix and B is $n \times p$ matrix with columns $\mathbf{b_1}, \mathbf{b_2}, ..., \mathbf{b_p}$, then the product AB is defined as the $m \times p$ matrix whose columns are $A\mathbf{b_1}, A\mathbf{b_2}, ..., A\mathbf{b_p}$. That is, $AB = \begin{bmatrix} A\mathbf{b_1} & ... & A\mathbf{b_p} \end{bmatrix}$

7. **Matrix Algebra**. Let A and B be $m \times n$ matrices, and let r and s be scalars. Then we have:

   i. $A + B = B + A$
   ii. $(A+B) + C = A + (B+C)$
   iii. $A + 0 = A$
   iv. $r(A+B) = rA + rB$
   v. $(r+s)A = rA + sA$
   vi. $r(sA) = (rs)A$

   Note that $A + B$ is not defined is A and B are not in the same shape.

8. More properties. Suppose that all sums and products are defined. Then we have:

   i. $A(BC) = (AB)C$
   ii. $A(B+C) = AB + AC$
   iii. $(B+C)A = BA + CA$
   iv. $r(AB) = (rA)B = A(rB)$

v. $IA = A = AI$

9. **Transpose** of a matrix: Given an $m \times n$ matrix A, the *transpose* of A is the matrix we get by interchanging its rows and columns.

10. Rules for Transposes:

    i. $(A^T)^T = A$
    ii. $(A+B)^T = A^T + B^T$
    iii. $(rA)^T = r(A^T)$
    iv. $(AB)^T = B^T A^T$

11. $\mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$.

L_{10}&L_{11}

1. **Inverse** of a matrix. A matrix A is called **invertible** if there exists a matrix C such that $AC = I$ and $CA = I$. The inverse of A is denoted as A^{-1}.

2. **Singular** matrix. a matrix that is not invertible is called a singular matrix.

3. Theorem. If A is an invertible $n \times n$ matrix, then for each $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

4. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

5. **Determinate**. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. $\det A = ad - bc$

6. Given a 2×2 matrix A, if the columns of A are linearly dependent, then A is NOT invertible.

7. Compute the inverse of A: solve the linear system $A\mathbf{x_i} = \mathbf{e_i}$ to get the ith column of A^{-1} for all $i \in [1, n]$.

8. Properties.

    i. If A is an invertible matrix, then A^{-1} is also invertible, and $(A^{-1})^{-1} = A$.
    ii. If A is an invertible matrix, then A^T is also invertible, and $(A^T)^{-1} = (A^{-1})^T$.
    iii. If A and B are $n \times n$ invertible matrices, then AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

9. **Invertible Matrix Theorem** (IMT). Let A be a square $n \times n$ matrix. Then the following statements are equivalent; that is, they are either all true or all false.

    i. A is an invertible matrix.
    iff ii. A^T is an invertible matrix.
    iff iii. The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.
    iff iv. A is row equivalent to the identity matrix.
    iff v. A has n pivot positions.
    iff vi. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
    iff vii. The columns of A form a linearly independent set.
    iff viii. The columns of A span $\mathbb{R}^n$.
    iff ix. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
    iff x. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

10. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S(T(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}$

11. Theorem. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix.

12. The state of the system at time k is a vector $\mathbf{x_k} \in \mathbb{R}^n$ (*state vector*), and $\mathbf{x_{k+1}} = T(\mathbf{x_k})$ for time $k = 0, 1, 2...$ where $T : \mathbb{R}^n \to \mathbb{R}^n$.

13. Definitions. Markov Chains.

    i. A *probability vector* is a vector of nonnegative entries that sums to 1.
    ii. A *stochastic matrix* is a square matrix of nonnegative values whose columns each sum to 1.
    iii. A *Markov chain* is a dynamical system whose state is a probability vector and which evolves according to a stochastic matrix.

