Matrix Algebra

Jan Tomek

November 30, 2022

Contents

1 Inner Product Space **5**

2 Orthogonal Basis and Gram-Schmidt Process **15**

3 Properties of Matrices **19**

4 Normal, Unitary and Simiar Matrices **23**

5 Positive/Negative (Semi)-Definite Matrices **25**

6 Woodbury and Sherman-Morrison **27**

7 Inverse and Determinant of Partitioned Symmetric Matrices **31**

8 Unitary Transform **35**

9 Eigenvalue Decomposition **39**

10 Generalized Eigenvalue Problem **47**

11 Rayleigh Quotient **51**

12 Normal Matrices and Diagonalizability **53**

13 Centering Matrix **57**

14 Matrix Usage Examples **61**
14.1 Polynomial equations 61
14.2 Lagrange interpolation polynomials 62

14.3 3D and 2D 66
14.3.1 Projection from 3D to 2D: 66
14.3.2 Model movement 67

Chapter 1

Inner Product Space

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node1.html`

- **Vector space**

  A *vector space* over a field F is a set V which closed under the following two operations of addition and scalar multiplication defined for its members (called *vectors*), i.e., the results of the operations are also members of V.

  1. The vector addition that maps any two vectors $\mathbf{x}, \mathbf{y} \in V$ to another vector $\mathbf{x} + \mathbf{y} \in V$ satisfiying the following properties:
     - Commutativity: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
     - Associativity: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.
     - Existence of zero: there is a vector $\mathbf{0} \in V$ such that: $\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$.
     - Existence of inverse: for any vector $\mathbf{x} \in V$, there is another vector $-\mathbf{x} \in V$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
  2. The scalar multiplication that maps a vector $\mathbf{x} \in \mathbf{V}$ and scalar $a \in F$ (F can be a real or complex fiels) to another vector $a\mathbf{x} \in V$ with the following properties:
     - $a(\mathbf{x} + \mathbf{y} = a\mathbf{x} + a\mathbf{y}$.
     - $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.
     - $ab\mathbf{x} = a(b\mathbf{x})$.
     - $\mathbf{1x} = \mathbf{x}$.

where $\mathbf{0} = [0, \dots, 0]^T$ and $\mathbf{1} = [1, \dots, 1]^T$ are two constant vectors.

A subset W of V is a subspace of V, denoted by $W \subseteq V$, if it is also a vector space, i.e., it is closed under the same operations defined in V:

1. The zero vector $\mathbf{0}$ of V must be in W (the zero vector is unique in V, which W must have).
2. For any $\mathbf{x}, \mathbf{y} \in W, \mathbf{x} + \mathbf{y} \in W$.
3. For any $\mathbf{x} \in W, a\mathbf{x} \in W$.

Listed below is a set of typical vector spaces:

- n-D vector space $\mathbb{R}^n$ or $\mathbb{C}^n$

  This space contains all n-D vectors expressed as an n-tuple, an ordered list of n elements (or components):

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T, \tag{1.1}$$

  which can be used to represent a discrete signal containing n samples. We will always represent a vector as a vertical or column vector, or the transpose of a horizontal or row vector. The space is denoted by either $\mathbb{C}^n$ if the elements are complex $x_i \in \mathbb{C}$, or $\mathbb{R}^n$ if they are real $x_i \in \mathbb{R}(i = 1, \dots, n)$.

  A subspace of $\mathbb{R}^n$ can be a $\mathbb{R}^m (m < n)$ that passes origin (zero). For example, any 2-D plane passing through the origin of a 3-D space is its subspace. However, if the 2-D plane does not pass through the origin, it is not a subspace. Also, as 3-D cube or sphere centered at the origin is not a subspace, as it is not closed under the operations of addition and scalar multiplication.

- A vector space can be defined to contain all $m \times n$ matrices composed of n m-D column vectors:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \tag{1.2}$$

where the i-th column is an m-D vector $\mathbf{a}^i = \begin{bmatrix} a_{1i} & \dots & a_{mi} \end{bmatrix}^T$. Such a matrix can be converted to an mn-D vector by cascading all of the column (or row) vectors.

- l^2 space:

  The concept of an n-D space $\mathbb{R}^n$ or $\mathbb{C}^n$ can be generalised by allowing the dimension n to become to infinity so that a vector in the space becomes a sequence $\mathbf{x} = \begin{bmatrix} \dots & x_i & \dots \end{bmatrix}^T$ for $0 \le i < \infty$ or $-\infty < i < \infty$. If all vectors are square summable, the space is denoted by l^2. All discrete energy signals are vectors in l^2.

- $\mathcal{L}^2$ space:

  A vector space can also be a set of real or complex valued continuous functions $x(t)$ defined over either a finite range such as $0 \le t < T$, or an infinite range $-\infty < t < \infty$. If all functions are square integrable, the space is denoted $\mathcal{L}^2$. All continuous energy signals are vectors in $\mathcal{L}^2$.

Note that the term "vector", generally denoted by $\mathbf{x}$ in the following, may be interpreted in two different ways. First, in the most general sense, it represents a member of a vector space, such as any of the vector spaces considered above, e.g. a function $\mathbf{x} = x(t) \in \mathcal{L}^2$ Second, in a more narrow sense, it can also represent a tuple of n elements, an n-D vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T \in \mathbb{C}^n$, where n may be infinity. It should be clear what a vector $\mathbf{x}$ from the context.

- **Linear independence**

  A set of vectors $\begin{Bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{Bmatrix}$ are linearly independent if none of them can be represented as a linear combination of the others.

  Theorem: For any set of linarly independent vectors $\begin{Bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{Bmatrix}$, if their linear combination is zero

  $$\sum_{j=1}^{n} c_j \mathbf{v}_j = 0 \tag{1.3}$$

  then all their coefficients must be zero $c_1 = \ldots = c_n = 0$, or $\mathbf{c} = \mathbf{0}$.

  Proof: Assume $c_k \neq 0$, then we would get

$$v_k = -\frac{1}{c_k} \sum_{j=1, i \neq k}^{n} c_j \mathbf{v}_j. \tag{1.4}$$

i.e., v_k is a linear combination of the remaining $n-1$ vectors, in contradiction with the assumption that they are independent.

In particular, in tj n-D Euclidean space, the n vectors can be written as $\mathbf{v}_j = \begin{bmatrix} v_{1j} & \dots & v_{nj} \end{bmatrix}^T$ $(j = 1, \dots, n)$, and their linear combination

$$\sum_{j=1}^{n} c_j \mathbf{v}_j = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \dots & v_{mn} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{V}\mathbf{c}. \tag{1.5}$$

For this to be a zero vector, i.e. for this homogeneous equation $\mathbf{Vc} = \mathbf{0}$ to hold, the coefficient vector $\mathbf{c}$ has to be zero, as $\mathbf{V}$ is a full rank matrix.

- **Convex combination**

  The *convex combination* of n vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ is their sum weighted by coefficients $\{c_1, \dots, c_n\}$ that add up to 1:

$$\textstyle\sum_{i=1}^{n} c_i \mathbf{x}_i, \quad \sum_{i=1}^{n} c_i = 1 \ . \tag{1.6}$$

  The *convex hull* of these points is the set of all their combinations. For example, the convex hull of three points in a plane is the triangle formed by these points as vertices, in which any point is a convex combination of the three vertices.

