Guide to MAT1120 - Linear Algebra

Gabriel Sigurd Cabrera

December 7, 2017

Contents

1 Spaces

We will define these spaces with respect to two operations; addition, and scalar multiplication. Let's us first define the following:

$$\vec{a} = (a_1, a_2, ..., a_n) \in \mathbb{R}^n,\ a_i \in \mathbb{R},\ i = 1, 2, ..., n \tag{1}$$

$$\vec{b} = (b_1, b_2, ..., b_n) \in \mathbb{R}^n,\ b_i \in \mathbb{R},\ i = 1, 2, ..., n \tag{2}$$

$$c \in \mathbb{R} \tag{3}$$

Then we can define these two operations:

Vector Addition

$$\vec{a} + \vec{b} = (a_1 + b_1, a_2 + b_2, ..., a_n + b_n) \tag{4}$$

Scalar Multiplication

$$c\vec{a} = (ca_1, ca_2, ..., ca_n) \tag{5}$$

1.1 Vector Spaces

Let $\mathbf{V}$ be a set where there are two defined operations (*addition* and *scalar multiplication*), and that we have the vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbf{V}$ and the scalars $c, d \in \mathbb{R}$.

Definition - The Axioms of a Vector Space Given the above, $\mathbf{V}$ is a **vector space** if and only if the following conditions are met:

1. $\vec{u} + \vec{v} \in \mathbf{V}$; closed under addition
2. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$; the commutative property of addition
3. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$; the associative property of addition
4. There exists a vector $\vec{0} \in \mathbf{V}$ such that $\vec{u} + \vec{0} = \vec{u}$ for all $\vec{u} \in \mathbf{V}$; the additive identity
5. There exists a vector $-\vec{u}$ such that $\vec{u} + (-\vec{u}) = \vec{0}$, for all $\vec{u} \in \mathbf{V}$; the additive inverse
6. $c\vec{u} \in \mathbf{V}$; closed under scalar multiplication
7. $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$; the first distributive property of scalar multiplication
8. $(c + d)\vec{u} = c\vec{u} + d\vec{u}$; the second distributive property of scalar multiplication
9. $(cd)\vec{u} = c(d\vec{u})$; the associative property of scalar multiplication
10. $1 \cdot \vec{u} = \vec{u}$ for all $\vec{u} \in \mathbf{V}$; the scalar multiplicative identity

Note The set of all real numbers $\mathbb{R}$ (as well as its generalization $\mathbb{R}^n$) and the set of all real polynomials $\mathbb{P}^n$ (where $p = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots c_nx^n$, $p \in \mathbb{P}^n$ and $x, c_i \in \mathbb{R}$, $i = 0, 1, 2, ..., n$) are both vector spaces, since they meet all of the above conditions.

Corollary Given that $\mathbf{V}$ is a vector space, that $\vec{u} \in \mathbf{V}$ and that $c \in \mathbb{R}$, we know that the following also holds:

1. $0 \cdot \vec{u} = \vec{0}$; the first scalar multiplicative zero
2. $c \cdot \vec{0} = \vec{0}$; the second scalar multiplicative zero
3. $-\vec{u} = (-1)\,\vec{u}$; the scalar multiplicative inverse

1.2 Subspaces

Let $\mathbf{V}$ be a *vector space*, where we have $\mathbf{H} \subset \mathbf{V}$ such that $\{\varnothing\} \neq \mathbf{H}$ (where $\{\varnothing\}$ is the empty set). In addition, we have the vectors $\vec{u}, \vec{v} \in \mathbf{H}$ and the scalar $c \in \mathbb{R}$

Definition Given the above, we have that the subset $\mathbf{H}$ is a **subspace** if the following conditions are met:

1. $\vec{0} \in \mathbf{H}$; the subset $\mathbf{H}$ contains a zero vector
2. $\vec{u} + \vec{v} \in \mathbf{H}$; $\mathbf{H}$ is *closed* under addition
3. $c\vec{u} \in \mathbf{H}$; $\mathbf{H}$ is *closed* under scalar multiplication

Notes

- A vector space is also a subspace
- A subspace is always a vector space

1.3 Null Spaces

Let A be an $m \times n$ matrix and let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

Definition Given the above, the **null space** of A is defined as:

$$\operatorname{nul}(A) \triangleq \left\{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\right\} \tag{6}$$

Notes Given the same conditions as above:

- Note that $A\vec{x} = \vec{0}$ is a *homogenous equation*; a system of m equations and n variables.
- $\operatorname{nul}(A)$ is a subspace of $\mathbb{R}^n$

Theorem $\operatorname{nul}(A)$ is a *subspace* of $\mathbb{R}^n$

1.4 Column Spaces

Let A be an $m \times n$ matrix consisting of the column vectors $\vec{a}_i \in \mathbb{R}^n$, where:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \tag{7}$$

and:

$$\vec{a}_i = \begin{bmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m,i} \end{bmatrix}, \text{ where } i = 1, 2, ..., n \tag{8}$$

Definition Given the above, the **column space** of A is defined as:

$$\text{col}(A) \triangleq \text{span}\,\{\vec{a}_1, \vec{a_2}, ..., \vec{a}_n\} \tag{9}$$

Notes Given the same conditions as above:

- Under "*normal*" circumstances, col(A) is a subspace of A
- If $m > n$, then col(A) cannot span over $\mathbb{R}^m$
- col(A) is a subspace of $\mathbb{R}^m$

1.5 Row Spaces

Let A be an $m \times n$ matrix, and let B be A in *row-echelon* form:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \sim B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix} \tag{10}$$

The **row space** of A is then defined as the set of all *row-vectors* in A, corresponding to the rows in B which contain pivot elements.

2 Linear Transformations

Let A be an $m \times n$ matrix; we also have $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. In addition, let $T(\vec{x}) = A\vec{x}$ be a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Finally, let $\mathbf{V}, \mathbf{W}$ be vector spaces such that we can redefine our function as $T : \mathbf{V} \rightarrow \mathbf{W}$

Definition Given the above, T is a **linear transformation** if the following conditions are met:

1. $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$
2. $T(c\vec{x}) = cT(\vec{x})$

2.1 The Kernel and Range

The *kernel* and *range* of a linear transformation are essentially the generalizations of the *null space* and *column space* for linear transformations that are not necessarily matrix-based. When a transformation is of the form $T = A\vec{x}$ (where A is an $m \times n$ matrix) such that $T : \mathbb{R}^n \to \mathbb{R}^m$, then $\ker(T) = \mathrm{nul}(A)$ and $\mathrm{range}(T) = \mathrm{col}(A)$

Definition of the Kernel Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces, and let T be a linear transformation such that $T : \mathbf{V} \to \mathbf{W}$. We also have $\vec{x} \in \mathbf{V}$.

Given the above, the **kernel** of T is defined as:

$$\ker(T) \triangleq \left\{\vec{x} \in \mathbf{V} \mid T(\vec{x}) = \vec{0}\right\} \tag{11}$$

Definition of the Range Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces, and let T be a linear transformation such that $T : \mathbf{V} \to \mathbf{W}$. We also have $\vec{x} \in \mathbf{V}$ and $\vec{y} \in \mathbf{W}$.

Given the above, the **range** of T is defined as:

$$\mathrm{range}(T) \triangleq \{\vec{y} \in \mathbf{W} \mid \exists\, \vec{x} \in \mathbf{V} \text{ such that } \vec{y} = T(\vec{x})\} \tag{12}$$

3 Linear Independence

Let $\mathbf{V}$ be a *vector space* with a set of vectors $S = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} \subset \mathbf{V}$. We also have a set of coefficients $\{c_1, c_2, ..., c_p\}$ where $c_i \in \mathbb{R}$

Definition Given the above, S is **linearly independent** if the following equation only has the trivial solution:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_p\vec{v}_p = \vec{0} \tag{13}$$

The trivial solution being $c_1 = c_2 = \cdots = c_p = 0$. We will now generalize this for any type of vector space:

Theorem 4 Let $S = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} \subset \mathbf{V}$, where $p > 1$ and $\vec{v}_1 \neq \vec{0}$. S is **linearly dependent** if and only if there exists a vector $\vec{v}_j$ with a $j > 1$ such that $\vec{v}_j$ is a linear combination of $\vec{v}_1, \vec{v}_2, ..., \vec{v}_{j-1}$

3.1 Bases

Let $\mathbf{V}$ be a vector space, and let $\mathbf{H}$ be a subspace of $\mathbf{V}$. Let $\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} \subset \mathbf{V}$.

Definition Given the above, $\mathfrak{B}$ is a **basis** for $\mathbf{H}$ if:

1. $\mathfrak{B}$ is a *linearly independent* set
2. $\operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} = \mathbf{H}$

Standard Basis There are many linear combinations one could use to create a basis for $\mathbb{R}^n$, however there is only one set known as the **standard basis**; if $\mathfrak{S}$ is our set, then it is the standard basis if and only if:

$$\mathfrak{S} = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, ..., \vec{e}_{n-1}, \vec{e}_n\} = \left\{ \begin{bmatrix}1\\0\\0\\\vdots\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\\vdots\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\\vdots\\0\\0\end{bmatrix}, \cdots, \begin{bmatrix}0\\0\\0\\\vdots\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\\vdots\\0\\1\end{bmatrix} \right\} \tag{14}$$

Theorem 5 Let $\mathbf{V}$ be a vector space, $S = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} \subset \mathbf{V}$ and $\mathbf{H} = \operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\}$

1. If any single vector $\vec{v}_k$ is a linear combination of the other $p-1$ vectors, then these $p-1$ vectors span the entirety of $\mathbf{H}$
2. If $\mathbf{H} \neq \left\{\vec{0}\right\}$, then there exists a subset of S which is a basis for $\mathbf{H}$

Finding the Basis of a Column Space Let A be an $m \times n$ matrix such that $A = [\vec{a}_1, \vec{a}_2, ..., \vec{a}_n]$ where $\vec{a}_i \in \mathbb{R}^m$. We know then that $\operatorname{col}(A) = \operatorname{span}\{\vec{a}_1, \vec{a}_2, ..., \vec{a}_n\}$. If we then take A and reduce it to a new matrix B in row-echelon form, then the set of all pivot columns in B will be the basis of $\operatorname{col}(B)$, and the corresponding columns in A will be the basis for $\operatorname{col}(A)$.

