Linear Algebra Study Guide

Contents

Linear Equations

System of Linear Equations & Vectors

Basic elementary row operations for solving a system of linear equations (which can be written as: $b = a_1x_1 + \cdots + a_nx_n$):

- Replace 1 equation by the sum of itself and a multiple of another equation.
- Interchange 2 equations.
- Multiply all the terms in an equation by a nonzero constant.

Properties of Echelon Form:

- All nonzero rows are above any rows of all zeros.
- Each leading entry of a row is in a column to the right of the leading entry of the row above it.
- All entries in a column below a leading entry are zeros.

➔ If a matrix in echelon form satisfies the following additional conditions, then it's in reduced echelon form:
  - The leading entry in each nonzero row is 1.
  - Each leading 1 is the only nonzero entry in its column.

➔ *Pivot position* is a position in a matrix that has the leading 1 → *pivot column* is column that has a pivot position.

Solving Linear System:

(1) Perform row reduction algorithm to the augmented matrix to obtain reduced echelon form.

(2) If the system is *consistent*, write the system of equations corresponding to the matrix obtained in step (1).

(3) Rewrite each nonzero equation from step (2) so that its one basic variable is expressed in terms of any free variables appearing in the equation → this is to obtain the *parametric* descriptions of the solution set.

Definitions Relating to Vectors:

- A set of all vectors with n entries is denoted by $\mathbb{R}^n$.
- Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_\mathrm{p}$ in $\mathbb{R}^n$ and given scalars $c_1, \ldots, c_p$, then the vector $\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ is called a ***linear combination*** of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ with *weights* $c_1, \ldots, c_p$.
- Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ (i.e. it's the set of all vectors that can be written in the form $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ with $c_1, \ldots, c_p$ scalars) → remember that the zero vector ($\mathbf{0}$) is part of all spans.
  - Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^3$, then Span$\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$ (i.e. a line through the origin) .
  - Let $\mathbf{u}$ & $\mathbf{v}$ be nonzero vectors in $\mathbb{R}^3$, then Span$\{\mathbf{u}, \mathbf{v}\}$ represents a plane in $\mathbb{R}^3$ that passes through the origin.
- Let A be an $m \times n$ matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$:

$$A\mathbf{x} = [\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$$

- Let A be an $m \times n$ matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and let B be an $n \times p$ matrix with column vectors $\mathbf{b}_1, \ldots, \mathbf{b}_p$:

$$AB = A[\mathbf{b}_1 \quad \ldots \quad \mathbf{b}_n] = [A\mathbf{b}_1 \quad \ldots \quad A\mathbf{b}_n]$$

Matrix Equations

General Definitions:

- Identity matrix of size $n \times n$ (denoted I_n) is a matrix with all 1's on the diagonal and 0's elsewhere.
- Matrix equation can be written in the form $A\mathbf{x} = \mathbf{b}$, which is equivalent to the vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$, which is also equivalent to the augmented matrix of a system of linear equation $[\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n \quad \mathbf{b}]$.

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

(1) For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot position in every ***row*** when it's written in reduced echelon form.

Homogeneous Linear Systems:

- A matrix equation is *homogeneous* if it can be written as $A\mathbf{x} = \mathbf{0}$, which ***always*** has at least one ***trivial*** solution ($\mathbf{x} = \mathbf{0}$).
- A consistent system can either have one unique solution or an infinite no. of solutions → since homogeneous system always has a trivial solution, it can only have nontrivial solutions if and only if there's an infinite no. of solutions.

Nonhomogeneous Linear Systems:

- Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and let $\mathbf{p}$ be a particular solution. Then the solution set of is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Note that in practice, it's much easier to just solve the augmented matrix and write the solution set as a vector equation.

Linear Dependence: A homogeneous system, described by an indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$, is linearly dependent if and only if the vector equation $x_1\mathbf{v}_1 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$ has nontrivial solutions (i.e. at least one of the x_i is nonzero).

- A set of 2 vectors is linearly dependent if one of them is a multiple of the other.
- If a set contains a zero vector, then the set is linearly dependent.
- If a set contains more vectors than there are entries in each vector, then the set is linearly dependent.
    - *Proof*: Let $A = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_p]$ (i.e. $n \times p$ matrix). Then the equation $A\mathbf{x} = \mathbf{0}$ corresponds to a system of n equations with p unknowns. If $p > n$, then there are more variables than equations, so there must be a free variable. Hence $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, making the set to be linearly dependent.

Linear Transformations → A transformation (or mapping) T is called a ***linear*** transformation if:

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$ & $\mathbf{v}$ in the domain of T.
- $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c & all $\mathbf{u}$ in the domain of T.

➔ Every matrix transformation $T(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$ is a linear transformation → matrix multiplication is a simple way to transform a set of points (represented by a set of vectors) into another set of points → useful in computer graphics.

Mapping Properties:

- A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be ***onto*** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\mathbf{x}$ in $\mathbb{R}^n$ (i.e. T is onto $\mathbb{R}^m$ when the range of T is the entire of the codomain $\mathbb{R}^m$).
- A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be ***one-to-one*** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\mathbf{x}$ in $\mathbb{R}^n$ → note that T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution (i.e. only a single vector can be mapped to $\mathbf{0}$) → let A be the standard matrix for T, then for $T(\mathbf{x}) = \mathbf{0}$ to have only trivial solution, the columns in A must be linearly independent.

Reflection over x- and y-axis	$\begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}$
Reflection over the line $y = \pm x$	$\begin{bmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{bmatrix}$
Reflection through the origin	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$
All points projected onto x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
All points projected onto y-axis	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
Horizontal resizing ($0 < k < 1$ for contraction, & $k > 1$ for expansion)	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical resizing ($0 < k < 1$ for contraction, & $k > 1$ for expansion)	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$
Horizontal shear	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

Suppose that we have a set of points on the xy-plane that together make up some form of shape (e.g. rectangle), then the table on the left describes some common linear transformations that can be applied to this set of points.

In order to find *standard* matrix A for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, it's best to study the effects that T has on the column vectors of the identity matrix $I_n = [\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n]$. If T has the mapping $\mathbf{e}_i \mapsto \hat{\mathbf{e}}_i$, then it's obvious that:

$$A = [T(\mathbf{e}_1) \quad \ldots \quad T(\mathbf{e}_n)] = [\hat{\mathbf{e}}_1 \quad \ldots \quad \hat{\mathbf{e}}_n]$$

Generalization of Linear Transformations:

- A linear transformation T from vector space V into vector space W maps each vector $\mathbf{x}$ in V to a unique vector in W
  - $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in V.
  - $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in V and all scalars c.
- ***Kernel*** of T is a set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$ → for matrix transformation, the kernel is the null space.
- ***Range*** of T is a set of all vectors in W of the form $T(\mathbf{x})$ → for matrix transformation, the range is the column space.

Matrix Algebra

Standard Matrix Theorems:

- $A(BC) = (AB)C$ (assosiative).
- $A(B + C) = AB + AC$ (left distributive law).
- $(B + C)A = BA + CA$ (right distributive law).
- $AB \neq BA$ (non-commutative).
- $I_n = A^0 = B^0 = C^0$ (for any $n \times n$ matrices A, B, and C).
- $A(B\mathbf{x}) = (AB)\mathbf{x}$, which can be proven using the properties of linear transformations as follow: Let A be $m \times n$ matrix, B be $n \times p$ matrix, and $\mathbf{x}$ in $\mathbb{R}^p$, then $A(B\mathbf{x}) = A\left(x_1\mathbf{b}_1 + \cdots + x_p\mathbf{b}_p\right) = x_1(A\mathbf{b}_1) + \cdots x_p\left(A\boldsymbol{b}_\text{p}\right) = (AB)\mathbf{x}$.
- Cancellation law generally doesn't hold for matrix algebra, i.e. if $AB = AC$, it's not always true that $B = C$.

Transposition of Matrix Theorems:

- $(A+B)^T = A^T + B^T$.
- $(AB)^T = B^T A^T$, which can be proven using the matrices' entries as follow:
  - $(AB)^T_{ij} = (AB)_{ji} = (A)_{j0}(B)_{0i} + \cdots + (A)_{jn}(B)_{ni} = (B)_{0i}(A)_{j0} + \cdots + (B)_{ni}(A)_{jn}$.
  - $(B^T A^T)_{ij} = (B^T)_{i0}(A^T)_{0j} + \cdots + (B^T)_{in}(A^T)_{nj} = (B)_{0i}(A)_{j0} + \cdots + (B)_{ni}(A)_{jn}$.

Invertible Matrix Theorems:

- If a ***square*** matrix A is invertible, then there must exist a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$.
- A^{-1} is ***unique*** for a particular matrix A → this can be proven using contradiction:
  - Let B be another inverse matrix of A different from A^{-1}, then $B = BI_n = B(AA^{-1}) = (BA)A^{-1} = I_nA^{-1} = A^{-1}$, which contradicts our assumption that B is different from A^{-1}.
- For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, if $\det A = ad - bc \neq 0$, then A is invertible with $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.
- $(AB)^{-1} = B^{-1}A^{-1}$ and $(A^T)^{-1} = (A^{-1})^T$.

General Algorithm to Find A^{-1}:

- *Fact*: If an elementary row operation is performed on an $n \times n$ matrix A, the resulting matrix can be written as EA, where $n \times n$ matrix E is created by performing the ***same*** row operation on I_n.
- Let P be an $n \times n$ matrix which can be reduced to I_n after the elementary row operation $\tilde{R}$. Let Q be an $n \times n$ matrix which is created by applying the elementary row operation $\tilde{R}$ on I_n. Then by the fact above, the result matrix after applying $\tilde{R}$ on P can be written as $EA = QP = I_n$ (similarly $PQ = I_n$) → P & Q are inverse of each other. Hence, if A is invertible, then it's possible to reduce from I_n to A^{-1} using the ***same*** row operations as when we reduce from A to I_n.

➔ To find A^{-1}, row reduce the augmented matrix $[A \quad I_n]$. If A is row equivalent to I_n, then $[A \quad I_n]$ can be reduced to $[I \quad A^{-1}]$. Otherwise, A does not have an inverse.

➔ Note that since $AA^{-1} = A[\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n] = I_n = [\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n]$, if it's only required to figure out one single column $\mathbf{x}_i$ in the inverse matrix A^{-1}, then it's much simpler to just solve the equation $A\mathbf{x}_i = \mathbf{e}_i$.

LU Factorization

LU factorization is used for computing optimization when solving a *series* of equations of the form $A\mathbf{x} = \mathbf{b}$ (all with the same matrix A). Let $A = LU$, then $A\mathbf{x} = (LU)\mathbf{x} = L(U\mathbf{x}) = L\mathbf{y} = \mathbf{b}$. By first solving for $\mathbf{y}$ (using L), and then use that to solve for $\mathbf{x}$ (using U), the computation would be much faster compared to row reduction algorithm for each of the equations.