- **Inner product**

  An *inner product* in a vector space V is a function that maps two vectors $\mathbf{x}, \mathbf{y} \in V$ to a scalar $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{C}$ or $\mathbb{R}$ and satisfies the following conditions:

  - Positive definiteness:

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0, \quad \langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0} \ . \tag{1.7}$$

- Conjugate symmetry:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}. \tag{1.8}$$

If the vector space is real, the inner product becomes *symmetric*:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle. \tag{1.9}$$

- Linearity in the first variable:

$$\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle. \tag{1.10}$$

where $a, b \in \mathbb{C}$. The linearity does not apply to the second variable unless the coefficients are real $a, b \in \mathbb{R}$:

$$\begin{aligned} &\langle \mathbf{x}, a\mathbf{y} + b\mathbf{z} \rangle = \overline{\langle a\mathbf{y} + b\mathbf{z}, \mathbf{x} \rangle} = \overline{a\langle \mathbf{y}, \mathbf{x} \rangle + b\langle \mathbf{z}, \mathbf{x} \rangle} = \overline{a}\langle \mathbf{x}\mathbf{y} \rangle + \overline{b}\langle \mathbf{x}\mathbf{z} \rangle \\ &\neq a\langle \mathbf{x}, \mathbf{y} \rangle + b\langle \mathbf{x}, \mathbf{z} \rangle. \end{aligned} \tag{1.11}$$

In the special case when $b = 0$, we have

$$\langle a\mathbf{x}, \mathbf{y} \rangle = a\langle \mathbf{x}, \mathbf{y} \rangle, \quad \langle \mathbf{x}, a\mathbf{y} \rangle = \overline{a}\langle \mathbf{x}, \mathbf{y} \rangle \ . \tag{1.12}$$

More generally we have

$$\langle \textstyle\sum_n c_n \mathbf{x}_n, \mathbf{y} \rangle = \sum_n c_n \langle \mathbf{x}_n, \mathbf{y} \rangle, \quad \langle \mathbf{x}, \sum_n c_n \mathbf{y}_n \rangle = \sum_n \overline{c}_n \langle \mathbf{x}, \mathbf{y}_n \rangle \ . \tag{1.13}$$

An *inner product* space is a vector space with inner product defined. In particular, when the inner product is defined, $\mathbb{C}^n$ is called a *unitary space* and $\mathbb{R}^n$ is called an *Euclidean space*.

Some examples of the inner product are listed below:

- In a n-D vector space, the inner product, also called the *dot product*, of two vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \overline{\mathbf{y}} = \mathbf{y}^* \mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \overline{y}_1 \\ \vdots \\ \overline{y}_n \end{bmatrix} = \sum_{i=1}^{b} x_i \overline{y}_i, \tag{1.14}$$

where $\mathbf{y}^* = \overline{y}^T$ is the conjugate transpose of $\mathbf{y}$.

- In a space of 2-D matrices containing $n \times m$ elements, the inner product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \bar{b}_{ij}. \tag{1.15}$$

When the column (or row) vectors of $\mathbf{A}$ and $\mathbf{B}$ are concatenated to form two mn-D vectors, their inner product takes the same form as that of two n-D vectors.

- In a function space, the inner product of two function vectors $\mathbf{x} = x(t)$ and $\mathbf{y} = y(t)$ us defined as

$$\langle x(t), y(t) \rangle = \int_a^b x(t)\overline{y(t)}dt = \overline{\int_a^b y(t)\overline{x(t)}dt} = \overline{\langle y(t), x(t) \rangle}. \tag{1.16}$$

- The covariance of two random variables x and y can be considered as an inner product

$$\langle x, y, \rangle = \sigma_{xy}^2 = E\left|(x - \mu_x)\overline{(y - \mu_y)}\right| = E(x\overline{y}) - \mu_x\overline{\mu}_y. \tag{1.17}$$

Specially when $\mu_x = \mu_y = 0$, we have

$$\langle x, y \rangle = E[x\overline{y}]. \tag{1.18}$$

- **Vector norm**

In general, the norm $||\mathbf{x}||$ of a vector $\mathbf{x} \in V$ is a non-negative real scalar that measures its size of length. $||\mathbf{x}|| = 0$ if and only $\mathbf{x} = 0$. There exist different deffinitions for vector norm, as shown later. A vector $\mathbf{x}$ is *normalized* (becomes a *unit* vector) if $||\mathbf{x}|| = 1$. Any given vector can be normalised when divided by its own norm $\mathbf{x}/||\mathbf{x}||$.

The most widely used *2-norm* of a vector is defined as

$$||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}. \tag{1.19}$$

The vector norm squared $\|\mathbf{x}\|^2$ can be considered as the energy of the vector. In particular, un an n-D unitary space, the 2-norm of a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T \in \mathbb{C}^n$ is:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\mathbf{x}^T \overline{\mathbf{x}}} = \sqrt{\mathbf{x}^* \mathbf{x}} = \left(\sum_{i=1}^{n} x_i \overline{x}_i \right)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} . \tag{1.20}$$

The total energy contained in this vector is its own norm squared:

$$\mathcal{E} = \|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^{n} |x_i|^2 . \tag{1.21}$$

Similarly, in a function space, the norm of a function vector $\mathbf{x} = x(t)$ is defined as:

$$\|\mathbf{x}\| = \left(\int_a^b x(t)\overline{x(t)}dt \right)^{1/2} = \left(\int_a^b |x(t)|^2 dt \right)^{1/2} , \tag{1.22}$$

where the lower and upper integral limits $a < b$ are two real numbers, which may be extended to all real values $\mathbb{R}$ in the entire real axis $-\infty < t < \infty$. This norm exists only if the integral converges to a finite value, i.e. $x(t)$ is an *energy signal* containing finite energy.

$$\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty. \tag{1.23}$$

All such functions $x(t)$ satisfiying the above are square-integrable, and they form a function space denoted by $\mathcal{L}^2(\mathbb{R})$.

- **Cauchy-Schwarz inequality**

  The Cauchy-Schwarz inequality holds for any two vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner product space V:

$$|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \le \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle; \quad \text{i.e.,} \quad 0 \le |\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\| \|\mathbf{y}\| \ . \tag{1.24}$$

  Proof:

If either $\mathbf{x}$ or $\mathbf{y}$ is zero, $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, the theorem holds (an equality). Otherwise, we consider the following inner product:

$$\langle \mathbf{x} - \lambda \mathbf{y}, \mathbf{x} - \lambda \mathbf{y} \rangle = ||\mathbf{x}||^2 - \overline{\lambda} \langle \mathbf{x}, \mathbf{y} \rangle - \lambda \langle \mathbf{y}, \mathbf{x} \rangle + |\lambda|^2 ||\mathbf{y}||^2 \geq 0, \tag{1.25}$$

where $\lambda \in \mathbb{C}$ is an arbitrary complex number, which can be assumed to be:

$$\lambda = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{y}||^2}, \quad \text{then } \overline{\lambda} = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{||\mathbf{y}||^2}, \quad |\lambda|^2 = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{||\mathbf{y}||^4}. \tag{1.26}$$

Substituting these into the previous equation we get:

$$||\mathbf{x}||^2 - \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{||\mathbf{y}||^2} \geq 0; \quad \text{i.e.,} \quad |\langle \mathbf{x}, \mathbf{y} \rangle| \leq ||\mathbf{x}|| ||\mathbf{y}||. \tag{1.27}$$

The equation holds only if $\mathbf{x} - \lambda \mathbf{y} = 0$ or $\mathbf{x} = \lambda \mathbf{y}$, i.e., the two vectors are linearly dependent.

- **Distance between two vectors**

  The distance $d(\mathbf{x}, \mathbf{y}$ between two vectors is a real constant that satisfies:

  - $d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$
  - $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
  - $d(\mathbf{x}, \mathbf{y} \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$

  In and inner product space in which the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ between any two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined, the distance between the two points can be defined as the norm of the difference between the two points:

$$d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| = \sqrt{\langle (\mathbf{x} - \mathbf{y}), (\mathbf{x} - \mathbf{y})}. \tag{1.28}$$

  The norm of a vector can now be seen as its dustance to the origin $\mathbf{0}$ of the space $||\mathbf{x}|| = ||\mathbf{x} - \mathbf{0}||$.

  A vector space V is a *metric space* if a distance (or *metric*) $d(\mathbf{x}, \mathbf{y})$ between any two vectors (or points) $\mathbf{x}$ and $\mathbf{y}$ is defined.