    That is, it is a probability vector $\mathbf{x_0}$ and a stochastic matrix $A \in \mathbb{R}^{n\times n}$ such that $\mathbf{x_{k+1}} = A\mathbf{x_k}$ for time $k = 0, 1, 2...$

14. If P is a stochastic matrix, then a steady-state vector (or equilibrium vector) of P is a probability vector $\mathbf{q}$ such that $P\mathbf{q} = \mathbf{q}$. Note that every stochastic matrix has a steady-state vector.

15. Solve a Markov Chain for its steady state is to find $\mathbf{q}$ such that $P\mathbf{q} = \mathbf{q}$:

    step 1: Solve the linear system $(P - I)\mathbf{q} = \mathbf{0}$
    step 2: Obtain a general solution. (free variable)
    step 3: Pick any specific solution and normalize (so that the entries add up to 1).

16. A stochastic matrix is regular if some matrix power P^k contains only strictly positive entries.

17. If P is an $n \times n$ regular stochastic matrix, then P has a unique steady-state vector $\mathbf{q}$. Further, if $\mathbf{x_0}$ is any initial state and $\mathbf{x_{k+1}} = P\mathbf{x_k}$ for $k = 0, 1, 2...$, then the Markov Chain $\{\mathbf{x_k}\}$ converges to $\mathbf{q}$ as $k \to \infty$. Note: Memoryless property of Markov Chain: it converges to its steady-state vector no matter what state the chain starts in.

$\mathbf{L_{12}}$&$\mathbf{L_{13}}$

1. *Factorization*. A factorization of a matrix A is an equation that expresses A as a product of two or more matrices $A = BC$.

2. Every elementary row operation on A can be performed by multiplying A by a suitable matrix.

3. *Elementary matrix* is one that is obtained by performing a single elementary row operation on the identity matrix.

4. **LU Factorization**. $A = LU$ where L is a lower triangular matrix and U is an upper triangular matrix. i.e $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$

5. Solve $A\mathbf{x} = \mathbf{b}$ is equivalent to solving $L(U\mathbf{x}) = \mathbf{b} \Rightarrow U\mathbf{x} = L^{-1}\mathbf{b} \Rightarrow \mathbf{x} = U^{-1}(L^{-1}\mathbf{b})$. With LU factorization, it is easier to solve for $\mathbf{x}$ because each matrix is triangular.

6. **3D Scaling**: $\begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. **3D Translation**: $\begin{bmatrix} 1 & 0 & 0 & h \\ 0 & 1 & 0 & k \\ 0 & 0 & 1 & m \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

    3D Rotation: $R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $R_y(\alpha) = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha & 0 \\ 0 & 1 & 0 & 0 \\ \sin\alpha & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $R_z(\alpha) = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 & 0 \\ -\sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

    Perspective Projection: $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d & 1 \end{bmatrix}$.