4 Coordinate Systems

Let $\mathbf{V}$ be a vector space with the basis $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$.

Theorem 7 Given the above, for each $\vec{x} \in \mathbf{V}$, there exists a unique $c_1, c_2, ..., c_n \in \mathbb{R}$ such that $\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots c_n\vec{b}_n$

The Coordinates of a Basis Assume we have the same vector space $\mathbf{V}$ and basis $\mathfrak{B}$ as above. Let $\vec{x} \in \mathbf{V}$ be defined as $\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots c_n\vec{b}_n$. We define the coefficients $c_1, c_2, ..., c_n$ as the **coordinates** of $\vec{x}$ with respect to the basis $\mathfrak{B}$. This can be written as follows:

$$[\vec{x}]_{\mathfrak{B}} = \begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}, \ [\vec{x}]_{\mathfrak{B}} \in \mathbb{R}^n \tag{15}$$

Finding the Standard Basis Assuming we have the same conditions as above, let us create a matrix $P_{\mathfrak{B}}$ whose elements consist of the basis $\mathfrak{B}$:

$$P_{\mathfrak{B}} = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{bmatrix} \tag{16}$$

The product of $[\vec{x}]_{\mathfrak{B}}$ (the coordinates of $\vec{x}$ with respect to $\mathfrak{B}$) and $P_{\mathfrak{B}}$ (the matrix consisting of the elements of the basis $\mathfrak{B}$) then gives us the standard basis matrix for $\vec{x}$:

$$[\vec{x}] = P_{\mathfrak{B}} [\vec{x}]_{\mathfrak{B}} \tag{17}$$

Theorem 8 Let $\mathbf{V}$ be a vector space, let the transformation function $T_{\mathfrak{B}} : \mathbf{V} \to \mathbb{R}^n$ be defined as $T_{\mathfrak{B}} (\vec{x}) = [\vec{x}]_{\mathfrak{B}}$

This **coordinate transformation** is a *linear transformation* which is both injective and surjective on $\mathbb{R}^n$

4.1 Isomorphism

Let $\mathbf{V}, \mathbf{W}$ be vector spaces, and let $T : \mathbf{V} \to \mathbf{W}$ be a linear transformation. Also assume that T is both *injective* and *surjective* to $\mathbf{W}$.

Definition Given the above, T is defined as being **isomorphic** between $\mathbf{V}$ and $\mathbf{W}$

Note The aforementioned coordinate transformation $T_{\mathfrak{B}} : \mathbf{V} \to \mathbb{R}^n$ is an isomorphism between $\mathbf{V}$ and $\mathbb{R}^n$. Note that this transformation is therefore dependent on the chosen basis $\mathfrak{B}$

5 Dimensions

Let $\mathbf{V}$ be a vector space with the basis $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$

Definition Given the above, the **dimension** of $\mathbf{V}$ is n (where n is the minimum number of vectors needed to span $\mathbf{V}$):

$$\dim (\mathbf{V}) = n \tag{18}$$

The following theorems confirm that n will be the same no matter the basis.

Theorem 9 Let $\mathbf{V}$ be a vector space with a basis $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$, then any set in $\mathbf{V}$ containing more than n vectors must be linearly dependent.

Theorem 10 Let $\mathbf{V}$ be a vector space with a basis consisting of n vectors, then any other basis of $\mathbf{V}$ will also consist of exactly n vectors.

5.1 Finite, Zero and Infinite Dimensions

Let $\mathbf{V}$ be a vector space spanned by a finite number of vectors - we define this "sort" of vector space as **finite dimensional**. If the basis for $\mathbf{V}$ is $\left\{\vec{0}\right\}$ (the null space), then $\mathbf{V}$ is **zero dimensional**. If $\mathbf{V}$ is finite dimensional and has a basis $\mathfrak{B} \neq \left\{\vec{0}\right\}$ with n elements, it is n-dimensional. If the basis for $\mathbf{V}$ has an infinite number of elements, then it is **infinite dimensional**.

Theorem 11 If $\mathbf{H}$ is a subspace in a *finite dimensional* vector space $\mathbf{V}$, then $\dim(\mathbf{H}) \leq \dim(\mathbf{V})$. In addition, if we have a *linearly independent* set $S = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_k\}$ in $\mathbf{H}$, then it is possible to extend S such that it becomes a basis for $\mathbf{H}$

Theorem 12 If $\mathbf{V}$ is a p-dimensional vector space, then *any* set of p *linearly independent* vectors in $\mathbf{V}$ will be a basis for $\mathbf{V}$.

Useful Rule If A is an $m \times n$ matrix, then the following is true:

$$\dim\left(\text{col}\left(A\right)\right) + \dim\left(\text{nul}\left(A\right)\right) = n \tag{19}$$

6 Rank

Definition Let A be an $m \times n$ matrix, then the rank of A is defined as:

$$\text{rank}(A) = \dim\left(\text{col}\left(A\right)\right) = \dim\left(\text{row}\left(A\right)\right) \tag{20}$$

rank(A) also has the property that $\text{rank}(A) \leq \min\left\{m, n\right\}$

Notes Let $\mathbf{V}, \mathbf{W}$ be finite dimensional vector spaces, let $T : \mathbf{V} \rightarrow \mathbf{W}$ be a linear transformation, and let A be an $m \times n$ matrix. Then:

- $\text{rank}(A) + \dim\left(\text{nul}\left(A\right)\right) = n$
- $\dim\left(\text{range}\left(T\right)\right) + \dim\left(\ker\left(T\right)\right) = \dim(\mathbf{V})$

6.1 Summary of Properties

If A is an $n \times n$ matrix, then the following are equivalent:

1. $\text{col}(A) = \mathbb{R}^n$
2. $\dim\left(\text{col}\left(A\right)\right) = n$
3. $\text{rank}(A) = n$
4. $\text{nul}(A) = \left\{\vec{0}\right\}$
5. $\dim\left(\text{nul}\left(A\right)\right) = 0$
6. $\det(A) \neq 0$

7 Change of Basis

Let $\mathbf{V}$ be a finite dimensional vector space. We define two of its bases $\mathfrak{B}$ and $\mathfrak{C}$ as:

$$\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\} \tag{21}$$

$$\mathfrak{C} = \{\vec{c}_1, \vec{c}_2, ..., \vec{c}_n\} \tag{22}$$

In addition, we have the vector $\vec{v} \in \mathbf{V}$ with the basis $\mathfrak{B}$ coordinates $[\vec{v}]_{\mathfrak{B}} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$

How do we find the coordinates for $\vec{v}$ in terms of the basis $\mathfrak{C}$? To do this, we need a transformation matrix $\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P}$, which gives us our formula:

$$[\vec{v}]_{\mathfrak{C}} = \underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P} [\vec{v}]_{\mathfrak{B}} \tag{23}$$

This can also be reversed:

$$[\vec{v}]_{\mathfrak{B}} = \underset{\mathfrak{B}\leftarrow\mathfrak{C}}{P} [\vec{v}]_{\mathfrak{C}} \tag{24}$$

In addition, the transformation matrices are each others' inverses:

$$\underset{\mathfrak{B}\leftarrow\mathfrak{C}}{P} = \left(\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P}\right)^{-1} \tag{25}$$

Finding a Transformation Matrix Let $\mathbf{V} \in \mathbb{R}^n$ and $\mathbf{W} \in \mathbb{R}^m$ be vector spaces. In addition to this, we have a linear transformation $T : \mathbf{V} \to \mathbf{W}$ such that $T(\vec{x}) = A\vec{x}$, where A is an $m \times n$ matrix and $\vec{x} \in V$.

Now, let $\mathfrak{B}$ and $\mathfrak{C}$ be bases for $\mathbf{V}$ and $\mathbf{W}$ respectively, such that $\vec{b}_i \in \mathfrak{B}$; if we want to find a transformation matrix $\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P}$ such that:

$$[\vec{v}]_{\mathfrak{C}} = \underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P} [\vec{v}]_{\mathfrak{B}} \tag{26}$$

Then we can write it as:

$$\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P} = \left[\left[T\left(\vec{b_1}\right)\right]_{\mathfrak{C}} \left[T\left(\vec{b_2}\right)\right]_{\mathfrak{C}} \cdots \left[T\left(\vec{b_n}\right)\right]_{\mathfrak{C}}\right] \tag{27}$$

This can be rewritten as:

$$\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P} = \left[\left[A\vec{b_1}\right]_{\mathfrak{C}} \left[A\vec{b_2}\right]_{\mathfrak{C}} \cdots \left[A\vec{b_n}\right]_{\mathfrak{C}}\right] \tag{28}$$

To find each individual column $\left[A\vec{b_i}\right]_{\mathfrak{C}}$, we must find a linear combination for the vector $A\vec{b_i} \in \mathbb{R}^m$ in terms of the basis $\mathfrak{C}$:

$$A\vec{b_i} = \beta_{i,1}\vec{c}_1 + \beta_{i,2}\vec{c}_2 + \cdots + \beta_{i,m}\vec{c}_m \tag{29}$$

The column we are looking for will be composed of our $\beta_{i,j}$ coefficients – to find them, solve for the generated augmented matrix. We are looking for an answer of the form:

$$\vec{\beta_i} = \begin{bmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \vdots \\ \beta_{i,m} \end{bmatrix} \tag{30}$$

Corollary for Theorem 8 Let $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$ be a basis for the vector space $\mathbf{V}$.

We also have the transformation $\mathbf{V} \to \mathbb{R}^n$ and the coordinate transformation $\vec{v} \to [\vec{v}]_{\mathfrak{B}}$. The transformation is linear, injective and surjective.