LU* Factorization Algorithm** (L for ***lower triangular matrix, U for ***upper*** triangular matrix):

$$A = \begin{bmatrix} \mathbf{-3} & 1 & 2 & -3 \\ \mathbf{9} & -3 & -4 & 10 \\ \mathbf{-12} & 4 & 12 & -5 \end{bmatrix} \sim \begin{bmatrix} -3 & 1 & 2 & -3 \\ 0 & 0 & \mathbf{2} & 1 \\ 0 & 0 & \mathbf{4} & 7 \end{bmatrix} \sim \begin{bmatrix} -3 & 1 & 2 & -3 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & \mathbf{5} \end{bmatrix} = U$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{-3} & 0 & 0 \\ \mathbf{9} & \mathbf{2} & 0 \\ \mathbf{-12} & \mathbf{4} & \mathbf{5} \end{bmatrix} \xrightarrow{\text{divide by the top pivot entries}} \begin{bmatrix} \mathbf{1} & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 \\ \mathbf{4} & \mathbf{2} & \mathbf{1} \end{bmatrix} = L$$

(1) Suppose that A can be row reduced to an *echelon form* U using just a sequence of elementary row replacement operations that adds a multiple of one row to another row ***below*** it.

(2) Then there must exist a sequence of unit lower elementary matrices $E_p, \dots, E_1$ such that $(E_p \dots E_1)A = U$. Hence, we have $A = (E_p \dots E_1)^{-1}U = LU$, where $L = (E_p \dots E_1)^{-1}$ can be proven to be a unit lower triangular matrix.

(3) As L & $(E_p \dots E_1)$ are inverse of each other, applying same elementary row operations on L will give identity matrix I.

➔ To compute U, reduce A to an echelon form by a sequence of row replacement operations (but only on the rows ***below*** the current rows). If that's not possible, then LU factorization doesn't exist for A.

➔ To compute L, place entries in L such that the *same sequence of row operations* reduces L to I (see the example).

➔ If U contains rows with only zero entries, these rows can be removed. But make sure the dimensions of L & U are correct.

Partitioned Matrices

From top-down view, partitioned matrix is a matrix whose entries are blocks (or submatrices). From bottom-up view, the entries of a partitioned matrix are simply scalars, and blocks are divided using separating lines.

$$A = \begin{bmatrix} \boxed{A}_{11} & \boxed{A}_{12} \\ \boxed{A}_{21} & \boxed{A}_{22} \end{bmatrix}, \text{ where } \boxed{A}_{11} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 5 & -2 \end{bmatrix}, \boxed{A}_{12} = \begin{bmatrix} 0 & -4 \\ 3 & -1 \end{bmatrix}, \boxed{A}_{21} = \begin{bmatrix} 0 & -4 & -2 \end{bmatrix}, \boxed{A}_{22} = \begin{bmatrix} 7 & -1 \end{bmatrix}$$

Let A & B be block matrices. We say that A & B are *conformable* for block multiplication AB (not BA), if the column partition of A matches the row partition of B. In other words, it must satisfy the following 2 rules:

- No. of columns in A must equal no. of rows in B.
- No. of columns in a submatrix in a row of A must equal no. of rows in the corresponding submatrix in a column of B.

➔ Let S_A & S_B be 2 tuples (i.e. ordered sets). S_A consists of no. of columns for all submatrices in a row of A, while S_B consists of no. of rows for all submatrices in a column of B. If $S_A = S_B$, then A & B are conformable for multiplication.

Determinants

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{1+n}a_{1n}\det(A_{1n})$$

$= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$ (cofactor expansion across the ith row)

$= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$ (cofactor expansion down the jth column)

$C_{ij} = (-1)^{i+j}\det(A_{ij})$ is called a ***cofactor*** (where A_{ij} is a matrix created by deleting ith row & jth column of A)

If $\tilde{R}$ is an ***elementary*** *row operation* performed on a square matrix A to get a new matrix B, then $\det(B) = r\det(A)$ where:

- $r = 1$ if $\tilde{R}$ is the operation where multiple of one row is added to another row.
- $r = -1$ if $\tilde{R}$ is the operation where two rows are interchanged.
- $r = k$ if $\tilde{R}$ is the operation where one row of A is multiplied by the constant k → $\det(A) = \det(B)/k$.

➔ Note that since $\det(A^T) = \det(A)$, this theorem also works for *elementary column operation*.

Matrix Determinant Theorems → Let A & B be $n \times n$ matrices, then:

- If all nonzero entries in A create a shape of a triangle, then $\det(A)$ is the product of the entries on the triangle's diagonal → proved by cofactor expansion along the column/row that has only one single nonzero entry.
- A is invertible if and only if $\det(A) \neq 0$.
  - *Proof:* Let A be reduced to an echelon-form matrix U by row reduction algorithm, then $\det(A) = r \det(U)$ ($r \neq 0$). As U is an upper-triangular matrix, its determinant is the product of all the diagonal entries. If A is not invertible, then A can't be reduced to I_n, which means that at least one of the diagonal entries in U is 0, making $\det(A) = 0$.
- $\det(A^T) = \det(A)$ → proved by induction and cofactor expansion of rows/columns.
- $\det(AB) = \det(A)\det(B)$ → *Proof:*
  - If either A or B is not invertible, then neither is AB → the theorem is true because both sides are zero.
  - If both A & B are invertible, then A can be row reduced to I_n, such that $A = E_k E_{k-1} \dots E_1 I_n$ (where E_i are elementary matrices corresponding to the row operations). Then by the relationship between determinants and elementary row operations, $\det(AB) = |AB| = |E_k \dots E_1 B| = |E_k| \dots |E_1||B| = |E_k \dots E_1||B| = |A||B| = \det(A)\det(B)$.
- In general, $\det(A + B) \neq \det(A) + \det(B)$.

➔ Notice that as the determinant itself can be defined recursively, many determinant theorems can be proved by induction.

Cramer's Rule: Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing the ith column with the vector $\mathbf{b}$, then the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by $x_i = \frac{\det A_i(\mathbf{b})}{\det A}$.

- ***Proof****:* Let $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$, and let the identity matrix $I = [\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n]$. If $A\mathbf{x} = \mathbf{b}$, then:

$$A \cdot I_i(\mathbf{x}) = A[\mathbf{e}_1 \quad \cdots \quad \mathbf{x} \quad \cdots \quad \mathbf{e}_n] = [A\mathbf{e}_1 \quad \cdots \quad A\mathbf{x} \quad \cdots \quad A\mathbf{e}_n] = [\mathbf{a}_1 \quad \cdots \quad \mathbf{b} \quad \cdots \quad \mathbf{a}_n] = A_i(\mathbf{b})$$

  ➔ Thus $\det\big(A_i(\mathbf{b})\big) = \det\big(A \cdot I_i(\boldsymbol{x})\big) = (\det A)(\det I_i(\mathbf{x}))$, where $\det\big(I_i(\mathbf{x})\big) = x_i$ (cofactor expansion along ith row).

- Let $\mathbf{x}$ & $\mathbf{e}_j$ be the jth column vectors of A^{-1} & I_n, respectively. Since $AA^{-1} = I_n$, thus $A\mathbf{x} = \mathbf{e}_j$. Then by Cramer's rule:

$$(A^{-1})_{ij} = x_i = \frac{1}{\det A} \cdot \det A_i(\mathbf{e}_j) = \frac{1}{\det A} \cdot (-1)^{i+j} \det(A_{ji}) = \frac{1}{\det A} \cdot C_{ji}$$

  ➔ $A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} = \frac{1}{\det A} \cdot \text{adjugate}(A) = \frac{1}{\det A} \cdot \text{adj}(A)$.

Determinants & Geometry:

- If A is a 2×2 matrix, the area of the parallelogram determined by the 2 column vectors of A is $|\det A|$.
- If A is a 3×3 matrix, the volume of the parallelepiped determined by the 3 column vectors of A is $|\det A|$.
- Let T be a linear transformation determined by the matrix A:
  - If A is 2×2, and S is a set of points in a parallelogram in $\mathbb{R}^2$, then $\{\text{area of } T(S)\} = |\det A| \cdot \{\text{area of } S\}$.
  - If A is 3×3, and S is a set of points in a parallelepiped in $\mathbb{R}^3$, then $\{\text{volume of } T(S)\} = |\det A| \cdot \{\text{volume of } S\}$.
  - ➔ This theorem actually holds whenever S represents a finite region (or shape) in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Vector Spaces

For a set V to be called a vector space, the following axioms must hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and for all scalars c & d:

1. Sum of $\mathbf{u}$ & $\mathbf{v}$ (i.e. $\mathbf{u}+\mathbf{v}$) is also in V.
2. Scalar multiple of $\mathbf{u}$ (i.e. $c\mathbf{u}$) is also in V.
3. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
4. $(\mathbf{u}+\mathbf{v})+\boldsymbol{w}=\mathbf{u}+(\mathbf{v}+\boldsymbol{w})$.
5. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.
6. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.
7. $c(d\mathbf{u})=(cd)\mathbf{u}$.
8. $1\mathbf{u}=\mathbf{u}$.
9. There's a *zero* vector $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$.
10. For each $\mathbf{u}$ in V, there's a vector $-\mathbf{u}$ in V such that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.

➔ Any sets that satisfy these axioms are vector spaces (e.g. this includes the set $\mathbb{P}_n$ of all polynomials with degree $\leq n$).

A *subspace* of a vector space V is a subset H of V that has 3 properties:

1. Zero vector ($\mathbf{0}$) of V is in H.
2. For each $\mathbf{u}$ & $\mathbf{v}$ in H, the sum $\mathbf{u}+\mathbf{v}$ is also in H.
3. For each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is also in H.

➔ In other words, only 3 properties of the vector space V has to be rechecked for the subspace H.

➔ If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in V, then $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a subspace of V.

$\text{Nul}(A)$ is called the ***null space*** of an $m \times n$ matrix A:

- The set of all solutions of the homogeneous equation $A\mathbf{x}=\mathbf{0}$ → $\text{Nul}\,A=\{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x}=\mathbf{0}\}$.
- $\text{Nul}(A)=\{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
- Note that $\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$.

$\text{Col}(A)$ is called the ***column space*** of an $m \times n$ matrix A:

- The set of all linear combinations of the columns of A → If $A=[\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$, then $\text{Col}(A)=\text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.
- $\text{Col}(A)=\mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ ***onto*** $\mathbb{R}^m$.
- Note that $\text{Col}(A)$ is a subspace of $\mathbb{R}^m$.

$\text{Row}(A)$ is called the ***row space*** of an $m \times n$ matrix A:

- The set of all linear combinations of the rows of A → If $A=\begin{bmatrix}\mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m\end{bmatrix}$, then $\text{Row}(A)=\text{Span}\{\mathbf{r}_1, \dots, \mathbf{r}_m\}$.
- Remember that row vectors should be written *horizontally*.

Basis: Let H be a subspace of a vector space V. Then an indexed set of vectors $\beta = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ in V is a ***basis*** *for* H if:

- $H = \text{Span}\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$.
- β is a linearly independent set, such that β contains no "unnecessary" vectors.

➔ Note that if β is linearly dependent, then to find the basis for H, we just need to remove all vectors in β that can be written as linear combinations of the remaining vectors. To do this, write β as a matrix in reduced echelon form, and then remove all ***non-pivot*** columns (since these columns can be written as linear combination of all the pivot columns).