7. Composition of matrices. $C = AB$ first apply B and then apply A.

$\mathbf{L_{14}}$&$\mathbf{L_{15}}$

1. **Subspace**. A subspace is any set H in $\mathbb{R}^n$ that has three properties:
   i. The zero vector is in H.
   ii. For each $\mathbf{u}$ and $\mathbf{v}$ that are in H, the sum $\mathbf{u}+\mathbf{v}$ is also in H.
   iii. For each $\mathbf{u}$ in H and each scalar c, $c\mathbf{u}$ is also in H.
2. **Column space**. The column space of a matrix A is the set $\mathrm{Col}(A)$ of all linear combinations of the columns of A. If $A = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n}\end{bmatrix}$, then $\mathrm{Col}(A)$ is the same as $\mathrm{Span}\{\mathbf{a_1}, \mathbf{a_2}, \dots, \mathbf{a_n}\}$
3. The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$.
4. **Null space**. The null space of a matrix A is the set $\mathrm{Nul}(A)$ of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.
5. The null space of an $m \times n$ matrix is a subspace of $\mathbb{R}^n$.
6. **Basis for subspace**. Let V be a vector space and $\beta = \{u_1, u_2, \dots, u_n\}$ be a subset of V. Then β is a basis for V if and only if each $v \in V$ can be uniquely expressed as a linear combination of vectors in β, that is can be expressed in the form: $v = a_1u_1 + a_2u_2 + \dots + a_nu_n$ for unique scalars $a_1, a_2, \dots, a_n$. i.e A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.
7. **Basis for null space**. Constructing a basis for the null space of A is equivalent to finding a parametric description of the solution of the equation $A\mathbf{x} = \mathbf{0}$.
8. **Basis for column space**. The pivot columns of a matrix A form a basis for the column space of A.
9. A basis provides a coordinate system for H. If we are given a basis for H, then each vector in H can be written in only one way as a linear combination of the basis vectors.
10. Suppose the set $\beta = \{\mathbf{b_1}, \mathbf{b_2}, \dots, \mathbf{b_p}\}$ is a basis for the subspace H. For each $\mathbf{x} \in H$, the coordinates of $\mathbf{x}$ relative to the basis β are the weights $c_1, \dots, c_p$ such that $\mathbf{x} = c_1\mathbf{b_1} + c_2\mathbf{b_2} + \dots + c_p\mathbf{b_p}$
11. $[\mathbf{x}]_\beta = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \in \mathbb{R}^p$ is called the coordinate vector of $\mathbf{x}$ (relative to β) or the β-coordinate vector of $\mathbf{x}$.
12. The dimension of a nonzero subspace H, denoted by $\dim(H)$, is the number of vectors in any basis for H.
13. The dimension of the zero subspace $\{\mathbf{0}\}$ is defined to be zero.
14. The vector space $M_{m\times n}(F)$ has dimension mn.
15. Find dimension of $\mathrm{Nul}(A)$ is equivalent to find the number of free variables in $A\mathbf{x} = \mathbf{0}$.
16. The rank of a matrix, denoted by $\mathrm{rank}(A)$, is the dimension of the column space of A: i.e number of pivot columns in A.
17. **The Rank Theorem**: If a matrix A has n columns, then $\mathrm{rank}(A) + \dim(\mathrm{Nul}(A)) = n$.
18. (Extension to IMT)
    xi. The columns of A form a basis for $\mathbb{R}^n$.
    xii. $\mathrm{Col}(A) = \mathbb{R}^n$.
    xiii. $\dim(\mathrm{Col}(A)) = n$
    xiv. $\mathrm{rank}(A) = n$

xv. $\mathrm{Nul}(A) = \{\mathbf{0}\}$

xvi. $\dim(\mathrm{Nul}(A)) = 0$

19. Let W be a subspace of a finite-dimensional vector space V. Then W is finite-dimensional and $\dim(W) \leq \dim(V)$. Moreover, if $\dim(W) = \dim(V)$, then $V = W$.

$\mathbf{L_{16}}$&$\mathbf{L_{17}}$

1. **Eigenvector**. An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ.

2. **Eigenvalue**. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$.

3. Such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.

4. λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

5. **Eigenspace**. Note that the set of all solution of the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ is the null space of the matrix $A - \lambda I$. So the set of all eigenvectors corresponding to a particular λ is a subspace of $\mathbb{R}^n$, and is called the **eigenspace** of A corresponding to λ. Computing the corresponding eigenspace can be done by constructing a basis for $\mathrm{Nul}(A - \lambda I)$.

6. Theorem. The eigenvalues of a triangular matrix are the entries on its main diagonal.

7. 0 is an eigenvalue of A if and only if A is not invertible.

8. Find the eigenvalues of A is equivalent to find all scalars λ such that $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution, and it is equivalent to finding all λ such that $A - \lambda I$ is not invertible, that is when $\det(A - \lambda I) = 0$.

9. Let A be an $n \times n$ matrix, then A is invertible if and only if

   xvii. The number 0 is not an eigenvalue of A.

   xviii. The determinant of A is not zero.

10. Properties. Determinants.

    i. $\det(AB) = (\det A)(\det B)$

    ii. $\det A^T = \det A$

    iii. If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A.

11. **Characteristic equation**. $\det(A - \lambda I) = 0$.

12. λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation.