- **Angle between two vectors**

  The *angle* between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as:

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}\right). \tag{1.29}$$

  Now the inner product of $\mathbf{x}$ and $\mathbf{y}$ can also be written as:

$$\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \leq \|\mathbf{x}\|\|\mathbf{y}\|. \tag{1.30}$$

  This is the Cauchy-Schwarz inequality. In particular:

  - If $\theta = 0, \cos\theta = 1$, then $\mathbf{x}$ and $\mathbf{y}$ are collinear or linearly dependent, and the inner product is maximised:

$$\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}\|\|\mathbf{y}\|. \tag{1.31}$$

    i.e., the Cauchy-Schwarz inequality becomes and equality.
  - If $0 < \theta < \pi/2$, $0 < \cos\theta < 1$, we get the Cauchy-Schwarz inequality:

$$\langle \mathbf{x}, \mathbf{y}\rangle < \|\mathbf{x}\|\|\mathbf{y}\|. \tag{1.32}$$

  - If $\theta = \pi/2$, $\cos\theta = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to each other, and the inner product is minimized:

$$\langle \mathbf{x}, \mathbf{y}\rangle = 0. \tag{1.33}$$

  Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* or *perpendicular* to each other, denoted by $\mathbf{x} \perp \mathbf{y}$, if their inner product is zero $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, i.e., the angle between them is $\theta = \cos^{-1} 0 = \pi/2$.

- **Projection**

  The *orthogonal projection* of a vector $\mathbf{x} \in V$ onto another vector $\mathbf{y} \in V$ is defined as a vector:

$$\mathbf{p}_y(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{y}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\langle \mathbf{y}, \mathbf{y}\rangle}\mathbf{y} = \|\mathbf{x}\|\cos\theta\frac{\mathbf{y}}{\|\mathbf{y}\|}, \tag{1.34}$$

where $\mathbf{y}/||\mathbf{y}||$ is the unit vector along the direction of $\mathbf{y}$. In particular, if $\mathbf{y}$ is normalized with $||\mathbf{y}|| = 1$, then:

$$\mathbf{p}_y(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{y} = ||\mathbf{x}|| \cos\theta \mathbf{y}. \tag{1.35}$$

Note that:

$$||\mathbf{p}_y(\mathbf{x})|| = ||\mathbf{x}|| \cos\theta. \tag{1.36}$$

If only the magnitude of the projection is of interest, the unit vector $\mathbf{y}/||\mathbf{y}||$ can be dropped:

$$p_y(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle / ||\mathbf{y}||. \tag{1.37}$$

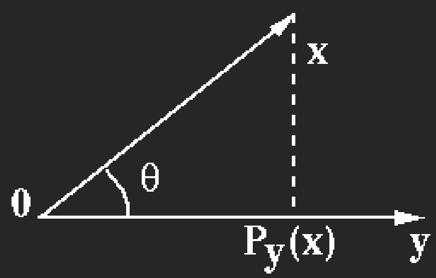

Figure 1.1: Projection of $\mathbf{x}$ onto $\mathbf{y}$.

- **Cauchy space**

  A sequence of points in a metric space $x_0, x_1, x_2, \ldots$ is a *Cauchy sequence* if it converges, i.e., for any $\epsilon > 0$, there exists an integer $N > 0$ so that the following is true for any $m, n > N$:

$$d(x_m, x_n) < \epsilon. \tag{1.38}$$

  If the limit of any Cauchy sequence of points in the space is also in the space, the space is called *complete*, referred to as a *Cauchy space*.

Chapter 2

Orthogonal Basis and Gram-Schmidt Process

From `http://fourier.eng.hmc.edu/e176/lectures/algebra/node2.html`

Assume n-D vector space $\mathbb{R}^n$ or $\mathbb{C}^n$ is spanned by a set of n independent basis vectors $\begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix}$, not necessarily orthogonal, so that any vector $\mathbf{x} \in \mathbb{R}^n$ can be represented as a linear combination of the basis vectors:

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{v}_i = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{V}\mathbf{c}, \tag{2.1}$$

where $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix}$ is an $m \times n$ matrix composed of the n vectors and the coefficients in $\mathbf{c} = \begin{bmatrix}c_1 & \dots & c_n\end{bmatrix}^T$ can be found by solving the lienar equation system to get $\mathbf{c} = \mathbf{V}^{-1}\mathbf{x}$ with complexity $O(n^3)$.

These linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ can be converted into a set of orthogonal vectors $\begin{Bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_n\end{Bmatrix}$ satisfying $\mathbf{u}_i^T \mathbf{u}_j = 0 (i \neq j)$ by the following *Gram-Schmidt process*:

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 \\
\mathbf{u}_2 &= \mathbf{v}_2 - \mathbf{p}_{\mathbf{u}_1}(\mathbf{v}_2) = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\right)\mathbf{u}_1 \\
\mathbf{u}_3 &= \mathbf{v}_3 - \mathbf{p}_{\mathbf{u}_1}(\mathbf{v}_3) - \mathbf{p}_{\mathbf{u}_2}(\mathbf{v}_3) = \mathbf{v}_3 - \left(\frac{\mathbf{v}_3^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\right)\mathbf{u}_1 - \left(\frac{\mathbf{v}_3^T\mathbf{u}_2}{\mathbf{u}_2^T\mathbf{u}_2}\right)\mathbf{u}_2 \\
&\cdots\cdots\cdots \\
\mathbf{u}_k &= \mathbf{v}_k - \sum_{i=1}^{k-1}\mathbf{p}_{\mathbf{u}_i}(\mathbf{v}_k) = \mathbf{v}_k - \sum_{i=1}^{k-1}\left(\frac{\mathbf{v}_k^T\mathbf{u}_i}{\mathbf{u}_i^T\mathbf{u}_i}\right)\mathbf{u}_i.
\end{aligned} \tag{2.2}$$

where $\mathbf{p}_{\mathbf{u}_i}(\mathbf{v}_k)$ is the projection of $\mathbf{v}_k$ onto $\mathbf{u}_i$:

$$\mathbf{p}_{\mathbf{u}_i}(\mathbf{v}_k) = \left(\frac{\mathbf{v}_k^T\mathbf{u}_i}{\mathbf{u}_i^T\mathbf{u}_i}\right)\mathbf{u}_i. \tag{2.3}$$

We see that $\mathbf{u}_2$ so obtained is indeed orthogonal to all $\mathbf{u}_1$:

$$\mathbf{u}_2^T\mathbf{u}_1 = \left[\mathbf{v}_2 - \left(\frac{\mathbf{v}_2^T\mathbf{u}_1}{\mathbf{u}_i^T\mathbf{u}_1}\right)\mathbf{u}_i\right]^T\mathbf{u}_1 = \mathbf{v}_2^T\mathbf{u}_1 - \frac{\mathbf{v}_2^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\mathbf{u}_1^T\mathbf{u}_1 = 0, \tag{2.4}$$

and, by mathematical induction, we can further prove that $\mathbf{u}_k^T\mathbf{u}_i = 0$ for all $i = 1, \ldots, k-1$. In other words, $\mathbf{u}_k$ is the component of $\mathbf{v}_k$ that is orthogonal to all previously found orthogonal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$.

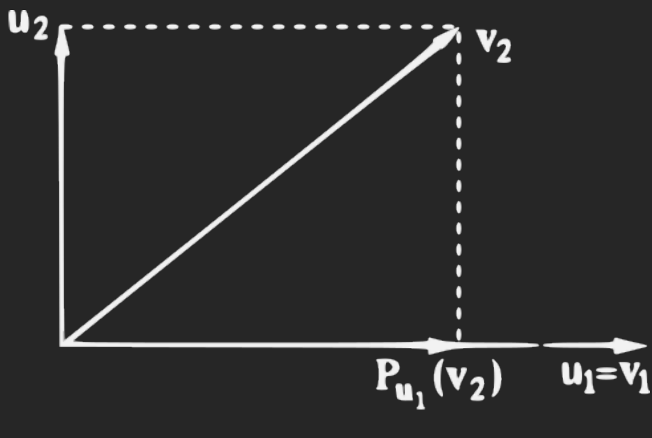

Figure 2.1: $\mathbf{u}_2 = \mathbf{v}_2 - \mathbf{p}_{\mathbf{u}_1}(\mathbf{v}_2)$

Given a set of orthogonal basis $\{\mathbf{u}_1 \quad \ldots \quad \mathbf{u}_n\}$ satisfying $\langle\mathbf{u}_i, \mathbf{u}_j\rangle = \mathbf{u}_i^T\mathbf{u}_j = 0$ for any $\mathbf{i} \neq \mathbf{j}$, we can again represent any given vector $\mathbf{x}$ as a linear combination of these bases as:

$$\mathbf{x} = \sum_{i=1}^{n} d_i \mathbf{u}_i. \tag{2.5}$$

The coefficients d_i can now be obtained easily. Premultiplying $\mathbf{u}_j^T$ on both sides of the equation above, we get:

$$\mathbf{u}_j^T \mathbf{x} = \mathbf{u}_j^T \left(\sum_{i=1}^{n} d_i \mathbf{u}_i \right) = \sum_{i=1}^{n} d_i \mathbf{u}_j^T \mathbf{u}_i = d_j \mathbf{u}_j^T \mathbf{u}_j. \tag{2.6}$$