In addition, we have the sets $S = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\} \subset \mathbf{V}$ and $S_{\mathfrak{B}} = \left\{[\vec{u}_1]_{\mathfrak{B}}, [\vec{u}_2]_{\mathfrak{B}}, ..., [\vec{u}_p]_{\mathfrak{B}}\right\} \subset \mathbb{R}^n$.

Given all these conditions, the following equivalencies can be deduced:

- The set S spans $\mathbf{V}$ $\iff$ the set $S_{\mathfrak{B}}$ spans $\mathbb{R}^n$
- The set S is *linearly independent* $\iff$ the set $S_{\mathfrak{B}}$ is *linearly independent*
- The set S is a basis for $\mathbf{V}$ $\iff$ the set $S_{\mathfrak{B}}$ is a basis for $\mathbb{R}^n$ $\iff$ $p = n$ and the matrix $[S_{\mathfrak{B}}] = \left[[\vec{u}_1]_{\mathfrak{B}}, [\vec{u}_2]_{\mathfrak{B}}, ..., [\vec{u}_p]_{\mathfrak{B}}\right]$ is an invertible $n \times n$ matrix $\iff$ $p = n$ and $\det(S_{\mathfrak{B}}) \neq 0$

One more equivalency If $\mathbf{S}$ is a vector space of dimension N, $S = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_n\}$ and $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$ is a basis, then the following statements are all equivalent:

- S is a basis
- The matrix $[S_{\mathfrak{B}}] = [[\vec{u}_1]_{\mathfrak{B}}, [\vec{u}_2]_{\mathfrak{B}}, ..., [\vec{u}_n]_{\mathfrak{B}}]$ is invertible
- rref($[S_{\mathfrak{B}}]$) has a pivot element in each column
- rref($[S_{\mathfrak{B}}]$) has a pivot element in each row

8 Markov Chains

Definition A **probability vector** is a vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ such that $x_i \geq 0$ and $x_1 + x_2 + \cdots + x_n = 1$

Definition A **stochastic matrix** P is an $n \times n$ matrix where *all* its columns are *probability vectors*. In addition, if all the matrix' elements are greater than zero ($p_{i,j} > 0$) then it is known as a **regular stochastic matrix**.

Definition The **steady-state vector** of a *stochastic matrix* P is a *probability vector* $\vec{q}$ such that $P\vec{q} = \vec{q}$. In other words, $\vec{q}$ is an *eigenvector* with an *eigenvalue* 1.

Theorem All *stochastic matrices* have a *steady-state vector*

Theorem If P is a **regular stochastic matrix**, then P has a unique *steady-state vector* $\vec{q}$. In addition, if $\vec{x}_0$ is an arbitrary *probability vector* and $\vec{x}_{k+1} = P\vec{x}_k$ where $k \geq 0$, then $\lim_{k\to\infty} \vec{x}_k = \vec{q}$.

9 Eigenvalues and Eigenvectors

Definition A vector $\vec{x} \in \mathbb{R}^n$ such that $\vec{x} \neq \vec{0}$ is called an **eigenvector** for an $n \times n$ matrix A if $A\vec{x} = \lambda\vec{x}$ for a $\lambda \in \mathbb{R}$. λ is called the corresponding **eigenvalue** for $\vec{x}$.

Equivalencies Assuming the previous conditions, the following statements are all equivalent:

- $A\vec{x} = \lambda\vec{x}$
- $A\vec{x} - \lambda\vec{x} = \vec{0}$
- $(A - \lambda I)\,\vec{x} = \vec{0}$

Geometric Interpretation If we have $A\vec{x} = \lambda\vec{x}$, having a $\lambda > 1$ implies that $\vec{x}$ is stretched out by a factor of λ.

$0 < \lambda < 1$ implies that $\vec{x}$ is shortened by a factor of λ .

If $\lambda < 0$, then the above will still hold, except $\vec{x}$ will also flip direction.

Finding our Eigenvalues and Eigenvectors To find these, we can use the following process:

1. Solve for λ in: $\det(A - \lambda I) = 0$. There may be multiple solutions.
2. Plug each λ from above into $A = \lambda I$ and solve for each system of equations to get your corresponding eigenvectors.

Theorem The *eigenvalues* of a triangular matrix are equal to the elements on its diagonal.

Theorem 2 Let A be an $n \times n$ matrix. If $\vec{v}_1, \vec{v}_2, ..., \vec{v}_r$ are *eigenvectors* for A, with the corresponding *eigenvalues* $\lambda_1, \lambda_2, ..., \lambda_r$ and these eigenvalues all differ from each other (no value is repeated), then $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_r\}$ is *linearly independent*

Row Operations When row reducing a matrix A, it is vital to realize that some row operations will change the determinant of the matrix, which in turn changes its *eigenvalues*. Specifically, if two rows are switched, then the determinant's sign will be flipped.

Now recall that all matrices can be rewritten in row-echelon form; let us consider the matrix U, which is the row-echelon form of A. If we want to find the determinant of U, we must simply find the product of all its diagonal elements (which we will call p) and note the number of row-switching operations that were performed in the reduction (which we will call r). We can then infer the following:

$$\det(U) = \begin{cases} p\,(-1)^r \text{, if } A \text{ is invertible} \\ 0 \text{, if } A \text{ is non-invertible} \end{cases} \tag{31}$$

9.1 Eigenspaces

Definition The **eigenspace** of an *eigenvector* λ is defined as the set of all solutions to the equation:

$$(A - \lambda I)\,\vec{x} = 0 \tag{32}$$

This becomes a subspace of $\mathbb{R}^n$

10 Characteristic Polynomials

The expression $\det(A - \lambda I)$ is known as the **characteristic polynomial** $P_{\text{A}}(\lambda)$ of an $n \times n$ matrix A. In other words:

$$P_{\text{A}}(\lambda) = \det(A - \lambda I) \tag{33}$$

$P_{\text{A}}(\lambda)$ is an n^{th} degree polynomial of λ whose roots are the *eigenvalues* of A

It is important to note that if two different matrices have identical characteristic polynomials, they will have the *same eigenvalues*, but *potentially different eigenspaces*. In addition, their *dimensions* may also differ.

Multiplicity Recall the properties of the characteristic polynomial of an $n \times n$ matrix A, or $P_{\text{A}}(\lambda)$; more specifically, recall the fact that its roots are the eigenvalues of A. If one of these roots repeats itself m times, then it is said to have a **multiplicity** of m.

More rigorously, if λ_1 is one of these aforementioned eigenvalues, we can factorize $P_{\text{A}}(\lambda)$ such that:

$$P_{\text{A}}(\lambda) = (\lambda_1 - \lambda)^m\, g(\lambda) \tag{34}$$

Where $n \geq m \geq 1$ and g is a polynomial of degree $m - n$ such that $g(\lambda) \neq 0$, and m is the *multiplicity* of λ_1

10.1 Similarity

Here is a question: when can two arbitrary matrices $A^{n\times n}$ and $B^{n\times n}$ have the same characteristic polynomial? The answer will come in theorem 12, but first we must define the term *similar*.

Definition Given the conditions above, A is **similar** B if there exists an *invertible* matrix $P^{n\times n}$ such that:

$$A = PBP^{-1} \iff B = P^{-1}AP = QBQ^{-1} \tag{35}$$

Where $Q = P^{-1}$

If we have a third matrix $C^{n\times n}$ and A is similar to B, and B is similar to C, then A is similar to C.

Theorem 12 If $A^{n\times n}$ and $B^{n\times n}$ are *similar*, then A and B have the *same* characteristic polynomial.

Note that this is a *one-way implication*!

11 Diagonalization

Let us begin with a question: which $n \times n$ matrices can we diagonalize? This is explained in the following theorem:

Theorem 5 A matrix $A^{n\times n}$ can be **diagonalized** *if and only if* A has n *linearly independent* eigenvectors.

Explanation In other words, $A = PDP^{-1}$ *if and only if* the matrix $P^{n\times n}$ has linearly independent columns which are also eigenvectors for A. In addition, $D^{n\times n}$ must be a diagonal matrix whose diagonal elements are eigenvalues for A such that they correspond to the eigenvector columns in P in the *exact same order*.

Theorem 6 An $n \times n$ matrix with n *different* eigenvalues can be diagonalized .

Note that this is not a necessary case for diagonalization, but rather one of several opportunities that allow for it.

Theorem 7 If the matrix $A^{n\times n}$ has p distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$, then:

1. For $1 \leq k \leq p$, the *dimension* of the eigenspace for λ_k is less than or equal to the multiplicity of λ_k
2. A can be diagonalized if and only if the *sum* of the dimensions of its eigenspaces is equal to n; this is only the case when $P_{\text{A}}(\lambda)$ can be factored out into n linear factors, and the dimensions of the eigenspaces for each λ_k is equal to the multiplicity of each λ_k
3. If A can be diagonalized, and $\mathfrak{B}_k$ is a basis for the eigenspace of each λ_k, then $\mathfrak{B}_1, \mathfrak{B}_2, ..., \mathfrak{B}_p$ together create a basis for $\mathbb{R}^n$

12 Inner Products, The Norm and Orthogonality

Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$, $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$ and $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n$

Definition We define the **inner product** (or dot product) for two vectors as:

$$\vec{u} \cdot \vec{v} \triangleq u_1v_1 + u_2v_2 + \cdots + u_nv_n \tag{36}$$

This can also be written as a matrix product:

$$\vec{u} \cdot \vec{v} = [\vec{u}]^T [\vec{v}] \tag{37}$$

Properties of the Inner Product

1. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$; the commutative property
2. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$; the distributive property
3. $(c\vec{u}) \cdot \vec{v} = c\,(\vec{u} \cdot \vec{v})$; the associative property
4. $\vec{u}^2 = \vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0 \iff \vec{u} = \vec{0}$

Definition The **norm** (or length) of a vector $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$ is defined as:

$$\|\vec{u}\| \triangleq \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} \tag{38}$$