Conversions between Coordinate Systems:

- Let $\mathbf{x} = P_\beta[\mathbf{x}]_\beta \Leftrightarrow [\mathbf{x}]_\beta = P_\beta^{-1}\mathbf{x}$, where $\beta = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ species the β coordinate system, and $P_\beta = [\mathbf{b}_1 \quad \dots \quad \mathbf{b}_n]$ is called the *changed-of-coordinates matrix*.
  - Linear transformation $[\mathbf{x}]_\beta \mapsto \mathbf{x}$ (through P_β) converts a vector from β coordinates to standard coordinates.
  - Linear transformation $\mathbf{x} \mapsto [\mathbf{x}]_\beta$ (through P_β^{-1}) converts a vector from standard coordinates to β coordinates.
  - ➔ Unless P_β is a square matrix, standard and β coordinates will describe the same vector with different no. of entries.

If a vector space V has the ***dimension*** $\dim(V) = n$, then every basis of V must consist of exactly n vectors → *Proof*:

- Let $\beta = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for V, then any set in V containing more than n vectors is linearly dependent.
  - Let $A = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p]$ (where $p > n$). Transforming $\mathbf{u}_i$ from standard coordinates to β coordinates would give a vector with n entries. Hence matrix A becomes a new matrix A' (which has $n \times p$ dimension).
  - As $p > n$, the equation $A'\mathbf{x} = \mathbf{0}$ must have non-trivial solutions. Thus column vectors of A' are linearly dependent. Since $\mathbf{u}_i$ can *equivalently* be described by standard & β coordinates, the set $\{\mathbf{u}_1, \dots \mathbf{u}_p\}$ must be linearly dependent.
- Let β_1 (with n vectors) & β_2 (with m vectors) be the basis of V. Since both β_1 & β_2 are linearly independent, it must be the case that $m \le n$ and $n \le m$ → $m = n = \dim(V)$.

Some Theorems about Dimensions & Basis:

- Let $\dim(V) = n$ (where $n \ge 1$). Then:
  - Any linearly independent set of exactly n elements in V is automatically a basis for V.
  - Any set of exactly n elements that spans V is automatically a basis for V.
- If H is a subspace of V, then $\dim(H) \le \dim(V)$.
- Let A & B be 2 matrices that are row equivalent, then $\text{Row}(A) = \text{Row}(B)$. If B is in echelon form, the nonzero rows in B forms a basis for $\text{Row}(B)$ as well as a basis for $\text{Row}(A)$. Thus, $\dim(\text{Row}\, A) = \dim(\text{Row}\, B)$ → *Proof*:
  - If $A \xrightarrow{\text{row operations}} B$, then each row in B is a linear combination of the rows in A. It then follows that the linear combination of all the rows in B is also a linear combination of the rows in A. Thus $\text{Row}(A) \subseteq \text{Row}(B)$. Similar logic applies in the reverse case when $B \xrightarrow{\text{row operations}} A$, leading to $\text{Row}(B) \subseteq \text{Row}(A)$ → So $\text{Row}(A) = \text{Row}(B)$.
  - If B is in echelon form, then nonzero rows in B must be pivot columns in B^T, thus forming a basis for $\text{Col}(B^T)$. Since $\text{Row}(B) = \text{Col}(B^T)$ by definition, so this is also a basis for $\text{Row}(B)$.

Some Theorems about Dimensions & Rank:

- Rank of a matrix A is the dimension of $\mathrm{Col}(A)$, i.e. $\mathrm{rank}(A) = \dim(\mathrm{Col}\, A)$
- $\dim(\mathrm{Col}\, A)$ equals no. of pivot columns in A.
- $\dim(\mathrm{Nul}\, A)$ equals no. of free variables in the solution of the equation $A\mathbf{x} = \mathbf{0}$.

➔ If a matrix A has n columns, then $\mathrm{rank}(A) + \dim(\mathrm{Nul}\, A) = \dim(\mathrm{Col}\, A) + \dim(\mathrm{Nul}\, A) = n$.

➔ Since an $n \times n$ matrix A is invertible if A can be row reduced to I_n, so from all the theorems above, it can be deduced that A is invertible if and only if $\mathrm{Col}\, A = \mathbb{R}^n \Leftrightarrow \mathrm{rank}(A) = \dim(\mathrm{Col}\, A) = n \Leftrightarrow \dim(\mathrm{Nul}\, A) = 0 \Leftrightarrow \mathrm{Nul}(A) = \{\mathbf{0}\}$.

➔ Let A be an $m \times n$ matrix (with B as an echelon of A), then $\dim(\mathrm{Row}\, A) = \dim(\mathrm{Col}\, A) = \dim(\mathrm{Col}\, A^T)$. This is because: (1) $\mathrm{rank}(A)$ equals no. of pivot columns in B, and (2) each pivot entry must belong to a nonzero row in B that makes up the basis of $\mathrm{Row}(A)$. Hence, $\dim(\mathrm{Row}\, A) = \{\text{no. of pivot columns}\} = \mathrm{rank}(A) \rightarrow \dim(\mathrm{Row}\, A) + \dim(\mathrm{Nul}\, A) = n$.

Change of Basis: Let $[\boldsymbol{\alpha}] = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$ and $[\boldsymbol{\beta}] = [\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n]$ be 2 matrices describing 2 basis for a vector space V. Let $\mathbf{x}$ be an arbitrary vector in V. Then there is a unique $n \times n$ matrix $P_{\alpha\to\beta} = \big[[\mathbf{a}_1]_\beta \quad \cdots \quad [\mathbf{a}_n]_\beta\big]$, where $\mathbf{a}_i = [\boldsymbol{\beta}][\mathbf{a}_i]_\beta$, that linearly converts the vector $\mathbf{x}$ from $[\mathbf{x}]_\alpha = \begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{bmatrix}$ to $[\mathbf{x}]_\beta = \begin{bmatrix}\beta_1\\ \vdots \\ \beta_n\end{bmatrix}$. This is because:

$$\begin{aligned}\mathbf{x} &= \beta_1\mathbf{b}_1 + \cdots + \beta_n\mathbf{b}_n = [\boldsymbol{\beta}][\mathbf{x}]_\beta \\ &= \alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n = [\boldsymbol{\alpha}][\mathbf{x}]_\alpha \\ &= \alpha_1\big([\boldsymbol{\beta}][\mathbf{a}_1]_\beta\big) + \cdots + \alpha_n\big([\boldsymbol{\beta}][\mathbf{a}_n]_\beta\big) = [\boldsymbol{\beta}]\big[[\mathbf{a}_1]_\beta \quad \cdots \quad [\mathbf{a}_n]_\beta\big][\mathbf{x}]_\alpha = [\boldsymbol{\beta}]P_{\alpha\to\beta}[\mathbf{x}]_\alpha \\ &\Leftrightarrow [\mathbf{x}]_\beta = P_{\alpha\to\beta}[\mathbf{x}]_\alpha \Leftrightarrow [\mathbf{x}]_\alpha = \left(P_{\alpha\to\beta}\right)^{-1}[\mathbf{x}]_\beta\end{aligned}$$

Eigenvalues & Eigenvectors

General Definitions:

- If there is a nontrivial solution to the equation $A\mathbf{x} = \lambda\mathbf{x}$, then the scalar λ is called an **eigenvalue** to the *square* matrix A.
- A nontrivial solution to the equation $A\mathbf{x} = \lambda\mathbf{x}$ is called an **eigenvector** corresponding to the eigenvalue λ.
- Since $A\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}$. Thus the **eigenspace** $\mathrm{Null}(A - \lambda I)$ contains all eigenvectors corresponding to λ. Note that eigenvalue-eigenvector is a one-to-many-relationship!
- For $(A - \lambda I)\mathbf{x} = \mathbf{0}$ to have nontrivial solutions, then $(A - \lambda I)$ must ***not*** be invertible, i.e. $\det(A - \lambda I) = 0$. This is called the **characteristic equation** of A. Solving this equation for λ should give all eigenvalues for the matrix A.
- The **multiplicity** of a particular eigenvalue λ is the no. of times λ appears as a solution to the characteristic equation.

Similarity: Matrix A is ***similar*** to B if there is an invertible matrix P such that $A = PBP^{-1} \Leftrightarrow B = P^{-1}AP$. If A & B are similar, then they have the same characteristic equation and hence the same eigenvalues.

- Let $B = P^{-1}AP$, then $B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(AP - \lambda P) = P^{-1}(A - \lambda I)P$
- Thus $\det(B - \lambda I) = \det(P^{-1}(A - \lambda I)P) = \det(P^{-1}P)\det(A - \lambda I) = \det(I)\det(A - \lambda I) = \det(A - \lambda I)$.
- Hence $\det(B - \lambda I) = \det(A - \lambda I) = 0$ is the characteristic equation for both A & B.

➔ Matrix A is called **diagonalizable** if A is *similar* to a diagonal matrix D such that $A = PDP^{-1}$ for an invertible matrix P

Some Theorems about Eigenvalues/Eigenvectors:

- Dimension of the eigenspace corresponding to an eigenvalue λ must be less than or equal to the multiplicity of λ.
- Matrix A is invertible if and only if zero is ***not*** an eigenvalue of A.
  - If $\lambda = 0$ is an eigenvalue of A, then $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$ must have nontrivial solutions, meaning that A is ***not*** invertible.
- Eigenvalues of a triangular matrix A are entries on its main diagonal.
  - Let λ be any entry on the diagonal of A. Then at least one of the entries on the diagonal of $(A - \lambda I)$ is zero. This means at least 1 row of $(A - \lambda I)$ has no pivot (remember that A is triangular), and thus $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has non-trivial solutions. Conclusion is that λ must be an eigenvalue of A.
- If $\mathbf{v}_1, \dots, \mathbf{v}_n$ are eigenvectors corresponding to ***distinct*** eigenvalues $\lambda_1, \dots, \lambda_n$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent.
  - *Assumption*: Suppose that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly dependent. Then at least one of the vectors (let this be $\mathbf{v}_{p+1}$) can be written as a linear combination of some other vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ that are linearly independent. In other words, it's possible to write $\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ (where $c_i \in \mathbb{R}^n$).

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

$$\Leftrightarrow \lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_{p+1}\mathbf{v}_1 + \cdots + c_p\lambda_{p+1}\mathbf{v}_p \;\left|\; \begin{array}{l} \Leftrightarrow A\mathbf{v}_{p+1} = c_1A\mathbf{v}_1 + \cdots + c_pA\mathbf{v}_p \\ \Leftrightarrow \lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p \end{array}\right.$$

$$\Leftrightarrow \mathbf{0} = c_1(\lambda_{p+1} - \lambda_1)\mathbf{v}_1 + \cdots + c_p(\lambda_{p+1} - \lambda_p)\mathbf{v}_p$$

  - Since λ_i are all distinct and $\mathbf{v}_1, \dots, \mathbf{v}_p$ are linearly independent vectors, then $c_i = 0$, which means that $\mathbf{v}_{p+1} = \mathbf{0}$. This is a *contradiction* because the eigenvectors $\mathbf{v}_i$ can <u>never</u> be $\mathbf{0}$. It follows that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent.