Note: $\sum_{i=1}^{n} d_i \mathbf{u}_j^T \mathbf{u}_i = d_j \mathbf{u}_j^T \mathbf{u}_j$ because $\mathbf{u}_i^T \mathbf{u}_j \neq 0 \iff i = j$ *(Tomek)*

Solving for d_j we get:

$$d_j = \frac{\mathbf{u}_j^T \mathbf{x}}{\mathbf{u}_j^T \mathbf{u}_j}, \qquad j = 1, \dots, n. \tag{2.7}$$

As the n coefficients are decoupled, each of them can be obtained separately with linear computational complexity $O(n)$ with total complexity $I(n^2$ for all n of them. The complexity is reduced to $O(n^2)$ from $O(n^3)$ for solving the equation system $mathbf{c} = mathbf{V}^{-1} mathbf{b}$, needed in the case where the basis is not orthogonal. This is the reason why orthogonal bases are preferred in general. Now the vector $mathbf{x}$ can be written as:

$$\mathbf{x} = \sum_{i=1}^{n} d_i \mathbf{u}_i = \sum_{i=1}^{n} \left(\frac{\mathbf{u}_j^T \mathbf{x}}{\mathbf{u}_j^T \mathbf{u}_j} \right) \mathbf{u}_i = \sum_{i=1}^{n} \mathbf{p}_{\mathbf{u}_i}(\mathbf{x}), \tag{2.8}$$

where the ith term of the summation above is simply the projection $\mathbf{p}_{\mathbf{u}_i}(\mathbf{x})$ of $\mathbf{x}$ onto the ith basis of $\mathbf{u}_i$.
The concept of orthogonal vectors satisfying $\mathbf{u}^T \mathbf{v} = 0$ can be generalised to *conjugate vectors* that satisfy $\mathbf{u}^T \mathbf{A} \mathbf{v} = 0$ with respect to a symmetric matrix $\mathbf{A} = \mathbf{A}^T$. This can be expressed in the form of inner product:

$$\mathbf{u}^T \mathbf{A} \mathbf{v} = (\mathbf{A}\mathbf{v})^T \mathbf{u} = \mathbf{v}^T \mathbf{A} \mathbf{u}. \tag{2.9}$$

Conjugate vectors with respect to $\mathbf{A}$ can also be considered as orthogonal to each other with respect to $\mathbf{A}$. Two orthogonal vectors satisfying $\mathbf{u}^T \mathbf{v} = 0$ can be considered as a special case of conjugate vectors with respect to $\mathbf{A} = \mathbf{I}$.

Based on this generalized orthogonality, we can also define the projection of $\mathbf{v}$ onto $\mathbf{u}$ with respect to $\mathbf{A}$ as:

$$\mathbf{p_u}(\mathbf{v}) = \left(\frac{\mathbf{u}^T\mathbf{A}\mathbf{v}}{\mathbf{u}^T\mathbf{A}\mathbf{u}}\right)\mathbf{u}. \tag{2.10}$$

Given a set of A-conjugate basis vectors $\{d_1 \quad \dots \quad d_n\}$ satisfying $\mathbf{d}_i^T\mathbf{A}\mathbf{d}_j = 0$ for any $\mathbf{i} \neq \mathbf{j}$, we can represent any $\mathbf{x}$ as:

$$\mathbf{x} = \sum_{i=1}^{n} a_i\mathbf{d}_i. \tag{2.11}$$

The coefficients a_i can be obtained by pre-multiplying $\mathbf{d}_j^T\mathbf{A}$ on both sides:

$$\mathbf{d}_j^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} a_i\mathbf{d}_j^T\mathbf{A}\mathbf{d}_i = a_j\mathbf{d}_j^T\mathbf{A}\mathbf{d}_j. \tag{2.12}$$

Solving for a_i we get:

$$a_j = \frac{\mathbf{d}_j^T\mathbf{A}\mathbf{x}}{\mathbf{d}_j^T\mathbf{A}\mathbf{d}_j}, \qquad j = 1, \dots, n. \tag{2.13}$$

and $\mathbf{x}$ is now represented as the sum of its A-projections onto the n basis vectors $\{d_1 \quad \dots \quad d_n\}$:

$$\mathbf{x} = \sum_{i=1}^{n} a_i\mathbf{d}_i = \sum_{i=1}^{n}\left(\frac{\mathbf{s}_i^T\mathbf{A}\mathbf{x}}{\mathbf{d}_i^T\mathbf{A}\mathbf{d}_i}\right)\mathbf{d}_i = \sum_{i=1}^{n}\mathbf{p}_{\mathbf{d}_i}(\mathbf{x}). \tag{2.14}$$

A set of independent basis vectors $\{v_1 \quad \dots \quad v_n\}$ can now be also converted to an orthogonal basis with respect to a symmetric vector $\mathbf{A}$ by a Gram-Schmidt process:

$$\mathbf{d}_k = \mathbf{v}_k - \sum_{i=1}^{k-1}\mathbf{p}_{\mathbf{d}_i}(\mathbf{v}_k) = \mathbf{v}_k - \sum_{i=1}^{k-1}\left(\frac{\mathbf{v}_k^T\mathbf{A}\mathbf{d}_i}{\mathbf{d}_i^T\mathbf{A}\mathbf{d}_i}\right)\mathbf{d}_i, \qquad (k = 1, \dots, n). \tag{2.15}$$

Chapter 3

Properties of Matrices

Rank, Trace, Determinant, Transpose and Inverse of Matrices
From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node3.html`
Let $\mathbf{A}$ be an $m \times n$ square matrix:

$$\mathbf{A} = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}_{m \times n} \tag{3.1}$$

where:

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}, \qquad (j = 1, \dots, n) \tag{3.2}$$

is the j^{th} column vector and

$$\begin{bmatrix} a_{i1} & \dots & a_{in} \end{bmatrix}, \qquad (i = 1, \dots, m) \tag{3.3}$$

is the i^{th} row vector. If $m = n$, $\mathbf{A}$ is a *square matrix*. In particular, if all entries of a square matrix are zero except those along the diagonal, it is a *diagonal matrix*. Morover, if the diagonal entries of a diagonal matrix are all one, it is the *identity matrix*:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}. \tag{3.4}$$

- **Rank**

  The m row vectors span the *row space* of $\mathbf{A}$ and the n columns span the *column space* of $\mathbf{A}$. The *rank* of each space is its dimension, the number of independent vectors in the space. The row and column spaces have the same rank, which is also the rank of matrix $\mathbf{A}$, i.e.:

  $$r = rank(\mathbf{A}) \leq \min(m, n). \tag{3.5}$$

  In other words, the rank of matrix $\mathbf{A}$ is the number of its independent rows or columns.

- **Transpose**

  The *transpose* $\mathbf{x}^T$ of a column vector $\mathbf{x}$ is a row vector:

  $$\mathbf{x}^T = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^T = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}. \tag{3.6}$$

  The transpose $\mathbf{A}^T$ of a matrix $\mathbf{A}$ is obtained by switching the position of elements a_{ij} and a_{ji}, for all $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$. In other words, the i^{th} column of $\mathbf{A}$ becomes the i^{th} row of $\mathbf{A}^T$.

  The properties of the transpose:

  - $(\mathbf{A}^T)^T = \mathbf{A}$,
  - $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$,
  - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^{\mathbf{T}}$,
  - $\mathbf{A}^T$ and $\mathbf{A}$ have the same eigenvalues and eigenvectors.

  If $\mathbf{A} = \mathbf{A}^T$, it is a *symmetric matrix*.