If a constant $c \in \mathbb{R}^n$ is within the norm operator, then it can be taken be out as follows:

$$\|c\vec{u}\| = c\,\|\vec{u}\| \tag{39}$$

Definition A **unit vector** $\hat{x}$ is any vector with a unit length, such that:

$$\|\hat{x}\| = 1 \tag{40}$$

Finding the Unit Vector Given an arbitrary vector $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$, its unit vector $\hat{u}$ can be found as follows:

$$\hat{u} = \frac{\vec{u}}{\|\vec{u}\|} \tag{41}$$

Definition Suppose we wish to find the **distance** between the endpoints of two arbitrary vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$; to accomplish this, then we can use the following:

$$\mathrm{dist}(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\| \tag{42}$$

Orthogonality To determine whether or not two arbitrary vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ are orthogonal to each other, we can use the inner product:

$$\vec{u} \perp \vec{v} \iff \vec{u} \cdot \vec{v} = 0 \tag{43}$$

It is also possible to use the distances between the vectors, as is shown in the next theorem:

Theorem 2 Two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ are orthogonal to each other if and only if $\|\vec{u} + \vec{v}\| = \|\vec{u} - \vec{v}\|$

12.1 The Orthogonal Complement

Let $W = \text{span}\left(\{\vec{w}_1, \vec{w}_2, ..., \vec{w}_r\}\right) \subset \mathbb{R}^n$ be a subspace

Definition Given the above, we define the **orthogonal complement** of W $\left(W^{\perp}\right)$ as:

$$W^{\perp} = \{\vec{u} \in \mathbb{R}^n \mid \vec{u} \text{ is orthogonal to all } \vec{w} \in W\} \tag{44}$$

Notes We can derive the following properties from the above:

- $\vec{u} \in W^{\perp} \iff \vec{u} \perp \vec{w} \; \forall \; \vec{w} \in W$
- $L = W^{\perp} \iff L^{\perp} = W$
- $W^{\perp} \cap W = \left\{\vec{0}\right\}$

Theorem 3 Let A be an $m \times n$ matrix, then we have that:

$$(\text{row}(A))^{\perp} = \text{nul}\,(A) \tag{45}$$

$$(\text{col}(A))^{\perp} = \text{nul}\left(A^T\right) \tag{46}$$

The Cosine Rule If we have two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, then:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \, \|\vec{v}\| \cos(\theta) \iff \cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|} \tag{47}$$

12.2 The Cauchy-Schwartz Inequality

Since $\cos(\theta) \in [-1, 1]$, and given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, the cosine rule above gives us that:

$$\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|} \in [-1, 1] \tag{48}$$

13 Orthogonal Sets

Let $S = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\} \subset \mathbb{R}^n$

Definition Given the above, S is an **orthogonal set** if $\vec{u}_i \cdot \vec{u}_j = 0 \;\; \forall \;\; i \neq j$

Theorem Let $S\left\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\right\} \subset \mathbb{R}^n$ be an *orthogonal set* where $\vec{u}_i \neq \vec{0} \;\; \forall \;\; i$; then we know that S is linearly independent and a basis for span $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\}$

Definition An **orthogonal basis** for a subspace **W** is a basis as well as an *orthogonal set*

Theorem 5 Let $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\}$ be an *orthogonal basis* for a subspace $\mathbf{W} \subset \mathbb{R}^n$, let $\vec{y} \in \mathbf{W}$ and let $c \in \mathbb{R}$. We then have that:

$$\vec{y} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_p\vec{u}_p \text{ such that } c_i = \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \tag{49}$$

Definition Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ such that $\vec{u} \neq \vec{0}$, and let $L = \text{span}\{\vec{u}\}$. The **orthogonal projection** of our vector $\vec{v}$ on L (or on the span of $\vec{u}$) is defined as:

$$\text{proj}_{\text{L}}(\vec{v}) \triangleq \frac{\vec{v} \cdot \vec{u}}{\vec{u} \cdot \vec{u}}\vec{u} = \hat{y} \tag{50}$$

Definition A set $S = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\} \subset \mathbb{R}^n$ is called **orthonormal** if:

- S is *orthogonal*
- $\|\vec{u}_i\| = 1, \quad \forall\, i$

Theorem 6 If we have the matrix $U^{m\times n} = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n]$ where $\vec{u}_i \in \mathbb{R}^m$ and $\vec{u}_1, \vec{u}_2, ..., \vec{u}_n$ are columns, then U has *orthonormal columns* if and only if $U^TU = I^{n\times n}$, and U is called an **orthogonal matrix**.

Theorem 7 Let $U^{m\times n}$ be an *orthogonal matrix*, and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. We then have the following:

1. $\|U\vec{x}\| = \|\vec{x}\|$
2. $(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot \vec{y}$
3. $(U\vec{x}) \cdot (U\vec{y}) = 0 \iff \vec{x} \cdot \vec{y} = 0$

Theorem 9 Let $\mathbf{W}$ be a subspace of $\mathbb{R}^n$, let $\vec{y} \in \mathbb{R}^n$ and let $\hat{y} = \text{proj}_{\text{W}}(\vec{y})$. We then have that $\hat{y}$ is the element in $\mathbf{W}$ that is closest to $\vec{y}$, or in other words:

$$\text{dist}(\vec{y}, \hat{y}) < \text{dist}(\vec{y}, \vec{v}) \text{ for any } \vec{v} \in \mathbf{W},\ \vec{v} \neq \hat{y} \tag{51}$$

Theorem Let $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\}$ be an *orthonormal basis* for $W \subset \mathbb{R}^n$ with $\vec{y} \in \mathbb{R}^n$, then we have that:

$$\text{proj}_{\text{W}}(\vec{y}) = (\vec{y} \cdot \vec{u}_1)\,\vec{u}_1 + (\vec{y} \cdot \vec{u}_2)\,\vec{u}_2 + \cdots + (\vec{y} \cdot \vec{u}_p)\,\vec{u}_p \tag{52}$$

Now, let $U^{n\times p} = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_p]$, then we can rewrite the above:

$$\text{proj}_{\text{W}}(\vec{y}) = \left(UU^T\right)\vec{y} \tag{53}$$

Generalization If we now take the projection of a vector $\vec{y}$ over a set $U = \{u_1, u_2, ..., u_n\}$, then we find the entire projection with the following sum:

$$\text{proj}_{\text{U}}(\vec{y}) = \sum_{i=1}^{n} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i}\vec{u}_i \tag{54}$$

13.1 Gram-Schmidt Orthogonalization

Let $\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_p\}$ be a basis for $W = \text{span}\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_p\} \subset \mathbb{R}^n$
Below is an algorithm that can be used to find an *orthogonal basis* $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\}$ for W:

1. $\vec{v}_1 = \vec{x}_1$
2. $\vec{v}_2 = \vec{x}_2 - \text{proj}_{\text{span}\{\vec{v}_1\}}(\vec{x}_2)$
3. $\vec{v}_3 = \vec{x}_3 - \text{proj}_{\text{span}\{\vec{v}_1\}}(\vec{x}_3) - \text{proj}_{\text{span}\{\vec{v}_2\}}(\vec{x}_3)$
4. $\vdots$
5. $\vec{v}_p = \vec{x}_p - \text{proj}_{\text{span}\{\vec{v}_1\}}(\vec{x}_p) - \text{proj}_{\text{span}\{\vec{v}_2\}}(\vec{x}_p) - \cdots - \text{proj}_{\text{span}\left\{\vec{v}_{(p-1)}\right\}}(\vec{x}_p)$

To make this an *orthonormal basis*, simply divide each $\vec{v}_i$ by $\|\vec{v}_i\|$

13.2 QR Factorization

Theorem 12 Let A be an $m \times n$ matrix with linearly independent columns. It is then possible to write the following:

$$A = QR \tag{55}$$

Where Q is an $m \times n$ orthogonal matrix and R is an $n \times n$ upper triangular matrix with all diagonal elements $r_{i,i} > 0$

Definition Let A be an $m \times n$ matrix with $m > n$, and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then, a **least squares** solution of the equation $A\vec{x} = \vec{y}$ is:

$$\hat{x} \in \mathbb{R}^n \text{ such that } \|\vec{y} - A\hat{x}\| \leq \|\vec{y} - A\vec{x}\|, \quad \forall\, \vec{x} \tag{56}$$

It may seem like a daunting task to find $\hat{x}$, however we are fortunate in that the following theorem will help us find our desired solution:

Theorem 3 The set of *least squares solutions* of $A\vec{x} = \vec{y}$ is equal to the set of solutions of:

$$\left(A^T A\right)\vec{x} = A^T\vec{y} \tag{57}$$

The above is known as the **normal equation**
Note that $\left(A^T A\right)$ is an $n \times n$ matrix, and that $A^T\vec{y}$ is a vector in $\mathbb{R}^n$

Note Assume that A has a *QR factorization* $A = QR$; to help us solve the equation given in the above theorem, we have that:

$$A^T A\vec{x} = A^T\vec{y} \iff R^T R\vec{x} = R^T Q^T\vec{y} \iff R\vec{x} = Q^T\vec{y} \tag{58}$$

Since R is a triangular matrix, this simplifies our problem significantly.

Terms In the above, we have that:

- A is the **design matrix**
- $\vec{y}$ is the **observation vector**
- $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$ is the **parameter vector**

13.3 Usage

Here are a few generalized situations - note that these are meant to be read in order, and that each $\alpha_i \in \mathbb{R}$ represents a constant coefficient, and each β_i represents an arbitrary function of some variable:

Linear Approximation Suppose we wish to approximate a data set (of n points) with a linear function of the form $y(x) = \alpha_0 + \alpha_1 x$; in essence, we wish for the following to be as small as possible:

$$S = \sum_{i=1}^{n} (y_i - (\alpha_0 + \alpha_1 x_i))^2 \tag{59}$$

We cannot control our y_i values, so we must find a α_0 and α_1 that minimizes our S.

We can reinterpret the above as a formula of two vectors, such that:

$$D = \operatorname{dist}(\vec{y}, \hat{y}) = \|\vec{y} - \hat{y}\| \tag{60}$$

Where we wish to minimize D; we also have that $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and $\hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1 x_1 \\ \alpha_0 + \alpha_1 x_2 \\ \vdots \\ \alpha_0 + \alpha_1 x_n \end{bmatrix}$.