Diagonalization Theorem: «An $n \times n$ matrix A is diagonalizable (i.e. $A = PDP^{-1}$)» <u>if and only if</u> «columns of P are 'n' linearly independent eigenvectors of A and the diagonal entries of D are eigenvalues that correspond to eigenvectors in P»

- Let S_1 & S_2 represent the 2 parts of the theorem. To prove that $S_1 \Leftrightarrow S_2$, let $P = [\mathbf{p}_1 \quad \cdots \quad \mathbf{p}_n]$ and $D = [\boldsymbol{\lambda}_1 \quad \dots \quad \boldsymbol{\lambda}_n]$, where $\boldsymbol{\lambda}_i$ is the ith column of the matrix $\lambda \quad I$ (for some scalar λ_i).
- *Proof for $S_1 \Rightarrow S_2$*: Suppose $AP = PD$, then $[A\mathbf{p}_1 \quad \dots \quad A\mathbf{p}_n] = [\mathbf{p}_1 \quad \cdots \quad \mathbf{p}_n][\boldsymbol{\lambda}_1 \quad \dots \quad \boldsymbol{\lambda}_n] = [\lambda_1\mathbf{p}_1 \quad \dots \quad \lambda_n\mathbf{p}_n]$, implying $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$. But P is invertible, so $\mathbf{p}_i \neq \mathbf{0}$. Hence $\mathbf{p}_i$ is an eigenvector corresponding to the eigenvalue λ_i.
- *Proof for $S_1 \Leftarrow S_2$*: Suppose $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$ (where $\mathbf{p}_i \neq \mathbf{0}$), then:
  - $PDP^{-1} = [\mathbf{p}_1 \quad \cdots \quad \mathbf{p}_n][\boldsymbol{\lambda}_1 \quad \dots \quad \boldsymbol{\lambda}_n]P^{-1} = [\lambda_1\mathbf{p}_1 \quad \dots \quad \lambda_n\mathbf{p}_n]P^{-1} = [A\mathbf{p}_1 \quad \dots \quad A\mathbf{p}_n]P^{-1} = APP^{-1} = AI = A$

➔ Due to linear independence, the set $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ forms a basis for $\mathbb{R}^n$. This is called the **eigenvector basis** of $\mathbb{R}^n$.

➔ Let $A = PDP^{-1}$, then $A^2 = PDP^{-1}PDP^{-1} = PD^2P^{-1}$. In general, $A^k = PD^kP^{-1}$. Note that computing D^k is very easy because D^k is a diagonal matrix whose diagonal entries are the kth power of the diagonal entries in D.

Eigenvectors & First-Order Recurrence Relation $\mathbf{x}_{k+1} = A\mathbf{x}_k \Leftrightarrow \mathbf{x}_k = A^k\mathbf{x}_0$ (for $k \in \mathbb{N}$):

- General solution is $\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n$ (where $\mathbf{v}_i$ is an eigenvector corresponding to an eigenvalue λ_i, and c_i can be determined using the initial condition $\mathbf{x}_0$). This can be proved by induction, in which the inductive step is: $A\mathbf{x}_k = A\sum_{i=1}^{n}\left(c_i(\lambda_i)^k\mathbf{v}_i\right) = \sum_{i=1}^{n}\left(c_i(\lambda_i)^kA\mathbf{v}_i\right) = \sum_{i=1}^{n}\left(c_1(\lambda_i)^k\lambda_i\mathbf{v}_i\right) = \sum_{i=1}^{n}\left(c_i(\lambda_i)^{k+1}\mathbf{v}_i\right) = \mathbf{x}_{k+1}$.
- Note that if $|\lambda_p| \leq 1$ for some p, then $c_p(\lambda_p)^k\mathbf{v}_p$ will have no effect on $\mathbf{x}_k$ as $k \to \infty$, because $\lim_{k\to\infty}\left(c_p(\lambda_p)^k\mathbf{v}_p\right) = \mathbf{0}$.

Markov Chain:

- *Probability* vector is a vector with nonnegative entries that add up to 1.
- *Stochastic* matrix is a square matrix whose columns are probability vectors.
- *Markov Chain* is a sequence of probability vectors $\{\mathbf{x}_i\}$ with a stochastic matrix P that satisfies the first-order recurrence relation $\mathbf{x}_k = P^k\mathbf{x}_0$. Thus the general solution is: $\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n$.
- If P is a regular stochastic matrix and $P\mathbf{q} = \mathbf{q}$, then the probability vector $\mathbf{q}$ is called a *steady-state vector*. It can be shown that $|\lambda_i| \le 1$ (note that any entry m in P must satisfy $|m| \le 1$). Hence, the Markov chain described by P will always converge to a steady-state vector $\mathbf{q}$ when $k \to \infty$.

Matrix Representation of Linear Transformation: Let $T : V \to W$ be linear transformation from V to W (which can have different dimensions). Suppose that $[\boldsymbol{\alpha}] = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_m]$ is a basis for V, and $[\boldsymbol{\beta}] = [\mathbf{b}_1 \quad \dots \quad \mathbf{b}_n]$ is a basis for W, then:

- $T(\mathbf{x}) = A\mathbf{x}$, where A is called the <u>*standard matrix for T*</u>.
- $[T(\mathbf{x})]_\beta = M[\mathbf{x}]_\alpha$, where $M = \big[[T(\mathbf{a}_1)]_\beta \quad \dots \quad [T(\mathbf{a}_m)]_\beta\big]$ is called the <u>*matrix representation for T relative to α & β*</u>. To prove that this is true, let $\mathbf{x} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_m\mathbf{a}_m$ (for some scalars α_i), it then follows that:

$$\begin{aligned} T(\mathbf{x}) &= T(\alpha_1\mathbf{a}_1 + \cdots + \alpha_m\mathbf{a}_m) = \alpha_1 T(\mathbf{a}_1) + \cdots + \alpha_m T(\mathbf{a}_m) \\ \Leftrightarrow [T(\mathbf{x})]_\beta &= [\alpha_1 T(\mathbf{a}_1) + \cdots + \alpha_m T(\mathbf{a}_m)]_\beta = \alpha_1[T(\mathbf{a}_1)]_\beta + \cdots + \alpha_m[T(\mathbf{a}_m)]_\beta \\ &= \big[[T(\mathbf{a}_1)]_\beta \quad \dots \quad [T(\mathbf{a}_m)]_\beta\big][\mathbf{x}]_\alpha = M[\mathbf{x}]_\alpha \end{aligned}$$

- If $V = W$ and $[\boldsymbol{\alpha}] = [\boldsymbol{\beta}]$, then $[T(\mathbf{x})]_\beta = [T]_\beta[\mathbf{x}]_\beta$, where $[T]_\beta$ is called the <u>*β-matrix for T*</u>. If the standard matrix A can be written as $A = PDP^{-1}$ (where $P = [\boldsymbol{\beta}]$ and D doesn't have to be a diagonal matrix), then $[T]_\beta = D$. Here's the proof:
  - Since $P = [\boldsymbol{\beta}]$, thus $\mathbf{x} = P[\mathbf{x}]_\beta \Leftrightarrow P^{-1}\mathbf{x} = [\mathbf{x}]_\beta$, and $T(\mathbf{x}) = P[T(\mathbf{x})]_\beta \Leftrightarrow P^{-1}T(\mathbf{x}) = [T(\mathbf{x})]_\beta$.
  - Thus $AP = PD \Leftrightarrow AP[\mathbf{x}]_\beta = A\mathbf{x} = T(\mathbf{x}) = PD[\mathbf{x}]_\beta \Leftrightarrow P^{-1}T(\mathbf{x}) = [T(\mathbf{x})]_\beta = D[\mathbf{x}]_\beta$. Hence, $[T]_\beta = D$.

Complex Eigenvalues: Let $\bar{z}$ be the complex conjugate of z, and $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ (where a & b are real and not both zero).

- With $c, \mathbf{x}, U, V$ all being complex, then $\overline{c\mathbf{x}} = \bar{c}\,\bar{\mathbf{x}}$, $\overline{cU} = \bar{c}\,\bar{U}$, $\overline{U\mathbf{x}} = \bar{U}\,\bar{\mathbf{x}}$, $\overline{UV} = \bar{U}\,\bar{V}$.
- $\lambda = a - bi = r(\cos\varphi + i\sin\varphi)$ and $\lambda^* = a + bi = r(\cos\varphi - i\sin\varphi) = r(\cos(-\varphi) + i\sin(-\varphi))$ are eigenvalues of matrix C → the transformation $\mathbf{x} \mapsto C\mathbf{x}$ rotates $\mathbf{x}$ by angle φ with the scaling of $r = |\lambda|$.

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = r\begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = r\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

- Let A be any 2×2 *real* matrix, then $\lambda = a - bi$ is an eigenvalue for <u>***both***</u> A & C, with the corresponding eigenvector $\mathbf{v} = \mathrm{Re}(\mathbf{v}) + i\,\mathrm{Im}(\mathbf{v})$. Also, $A = PCP^{-1} \Leftrightarrow AP = PC = [\mathbf{q}_1 \quad \mathbf{q}_2]$, where $P = [\mathrm{Re}(\mathbf{v}) \quad \mathrm{Im}(\mathbf{v})]$ → proof:
  - $A\mathbf{v} = \lambda\mathbf{v} = (a - bi)(\mathrm{Re}(\mathbf{v}) + i\,\mathrm{Im}(\mathbf{v})) \Leftrightarrow A\,\mathrm{Re}(\boldsymbol{v}) = a\,\mathrm{Re}(\mathbf{v}) + b\,\mathrm{Im}(\mathbf{v})$, and $A\,\mathrm{Im}(\mathbf{v}) = -b\,\mathrm{Re}(\mathbf{v}) + a\,\mathrm{Im}(\mathbf{v})$.
  - $AP = [A\,\mathrm{Re}(\mathbf{v}) \quad A\,\mathrm{Im}(\mathbf{v})]$ → $\mathbf{q}_1 = A\,\mathrm{Re}(\mathbf{v})$, and $\mathbf{q}_2 = A\,\mathrm{Im}(\mathbf{v})$.
  - $PC = \left[P\begin{pmatrix} a \\ b \end{pmatrix} \quad P\begin{pmatrix} -b \\ a \end{pmatrix}\right]$ → $\mathbf{q}_1 = a\,\mathrm{Re}(\mathbf{v}) + b\,\mathrm{Im}(\mathbf{v}) = A\,\mathrm{Re}(\mathbf{v})$, and $\mathbf{q}_2 = -b\,\mathrm{Re}(\mathbf{v}) + a\,\mathrm{Im}(\mathbf{v}) = A\,\mathrm{Im}(\mathbf{v})$.

➔ Note that if we have $C = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, then we must use $\bar{\lambda} = a + bi$ and its corresponding eigenvector $\bar{\mathbf{v}}$ instead.

Differential Equations (DE)

Let $\mathbf{x}'(t) = A\mathbf{x}(t)$ be a DE, where A is an $n \times n$ matrix and all entries in $\mathbf{x}(t) = [x_1(t) \quad \dots \quad x_n(t)]^T$ are functions of t.

(1) Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be 2 solutions such that $\mathbf{x}_1{}' = A\mathbf{x}_1$ and $\mathbf{x}_2{}' = A\mathbf{x}_2$, then $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2$ is also a solution because:

$$\mathbf{x}' = (c_1\mathbf{x}_1 + c_2\mathbf{x}_2)' = c_1\mathbf{x}_1{}' + c_2\mathbf{x}_2{}' = c_1(A\mathbf{x}_1) + c_2(A\mathbf{x}_2) = A(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = A\mathbf{x}$$

(2) Let λ_i be an eigenvalue of A such that $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, then $\mathbf{x}_i(t) = \mathbf{v}_i e^{\lambda_i t}$ is a solution to the DE because:

$$\mathbf{x}_i{}'(t) = \frac{d}{dt}\left(\mathbf{v}_i e^{\lambda_i t}\right) = \lambda_i \mathbf{v}_i e^{\lambda_i t} = A\mathbf{v}_i e^{\lambda_i t} = A\mathbf{x}_i(t)$$

➔ From (1) & (2), the general solution must be $\mathbf{x}(t) = \sum_i (c_i\mathbf{x}_i) = \sum_i \left(c_i\mathbf{v}_i e^{\lambda_i t}\right)$, where c_i depends on initial conditions.