13. **Characteristic polynomial**. For any $n \times n$ matrix A, $\det(A - \lambda I)$ is a polynomial of degree n, called the **characteristic polynomial** of A.

14. **Similarity**. If A and B are $n \times n$ matices, then A **is similar to B** if there is an invertible matrix P such that $P^{-1}AP = B$, or equivalently, $A = PBP^{-1}$.

15. If A is similar to B, then B is similar to A.

16. **Similarity transformation**. Changing A into B is called similarity transformation.

17. If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial, and hence the same eigenvalues (with the same multiplicities).

18. Theorem. The largest eigenvalue of a Markov Chain is 1.

19. **Determinant**. *Click for MIT Notes Link*

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

$$\begin{aligned} \det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} &= a \cdot \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \cdot \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \\ &= aei + bfg + cdh - ceg - bdi - afh \end{aligned}$$

   i. $\det I_n = 1$.

   ii. If you exchange two rows of a matrix, you reverse the sign of its determinant from positive to negative or from negative to positive.
   Extension: Let A be an $n \times n$ matrix, and let U be any echelon form obtained from A by row replacements and row interchanges. Let r be the number of row interchanges, then the determinant of A is $(-1)^r$ times the product of the diagonal entries in U.

$$\det A = \begin{cases} (-1)^r \cdot \text{product of povots in U} & \text{if } A \text{ is invertible} \\ 0 & \text{if } A \text{ is not invertible} \end{cases}$$

   iii. If we multiply one row of a matrix by t, the determinant is multiplied by t.

$$\det \begin{bmatrix} ta & tb \\ c & d \end{bmatrix} = t \cdot \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

   iv. The determinant behaves like a linear function on the rows of the matrix:

$$\det \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{bmatrix} = \det \begin{bmatrix} a_1 & b_1 \\ c & d \end{bmatrix} + \det \begin{bmatrix} a_2 & b_2 \\ c & d \end{bmatrix}$$

   v. Property i gives $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1$, and property ii gives $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1$

L_{18}&L_{19}

1. Given a square matrix A, the factorization of A is of the form $A = PDP^{-1}$, where D is a diagonal matrix. Note that A and D are similar. This factorization allows:

   i. represent A in a form that exposes the properties of A.

   ii. represent A^k in an easy to use form

   iii. compute A^k efficiently with large values of k.

2. **Powers of diagonal matrix**. If $D = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, then $D^k = \begin{bmatrix} a^k & 0 \\ 0 & b^k \end{bmatrix}$

3. For $k \geq 1$, $A^k = PD^kP^{-1}$, provided that A is diagonalizable.

4. **Diagonalizable matrix**. A square matrix A is **diagonalizable** if A is similar to a diagonal matrix. That is, if there exists some invertible matrix P such that $A = PDP^{-1}$ and D is a diagonal matrix.

5. Theorem. **The Diagonalization Theorem**. An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

   In fact, $A = PDP^{-1}$ with D a diagonal matrix if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

   In other word, A is diagonalizable if and only there are enough eigenvectors to form a basis of $\mathbb{R}^n$.

   We call such a basis an **eigenvector basis** or an **eigenbasis** of $\mathbb{R}^n$.

6. **Diagonalize an $n \times n$ matrix**.

   step i. Find the eigenvalues of A.

   step ii. Find n linealry independent eigenvectors of A.

   step iii. Construct P from the vectors.

   step iv. Construct D from the corresponding eigenvalues.

7. Theorem. An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

8. **Random Walk.** Let G be a graph. A random walk models the movement of an object on G.

9. Random walk with absorbing boundaries: the walker remains fixed at the location.

10. Random walk with reflecting boundaries: the walker bounces back one unit when an endpoint is reached.

11. Please consult *https://www.cs.bu.edu/fac/crovella/cs132-book/L19PageRank.html* for examples on random walk on directed and undirected graph: transition matrix, steady-state probability vector (undirected), computing PageRank, etc.