This can in turn be rewritten as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \vec{y} \tag{61}$$

Where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}$

It is now simply a matter of solving for $\vec{y}$ using the *method of least squares*

Polynomial Approximation Now, suppose we wish to approximate another data set (of n points) with a m^{th} degree polynomial of the form $y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_m x^m$; we can generalize the previous example as follows (we still wish to minimize S):

$$S = \sum_{i=1}^{n} \left(y_i - \left(\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \cdots + \alpha_m x_i^m\right)\right)^2 \tag{62}$$

We must now find a set $\{\alpha_0, \alpha_1, ..., \alpha_m\}$ that minimizes our S.

We can once again interpret the above as a formula of two vectors where we wish to minimize a value D:

$$D = \text{dist}\,(\vec{y}, \hat{y}) = \|\vec{y} - \hat{y}\| \tag{63}$$

Where:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \cdots + \alpha_m x_1^m \\ \alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2 + \cdots + \alpha_m x_2^m \\ \vdots \\ \alpha_0 + \alpha_1 x_n + \alpha_2 x_n^2 + \cdots + \alpha_m x_n^m \end{bmatrix} \tag{64}$$

This can once again be rewritten as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \vec{y} \tag{65}$$

Where:

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \tag{66}$$

Finally, we can solve for $\vec{y}$ using the *method of least squares*

General Approximation Let us now get a generalize this completely - suppose we once again have a data set consisting of n points, but this time we wish to approximate it with $y(x) = \alpha_0 + \alpha_1\beta_1(x) + \alpha_2\beta_2(x) + \cdots + \alpha_m\beta_m(x)$. As previously, we wish to minimize the following S:

$$S = \sum_{i=1}^{n} \left(y_i - (\alpha_0 + \alpha_1\beta_1(x_i) + \alpha_2\beta_2(x_i) + \cdots + \alpha_m\beta_m(x_i))\right)^2 \tag{67}$$

We can rewrite this as a function D:

$$D = \text{dist}\,(\vec{y}, \hat{y}) = \|\vec{y} - \hat{y}\| \tag{68}$$

Where:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1\beta_1(x_1) + \alpha_2\beta_2(x_1) + \cdots + \alpha_m\beta_m(x_1) \\ \alpha_0 + \alpha_1\beta_1(x_2) + \alpha_2\beta_2(x_2) + \cdots + \alpha_m\beta_m(x_2) \\ \vdots \\ \alpha_0 + \alpha_1\beta_1(x_n) + \alpha_2\beta_2(x_n) + \cdots + \alpha_m\beta_m(x_n) \end{bmatrix} \tag{69}$$

Once again, this can be written as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \vec{y} \tag{70}$$

Where:

$$A = \begin{bmatrix} 1 & \beta_1(x_1) & \beta_2(x_1) & \cdots & \beta_m(x_1) \\ 1 & \beta_1(x_2) & \beta_2(x_2) & \cdots & \beta_m(x_2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \beta_1(x_n) & \beta_2(x_n) & \cdots & \beta_m(x_n) \end{bmatrix} \tag{71}$$

Finally, we can solve for $\vec{y}$ using the *method of least squares*

14 Symmetrical Matrices

Defintion An $n \times n$ matrix A is considered **symmetrical** if $A = A^T$; note that this means that A absolutely must be a square matrix!

Note *Symmetrical matrices* can always be *diagonalized*

Theorem 1 If A is *symmetrical*, then two *eigenvectors* from different *eigenspaces* will always be orthogonal to each other

Definition A matrix A is **orthogonally diagonizable** if there exists an *orthogonal matrix* P and a *diagonal matrix* D such that:

$$A = PDP^T = PDP^{-1} \tag{72}$$

Note The following statements are equivalent:

- The matrix A is *orthogonally diagonizable*
- The matrix A is symmetrical

Theorem 3 For an $n \times n$ symmetrical matrix A, we have that:

1. A has n real *eigenvalues*, counting multiplicities
2. The dimensions of the *eigenspace* for each *eigenvalue* λ equals the *multiplicity* of λ as a root of the *characteristic equation*
3. The *eigenspaces* are mutually orthogonal, in the sense that *eigenvectors* corresponding to different *eigenvalues* are orthogonal; in other words, if $\vec{v}_1$ and $\vec{v}_2$ are vectors from different *eigenspaces*, then $\vec{v}_1 \cdot \vec{v}_2 = 0$
4. A is *orthogonally diagonizable*

14.1 Spectral Decomposition

Definition Let A be an $n \times n$ *symmetrical matrix*, let P be an *orthogonal matrix* such that $P = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n]$ and let D be a *diagonalized matrix* such that:

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{73}$$

We know that $A = PDP^T$, so let us rewrite this expression:

$$A = PDP^T = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \tag{74}$$

With further manipulation, we are left with the **spectral decomposition** of A:

$$A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \lambda_2 \vec{u}_2 \vec{u}_2^T + \cdots + \lambda_n \vec{u}_n \vec{u}_n^T \tag{75}$$

Usage Given the same conditions as above, the expression $A\vec{x}$ can (using *spectral decomposition*) be rewritten as:

$$A\vec{x} = \lambda_1 \operatorname{proj}_{\operatorname{span}(\vec{u}_1)}(\vec{x}) + \lambda_2 \operatorname{proj}_{\operatorname{span}(\vec{u}_2)}(\vec{x}) + \cdots \lambda_n \operatorname{proj}_{\operatorname{span}(\vec{u}_n)}(\vec{x}) \tag{76}$$

15 Quadratic Forms

Definition Let A be an $n \times n$ symmetrical matrix such that $\vec{x} \in \mathbb{R}^n$ We then define the **quadratic form** of A as:

$$Q(\vec{x}) \triangleq (\vec{x})^T A\vec{x} \in \mathbb{R} \tag{77}$$

Generally, we have that:

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 = [x_1, x_2] \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{78}$$

Generally, a homogeneous n^{th} degree polynomial can be written in the form:

$$\vec{x}^T A\vec{x} \mid A \text{ is symmetrical} \tag{79}$$

We wish to write $Q(\vec{x}) = \vec{x}^T A\vec{x}$ as simply as possible; we can accomplish this through a change of variables - this is a good solution, since it avoids dealing with products like $x_i x_j$ where $i \neq j$

We can show this with an example - let A be a 2×2 matrix:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \tag{80}$$

We also have a matrix P, such that:

$$Q = P^T AP \tag{81}$$

This gives us:

$$Q(x_1, x_2) = [x_1, x_2]\, A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = ax_1^2 + bx_1x_2 + cx_2^2 \tag{82}$$

We can define $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$; this gives us then:

$$Q(x_1, x_2) = (P\vec{y})^T A (P\vec{y}) = dy_1^2 + ey_2^2 \tag{83}$$

And our product is removed

Generalization We have $Q(\vec{x}) = (\vec{x})^T A\vec{x}$, where A is a symmetrical $n \times n$ matrix; in addition, let $P = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n]$ be an *orthogonal matrix*. We can then get the following:

$$D = \begin{bmatrix} \lambda_{1,1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \lambda_{2,2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \lambda_{3,3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{n-2,n-2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \lambda_{n,n} \end{bmatrix} \tag{84}$$

We then set $\vec{x} = P\vec{y}$ and $\vec{x}^T A\vec{x} = (P\vec{y})^T A (P\vec{y})$ and end up with:

$$\vec{x}^T A\vec{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{85}$$

Theorem 4 If A is a *symmetrical* $n \times n$ matrix, then there exists an *orthogonal matrix* P such that a variable change $\vec{x} = P\vec{y}$ transforms $\vec{x}^T A\vec{x}$ into $\vec{y}^T D\vec{y}$, where D is a *diagonal matrix*. That is to say, the *quadratic form* ends up without a product of the form $x_i x_j$ where $i \neq j$.

Geometric Interpretation Let $Q(x_1, x_2) = 1$ then we have that:

- If $A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, then $\vec{x}^T A\vec{x} = 1 \implies Q(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2 = 1$ and we end up with an ellipse
- If $B = \begin{bmatrix} \lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}$, then $\vec{x}^T B\vec{x} = 1 \implies Q(x_1, x_2) = \lambda_1 x_1^2 - \lambda_2 x_2^2 = 1$ and we end up with a hyperbola
- If $C = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$, then $\vec{x}^T C\vec{x} = 1 \implies Q(x_1, x_2) = \lambda x_1^2 + \lambda x_2^2 = 1$ and we end up with a circle of radius $\frac{\sqrt{\lambda}}{\lambda}$

- If $D = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, then $\vec{x}^T D\vec{x} = 1 \implies Q(x_1, x_2) = -\lambda x_1^2 - \lambda x_2^2$ and we end up with a null value

Definition Let A be a symmetrical $n \times n$ matrix. We can then conclude that:

1. **Positive definite** if $Q\vec{x} > 0 \;\; \forall \;\; \vec{x} \neq 0$; for example, $2x_1^2 + x_2^2 = Q_1(x_1, x_2)$
2. **Negative definite** if $Q\vec{x} < 0 \;\; \forall \;\; \vec{x} \neq 0$; for example, $-2x_1^2 - x_2^2 = Q_2(x_1, x_2)$
3. **Indefinite** if $Q\vec{x}$ has both positive and negative values; for example $2x_1^2 - x_2^2 = Q_3(x_1, x_2)$

Theorem Let A and Q be the same as above; then:

1. Q is *positive definite* if and only if A has *strictly positive* eigenvalues
2. Q is *negative definite* if and only if A has *strictly negative* eigenvalues
3. Q is *indefinite* if and only if A has *both positive and negative* eigenvalues

16 Optimization with Parameters

Let $Q(\vec{x}) = \vec{x}^T A\vec{x}$ where A is a symmetrical $n \times n$ matrix and $\vec{x} \in \mathbb{R}^n$. How can we find the maximum or minimum value of $Q(\vec{x})$ when $\|\vec{x}\| = 1$?