Let $\mathbf{x}'(t) = A\mathbf{x}(t)$ be a DE, where A has complex eigenvalues $\lambda = a + bi$ and $\bar{\lambda} = a - bi$.

- By using Euler's formula, the 2 ***complex*** solutions are:
  - $\mathbf{x}(t) = \mathbf{v}e^{\lambda t} = \mathbf{v}e^{(a+bi)t} = \mathbf{v}e^{at}e^{i(bt)} = \mathbf{v}e^{at}\operatorname{cis}(bt)$
  - $\overline{\mathbf{x}(t)} = \bar{\mathbf{v}}e^{\bar{\lambda} t} = \bar{\mathbf{v}}e^{(a-bi)t} = \bar{\mathbf{v}}e^{at}e^{i(-bt)} = \bar{\mathbf{v}}e^{at}\operatorname{cis}(-bt)$
- The 2 ***real*** solutions are $\mathbf{x}_1(t) = \operatorname{Re}(\mathbf{x}) = (\mathbf{x} + \bar{\mathbf{x}})/2$ and $\mathbf{x}_2(t) = \operatorname{Im}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})/2i$, because both of them are linear combinations of the 2 solutions $\mathbf{x}$ and $\bar{\mathbf{x}}$. They can also be derived as follow:
  - $\mathbf{x}_1(t) = \operatorname{Re}(\mathbf{x}) = \operatorname{Re}(\mathbf{v}e^{at}\operatorname{cis} bt) = \operatorname{Re}(\mathbf{v})\, e^{at}\cos(bt) - \operatorname{Im}(\mathbf{v})\, e^{at}\sin(bt)$
  - $\mathbf{x}_2(t) = \operatorname{Im}(\mathbf{x}) = \operatorname{Im}(\mathbf{v}e^{at}\operatorname{cis} bt) = \operatorname{Re}(\mathbf{v})\, e^{at}\sin(bt) + \operatorname{Im}(\mathbf{v})\, e^{at}\cos(bt)$

Decoupled DE: Let $I_n = [\mathbf{e}_1 \quad \dots \quad \mathbf{e}_n]$ and $A = PDP^{-1}$, where $P = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n]$ and $D = \operatorname{diag}[\lambda_1 \quad \dots \quad \lambda_n]$.

- $\mathbf{y}'(t) = D\mathbf{y}(t)$ is a *decoupled* DE because the derivative $y_i{}'(t) = \lambda y_i(t)$ only depends on the function $y_i(t)$ itself. Working with decoupled DE is easy because the solution will be $\mathbf{y}(t) = \sum_i (c_i\mathbf{x}_i) = \sum_i \left(c_i\mathbf{e}_i e^{\lambda_i t}\right) = \sum_i \left(c_i e^{\lambda_i t}\right)$.
- Let $\mathbf{x}'(t) = A\mathbf{x}(t)$ with $\mathbf{x}(t) = P\mathbf{y}(t)$, where P is a change-of-variable matrix, then:

$$\left.\begin{array}{r} \mathbf{x}' = (P\mathbf{y})' = P\mathbf{y}' \\ \mathbf{x}' = A\mathbf{x} = A(P\mathbf{y}) = (PDP^{-1})P\mathbf{y} = PD\mathbf{y} \end{array}\right\} \Leftrightarrow P\mathbf{y}' = PD\mathbf{y} \Leftrightarrow \mathbf{y}' = D\mathbf{y}$$

  ➔ By using $\mathbf{x}(t) = P\mathbf{y}(t)$, we have decoupled the DE, making it easier to work with.

Iterative Estimates for Eigenvalues

Preparation for Power Method (PM): Let $A = PDP^{-1}$, where $D = \operatorname{diag}[\lambda_1 \quad \dots \quad \lambda_n]$ (such that $|\lambda_1| > |\lambda_i|$ for all $i \neq 1$), and $P = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n]$. Note that P represents an orthogonal basis for $\mathbb{R}^n$. So for any $\mathbf{x} \in \mathbb{R}^n$, there exists scalars c_i such that $\mathbf{x} = \sum_i c_i\mathbf{v}_i \Leftrightarrow A^k\mathbf{x} = \sum_i c_i A^k\mathbf{v}_i = \sum_i c_i\lambda_i{}^k\mathbf{v}_i$. Dividing both sides by $\lambda_1{}^k$, we have:

$$\left\{\frac{1}{\lambda_1{}^k}A^k\mathbf{x}\right\} = \left\{c_1\mathbf{v}_1 + \sum_{i=2}^{n} c_i\mathbf{v}_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\right\} \Leftrightarrow \lim_{k\to\infty}\left\{\frac{1}{\lambda_1{}^k}A^k\mathbf{x}\right\} = c_1\mathbf{v}_1 + 0 = c_1\mathbf{v}_1$$

➔ For $k \to \infty$, we have $A^k\mathbf{x} = \left(c_1\lambda_1{}^k\right)\mathbf{v}_1 = c\mathbf{v}$. Note that $c = c_1\lambda_1{}^k$ is a constant, which means that both $\mathbf{v}$ & $\mathbf{v}_1$ are eigenvectors associated with λ_1 (i.e. $A\mathbf{v} = \lambda_1\mathbf{v}$). By choosing c wisely such that the largest entry in $\mathbf{v}$ is 1, then λ_1 will be the largest entry in the vector $A\mathbf{v}$.

Preparation for Inverse Power Method (IPM): Unlike PM, the IPM is able to compute any eigenvalues of a matrix. Let $A = (B - \alpha I_n)^{-1}$. Then $\lambda_i = 1/(\Lambda_i - \alpha) \Leftrightarrow \Lambda_i = \alpha(1/\lambda_i)$ where λ_i & Λ_i are eigenvalues for A & B. Proof:

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \Leftrightarrow \mathbf{v}_i/\lambda_i = A^{-1}\mathbf{v}_i = (B - \alpha I_n)\mathbf{v}_i = B\mathbf{v}_i - \alpha\mathbf{v}_i \Leftrightarrow B\mathbf{v}_i = (\alpha + 1/\lambda_i)\mathbf{v}_i = \Lambda_i \mathbf{v}_i$$

➔ For a particular i, if we choose $\alpha \approx \Lambda_i$, then λ_i will be huge, thus becoming the dominant eigenvalue λ_1 of A.

Problem to be solved by PM: Let A be an $n \times n$ diagonalizable matrix. Use PM to estimate the 'largest' eigenvalue λ_1 of A (i.e. under the condition that $|\lambda_1| > |\lambda_i|$ for all $i \neq 1$).

Problem to be solved by IPM: Let $B = A^{-1} + \alpha I_n \Leftrightarrow A = (B - \alpha I_n)^{-1}$. Use IPM to estimate *any* eigenvalue Λ_i of B (where α is the initial estimate to which Λ_i is closest to, and $|\Lambda_i| \neq |\Lambda_j|$ for all $i \neq j$). Note that since $\Lambda_i = \alpha + (1/\lambda_1)$, thus estimating λ_1 will help us find Λ_i, which means that applying IPM on B is equivalent to applying PM on A.

Algorithm for PM and IPM:

(1) Select an initial vector $\mathbf{x}_0$ whose largest entry is 1. For IPM, also select an initial estimate α closest to Λ_i.

(2) For $k = 0, 1, ...,$
   - Compute $\mathbf{y}_k = A\mathbf{x}_k$. For IPM, it's actually easier to solve for $\mathbf{y}_k$ with this equation: $A^{-1}\mathbf{y}_k = (B - \alpha I_n)\mathbf{y}_k = \mathbf{x}_k$.
   - Let μ_k be the entry in $\mathbf{y}_k$ with the largest absolute value, and $\gamma_k = \alpha + (1/\mu_k)$.
   - Compute $\mathbf{x}_{k+1} = (1/\mu_k)\mathbf{y}_k$ for the next iteration.

(3) As k increases, the sequence $\{\mu_k\}$ will approach the dominant eigenvalue λ_1 of A, and the sequence $\{\mathbf{x}_k\}$ will approach corresponding eigenvector. On the other hand, the sequence $\{\gamma_k\}$ will approach the eigenvalue Λ_i of B.

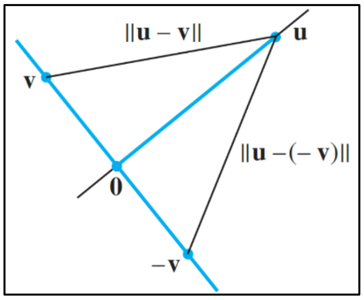
Orthogonality

An **inner product** on an **inner product space** V is a *function* that maps each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ to a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ (note that the dot product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$ is a special case of an inner product), such that the following axioms are satisfied:

- $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ (commutative property)
- $\langle \mathbf{u} + \mathbf{v},\ \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ (distributive property)
- $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
- $\langle \mathbf{u}, \mathbf{v} \rangle \geq 0$, where $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Length of a vector $\mathbf{v}$ is defined as $|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \Leftrightarrow |\mathbf{v}|^2 = \langle \mathbf{v}, \mathbf{v} \rangle$. If the inner product is a dot multiplication, then we have $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{{v_1}^2 + \cdots + {v_n}^2}$. ***Unit*** *vector* of $\mathbf{v}$ can be obtained by $\bar{\mathbf{v}} = \mathbf{v}/|\mathbf{v}|$, and the ***distance*** between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is:

$$\text{dist}(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}| = |\mathbf{v} - \mathbf{u}| = \sqrt{|\mathbf{u}|^2 + |\mathbf{v}|^2 - 2(\mathbf{u} \cdot \mathbf{v})}$$



2 vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ or in $\mathbb{R}^3$ are perpendicular to each other if $\text{dist}(\mathbf{u}, \mathbf{v}) = \text{dist}(\mathbf{u}, -\mathbf{v})$. This also extends to $\mathbb{R}^n$, thus 2 vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are perpendicular if and only if:

$(\text{dist}(\mathbf{u}, \mathbf{v}))^2 = |\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2(\mathbf{u} \cdot \mathbf{v})$

$= (\text{dist}(\mathbf{u}, -\mathbf{v}))^2 = |\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2(\mathbf{u} \cdot \mathbf{v})$

$\Leftrightarrow 2(\mathbf{u} \cdot \mathbf{v}) = -2(\mathbf{u} \cdot \mathbf{v}) \Leftrightarrow \mathbf{u} \cdot \mathbf{v} = 0$ (orthogonality theorem)

$\Leftrightarrow |\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$ (the Pythagorean theorem)

Let Q be a subspace of $\mathbb{R}^n$, then a vector $\mathbf{v}$ is *orthogonal* to Q if and only if $\mathbf{v}$ is orthogonal to every vectors in Q. The set of all vectors that are orthogonal to Q is called the ***orthogonal complement*** of Q and is denoted by $Q^\perp$.