L_{20}&L_{21}

1. Properties of inner product.

   i. Inner product is symmetric. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$

   ii. Inner product is linear in each term.

$$(\mathbf{u}+\mathbf{v})^T\mathbf{w} = \mathbf{u}^T\mathbf{w} + \mathbf{v}^T\mathbf{w}$$
$$(c\mathbf{u})^T\mathbf{v} = c(\mathbf{u}^T\mathbf{v}) = \mathbf{u}^T(c\mathbf{v})$$

   iii. Inner product of a vector with itself is never negative. That is, $\mathbf{u}^T\mathbf{u} \geq 0$, and $\mathbf{u}^T\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

2. **Norm of a vector**. The *norm* of $\mathbf{v}$ is the non-negative scalar $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T\mathbf{v}} = \sqrt{\sum_{i=1}^n v_i^2}$

3. For any scalar c, $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.

4. **Cauchy-schwarz Inequality**. $|\mathbf{u}^T\mathbf{v}| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|$

5. **Triangle Inequality**. $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

6. **Normalize a vector**. Normalized vector $\mathbf{u}$ for $\mathbf{v}$ is a unit vector and is defined as $\mathbf{u} = \dfrac{\mathbf{v}}{\|\mathbf{v}\|}$.

7. **Distance**. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the *distance* between the two vectors, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is, $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2 + \ldots + (v_n - u_n)^2}$

8. **Orthogonal vectors**. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are *orthogonal* to each other if $\mathbf{u}^T\mathbf{v} = 0$

9. **Angle between vectors**. $\mathbf{u}^T\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$

10. **Orthogonal Set**. A set of vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ in $\mathbb{R}^n$ is an *orthogonal set* if each pair of distinct vectors from the set is orthogonal. i.e. $\mathbf{u_i}^T\mathbf{u_j} = 0$ whenever $i \neq j$

11. Theorem. If $S = \{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent. (proof in lecture notes)

12. **Orthogonal basis**. An *orthogonal basis* for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

13. Let $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y} \in W$, the weights of the linear combination $\mathbf{y} = c_1\mathbf{u_1} + c_2\mathbf{u_2} + \ldots + c_p\mathbf{u_p}$ are given by $c_j = \frac{\mathbf{y}^T\mathbf{u_j}}{\mathbf{u_j}^T\mathbf{u_j}}$ for $j = 1, \ldots, p$.

14. **Orthogonal projection**. $\text{proj}_L \mathbf{y} = \hat{\mathbf{y}} = \frac{\mathbf{y}^T\mathbf{u}}{\mathbf{u}^T\mathbf{u}}\mathbf{u}$ is the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ (or onto subspace L). $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ is called the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$.

15. Let $\{\mathbf{u_1}, ..., \mathbf{u_p}\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. We can decomposite $\mathbf{y} = c_1\mathbf{u_1} + c_2\mathbf{u_2} + ... + c_p\mathbf{u_p}$ into a sum of orthogonal projections onto one-dimensional subspaces. That is,

$$\mathbf{y} = \frac{\mathbf{y}^T\mathbf{u_1}}{\mathbf{u_1}^T\mathbf{u_1}}\mathbf{u_1} + \frac{\mathbf{y}^T\mathbf{u_2}}{\mathbf{u_2}^T\mathbf{u_2}}\mathbf{u_2} + ... + \frac{\mathbf{y}^T\mathbf{u_p}}{\mathbf{u_p}^T\mathbf{u_p}}\mathbf{u_p}$$

16. **Orthonormal Set**. A set of vectors $\{\mathbf{u_1}, ..., \mathbf{u_p}\}$ in $\mathbb{R}^n$ is an *orthonormal set* if it is an orthogonal set of unit vectors.

17. Theorem. An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

18. Theorem. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

    i. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
    ii. $(U\mathbf{x})^T(U\mathbf{y}) = \mathbf{x}^T\mathbf{y}$
    iii. $(U\mathbf{x})^T(U\mathbf{y}) = 0$ if and only if $\mathbf{x}^T\mathbf{y} = 0$

L_{22}&L_{23}

1. *General lease-squares problem.* To find an $\mathbf{x}$ that minimize $\|A\mathbf{x} - \mathbf{b}\|$ (or $\|A\mathbf{x} - \mathbf{b}\|^2$)

2. **Least square solution**. If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a least squares solution of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that $\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$ for all $\mathbf{x} \in \mathbb{R}^n$. Or equivalently, $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|$.