Generalized Process We assume the above conditions still apply, and that $\vec{y} \in \mathbb{R}^n$; we wish to find an orthogonal matrix $P = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n]$ such that $\vec{x} = P\vec{y}$ as well as a diagonal matrix:

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \text{ such that } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{86}$$

This gives us that:

$$Q(\vec{x}) = \vec{y}^T P^T AP\vec{y} = \vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{87}$$

Note that if $\vec{x} = P\vec{y}$, then we have that:

$$\|\vec{x}\|^2 = \vec{x}^T\vec{x} = \vec{y}^T P^T P\vec{y} = \vec{y}^T\vec{y} = \|\vec{y}\|^2 \tag{88}$$

This implies that $\max\{Q(x)\} = \max\left\{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2\right\}$.
We also know that $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \leq \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 = \lambda_1$
This means that our maximum ends up being λ_1 such that $Q(\pm\vec{u}_1) = \lambda_1$, with a minimum of λ_n such that $Q(\pm\vec{u}_n) = \lambda_n$

Theorem 8 Let $Q(\vec{x}) = \vec{x}^T A \vec{x}$ where A is a symmetrical $n \times n$ matrix and $\vec{x} \in \mathbb{R}^n$. Let $A = PDP^T$ such that $P = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_n]$ and:

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \text{ such that } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{89}$$

For $k = 1, ..., n$, the maximum of $Q(\vec{x})$ when $\|\vec{x}\| = 1$ and $\vec{x}^T \vec{u}_1 = \cdots = \vec{x}^T \vec{u}_{k-1} = 0$ is equal to the eigenvalue λ_k where $Q(\pm \vec{u}_k) = \lambda_k$.

17 Singular-Value Decomposition

Assuming that A is a symmetrical $n \times n$ matrix, we can write that $A = PDP^T$ where D is a diagonal matrix and P is an orthogonal matrix. This type of decomposition makes it easy to run calculations with A.

Now, let A be an $m \times n$ matrix – we can write the **singular-value decomposition** of A (or its **SVD**) as $A = U\Sigma V^T$, where U is an orthogonal $m \times m$ matrix, V is an orthogonal $n \times n$ matrix and Σ is an $m \times n$ matrix of the form $\Sigma = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right]$.

We also have that $r = \text{rank}(A) = \dim(\text{col}(A))$ and that D is an $r \times r$ diagonal matrix of the form:

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} \text{ such that } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \tag{90}$$

Finding the SVD Let $V = [\vec{v}_1, \vec{v}_2, ..., \vec{v}_n]$, where each $\vec{v}_i$ is an eigenvector for $A^T A$ and let:

$$U = \left[\frac{A\vec{v}_1}{\sigma_1}, \frac{A\vec{v}_2}{\sigma_2}, ..., \frac{A\vec{v}_r}{\sigma_r}\right] \tag{91}$$

Which can alternatively be written as:

$$U = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_r] \tag{92}$$

If $r < m$, use *Gram-Schmidt Orthogonalization* to extend U into an orthonormal basis for $\mathbb{R}^m$ such that we end up with:

$$U = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_r, \vec{u}_{r+1}, ..., \vec{u}_m] \tag{93}$$

Note that $\|\vec{u}_i\| = 1$ since $\vec{u}_i = \frac{A(\vec{v}_i)}{\sigma_i}$ and $\|A(\vec{v}_i)\| = \sigma_i$ and that $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_r\}$ is orthogonal because $\{A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_r\}$ is orthogonal.

Finally, we will need the matrix Σ to complete the SVD:

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right] \tag{94}$$

Explanation The reason this is possible is due to the following:

First, we have that:

$$AV = [A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_n] = [\sigma_1\vec{u}_1, \sigma_2\vec{u}_2, ..., \sigma_r\vec{u}_r, 0, ..., 0] \tag{95}$$

Next, we can find that equation 95 is equivalent to the following:

$$U\Sigma = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right] = [\sigma_1\vec{u}_1, \sigma_2\vec{u}_2, ..., \sigma_r\vec{u}_r, 0, ..., 0] \tag{96}$$

This means that:

$$AV = U\Sigma \tag{97}$$

Since V is an orthogonal matrix, it is invertible with the added property that $V^{-1} = V^T$; as a result, we have that:

$$A = U\Sigma V^{-1} = U\Sigma V^T \tag{98}$$

Theorem 9 Assuming all the above conditions still apply, we have that $\{A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_r\}$ is an orthogonal basis for $\text{col}\,(A)$ and that $\text{rank}\,(A) = r$

The Basis of row(A) Let $A = U\Sigma V^T$ be the *singular-value decomposition* of an $m \times n$ matrix A; we also have that $U = [\vec{u}_1, \vec{u}_2, ...\vec{u}_r, \vec{u}_{r+1}, ..., \vec{u}_n]$ such that $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$ for $i \leq r$ and $\sigma_i < \|A\vec{v}_i\|$.

From theorem 9, we know that $\{A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_r\}$ is an orthogonal basis for $\text{col}\,(A)$.

We also have that:

$$[\vec{u}_1, \vec{u}_2, ..., \vec{u}_r] \text{ is an orthonormal basis for } \text{col}\,(A) \quad | \quad r = \text{rank}\,(A)\,,\ A\vec{v}_i = 0 \text{ for } i > r \tag{99}$$

This implies that $\{\vec{u}_{r+1}, ..., \vec{v}_n\}$ becomes an orthonormal basis for $\text{nul}\,(A) = (\text{row}\,(A))^\perp$

Next, we know that:

$$(\text{nul}\,(A))^\perp = (\text{span}\,\{\vec{v}_{r+1}, ..., \vec{v}_n\})^\perp = \{\vec{v}_1, ..., \vec{v}_r\} \tag{100}$$

Finally, since $\{\vec{v}_1, ..., \vec{v}_n\}$ is an orthonormal basis, we have that:

$$(\text{nul}\,(A))^\perp = \left((\text{row}\,(A))^\perp\right)^\perp = \text{row}\,(A) \tag{101}$$

Where $\{\vec{v}_1, ..., \vec{v}_r\}$ is an orthonormal basis for $\text{row}\,(A)$

17.1 The Matrix of a Linear Transformation

[1] Let $\mathbf{V}$ be an n-dimensional vector space, let $\mathbf{W}$ be an m-dimensional vector space and let T be an arbitrary linear transformation $T : \mathbf{V} \to \mathbf{W}$. To associate a matrix with T, choose ordered bases $\mathfrak{B}$ and $\mathfrak{C}$ for $\mathbf{V}$ and $\mathbf{W}$, respectively.

Given any $\vec{x}$ in $\mathbf{V}$, the coordinate vector $[\vec{x}]_{\mathfrak{B}}$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[T(\vec{x})]_{\mathfrak{C}}$ is in $\mathbb{R}^m$. In addition, let $\left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$ be the basis $\mathfrak{B}$ for $\mathbf{V}$

This gives us that:

$$[T(\vec{x})]_{\mathfrak{C}} = M[\vec{x}]_{\mathfrak{B}} \tag{102}$$

Where:

$$M = \left[\left[T\left(\vec{b}_1\right)\right]_{\mathfrak{C}}, \left[T\left(\vec{b}_2\right)\right]_{\mathfrak{C}}, ..., \left[T\left(\vec{b}_n\right)\right]_{\mathfrak{C}}\right] \tag{103}$$

The matrix M is a matrix representation of T, called the **matrix for T relative to the bases $\mathfrak{B}$ and $\mathfrak{C}$**

Properties We have that:

1. T is injective $\iff$ T_{M} is injective
2. T is surjective on $\mathbf{W}$ $\iff$ T_{M} is surjective on $\mathbb{R}^m$
3. T is isomorphic $\iff$ T_{M} is isomorphic

Corollary We have a linear transformation $T_m : \mathbb{R}^n \to \mathbb{R}^m$ such that $T_{\text{M}}(\vec{v}) = M\vec{v}$ (where M is an $m \times n$ matrix) and we also have that $\text{nul}(M) = \text{rank}(M)$. If we wish to make a similar equation of the form $T : \mathbf{V} \to \mathbf{W}$, where $\mathbf{V}, \mathbf{W}$ are subspaces.

We know as a result of the above that $n = \dim(\text{nul}(M)) + \text{rank}(M)$. This implies then that:

$$n = \dim(\mathbf{V}) = \dim(\ker(T)) + \dim(T(\mathbf{V})) \tag{104}$$

Special Case Let $n = m$, $\mathbf{V} = \mathbf{W}$, and $\mathfrak{B} = \mathfrak{C}$ such that $T : \mathbf{V} \to \mathbf{V}$. If we denote M with $[T]_{\mathfrak{B}}$ then we have that:

$$[T]_{\mathfrak{B}} = [[T(b_1)]_{\mathfrak{B}}, [T(b_2)]_{\mathfrak{B}}, ..., [T(b_n)]_{\mathfrak{B}}] \tag{105}$$

Corollary Assume that $T : \mathbf{V} \to \mathbf{V}$ is injective, surjective and isomorphic; if we then define a $T^{-1} : \mathbf{V} \to \mathbf{V}$ such that $T^{-1}(\vec{x}) = \vec{y} \iff T(\vec{y}) = \vec{x}$ we have that:

$$\left[T^{-1}\right]_{\mathfrak{B}} = \left([T]_{\mathfrak{B}}\right)^{-1} \tag{106}$$

Similarity Recall that $[\vec{v}]_{\mathfrak{B}} = \underset{\mathfrak{B}\leftarrow\mathfrak{C}}{P}[\vec{v}]_{\mathfrak{C}}$. A consequence of this is as follows:

$$[T]_{\mathfrak{C}} = \underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P}[T]_{\mathfrak{B}}\left(\underset{\mathfrak{C}\leftarrow\mathfrak{B}}{P}\right)^{-1} \tag{107}$$

This also means that $[T]_{\mathfrak{B}}$ and $[T]_{\mathfrak{C}}$ are similar

[1] Linear Algebra and its Applications 5th edition, p.307

18 Complex Eigenvalues

Let A be a real $n \times n$ matrix with an n$^{\text{th}}$ degree characteristic polynomial with real coefficients:

$$P_{\text{A}}(\lambda) = \det(A - \lambda I) \tag{108}$$

In addition to this, P_{A} may have **complex roots**.