- **Conjugate transpose**

  The *conjugate transpose* of a matrix $\mathbf{A}$, denoted by $\mathbf{A}^*$, is by taking the complex conjugate of its transpose:

  $$\mathbf{A}^* = (\overline{\mathbf{A}})^T = \overline{\mathbf{A}^T}, \tag{3.7}$$

  i.e., $(\mathbf{A}^*)_{ij} = \overline{\mathbf{A}}_{ji}$, the ij^{th} entry of $\mathbf{A}^*$ is the complex conjugate of the ji^{th} entry of $\mathbf{A}$.

  The properties of the conjugate transpose:

  - $(\mathbf{A}^*)^* = \mathbf{A}$
  - $(c\mathbf{A})^* = \bar{c}\mathbf{A}^*$
  - $(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$
  - $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$
  - $\det(\mathbf{A}^*) = (\det \mathbf{A})^* = \overline{\det \mathbf{A}}$
  - $tr(\mathbf{A}^*) = (tr\mathbf{A})^* = \overline{tr\mathbf{A}}$
  - $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$
  - The eigenvalues of $\mathbf{A}^*$ are the complex conjugates of the eigenvalues of $\mathbf{A}$. (But their eigenvectors are not related.)
  - $\langle \mathbf{Ax}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}^*\mathbf{y} \rangle$

  Proof:

  $$\langle \mathbf{Ax}, \mathbf{y} \rangle = (\mathbf{Ax})^*\mathbf{y} = \mathbf{x}^*\mathbf{A}^*\mathbf{y} = \mathbf{x}^*(\mathbf{A}^*\mathbf{y}) = \langle \mathbf{x}, \mathbf{A}^*\mathbf{y} \rangle. \tag{3.8}$$

  If $\mathbf{A} = \mathbf{A}^*$, it is a *Hermitian matrix.* A real Hermitian matrix is a symmetric matrix. The inverse of an invertible Hermitian matrix is also Hermitian, i.e., if $\mathbf{A} = \mathbf{A}^*$, then $(\mathbf{A}^{-1})^* = \mathbf{A}^{-1}$.

- **Trace**

  The *trace* of a square matrix A is the sum of its diagonal elements:

  $$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{3.9}$$

  The properties of the trace:

- $tr(c\mathbf{A}) = c\,tr(\mathbf{A})$
- $tr(\mathbf{A}^T) = tr(\mathbf{A})$
- $tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{B}+\mathbf{A})$
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$
- $tr(\mathbf{A}^T\mathbf{B}) = tr(\mathbf{AB}^T)$
- $tr(\mathbf{R}^{-1}\mathbf{AR}) = tr(\mathbf{R}^{-1}(\mathbf{AR})) = tr((\mathbf{AR})\mathbf{R}^{-1}) = tr(\mathbf{A})$

- **Determinant**

  The *determinant* of a square matrix $\mathbf{A}$ is denoted by $\det(\mathbf{A}) = |\mathbf{A}|$, and $\det(\mathbf{A}) \neq 0$ if and only if it is full rank, i.e., $rank(\mathbf{A}) = n$.

  The properties of the determinant:

  - $\det(\mathbf{I}) = 1$,
  - $\det(\mathbf{A}^T) = \det(\mathbf{A})$,
  - $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$,
  - $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$,
  - $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$.

- **Inverse**

  If $\mathbf{A}$ is an $m\times n$ full rank square matrix with $m = n = r$, then there exists an *inverse matrix* $\mathbf{A}^{-1}$ that satisfies $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse $\mathbf{A}^{-1}$ does not exist if $\mathbf{A}$ is not square ($m \neq n$) or full rank ($r < m = n$).

  The properties of the inverse:

  - $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
  - $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$,
  - $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$,
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

  For example, the solution of a linear equation system $\mathbf{Ax} = \mathbf{b}$ of m equations and $n = m$ variables can be obtained as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Chapter 4

Normal, Unitary and Simiar Matrices

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node4.html`

- **Commutative matrices**

  Two square matrices $\mathbf{A}$ and $\mathbf{B}$ *commute* if $\mathbf{AB} = \mathbf{BA}$.

  Obviously all diagonal matrices commute.

- **Normal matrix**

  A square matrix $\mathbf{A}$ is *normal* if it commutes with its conjugate transpose: $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$. If $\mathbf{A}^* = \mathbf{A}^T$ is real, then $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$.

  Obviously unitary matrices ($\mathbf{A}^* = \mathbf{A}^{-\mathbf{1}}$), Hermitian matrices ($\mathbf{A}^* = \mathbf{A}$), and skew-Hermitian matrices ($\mathbf{A}^* = -\mathbf{A}$) are all normal. But there exit normal matrices that are not Hermitian.

- **Unitary matrix**

  $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n \end{bmatrix}$ is a *unitary matrix* if its conjugate transpose is equal to its inverse $\mathbf{U}^* = \mathbf{U}^{-1}$, i.e., $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. When a unitary matrix $\overline{\mathbf{U}} = \mathbf{U}$ is real, it becomes an *orthogonal matrix*, $\mathbf{U}^T = \mathbf{U}^{-1}$.

  The column (or row) vectors of a unitary matrix $\mathbf{A}$ are *orthonormal*, i.e., they are both orthogonal and normalized:

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbf{u}_i^T \mathbf{u}_j = \mathbf{u}^* \mathbf{u}_i = \sum_k u_{ik} \overline{u}_{jk} = \delta_{ij} \stackrel{\triangle}{=} \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} . \tag{4.1}$$

Let λ and $\mathbf{v}$ be an eigenvalue and the corresponding eigenvector of $\mathbf{U}$, i.e., $\mathbf{U}\mathbf{v} = \lambda \mathbf{v}$, then we have:

$$\mathbf{v}^* \mathbf{v} = \mathbf{v}^* \mathbf{U}^* \mathbf{U} \mathbf{v} = (\mathbf{U}\mathbf{v})^* \mathbf{U}\mathbf{v} = \lambda^2 \mathbf{v}^* \mathbf{v}. \tag{4.2}$$

We see that $\lambda^{2=1}$, i.e., $|\lambda| = 1$.

A Hermitian matrix $\mathbf{A}$ can be converted to a diagonal matrix $\mathbf{\Lambda}$ (or diagonalized) by a particulat unitary matrix $\mathbf{U}$:

$$\mathbf{U}^* \mathbf{A} \mathbf{U} = \mathbf{\Lambda} = diag \begin{bmatrix} \lambda_1 & \dots & \lambda_n \end{bmatrix}, \tag{4.3}$$

where $\mathbf{\Lambda}$ is a diagonal matrix, i.e., all its off diagonal elements are 0.

- **Similar matrices**

  Two matrices $\mathbf{A}$ and $\mathbf{B}$ are *similar matrices* if they are related by $\mathbf{A} = \mathbf{P}^{-1} \mathbf{B} \mathbf{P}$. Similar matrices have the same determinant, trace and eigenvalues.

Chapter 5

Positive/Negative (Semi)-Definite Matrices

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node5.html`
Given a Hermitian matrix $\mathbf{A} = \mathbf{A}^*$ and any non-zero vector $\mathbf{x} \neq 0$, we can construct a quadratic form $\mathbf{x}^*\mathbf{A}\mathbf{x}$. The matrix $\mathbf{A}$ is said to be:

- *positive definite* $\mathbf{A} > 0$, if $\mathbf{x}^*\mathbf{A}\mathbf{x} > 0$,
- *positive semi-definite* $\mathbf{A} \geq 0$, if $\mathbf{x}^*\mathbf{A}\mathbf{x} \geq 0$,
- *negative definite* $\mathbf{A} < 0$, if $\mathbf{x}^*\mathbf{A}\mathbf{x} < 0$,
- *negative semi-definite* $\mathbf{A} \leq 0$, if $\mathbf{x}^*\mathbf{A}\mathbf{x} \leq 0$,
- *indefinite* if there exists $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0 < \mathbf{y}^T\mathbf{A}\mathbf{y}$.

For example, consider the covariance matrix of a random vector $\mathbf{x}$:

$$\mathbf{\Sigma}_x = E[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^*]. \tag{5.1}$$

The corresponding quadratic form is:

$$\begin{aligned} \mathbf{v}^*\mathbf{\Sigma}_x\mathbf{v} &= \mathbf{v}^* E[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^*]\mathbf{v} \\ &= E[\mathbf{v}^*(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^*\mathbf{v}] = E(s^2) \geq 0, \end{aligned} \tag{5.2}$$

where $s = \mathbf{v}^*(\mathbf{x} - \mathbf{m}_x)$ is a scalar. Therefore $\mathbf{\Sigma}_x$ is positive semi-definite.