3. Geometrically, it is to find the closest point in Col(A) to $\mathbf{b}$.

4. *Solve the General Least Square Problem.*

    - **The Orthogonal Decomposition Theorem**. Let $W \subseteq \mathbb{R}^n$. Then $\forall \mathbf{y} \in \mathbb{R}^n$, $\exists \hat{\mathbf{y}} \in W$, and $\mathbf{z}$ orthogonal to every vector in W, such that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$. That is, $\hat{\mathbf{y}} = \text{proj}_W(\mathbf{y})$, and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.
    - **The Best Approximation Theorem**. Let $W \subseteq \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. Let $\hat{\mathbf{y}} = \text{proj}_W(\mathbf{y})$. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$. That is, $\|\mathbf{y} - \hat{\mathbf{y}}\|$ is minimal, and $\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{v}\| \;\; \forall \mathbf{v} \in W$. (proof in lecture notes)
    - Finding $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|$ is equivalent to finding the projection of $\mathbf{b}$ onto Col(A), i.e, $\hat{\mathbf{b}} = \text{proj}_{\text{Col}(A)}\mathbf{b}$, and solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$

5. Theorem. The set of least squares solutions of $A\mathbf{x} = \mathbf{b}$ is equal to the (nonempty) set of solutions of the **normal equations** $A^TA\mathbf{x} = A^T\mathbf{b}$.

    Recap (general steps) and proof:

    - Goal: find $\hat{\mathbf{x}}$ to minimize $\|A\mathbf{x} - \mathbf{b}\|$.
    - Orthogonal Decomposition. $\mathbf{b} = \hat{\mathbf{x}} + \mathbf{z}$, where $\hat{\mathbf{x}} \in \text{Col}(A)$ and $\mathbf{z} = \mathbf{b} = \hat{\mathbf{b}}$ is orthogonal to Col(A).
    - Best Approximation. $\hat{\mathbf{b}} = \text{proj}_{\text{Col}(A)}(\mathbf{b})$ is the closest point in Col(A) of $\mathbf{b}$.
    - Solve. (Assume such $\hat{\mathbf{x}}$ exists.) To find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

$$\begin{aligned}
A\hat{\mathbf{x}} &= \hat{\mathbf{b}} \\
A\hat{\mathbf{x}} - \mathbf{b} &= \hat{\mathbf{b}} - \mathbf{b} \\
A^TA\hat{\mathbf{x}} - A^T\mathbf{b} &= A^T(\hat{\mathbf{b}} - \mathbf{b}) \\
A^TA\hat{\mathbf{x}} - A^T\mathbf{b} &= 0 && \text{Since } \hat{\mathbf{b}} - \mathbf{b} \text{ is orthogonal to Col}(A) \\
A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \\
\hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} && \text{Solve for } \hat{\mathbf{x}}
\end{aligned}$$

6. When $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ has multiple solutions, the columns of A are linearly independent.

7. Theorem. Let A be an $m \times n$ matrix. The following statements are equivalent:

    i. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution for each $\mathbf{b} \in \mathbb{R}^m$.

    ii. The columns of A are linearly independent.

    iii. The matrix A^TA is invertible.

    When these statements are true, the least squares solution $\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$. Since $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, we have $\hat{\mathbf{b}} = A(A^TA)^{-1}A^T\mathbf{b}$

8. **Linear Models**.

    i. The independent variables are collected into a matrix X, called **design matrix**.

    ii. The dependent variables are collected into an **observation** vector $\mathbf{y}$.

    iii. The parameters of the model are collected into a **parameter** vector β.

    Goal: Find the linear parameters β for some $\mathbf{y}$ to model the data X, and minimize $\|X\beta - \mathbf{y}\|$.

9. A least squares problem for linear model: $X\beta = \mathbf{y}$, where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$