Now, let $\mathbb{C}$ be the complex numbers, and let $\mathbb{C}^n$ be an n-vector of the $\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ n \end{bmatrix}$, $z_i \in \mathbb{C}$ and $i = 1, 2, ..., n$; we also have that $\mathbb{R}^n \subset \mathbb{C}^n$.

Definition Given the above, we have that:

$$T_{\text{A}} : \mathbb{C}^n \to \mathbb{C}^n \text{ whereby } T_{\text{A}}(\vec{z}) = A\vec{z} \tag{109}$$

If $\lambda \in \mathbb{C}$, we can show that there exists a $\vec{z} \neq \vec{0}$ such that $A\vec{z} = \lambda\vec{z}$ *if and only if* $\det(A - \lambda I) = 0 \iff P_{\text{A}}(\lambda) = 0$. We then have that $\vec{z}$ is an *eigenvector* of λ.

Imaginary and Real Parts Let $\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}$; for $j = 1, 2, ..., n$ we have that:

$$z_j = x_j + iy_j \text{ where } x_j = \Re(z_j) \text{ and } y_j = \Im(z_j) \tag{110}$$

This means that we can divide $\vec{z}$ into:

$$\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + i \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \vec{x} + i\vec{y} \tag{111}$$

Where $\vec{x} = \Re(\vec{z})$ and $\vec{y} = \Im(\vec{z})$.

Definition Let A be a real-valued $n \times n$ matrix with a complex eigenvalue λ. For a matrix B with complex elements, we have that:

$$A\overline{\vec{z}} = \overline{\lambda\vec{z}} \tag{112}$$

Where $\vec{z} = \Re(\vec{z}) + i\Im(\vec{z})$ and $\overline{\vec{z}} = \Re(\vec{z}) - i\Im(\vec{z})$.
This can be interpreted as saying that $\overline{\vec{z}}$ is an eigenvector with an eigenvalue $\overline{\lambda}$.

18.1 Usage of Eigenvalues

Let A be an $n \times n$ matrix, and let $\vec{x}_0$ be a vector such that $\vec{x}_0 \in \mathbb{R}^n$ and $\vec{x}_0 \neq \vec{0}$; we can create a discrete and dynamic system where:

$$\vec{x}_{k+1} = A\vec{x}_k = A^k\vec{x}_0 \tag{113}$$

Let us now assume that A is diagonizable (with only *real* eigenvalues) and that it has a basis of eigenvectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ with corresponding eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ such that $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n|$. Then, we can *always* write that:

$$\vec{x}_0 = c_1\vec{v}_1 + c_2\vec{v}_2 + ... + c_n\vec{v}_n \tag{114}$$

Note that we also know that:

$$A^k\vec{v}_j = \lambda^k\vec{v}_j \implies \vec{x}_k = c_1\lambda_1^k\vec{v}_1 + c_2\lambda_2^k\vec{v}_2 + ... + c_n\lambda_n^k\vec{v}_n \tag{115}$$

18.2 Disconnected Systems

Let A be an $n \times n$ matrix. Assume that A is diagonizable and has the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ with an eigenvector basis $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$. Let $P = [\vec{v}_1, \vec{v}_2, ..., \vec{v}_n]$ and let us perform a coordinate shift $\vec{x} = P\vec{y}$, $\vec{y} = P^{-1}\vec{x}$.

Given the system $\vec{x}_{k+1} = A\vec{x}_k$, we also have that:

$$\vec{x}_{k+1} = P\vec{y}_{k+1} = A\vec{x}_k = AP\vec{y}_k \implies \vec{y} = P^{-1}AP\vec{y}_k \tag{116}$$

Where:

$$P^{-1}AP = D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{117}$$

We then have that:

$$\vec{y}_{k+1} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \vec{y}_k \tag{118}$$

Since $\vec{y}_k = \begin{bmatrix} y_{k,1} \\ y_{k,2} \\ \vdots \\ y_{k,n} \end{bmatrix}$, we have that:

$$\begin{bmatrix} y_{k+1,1} \\ y_{k+1,2} \\ \vdots \\ y_{k+1,n} \end{bmatrix} = \begin{bmatrix} \lambda_1 y_{k,1} \\ \lambda_2 y_{k,2} \\ \vdots \\ \lambda_n y_{k,n} \end{bmatrix} \text{ where } \vec{y}_k = \begin{bmatrix} \lambda_1^k y_{0,1} \\ \lambda_2^k y_{0,2} \\ \vdots \\ \lambda_n^k y_{0,n} \end{bmatrix} \tag{119}$$

We then have a **decoupled system**:

$$y_{k+1,j} = \lambda_j y_{k,j} \implies y_{k,j} = \lambda_j^k y_{0,j} \tag{120}$$

Note If $|\lambda_j| > 1$ for all j, and $\vec{x}_0 \neq \vec{0}$ (also implying $\vec{y}_0 \neq \vec{0}$) then it *must* be the case that:

$$\lim_{k\to\infty} \|\vec{y}_k\| = \infty \implies \lim_{k\to\infty} \|\vec{x}_k\| = \lim_{k\to\infty} \|P\vec{y}_k\| = \infty \tag{121}$$

Attractors and Repellers If we have a system of the form $\vec{y}_k = |\lambda|^k M\vec{y}_0$, then:

- If $|\lambda| > 0$, $\vec{0}$ is a repeller
- If $|\lambda| < 0$, $\vec{0}$ is an attractor

Application to Differential Equations Let A be an $n \times n$ diagonal matrix. If we have a system of differential equations:

$$\vec{x}(t) = A\vec{x}(t) \tag{122}$$

Where:

$$x_i' = \lambda_i x_i \iff x_i = c_i e^{\lambda_i t} \tag{123}$$

And:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} c_1 e^{\lambda_1 t}\vec{e}_1 + c_2 e^{\lambda_2 t}\vec{e}_2 + \cdots + c_n e^{\lambda_n t}\vec{e}_n \tag{124}$$

Then we have that:

$$\vec{x}(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t}\hat{e}_i \iff |\vec{x}(t)|^2 = \sum_{i=1}^{n} c_i^2 e^{2\lambda_i t} \tag{125}$$

- If $\lambda_i > 0$ for all i and $c_i \neq 0$ for *at least* one i, then we have that $|\vec{x}(t)|^2 \to \infty$ as $t \to \infty$. Here, we have that $\vec{0}$ is a *repeller*.
- If $\lambda_i < 0$ for all i and $c_i \neq 0$ for *at least* one i, then we have that $|\vec{x}(t)|^2 \to 0$ as $t \to \infty$. Here, we have that $\vec{0}$ is a *repeller*

Decoupled Systems If we have a case where $\vec{x}'(t) = D\vec{x}(t)$ where D is a diagonal matrix, then we call such a system *decoupled*

Assume now that $\vec{x}'(t) = A\vec{x}(t)$ such that A is *not* a diagonal matrix (but is still diagonizable). We can find a P such that:

$$P = [\vec{v}_1, \vec{v}_2, ..., \vec{v}_n] \tag{126}$$

Where $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ is a basis of eigenvectors, with the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$.

We have that $A = PDP^{-1}$ and:

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{127}$$

We then change variables such that $\vec{x} = P\vec{y}$ and $\vec{y} = P^{-1}\vec{x}$, with a solution: $\vec{x}(t)$. Then:

$$\vec{y}'(t) = \left(P^{-1}\vec{x}(t)\right) = P^{-1}\vec{x}(t) = P^{-1}A\vec{x}(t) = P^{-1}AP\vec{y}(t) = D\vec{y} \tag{128}$$

Where $\vec{y}'(t) = D\vec{y}(t)$. That is to say, $\vec{y}(t) = c_1\lambda_1 e^{\lambda_1 t}\vec{e}_1 + c_1\lambda_1 e^{\lambda_1 t}\vec{e}_1 + ... + c_1\lambda_1 e^{\lambda_1 t}\vec{e}_1$

So a fundamental set of solutions is:

$$e^{\lambda_1 t}\vec{v}_1, e^{\lambda_2 t}\vec{v}_2, ..., e^{\lambda_n t}\vec{v}_n \tag{129}$$

Solving for Decoupled Systems with Complex Eigenvalues Let A have a complex eigenvalue λ with a complex eigenvector $\vec{z}$ such that $A\vec{z} = \lambda\vec{z}$. Then, it is *always* the case that if we write$\vec{z}(t) = e^{\lambda t}\vec{z}$, then $A\vec{z}'(t) = A\vec{z}(t)$, where:

$$e^{\lambda t} = e^{at}e^{ibt} \text{ if } \lambda = a + bi \tag{130}$$

And:

$$\vec{z}' = \lambda e^{\lambda t}\vec{z} = e^{\lambda t}A\vec{z} = A\left(e^{\lambda t}\vec{z}\right) \tag{131}$$

We know that $\overline{\vec{z}}$ is an eigenvector for A with the eigenvalue $\overline{\lambda}$, and set $\overline{\vec{z}}(t) = e^{\overline{\lambda} t}\overleftarrow{\overline{z}}$, which becomes our *solution*.

In addition to the above, we also have the following two equivalent solutions:

$$\frac{1}{2}\left(\vec{z}(t) + \overline{\vec{z}}(t)\right) = \Re\left\{\vec{z}(t)\right\} \tag{132}$$

$$\frac{1}{2i}\left(\vec{z}(t) - \overline{\vec{z}}(t)\right) = \Im\left\{\vec{z}(t)\right\} \tag{133}$$

18.3 The Power Method

We have an $n \times n$ matrix A, (which is diagonalizable) with the eigenvales $\lambda_1, \lambda_2, ..., \lambda_n$ such that $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$.