- Let Q be spanned by the set $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$, then a vector $\mathbf{v}$ is in $Q^\perp$ if and only if $\mathbf{v}$ is orthogonal to every $\mathbf{w}_i$ → proof:
  - From the definition of orthogonality above, $\mathbf{v}$ is in $Q^\perp$ if and only if $\mathbf{v} \cdot (x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n) = 0$ (where x_i are arbitrary scalars). But this is only the case if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ (for all i).
- $Q^\perp$ is also a subspace of $\mathbb{R}^n$ → proof:
  - Let $\mathbf{u}$ and $\mathbf{v}$ be 2 vectors in $Q^\perp$, then $\mathbf{u} \cdot \mathbf{w}_i = \mathbf{v} \cdot \mathbf{w}_i = 0 \Leftrightarrow (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w}_i = 0$, which means $(\mathbf{u} + \mathbf{v}) \in Q^\perp$.
  - Let c be a scalar and $\mathbf{v} \in Q^\perp$, then $c(\mathbf{v} \cdot \mathbf{w}_i) = 0 \Leftrightarrow (c\mathbf{v}) \cdot \mathbf{w}_i = 0$, which means $c\mathbf{v} \in Q^\perp$.

Let A be an $m \times n$ matrix, then $(\operatorname{Row} A)^\perp = \operatorname{Nul}(A) \Leftrightarrow (\operatorname{Row} A^T)^\perp = \operatorname{Nul}(A^T) \Leftrightarrow (\operatorname{Col} A)^\perp = \operatorname{Nul}(A^T)$. To prove this, let $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$ be the ***row*** vectors of A, then a vector $\mathbf{x} \in \mathbb{R}^n$ is in $\operatorname{Nul}(A)$ if and only if:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1\mathbf{x} \\ \vdots \\ \mathbf{a}_m\mathbf{x} \end{bmatrix} = \mathbf{0} \Leftrightarrow \mathbf{a}_i\mathbf{x} = 0 \Leftrightarrow \mathbf{x} \in (\operatorname{Row} A)^\perp$$

If Q is a subspace of $\mathbb{R}^n$, then $\dim(Q) + \dim(Q^\perp) = n$. To prove this, let A be a matrix representing a basis for Q. Then we have $Q = \operatorname{Col}(A) = \operatorname{Row}(A^T) \Leftrightarrow Q^\perp = (\operatorname{Row} A^T)^\perp = \operatorname{Nul}(A^T)$. Since A^T has exactly n columns, it then follows that $\dim(Q) + \dim(Q^\perp) = \dim(\operatorname{Row} A^T) + \dim(\operatorname{Nul} A^T) = n$ (by the rank-nullity theorem).

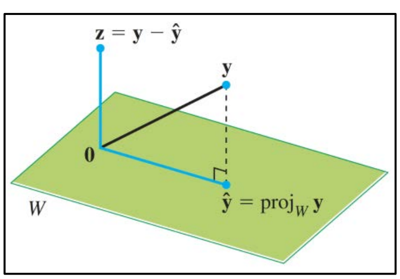
Orthogonal AND Orthonormal Sets

$S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an orthogonal set if all vectors in S are orthogonal to each other (i.e. $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for every $i \neq j$) → If $\mathbf{u}_k \neq 0$ for every k, then S is *linearly independent* and hence is an *orthogonal basis* of the subspace spanned by S. Let $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$, to prove that S is linearly independent, we just need to show that $c_k = 0$ for all $k \in \{1, \dots, p\}$:

$$0 = \mathbf{u}_k(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) = c_1(\mathbf{u}_k \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_k \cdot \mathbf{u}_p) = c_k(\mathbf{u}_k \cdot \mathbf{u}_k) \Leftrightarrow c_k = 0$$

Orthogonal basis is very useful because it gives us an easy method to solve a non-homogenous equation without having to do row-reduction algorithm. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace V of $\mathbb{R}^n$. Then for each vector $\mathbf{y}$ in V:

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p \Leftrightarrow \mathbf{u}_k \cdot \mathbf{y} = c_1(\mathbf{u}_k \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_k \cdot \mathbf{u}_p) = c_k(\mathbf{u}_k \cdot \mathbf{u}_k) \Leftrightarrow c_k = \frac{\mathbf{y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k}$$



Let $\hat{\mathbf{y}} = \alpha\mathbf{u}$ be the orthogonal projection of vector $\mathbf{y}$ onto $\mathbf{u}$ (where α is just some constant). Then it's obvious from the figure that the projection line $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to vector $\mathbf{u}$. Hence:

$$\mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u} = (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u}) = 0 \Leftrightarrow \hat{\mathbf{y}} = \alpha\mathbf{u} = \left(\frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u}$$

Note that $\mathbf{y} = \mathbf{z} + \hat{\mathbf{y}}$ is the sum of two orthogonal vectors (i.e. $\mathbf{z} \cdot \hat{\mathbf{y}} = 0$).

Orthogonal Decomposition Theorem: Let W be a subspace of $\mathbb{R}^n$ with $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ as an orthogonal basis, then each $\mathbf{y}$ in $\mathbb{R}^n$ can be ***uniquely*** decomposed to $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ (such that $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$), where:

$$\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y} = \sum_{k=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k}\mathbf{u}_k \Leftrightarrow \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$

- To prove that the decomposition is unique, assume that there is a 2nd decomposition such that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$. Then $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$. Since both W & $W^\perp$ are subspaces, so $\mathbf{v} = (\hat{\mathbf{y}} - \hat{\mathbf{y}}_1) \in W$ and $\mathbf{v} = (\mathbf{z}_1 - \mathbf{z}) \in W^\perp$, which means that $\mathbf{v} \cdot \mathbf{v} = \mathbf{0} \Leftrightarrow \mathbf{v} = \mathbf{0} \Leftrightarrow \hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and $\mathbf{z} = \mathbf{z}_1$.

Orthonormal Set: If S is an orthogonal set with only ***unit*** vectors, then S is an orthonormal basis for Span$\{S\}$ subspace.

- Columns of an $m \times n$ matrix U forms an orthonormal set if and only if $U^T U = I_n$ because:

$$U^T U = \begin{bmatrix} \mathbf{u}_1{}^T \\ \vdots \\ \mathbf{u}_n{}^T \end{bmatrix} [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n] = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n \end{bmatrix} = I_n$$

- Let $U^T U = I_n$ (i.e. U has orthonormal columns), then $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$, thus $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{x} \Leftrightarrow |U\mathbf{x}| = |\mathbf{x}|$. Proof is as follow: $(U\mathbf{x}) \cdot (U\mathbf{y}) = (U\mathbf{x})^T (U\mathbf{y}) = (\mathbf{x}^T U^T)(U\mathbf{y}) = \mathbf{x}^T (U^T U)\mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$
- Let $U = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p]$ be a matrix representing an orthonormal basis for W, then:

$$\begin{aligned} \hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y} &= (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p = (\mathbf{u}_1{}^T \mathbf{y})\mathbf{u}_1 + \cdots + (\mathbf{u}_p{}^T \mathbf{y})\mathbf{u}_p \\ &= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p] \begin{bmatrix} \mathbf{u}_1{}^T \mathbf{y} \\ \vdots \\ \mathbf{u}_p{}^T \mathbf{y} \end{bmatrix} = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p] \begin{bmatrix} \mathbf{u}_1{}^T \\ \vdots \\ \mathbf{u}_p{}^T \end{bmatrix} \mathbf{y} = UU^T \mathbf{y} \end{aligned}$$

Inner Product Inequalities:

- $|\langle \mathbf{u}, \mathbf{v} \rangle| \le |\mathbf{u}||\mathbf{v}|$ (Cauchy-Schwars inequality). To prove this, let $W = \text{Span}\{\mathbf{u}\}$, then:

$$|\mathbf{v}| \ge |\operatorname{proj}_W(\mathbf{v})| = \left| \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} \right| = \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{|\langle \mathbf{u}, \mathbf{u} \rangle|} |\mathbf{u}| = \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{|\mathbf{u}|^2} |\mathbf{u}| = \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{|\mathbf{u}|} \Leftrightarrow |\langle \mathbf{u}, \mathbf{v} \rangle| \le |\mathbf{u}||\mathbf{v}|$$

- $|\mathbf{u} + \mathbf{v}| \le |\mathbf{u}| + |\mathbf{v}|$ (Triangle inequality). Proof is as follow:

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 &= \langle (\mathbf{u} + \mathbf{v}), (\mathbf{u} + \mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle \\ &\le |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| \le |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2|\mathbf{u}||\mathbf{v}| = (|\mathbf{u}| + |\mathbf{v}|)^2 \end{aligned}$$

Gram-Schmidt Process is an algorithm to produce an orthogonal basis for a subspace of $\mathbb{R}^n$. Given a non-orthogonal basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, then we can deduce an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ recursively as follow:

- Let $W_1 = \text{Span}\{\mathbf{v}_1\}$, where $\mathbf{v}_1 = \mathbf{x}_1$.
- Let $W_2 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_2$ is orthogonal to $\mathbf{v}_1$, and is equal to:

$$\mathbf{v}_2 = \mathbf{x}_2 - \operatorname{proj}_{W_1}(\mathbf{x}_2) = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1$$

- Let $W_3 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_3$ is orthogonal to both $\mathbf{v}_1$ & $\mathbf{v}_2$, and is equal to:

$$\mathbf{v}_3 = \mathbf{x}_3 - \operatorname{proj}_{W_2}(\mathbf{x}_3) = \mathbf{x}_3 - \left\{ \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \right\}$$

- Repeat this process. Finally let $W = W_p = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, where $\mathbf{v}_p$ is orthogonal to all $\mathbf{v}_k$ ($k < p$), and is equal to:

$$\mathbf{v}_p = \mathbf{x}_p - \operatorname{proj}_{W_{p-1}}(\mathbf{x}_p) = \mathbf{x}_p - \left\{ \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \cdots + \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}} \mathbf{v}_{p-1} \right\}$$

➔ Note that $\mathbf{x}_k = \mathbf{v}_k + \operatorname{proj}_{W_{k-1}}(\mathbf{x}_k)$ for all $k \in \{1, \ldots, p\}$. Since W_{k-1} is *embedded* within W_k, so $\operatorname{proj}_{W_{k-1}}(\mathbf{x}_k) \in W_k$. It's also obvious that $\mathbf{v}_k \in W_k$. Both of these facts lead to the conclusion that $\mathbf{x}_k \in W_k = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$

QR Factorization: Let $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$ be an $m \times n$ matrix with linearly independent columns (i.e. $n \leq m$), then A can be factored as $A = QR$, where $Q = [\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_n]$ is an $m \times n$ matrix with columns that form an _orthonormal_ basis for $\text{Col}(A)$, and R is an $n \times n$ _upper triangular invertible matrix_. Here is the proof:

- Let $W = \text{Col}\, A$, and Q be computed by the Gram-Schmidt process on columns of A. Then $\mathbf{a}_k \in W_k = \text{Span}\{\mathbf{q}_1, \dots, \mathbf{q}_k\}$ for each $k \in \{1, \dots, n\}$. So there exists constants $r_{1k}, \dots, r_{kk}$ such that $\mathbf{a}_k = r_{1k}\mathbf{q}_1 + \cdots + r_{kk}\mathbf{q}_k$. Extending this to n-dim, we have $\mathbf{a}_k = Q\mathbf{r}_k$ where the *column* vector $\mathbf{r}_k$ is $\mathbf{r}_k = [r_{1k} \quad \cdots \quad r_{kk} \quad 0 \quad \cdots \quad 0]^T$.