As $\mathbf{A}$ is Hermitian, its eigenvalues $\lambda_1, \ldots, \lambda_n$ are real and its eigenvector matrix is unitary $\mathbf{V}^{-1} = \mathbf{V}^*$, by which it can be diagonalized:

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{V}^*\mathbf{A}\mathbf{V} = \mathbf{\Lambda}, \qquad \text{i.e.} \qquad \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*, \tag{5.3}$$

where vector $\mathbf{y} = \mathbf{V}^*\mathbf{x}$ is a unitary transform of vector $\mathbf{x}$:
For any $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$, we have:

$$\begin{aligned} \mathbf{x}^*\mathbf{A}\mathbf{x} &= \mathbf{x}^*(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^*)\mathbf{x} = (\mathbf{V}^*\mathbf{x})^*\mathbf{\Lambda}(\mathbf{V}^*\mathbf{x}) = \mathbf{y}^*\mathbf{\Lambda}\mathbf{y} = \textstyle\sum_{i=1}^n \lambda_i |y_i|^2 \\ &= \begin{cases} > 0 \text{ is positive definite} & \text{if } \lambda_i > 0 \\ \geq 0 \text{ is positive semi-definite} & \text{if } \lambda_i \geq 0 \\ < 0 \text{ is negative definite} & \text{if } \lambda_i < 0 \\ \leq 0 \text{ is negative semi-definite} & \text{if } \lambda_i \leq 0 \end{cases} \quad \text{(for all } i). \end{aligned} \tag{5.4}$$

Also, if some eigenvalues are positive and some others are negative, $\mathbf{x}^*\mathbf{A}\mathbf{x}$ may be either positive or negative depending on $\mathbf{x}$, i.e., $\mathbf{A}$ is indefinite.
As the eigenvalues of $\mathbf{A}^{-1}$ are $1/\lambda_i$, $i = (1, \ldots, n)$, $\mathbf{A}^{-1}$ and $\mathbf{A}$ share the same positive/negative definiteness.

Chapter 6

Woodbury and Sherman-Morrison

Woodbury Matrix Identity and Sherman-Morison Formula
From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node6.html`
The *Woodbury matrix identity* gives the inverse of $n \times n$ square matrix modified by a perturbation term $\mathbf{UBV}^T$:

$$(\mathbf{A} + \mathbf{UUV}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}. \tag{6.1}$$

The **proof** is straightforward:

$$\begin{array}{l} (\mathbf{A} + \mathbf{UUV}^T)^{-1}[\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}] \\ = [\mathbf{I} - \mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}] \\ +[\mathbf{UBV}^T\mathbf{A}^{-1} - \mathbf{UBV}^T\mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}] \\ = \mathbf{I} + \mathbf{UBV}^T\mathbf{A}^{-1} - [\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1} \\ +\mathbf{UBV}^T\mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}] \\ = \mathbf{I} + \mathbf{UBV}^T\mathbf{A}^{-1} - (\mathbf{U} + \mathbf{UBV}^T\mathbf{A}^{-1}\mathbf{U})(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1} \\ = \mathbf{I} + \mathbf{UBV}^T\mathbf{A}^{-1} - \mathbf{UB}(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})(\mathbf{B}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1} \\ = \mathbf{I} + \mathbf{UBV}^T\mathbf{A}^{-1} - \mathbf{UBV}^T\mathbf{A}^{-1} = \mathbf{I} \end{array} . \tag{6.2}$$

Consider some special cases:

- if $\mathbf{U} = \mathbf{V} = \mathbf{I}$, then we get:

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}. \tag{6.3}$$

- if $\mathbf{B} = \mathbf{I}$, and let $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_m\end{bmatrix}$ and $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_m\end{bmatrix}$, then we get the inverse of a rank-n modified matrix:

$$(\mathbf{A}+\mathbf{U}\mathbf{V}^T)^{-1} = \left(\mathbf{A} + \sum_{i=1}^{m} \mathbf{u}_i\mathbf{v}_i^T\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I}+\mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}. \tag{6.4}$$

- More specifically, when $m = 1$, $\mathbf{U} = \mathbf{u}$ and we get the inverse of rand-1 modified matrix:

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{6.5}$$

Proof I

We first show the following is an identity:

$$(\mathbf{I}+\mathbf{w}\mathbf{v}^T)^{-1} = \mathbf{I} - \frac{\mathbf{w}\mathbf{v}^T}{1+\mathbf{v}^T\mathbf{w}}. \tag{6.6}$$

Pre-multiplying $\mathbf{I}+\mathbf{w}\mathbf{v}^T$, the right side becomes $\mathbf{I}$ as wel as the left side:

$$\mathbf{I}+\mathbf{w}\mathbf{v}^T - \frac{\mathbf{w}\mathbf{v}^T+\mathbf{w}\mathbf{v}^T\mathbf{w}\mathbf{v}^T}{1+\mathbf{v}^T\mathbf{w}} = \mathbf{I}+\mathbf{w}\mathbf{v}^T - \frac{\mathbf{w}(1+\mathbf{v}^T\mathbf{w})\mathbf{v}^T}{1+\mathbf{v}^T\mathbf{w}} = \mathbf{I}+\mathbf{w}\mathbf{v}^T - \mathbf{w}\mathbf{v}^T = \mathbf{I}. \tag{6.7}$$

We next let $\mathbf{u} = \mathbf{A}^{-1}\mathbf{u}$, and the left side of the formula to be proven becomes:

$$\begin{aligned}(\mathbf{A}+\mathbf{u}\mathbf{v}^T)^{-1} &= (\mathbf{A}+\mathbf{A}\mathbf{w}\mathbf{v}^T)^{-1} = (\mathbf{A}(\mathbf{I}+\mathbf{w}\mathbf{v}^T))^{-1} = (\mathbf{I}+\mathbf{w}\mathbf{v}^T)^{-1}\mathbf{A}^{-1} \\ &= \left(\mathbf{I} - \tfrac{\mathbf{w}\mathbf{v}^T}{1+\mathbf{v}^T\mathbf{w}}\right)\mathbf{A}^{-1}\end{aligned}. \tag{6.8}$$

Substituting $\mathbf{w} = \mathbf{A}^{-1}\mathbf{u}$ we get the *Sherman-Morrison formula*:

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^T)^{-1} = \left(\mathbf{I} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\right)\mathbf{A}^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{6.9}$$

Proof II

This formula can also be proved based on the following problem. Assuming a linear equation system $\mathbf{A}\mathbf{y} = \mathbf{b}$ is solved to get $\mathbf{y} = \mathbf{A}^{-1}\mathbf{b}$, we want to solve this system:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)\mathbf{x} = \mathbf{b}. \tag{6.10}$$

We first pre-multiply both sides of this equation by $\mathbf{A}^{-1}$ to get:

$$\mathbf{A}^{-1}(\mathbf{A} + \mathbf{u}\mathbf{v}^T)\mathbf{x} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T)\mathbf{x} = \mathbf{x} + \mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{y}. \tag{6.11}$$

If we define $\mathbf{w} = \mathbf{A}^{-1}\mathbf{u}$ and $\alpha = \mathbf{v}^T\mathbf{x}$, the above equation can be written as:

$$\mathbf{x} + \mathbf{w}\alpha = \mathbf{y}. \tag{6.12}$$

Premultiplying both sides of this equation by $\mathbf{v}^T$, we get:

$$\mathbf{v}^T\mathbf{x} + \mathbf{v}^T\mathbf{w}\alpha = \mathbf{v}^T\mathbf{y}, \qquad i.e. \qquad \alpha + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}\alpha = \mathbf{v}^T\mathbf{y}. \tag{6.13}$$

Solving for α we get:

$$\alpha = \frac{\mathbf{v}^T y}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{6.14}$$

Substituting α into the previous equation for $\mathbf{x}$, we get

$$\mathbf{x} = \mathbf{y} - \mathbf{w}\alpha = \mathbf{y} - \frac{\mathbf{w}\mathbf{v}^T\mathbf{y}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{A}^{-1}\mathbf{b} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{b}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \left(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\right)\mathbf{b}. \tag{6.15}$$