We can then use the following algorithm to find λ_1 numerically; we begin with a basis of eigenvectors $\{\vec{v}_1, \vec{v}_2, ...\vec{v}_n\}$ where $\vec{x}_0 \in \mathbb{R}^n$, $\vec{x}_0 c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$. Also, assume that $c_1 \neq 0$.

Then, we can set up the following:

$$A^k\vec{x}_0 = c_1\lambda_1^k\vec{v}_1 + c_2\lambda_2^k\vec{v}_2 + \cdots + c_n\lambda_n^k\vec{v}_n \tag{134}$$

Where:

$$\frac{A^k\vec{x}_0}{\lambda_1^k} = c_1\frac{\lambda_1}{\lambda_1}^k\vec{v}_1 + c_2\frac{\lambda_2}{\lambda_1}^k\vec{v}_2 + \cdots + c_n\frac{\lambda_n}{\lambda_1}^k\vec{v}_n \tag{135}$$

Such that $\frac{\lambda_i}{\lambda_1} < 1$ and $i > 1$. This gives us that $\frac{A^k\vec{x}_k}{\lambda_1^k} \to c_1\vec{v}_1$ (that is to say, the direction of $A\vec{x}_0$ is the same as the direction of $c_1\vec{v}_1$, or the direction of $\vec{v}_1$)

The Algorithm

1. Start with an arbitrary $\vec{x}_0$ such that if $\vec{x}_0 = \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ \vdots \\ x_{0,k} \end{bmatrix}$. Then, we have that $\max\limits_j (x_{0,j}) = \max\limits_j |x_{0,j}| = 1$

2. For $k = 0, 1, ...$

3. Find $A\vec{x}_k = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

4. Let i be restricted such that $|x_i| = \max_j |x_j|$ and $\mu_k = x_i$

5. Set $x_{k+1} = \left(\frac{1}{\mu_k}\right) A\vec{x}_k$

Then, we will have that $\mu_k \to \lambda_1$ and $x_1 \to$ [eigenvector of λ_1]

19 The Inner Product Space

In $\mathbb{R}^n$, we have that the *dot product* is defined as:

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \tag{136}$$

1. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
2. $(\vec{u}_1 + \vec{u}_2) \cdot \vec{v} = \vec{u}_1 \cdot \vec{v} + \vec{u}_2 \cdot \vec{v}$
3. $c\vec{u} \cdot \vec{v} = c\,(\vec{u} \cdot \vec{v})$
4. $\vec{u} \cdot \vec{u} \geq 0$

Definition Let $\mathbf{V}$ be a vector space. An inner product in $\mathbf{V}$ is a function that for each pair $\vec{u}, \vec{v} \in \mathbf{V}$ returns a real number $\langle \vec{u}, \vec{v} \rangle$ and fulfills the following:

1. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
2. $\langle \vec{u}_1 + \vec{u}_2, \vec{v} \rangle = \langle \vec{u}_1, \vec{v} \rangle + \langle \vec{u}_2, \vec{v} \rangle$
3. $\langle c\vec{u}, \vec{v} \rangle = c\,\langle \vec{u}, \vec{v} \rangle$
4. $\langle \vec{u}, \vec{u} \rangle \geq 0$
5. $\langle \vec{u}, \vec{u} \rangle = 0 \iff \vec{u} = \vec{0} \in \mathbf{V}$

A *vector space* with an inner product is called an **inner product space**

Norm Let $\mathbf{V}$ be a vector space with an inner product, with $\vec{v} \in \mathbf{V}$. We can then define the norm of $\vec{v}$ as:

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle} \tag{137}$$

Orthogonalization Given a *finite dimensional* vector space $\mathbf{V}$ with an inner product, and given a basis $\mathfrak{B} = \left\{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\right\}$, we can use a *Gram-Schmidt* algorithm and make this basis orthogonal:

1. $\vec{v}_1 = \vec{b}_1$
2. $\vec{v}_2 = \vec{b}_2 - \frac{\left\langle \vec{b}_2, \vec{v}_1 \right\rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1$
3. $\vec{v}_3 = \vec{b}_3 - \frac{\left\langle \vec{b}_3, \vec{v}_1 \right\rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\left\langle \vec{b}_3, \vec{v}_2 \right\rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}\vec{v}_2$
4. $\vdots$

Projections If V is an *inner product space*, and W is an n-dimensional subspace with an orthogonal basis $\{\vec{q}_1, \vec{q}_2, ..., \vec{q}_n\}$ where $\dim(W) = n$ and $\vec{v} \in V$, then we can define the projection of $\vec{v}$ on w as:

$$\mathrm{proj}_{\mathrm{W}}(\vec{v}) = \frac{\langle \vec{v}, \vec{q}_1 \rangle}{\langle \vec{q}_1, \vec{q}_1 \rangle}\vec{q}_1 + \frac{\langle \vec{v}, \vec{q}_2 \rangle}{\langle \vec{q}_2, \vec{q}_2 \rangle}\vec{q}_2 + \cdots + \frac{\langle \vec{v}, \vec{q}_n \rangle}{\langle \vec{q}_n, \vec{q}_n \rangle}\vec{q}_n \tag{138}$$

Now, let us define $\hat{\vec{v}} \triangleq \mathrm{proj}_{\mathrm{W}}(\vec{v})$. With this, we have that $\hat{\vec{v}}$ is the vector in W that most closely approximates $\vec{v}$. This means that:

$$\left\| \vec{v} - \hat{\vec{v}} \right\| < \| \vec{v} - \vec{w} \| \text{ for all } \vec{w} \in W \text{ such that } \vec{w} \neq \hat{\vec{v}} \tag{139}$$

In addition to this, we have that if $\vec{v}, \vec{u} \in V$ such that $\vec{v} \perp \vec{u}$ (that is to say, $\langle \vec{v}, \vec{u} \rangle = 0$), then we have that:

$$\| \vec{v} + \vec{w} \|^2 = \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle = \langle \vec{v}, \vec{v} \rangle + 2\langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{w} \rangle \tag{140}$$

This gives us that $\vec{v} - \mathrm{proj}_{\mathrm{W}}(\vec{v}) \in W^{\perp}$. This then gives us that:

$$\| \vec{v} \|^2 = \| \mathrm{proj}_{\mathrm{W}}(\vec{v}) \|^2 + \| \vec{v} - \mathrm{proj}_{\mathrm{W}}(\vec{v}) \| \tag{141}$$

Note that this gives us that:

$$\| \vec{v} \| \leq \| \mathrm{proj}_{\mathrm{W}}(\vec{v}) \| \tag{142}$$

Theorem For $\vec{u}, \vec{v} \in V$ we have that:

$$|\langle \vec{u}, \vec{v} \rangle| \leq \| \vec{u} \| \, \| \vec{v} \| \tag{143}$$

Theorem 17 The *triangle inequality*

$$\| \vec{u} + \vec{v} \| \leq \| \vec{u} \| + \| \vec{v} \| \tag{144}$$

20 Appendix

20.1 Terminology

Codomain The **codomain** or **target set** of a function is the set Y into which all of the output of the function is constrained to fall. It is the set Y in the notation $f : X \to Y$. The codomain is also sometimes referred to as the **range** but that term is ambiguous as it may also refer to the image. [2]

Diagonal Matrices A matrix that is both upper and lower triangular is called a diagonal matrix.[3]

A diagonal matrix:

$$\begin{bmatrix} a_{1,1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & a_{2,2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & a_{3,3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-2,n-2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & a_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_{n,n} \end{bmatrix} \tag{145}$$

Injectivity An **injective** (or **1-1**) function is a function $f : A \to B$ with the following property. For every element b in the codomain B there is maximum one element a in the domain A such that $f(a) = b$. [4,5]

Isomorphy A function is **isomorphic** if it is both injective and surjective

Square Matrices A square matrix is a matrix with the same number of rows and columns. An $n \times n$ matrix is known as a square matrix of order n. Any two square matrices of the same order can be added and multiplied.[6]

Surjectivity A **surjective** (or **onto**) function is a function $f : A \to B$ with the following property. For every element b in the codomain B there is *at least one* element a in the domain A such that $f(a) = b$. This means that the range and codomain of f are the same set.[7,8]

---

[2] https://en.wikipedia.org/wiki/Codomain

[3] https://en.wikipedia.org/wiki/Triangular_matrix

[4] Weisstein, Eric. "Surjective function" (in English). From MathWorld–a Wolfram Web Resource. Retrieved January 2014.

[5] C.Clapham, J.Nicholson (2009). "Oxford Concise Dictionary of Mathematics, Onto Mapping" (in English). addison-Wesley. p. 568. Retrieved January 2014.

[6] https://en.wikipedia.org/wiki/Square_matrix

[7] Weisstein, Eric. "Surjective function" (in English). From MathWorld–a Wolfram Web Resource. Retrieved January 2014.

[8] C.Clapham, J.Nicholson (2009). "Oxford Concise Dictionary of Mathematics, Onto Mapping" (in English). addison-Wesley. p. 568. Retrieved January 2014.

Triangular Matrices A square matrix is called lower triangular if all the entries above the main diagonal are zero. Similarly, a square matrix is called upper triangular if all the entries below the main diagonal are zero. A triangular matrix is one that is either lower triangular or upper triangular.[9]

An upper triangular matrix:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n-2} & a_{1,n-1} & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,n-2} & a_{2,n-1} & a_{2,n} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,n-2} & a_{3,n-1} & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-2,n-2} & a_{n-2,n-1} & a_{n-2,n} \\ 0 & 0 & 0 & \cdots & 0 & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_{n,n} \end{bmatrix} \tag{146}$$

A lower triangular matrix:

$$\begin{bmatrix} a_{1,1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & \cdots & 0 & 0 & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{n-2,1} & a_{n-2,2} & a_{n-2,3} & \cdots & a_{n-2,n-2} & 0 & 0 \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \cdots & a_{n-1,n-2} & a_{n-1,n-1} & 0 \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n-2} & a_{n,n-1} & a_{n,n} \end{bmatrix} \tag{147}$$

[9] https://en.wikipedia.org/wiki/Triangular_matrix