➔ All the zeros in $\mathbf{r}_k$ explain why the matrix R must be upper triangular.

➔ R must be invertible because columns of A are linearly independent. To see why, let $R\mathbf{x} = \mathbf{0}$. Thus $Q(R\mathbf{x}) = A\mathbf{x} = \mathbf{0}$. Since columns of A are linearly independent, $\mathbf{x} \in \{\mathbf{0}\}$, which means $R\mathbf{x} = \mathbf{0}$ only has trivial solution.

➔ Since Q is orthonormal, $Q^TQ = I_n$. From $A = QR$, we can compute R by $Q^TA = Q^TQR = I_nR = R$.

Least Squares Problem

Let A be an $m \times n$ matrix, then a least-squares solution (LSS) to the equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that the approximation error $|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}|$ is as small as possible. If $A\mathbf{x} = \mathbf{b}$ is consistent, then the LSS has zero error. Since $A\hat{\mathbf{x}} \in \text{Col}(A)$, so $A\hat{\mathbf{x}}$ must be the closest point in $\text{Col}(A)$ to $\mathbf{b}$. Thus a straight-forward way to find LSS is to solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}} = \text{proj}_{\text{Col}\, A}(\mathbf{b})$.

Orthogonal Columns Method: Let $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where columns of $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$ are orthogonal to each other, then:

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}\, A}(\mathbf{b}) = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}\mathbf{a}_1 + \cdots + \frac{\mathbf{b} \cdot \mathbf{a}_n}{\mathbf{a}_n \cdot \mathbf{a}_n}\mathbf{a}_n$$

➔ The weights in the vector equation above make up the entries to the _unique_ LSS vector $\hat{\mathbf{x}}$. The solution is unique here, due to the orthogonal decomposition theorem. Note that since columns in A are orthogonal, they must also be linearly independent, which also implies that the LSS is unique (as will be shown later on).

Normal Equations Method: It's obvious that $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\text{Col}(A)$, which means that it's also orthogonal to each column vector of $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$. So for each $k \in \{1, \dots, n\}$, we have:

$$\mathbf{a}_k \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = {\mathbf{a}_k}^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \Leftrightarrow \begin{bmatrix} {\mathbf{a}_1}^T \\ \vdots \\ {\mathbf{a}_n}^T \end{bmatrix}(\mathbf{b} - A\hat{\mathbf{x}}) = A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \Leftrightarrow A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

➔ Solving the normal equations $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ will give a set of LSS to the equation $A\hat{\mathbf{x}} = \mathbf{b}$.

➔ If A^TA is invertible, then there is a **_unique_** LSS for every $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{b} = \mathbf{0}$, then the equation $(A^TA)\hat{\mathbf{x}} = A^T(A\hat{\mathbf{x}}) = \mathbf{0}$ only has trivial solution, which implies that $A\hat{\mathbf{x}} = \mathbf{0}$ only has trivial solution. Hence, A has linearly independent columns.

QR Factorization Method: Let $A = QR$ (where columns of A are linearly independent, columns of Q are orthonormal to each other, and R is invertible), then by using the normal equations, we have:

$A^TA\hat{\mathbf{x}} = (QR)^T(QR)\hat{\mathbf{x}} = R^T(Q^TQ)R\hat{\mathbf{x}} = R^TI_nR\hat{\mathbf{x}} = R^TR\hat{\mathbf{x}}$

$= A^T\mathbf{b} = (QR)^T\mathbf{b} = R^TQ^T\mathbf{b}$

$\Leftrightarrow R\hat{\mathbf{x}} = Q^T\mathbf{b} \Leftrightarrow \hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$ (which is _unique_ since columns of A are linearly independent)

Symmetric Matrices & Quadratic Forms

Orthogonal Matrix: An $n \times n$ matrix U has *orthonormal* columns if and only if U is invertible and $U^{-1} = U^T$ (such that $U^T U = U^{-1} U = U U^{-1} = U U^T = I_n$). Such a matrix is called an orthogonal matrix.

Symmetric Matrix: An $n \times n$ matrix A is symmetric if $A = A^T$. Below are *spectral theorems* for symmetric matrices:

- Matrix A is <u>*orthogonally diagonalizable*</u> if $A = PDP^T = PDP^{-1}$ (where P is an orthogonal matrix, and D is diagonal). This is only possible if and only if A is <u>*symmetric*</u>, because $A^T = (PDP^T)^T = P^{TT} D^T P^T = PD^T P^T = PDP^T = A$.
- If matrix A is symmetric, then any 2 eigenvectors from 2 different eigenspaces are orthogonal. To prove this, let $(\lambda_1, \mathbf{v}_1)$ and $(\lambda_2, \mathbf{v}_2)$ be 2 eigenvalue-eigenvector pairs of A, then:
  - $\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1{}^T A^T \mathbf{v}_2 = \mathbf{v}_1{}^T (A^T \mathbf{v}_2) = \mathbf{v}_1{}^T (A\mathbf{v}_2) = \mathbf{v}_1{}^T (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1{}^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$
  - $\Leftrightarrow (\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But since $\lambda_1 \neq \lambda_2$, so $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Spectral Decomposition: Let A be orthogonally diagonalizable with $P = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$, then spectral decomposition of A involves writing A as a sum of terms of the form $\lambda_i \mathbf{v}_i \mathbf{v}_i{}^T$ which are determined by the spectrum (eigenvalues) of A.

$$A = PDP^T = [\lambda_1 \mathbf{v}_1 \quad \cdots \quad \lambda_n \mathbf{v}_n] \begin{bmatrix} \mathbf{v}_1{}^T \\ \cdots \\ \mathbf{v}_n{}^T \end{bmatrix} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1{}^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n{}^T = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i{}^T$$

- Each term is an $n \times n$ matrix of rank 1, because every column of $\mathbf{v}_i \mathbf{v}_i{}^T$ is a multiple of $\mathbf{v}_i$.
- Each matrix $\mathbf{v}_i \mathbf{v}_i{}^T$ is called a *projection matrix* because $(\mathbf{v}_i \mathbf{v}_i{}^T)\mathbf{x} = \text{proj}_W(\mathbf{x})$ (where $W = \text{Span}\{\mathbf{v}_i\}$ and $\mathbf{x} \in \mathbb{R}^n$). To prove this, recall that $\text{proj}_W(\mathbf{x}) = \alpha \mathbf{v}_i$ where $\alpha = \mathbf{v}_i{}^T \mathbf{x}$, then:

$$\alpha \mathbf{v}_i \cdot (\mathbf{x} - \alpha \mathbf{v}_i) = \alpha \mathbf{v}_i{}^T \boldsymbol{x} - \alpha^2 \mathbf{v}_i \cdot \mathbf{v}_i = \alpha^2 - \alpha^2 \mathbf{v}_i{}^T \mathbf{v}_i = \alpha^2 (1 - \mathbf{v}_i{}^T \mathbf{v}_i) = 0$$

Quadratic Forms (QF)

Quadratic Form: $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ is a function (with the domain $\mathbf{x} \in \mathbb{R}^n$) that is determined by a symmetric matrix $A = [a_{ij}]$ (such that $a_{ij} = a_{ji}$). Let $\mathbf{x} = [x_1 \quad \cdots \quad x_n]^T$, then by expanding $\mathbf{x}^T A \mathbf{x}$, we can define $Q(\mathbf{x})$ mathematically as follow:

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=i}^{n} c_{ij} x_i x_j \quad \text{where} \begin{cases} c_{ij} = a_{ij}, & \text{for } i = j \\ c_{ij} = 2a_{ij} = 2a_{ji}, & \text{for } i \neq j \end{cases}$$

With the aid of principle axes theorem, we can define conditions for different types of QF. Let $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, then $Q(\mathbf{x})$ is:

- <u>*Indefinite*</u> if all the cases below are false for this QF.
- <u>*Positive definite*</u> if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ → this occurs if and only if $\lambda > 0$ (for <u>*all*</u> eigenvalues λ of A).
- <u>*Negative definite*</u> if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ → this occurs if and only if $\lambda < 0$ (for <u>*all*</u> eigenvalues λ of A).
- <u>*Positive semi-definite*</u> if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ → this occurs if and only if $\lambda \geq 0$ (for <u>*all*</u> eigenvalues λ of A).
- <u>*Negative semi-definite*</u> if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ → this occurs if and only if $\lambda \leq 0$ (for <u>*all*</u> eigenvalues λ of A).

➔ Note that all positive/negative definite QF are positive/negative semi-definite, but <u>***not***</u> vice versa!

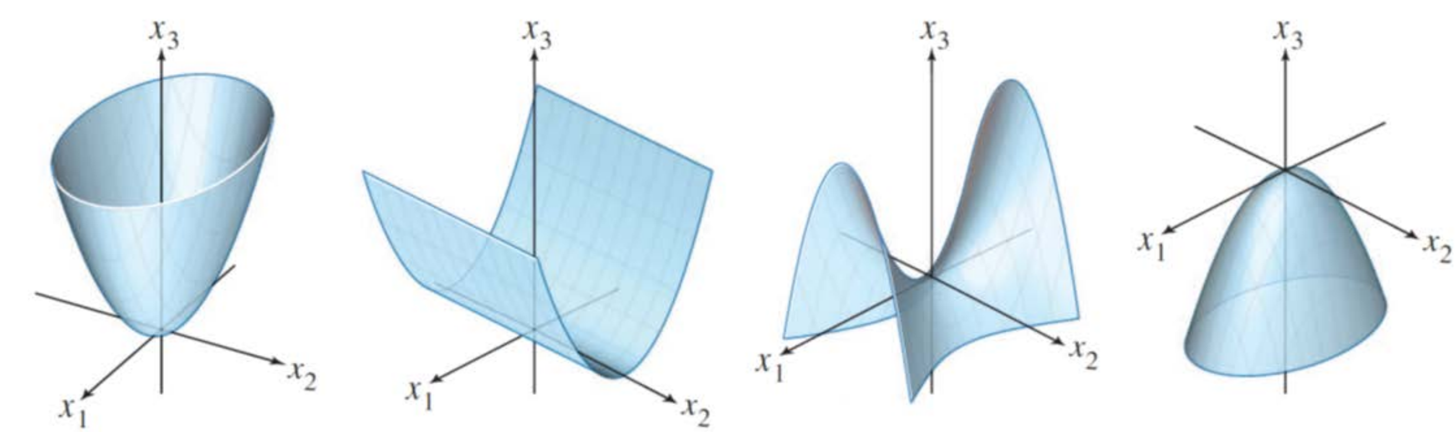
Principle Axes Theorem:

- It's tedious to work with cross-product terms $c_{ij}x_ix_j$ (where $i \neq j$). To avoid this, find a change-of-variable matrix P, that maps $\mathbf{x}$ to a new variable $\mathbf{y}$ (i.e. $\mathbf{x} = P\mathbf{y} \Leftrightarrow \mathbf{y} = P^{-1}\mathbf{x}$), such that $Q(\mathbf{y})$ doesn't have any cross-product terms.
- Matrix A must be orthogonally diagonalizable because it's symmetric, and thus $A = PDP^T \Leftrightarrow D = P^TAP$. This means that $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = (P\mathbf{y})^TA(P\mathbf{y}) = \mathbf{y}^T(P^TAP)\mathbf{y} = \mathbf{y}^TD\mathbf{y} = Q(\mathbf{y})$, where $Q(\mathbf{y})$ is a QF determined by D.