But solving the equation $(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathbf{T}})\mathbf{x} = \mathbf{b}$ we also get

$$\mathbf{x} = (\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1}\mathbf{b}. \tag{6.16}$$

we therefore have:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{6.17}$$

Chapter 7

Inverse and Determinant of Partitioned Symmetric Matrices

From `http://fourier.eng.hmc.edu/e176/lectures/algebra/node7.html`

- **Theorem 1 (Woodbury identity)**

$$(\mathbf{A}+\mathbf{CBD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}. \tag{7.1}$$

Proof:

$$\begin{aligned}
&(\mathbf{A}+\mathbf{CBD})[\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}]\\
&=(\mathbf{A}+\mathbf{CBD})\mathbf{A}^{-1}-(\mathbf{A}+\mathbf{CBD})\mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}\\
&=\mathbf{I}+\mathbf{CBDA}^{-1}-(\mathbf{C}+\mathbf{CBDA}^{-1}\mathbf{C})(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}\\
&=\mathbf{I}+\mathbf{CBDA}^{-1}-\mathbf{CB}(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}\\
&=\mathbf{I}+\mathbf{CBDA}^{-1}-\mathbf{CBDA}^{-1}=\mathbf{I}.
\end{aligned} \tag{7.2}$$

In particular, if $\mathbf{C}=\mathbf{D}=\mathbf{I}$, then we have

$$(\mathbf{A}+\mathbf{B})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}. \tag{7.3}$$

- **Theorem 2** (inverse of a partitioned symmetric matrix)

  Divide an $n \times n$ symmetric matrix $\mathbf{A}$ into four blocks:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}. \tag{7.4}$$

  The inverse matrix $\mathbf{B} = \mathbf{A}^{-1}$ can also be divided into four blocks:

$$\mathbf{B} = \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}. \tag{7.5}$$

  Here we assume the dimensionalities of these blocks are:

  - $\mathbf{A}_{11}$ and $\mathbf{B}_{11}$ are $p \times p$,
  - $\mathbf{A}_{22}$ and $\mathbf{B}_{22}$ are $q \times q$,
  - $\mathbf{A}_{12} = \mathbf{A}_{21}^T$ and $\mathbf{B}_{12} = \mathbf{B}_{21}^T$ are $p \times q$,

  with $p + q = n$. Then we have

$$\begin{aligned}
\mathbf{B}_{11} \quad &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \\
&= \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\
\mathbf{B}_{22} \quad &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) \\
&= \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\
\mathbf{B}_{21} \quad &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \\
\mathbf{B}_{12} \quad &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}
\end{aligned} \tag{7.6}$$

  Proof:

$$\begin{aligned}
\mathbf{I}_n \quad &= \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}\mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{12} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}
\end{aligned} \tag{7.7}$$

  Equate each of the four blocks to get:

$$\begin{array}{lll}
\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{12} = \mathbf{I}_p, & \text{or} & \mathbf{B}_{11} = \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{21} \\
\mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0}, & \text{or} & \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\
\mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} = \mathbf{0}, & \text{or} & \mathbf{B}_{21} - \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11} \\
\mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I}_q, & \text{or} & \mathbf{B}_{22} = \mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{12}
\end{array} \tag{7.8}$$

Plug $\mathbf{B}_{21}$ into $\mathbf{B}_{11}$ to get:

$$\mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}. \tag{7.9}$$

Solve for $\mathbf{B}_{11}$ to get:

$$\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}. \tag{7.10}$$

Applying theorem 1 to this expression, we also get the other expression in the theorem. Similarily we can get:

$$\begin{aligned} \mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \\ \mathbf{B}_{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \\ \mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}^{-1})^{-1} \end{aligned} \tag{7.11}$$

- **Theorem 3** (Determinant of a partitioned symmetric matrix)

$$\begin{aligned} |\mathbf{A}| &= \left|\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{22} & \end{bmatrix}\right| \\ &= |\mathbf{A}_{22}||\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| = |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| \end{aligned} \tag{7.12}$$

Proof:

$$\begin{aligned} A = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21}7\mathbf{A}_{22} & \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ 0 & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \end{bmatrix} \end{aligned} \tag{7.13}$$

The theorem is proved as we also know that

$$|\mathbf{BC}| = |\mathbf{B}||\mathbf{C}| \tag{7.14}$$

and

$$\begin{vmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{X} & \mathbf{C} \end{vmatrix} = \begin{vmatrix} \mathbf{B} & \mathbf{X} \\ \mathbf{0} & \mathbf{C} \end{vmatrix} = |\mathbf{B}||\mathbf{C}|. \tag{7.15}$$

Chapter 8

Unitary Transform

From `http://fourier.eng.hmc.edu/e176/lectures/algebra/node8.html`
Given any *unitary matrix* $\mathbf{A}$ satisfying $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A} = \mathbf{I}$, we can define a *unitary transform* of a vector $\mathbf{x} = [x_1, \ldots, x_n]^T$:

$$\begin{cases} \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \ldots \\ \mathbf{y}_n \end{bmatrix} = \mathbf{A}^*\mathbf{x} = \begin{bmatrix} \bar{a}_1^T \\ \vdots \\ \bar{a}_n^T \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, & \text{(forward transform)} \\ \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \ldots \\ \mathbf{x}_n \end{bmatrix} = \mathbf{A}\mathbf{y} = \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n y_i a_i. & \text{(inverse transform)} \end{cases} \tag{8.1}$$

When $\mathbf{A} = \overline{\mathbf{A}}$ is real, $\mathbf{A}^{-1} = \mathbf{A}^T$ is an orthogonal matrix and the corresponding transform is an orthogonal transform.
The first equation above is the *forward transform* and can be written in component form as:

$$y_i = \bar{\mathbf{a}}_i^T \mathbf{x} = \langle \mathbf{x}, \mathbf{a}_i \rangle = \sum_{j=1}^n x_j \bar{a}_{ij}, \quad (i = 1, \ldots, n) \ . \tag{8.2}$$

The transform coefficient is an inner product $y_i = \langle \mathbf{x}, \mathbf{a}_i \rangle$, representing the projection of vector $\mathbf{x}$ onto the ith column vector $\mathbf{a}_i$ of the transform matrix $\mathbf{A}$. The second equation is the *inverse transform* and can also be written in component form as:

$$x_j = \sum_{i=1}^n a_{ji} y_i, \quad (j = 1, \ldots, n) \ . \tag{8.3}$$

By this transform, vector $\mathbf{x}$ is represented as a linear combination (weighted sum) of the n column vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ of matrix $\mathbf{A}$. Geometrically, $\mathbf{x}$ is a point in the n-dimensional space spanned by these n orthonormal basis vectors. Each coefficient (coordinate) y_i is the projection of $\mathbf{x}$ onto the corresponding basis vector $\mathbf{a}_i$.

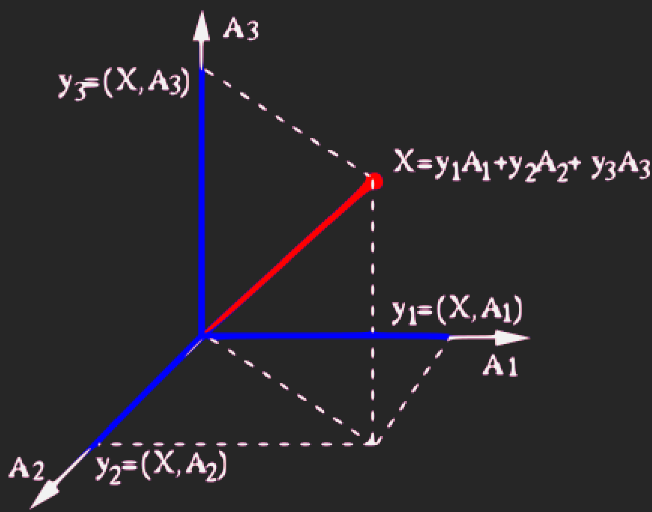

Figure 8.1: Each coefficient (coordinate) y_i is the projection of $\mathbf{x}$ onto the corresponding basis vector $\mathbf{a}_i$

As the n-dimensional space can be spanned by the column vectors of *any* $n \times n$ unitary (orthogonal) matrix, a vector $\mathbf{x}$ in the space can be represented by any of such matrices, each defining a different transform.