➔ Note that P is a change-of-basis matrix. Its columns are called the *principle axes* (as they make up an orthonormal basis).

Geometric View of Principle Axes:

- Let $Q(\mathbf{x})$ & $Q(\mathbf{y})$ be related by the relation $\mathbf{x} = P\mathbf{y}$. Then their graphs will have exactly the *same* shape, but will have different orientations. Graph of $Q(\mathbf{x})$ will be relative to the standard basis of $\mathbb{R}^n$, while the graph of $Q(\mathbf{y})$ will be relative to the orthonormal basis of $\mathbb{R}^n$ that are made up of the columns of P.
- Let $Q(\mathbf{x})$ be a quadratic form with no cross-product terms ($\mathbf{x} \in \mathbb{R}^2$), and c be a constant. Then $Q(\mathbf{x}) = c$ will be a:
  - o Cross-section with the shape of an <u>*ellipse*</u> as in cases (a) & (d), if and only if $Q(\mathbf{x})$ is positive/negative <u>*definite*</u>.
  - o Cross-section with the shape of a <u>*hyperbola*</u> as in the case (c), if and only if $Q(\mathbf{x})$ is <u>*indefinite*</u>.



(a) $z = 3x_1^2 + 7x_2^2$ (b) $z = 3x_1^2$ (c) $z = 3x_1^2 - 7x_2^2$ (d) $z = -3x_1^2 - 7x_2^2$

Let $\mathbf{x} = P\mathbf{y}$ such that $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$ and $Q(\mathbf{y}) = \mathbf{y}^TD\mathbf{y}$ (where P is an orthogonal matrix). Assume that columns of D & P are arranged in an 'orderly' fashion such that $\lambda_i \geq \lambda_j$ (for every $i > j$), i.e. $\max(\lambda) = \lambda_1$ and $\min(\lambda) = \lambda_n$. Let there now be a **constraint** on these QF such that $|\mathbf{x}| = |P\mathbf{y}| = |\mathbf{y}| = 1$, then:

$$\left\{Q(\mathbf{y}) = \sum_{i=1}^{n} \lambda_i {y_i}^2\right\} \leq \left\{M = \max\big(Q(\mathbf{y})\big) = \lambda_1 \sum_{i=1}^{n} {y_i}^2 = \lambda_1|\mathbf{y}|^2 = \lambda_1\right\}, \text{and}$$

$$\left\{Q(\mathbf{y}) = \sum_{i=1}^{n} \lambda_i {y_i}^2\right\} \geq \left\{m = \min\big(Q(\mathbf{y})\big) = \lambda_n \sum_{i=1}^{n} {y_i}^2 = \lambda_n|\mathbf{y}|^2 = \lambda_n\right\}$$

- Since $\mathbf{x} = P\mathbf{y}$ is one-to-one, so $Q(\mathbf{x})$ & $Q(\mathbf{y})$ is one-to-one. Hence, $\lambda_1 \leq Q(\mathbf{y}) \leq \lambda_n$ also implies that $\lambda_1 \leq Q(\mathbf{x}) \leq \lambda_n$.
- Let $I_n = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]$, and $P = [\mathbf{p}_1 \ \cdots \ \mathbf{p}_n]$ (where $\mathbf{p}_i$ is the <u>***unit***</u> eigenvector corresponding to λ_i). Then:
  - o $Q(\mathbf{y}) = M$ when $\mathbf{y} = [y_1 \ \cdots \ y_n]^T = \mathbf{e}_1 \rightarrow \max\big(Q(\mathbf{x})\big)$ occurs when $\mathbf{x} = P\mathbf{y} = P\mathbf{e}_1 = [\mathbf{p}_1 \ \cdots \ \mathbf{p}_n]\mathbf{e}_1 = \mathbf{p}_1$.
  - o $Q(\mathbf{y}) = m$ when $\mathbf{y} = [y_1 \ \cdots \ y_n]^T = \mathbf{e}_n \rightarrow \min\big(Q(\mathbf{x})\big)$ occurs when $\mathbf{x} = P\mathbf{y} = P\mathbf{e}_n = [\mathbf{p}_1 \ \cdots \ \mathbf{p}_n]\mathbf{e}_n = \mathbf{p}_n$.

- Let's now add an **additional constraint** such that $\mathbf{x}^T\mathbf{p}_1 = \mathbf{x} \cdot \mathbf{p}_1 = 0$.
  - o Under this new constraint, $\max\big(Q(\mathbf{x})\big)$ cannot occur when $\mathbf{x} = \mathbf{p}_1$ since $\mathbf{x}^T\mathbf{p}_1 = \mathbf{p}_1{}^T\mathbf{p}_1 \neq 0$. Consequently, we have $\max\big(Q(\mathbf{x})\big) = \lambda_2$ (the second largest eigenvalue), which occurs at $\mathbf{x} = \mathbf{p}_2$ (or correspondingly, $\mathbf{y} = \mathbf{e}_2$).
  - o Note that this doesn't change where $\min\big(Q(\mathbf{x})\big)$ occurs, as $\mathbf{x}^T\mathbf{p}_1 = \mathbf{p}_n{}^T\mathbf{p}_1 = 0$ which still satisfies all constraints.

➔ *Generalization*: Let $k \in \{1, \dots, n\}$ with constraints $|\mathbf{x}| = 1, \quad \mathbf{x}^T\mathbf{p}_1 = 0, \quad \dots, \quad \mathbf{x}^T\mathbf{p}_{k-1} = 0$. Then $\max\big(Q(\mathbf{x})\big) = \lambda_k$ is attained at $\mathbf{x} = \mathbf{p}_k$, and $\min\big(Q(\mathbf{x})\big) = \lambda_n$ is attained at $\mathbf{x} = \mathbf{p}_n$.

Let A be an $m \times n$ matrix and $Q(\mathbf{x}) = \mathbf{x}^T(A^TA)\mathbf{x}$, where $A^TA = PDP^T$ is symmetric and orthogonally diagonalizable with $D = \text{diag}[\lambda_1 \quad \cdots \quad \lambda_n]$. Then $\max(|A\mathbf{x}|) = \max(|A\mathbf{x}|^2) = \max\big((A\mathbf{x})^T(A\mathbf{x})\big) = \max(\mathbf{x}^T(A^TA)\mathbf{x}) = \max\big(Q(\mathbf{x})\big) = \lambda_1$.

Singular Value Decomposition (SVD)

Singular Values: Let A be an $m \times n$ matrix and $A^TA = PDP^T$ (with $P = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$ and $D = [\lambda_1 \quad \cdots \quad \lambda_n]$ sorted such that $\lambda_i \geq \lambda_j$ for all $i > j$). Then $0 \leq |A\mathbf{v}_i|^2 = \mathbf{v}_i{}^T(A^TA)\mathbf{v}_i = \mathbf{v}_i{}^T\lambda_i\mathbf{v}_i = \lambda_i\mathbf{v}_i{}^T\mathbf{v}_i = \lambda_i$ for all $i \in \{1, \dots, n\}$. Singular values of matrix A are thus defined as $\sigma_i = \sqrt{\lambda_i} = |A\mathbf{v}_i|$ with the inequality $\sigma_i \geq 0$.

Single Value Decomposition: Let $r = \text{rank}(A)$ (with A being $m \times n$), then the SVD of A is $A = U\Sigma V^T$, where:

- $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ is an $m \times n$ matrix with $D = \text{diag}[\sigma_1 \quad \cdots \quad \sigma_r]$, where $\sigma_1, \dots, \sigma_r$ are positive singular values of A.
- $U = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m]$ is an $m \times m$ orthogonal matrix (representing an orthonormal basis for $\mathbb{R}^m$), where for $1 \leq i \leq r$, we have $\mathbf{u}_i = A\mathbf{v}_i/|A\mathbf{v}_i| = A\mathbf{v}_i/\sigma_i \Leftrightarrow A\mathbf{v}_i = \sigma_i\mathbf{u}_i$. Now, for $r < j \leq m$, we have $\mathbf{u}_i \cdot \mathbf{u}_j = 0$. So in order to compute $\mathbf{x} = \mathbf{u}_j$ for <u>*each*</u> of the js, we must solve the system of linear equations $\mathbf{u}_1{}^T\mathbf{x} = 0, \quad \dots, \quad \mathbf{u}_r{}^T\mathbf{x} = 0$.
- $V = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$ is an $n \times n$ orthogonal matrix, where $\mathbf{v}_i$ are called the *right singular vectors* of A which equal to the <u>*unit*</u> eigenvectors of A^TA, corresponding to the eigenvalues $\lambda_i = \sigma_i{}^2$.

➔ SVD works because $VV^T = VV^{-1} = I_n$, and $U\Sigma = [\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}] = AV$. Thus, $U\Sigma V^T = AVV^T = A$.

➔ Note that $A^T = (U\Sigma V^T)^T = V\Sigma^TU^T$. This allows a quick conversion between the SVD of A and the SVD of A^T.

Let $A = U\Sigma V^T$ be an $m \times n$ matrix with $A^TA = PDP^T$, where $P = V = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$ and $D = \text{diag}[\lambda_1 \quad \cdots \quad \lambda_n]$, then:

(1) $S = \{\mathbf{u}_1, \dots, \mathbf{u}_r\} = \{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is orthogonal basis for $\text{Col}(A)$, and hence $\text{rank}(A) = \dim(\text{Col}\, A) = r$ ➔ proof:
  - o $\mathbf{v}_i{}^T\mathbf{v}_j = \mathbf{v}_i{}^T\lambda_i\mathbf{v}_j = \mathbf{v}_i{}^T(A^TA)\mathbf{v}_j = (A\mathbf{v}_i)^T(A\mathbf{v}_i) = \mathbf{u}_i{}^T\mathbf{u}_j = 0$ for all $i \neq j$ ➔ S is an orthogonal basis for $\text{Span}(S)$.
  - o It's obvious that for $1 \leq i \leq r < j \leq n$, we have $\{\mathbf{u}_i = A\mathbf{v}_i\} \in \text{Col}(A)$ and $A\mathbf{v}_j = \lambda_j\mathbf{v}_j = \mathbf{0}$. Since the set S is linearly independent, so $\text{Col}(A) = c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r + c_{r+1}A\mathbf{v}_{r+1} \dots + c_nA\mathbf{v}_n = c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r$.

(2) $S = \{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ is orthogonal basis for $Q^\perp = (\text{Col}\, A)^\perp = \text{Nul}(A^T)$, because both Q & $Q^\perp$ are subspaces of $\mathbb{R}^m$.

(3) $S = \{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is orthogonal basis for $\text{Nul}(A)$, because $\dim(\text{Nul}\, A) = n - r$, and $A\mathbf{v}_i = \mathbf{0}$ for $r \leq i < n$.

(4) $S = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is orthogonal basis for $Q^\perp = (\text{Nul}\, A)^\perp = \text{Col}(A^T) = \text{Row}(A)$, since both Q & $Q^\perp$ are subspaces of $\mathbb{R}^n$.

➔ $\dim(\text{Nul}\, A) = n - r = 0$ only occurs when $r = n$ ➔ A is invertible if and only if it has n nonzero singular values.