Examples:

- When $\mathbf{A} = \mathbf{I} = [\dots, \mathbf{e_i}, \dots]$ is an identity matrix, we have

$$\mathbf{x} = \sum_{i=1}^{n} y_i \mathbf{a}_i = \sum_{i=1}^{n} x_i \mathbf{e}_i \tag{8.4}$$

  where $\mathbf{e}_i = [0, \dots, 0, 1, 0, \dots, 0]^T$ is the ith column of $\mathbf{I}$ with the ith element equal 1 and all other 0.

- When $a_{mn} = w[m, n] = e^{-j2\pi mn/N}$, the corresponding transform is discrete Fourrier transform. The nth column vector $\mathbf{w}_n$ if the transform matrix $\mathbf{W} = [\mathbf{w}_0, \ldots, \mathbf{w}_{N-1}]$ represents a sinusoid of frequency nf_0, and the corresponding complex coordinate $y_n = (\mathbf{x}, \mathbf{w}_n)$ represents the magnitude $|y_n|$ and phase $\angle y_n$ of this nth frequency component. The Fourrier transform $\mathbf{y} = \mathbf{W}\mathbf{x}$ represents a rotation of the coordinate system.

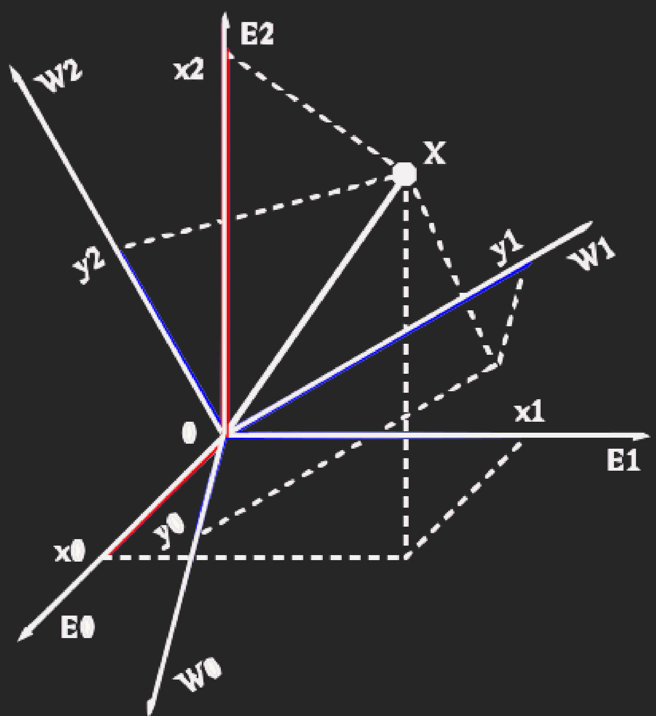

Figure 8.2: The Fourrier transform $\mathbf{y} = \mathbf{W}\mathbf{x}$ represents a rotation of the coordinate system.

A unitary (orthogonal if real) transform $\mathbf{y} = \mathbf{A}\mathbf{x}$ can be interpreted geometrically as the rotation vector $\mathbf{x}$ about the origin, or equivalently, the representation of the same vector in a rotated coordinate system. A unitary transform $\mathbf{y} = \mathbf{A}\mathbf{x}$ does not change the inner product. Let $\mathbf{u} = \mathbf{A}^*\mathbf{x}$ and $\mathbf{v} = \mathbf{A}^*\mathbf{y}$ be the unitary transforms of vectors $\mathbf{x}$ and $\mathbf{y}$, then th inner product of these two vectors is preserved:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{A}^* \mathbf{u}, \mathbf{A}^* \mathbf{v} \rangle = \mathbf{x}^* \mathbf{A} \mathbf{A}^* \mathbf{y} = \mathbf{x}^* \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle \tag{8.5}$$

In particular, if $\mathbf{x} = \mathbf{y}$, we see that he vector norm is preserved by the unitary transform:

$$||\mathbf{u}||^2 = \langle \mathbf{u}, \mathbf{u} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle = ||\mathbf{x}||^2 \tag{8.6}$$

This is the Parseval's identity that indicates that the notm (length) of a vector is preserved under any unitary transform. If $\mathbf{X}$ is interpreted as a signal, then its length $||\mathbf{x}||^2 = ||\mathbf{y}||^2$ represents the total energy of information contained in the signal, which is conserved during any unitary transform. However, some other features of the signal may change, e.g., the signal may be decorrelated and its total energy redistributed among its components after the transform, which may be desirable in many applications.

If $\mathbf{x}$ is a random vector with mean vector $\mathbf{m}_x$ and covariance matrix $\mathbf{\Sigma}_x$:

$$\mathbf{m}_x = E(\mathbf{x}), \quad \mathbf{\Sigma}_x = E(\mathbf{x}\mathbf{x}^*) - \mathbf{m}_x \mathbf{m}_x^* \;, \tag{8.7}$$

then its transform $\mathbf{y} = \mathbf{A}^* \mathbf{x}$has the following mean vector and covariance matrix:

$$\begin{aligned} \mathbf{m}_y &= E(\mathbf{y}) = E(\mathbf{A}^* \mathbf{x}) = \mathbf{A}^* E(\mathbf{x}) = \mathbf{A}^* \mathbf{m}_x \\ \mathbf{\Sigma}_y &= E(\mathbf{y}\mathbf{y}^*) - \mathbf{m}_y \mathbf{m}_y^* = E[(\mathbf{A}^* \mathbf{x})(\mathbf{A}^* \mathbf{x})^*] - (\mathbf{A}^* \mathbf{m}_x)(\mathbf{A}^* \mathbf{m}_x)^* \\ &= E[\mathbf{A}^* (\mathbf{x}\mathbf{x}^*) \mathbf{A}] - \mathbf{A}^* \mathbf{m}_x \mathbf{m}_x^* \mathbf{A} = \mathbf{A}^* [E(\mathbf{x}\mathbf{x}^*) - \mathbf{m}_x \mathbf{m}_m^*] \mathbf{A} \\ &= \mathbf{A}^* \mathbf{\Sigma}_x . \end{aligned} \tag{8.8}$$

In general the unitary transform of any square matrix $\mathbf{A}$ by a unitary matrix $\mathbf{R}$ is

$$\mathbf{B} = \mathbf{R}^* \mathbf{A} \mathbf{R}. \tag{8.9}$$

Chapter 9

Eigenvalue Decomposition

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node9.html`
For any $n \times n$ square matrix $\mathbf{A}$, if there exists a vector $\mathbf{v}$ such that the following *eigenequation* holds:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \tag{9.1}$$

then λ and $\mathbf{v}$ are called the *eigenvalue* and *eigenvector* of matrix $\mathbf{A}$, respectively. In other words, the linear transformation of vector $\mathbf{v}$ by $\mathbf{A}$ has the same effect of scaling the vector by factor λ. (Note that for an $n \times n$ non-square matrix $\mathbf{A}$ with $m \neq n$, $\mathbf{A}\mathbf{v}$ is an m-D vector but $\lambda\mathbf{v}$ is an n-D vector, i.e., no eigenvalues and eigenvectors are defined.)
Given $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, we also have $\mathbf{A}c\mathbf{v} = \lambda c\mathbf{v}$ for any scalar constant c, i.e., the eigenvector $\mathbf{v}$ is not unique but upt ot any scaling factor. For the uniqueness of $\mathbf{v}$, we typically keep it normalised so that $||v|| = 1$.
To obtain λ, we rewrite the above equation as

$$(\lambda \mathbf{I} - \mathbf{A})\mathbf{v} = 0. \tag{9.2}$$

For this homogeneous equation system to have non-zero solutions for $\mathbf{v}$, the determinant of its coefficient matrix has to be zero:

$$\det(\lambda \mathbf{I} - \mathbf{A}) = |\lambda \mathbf{I} - \mathbf{A}| = 0 \tag{9.3}$$

This is the *characteristic polynomial equation* of matrix $\mathbf{A}$. Solving this n^{th} order equation of λ we get n eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. Substituting each λ_i back into the homogeneous equation system, we get the corresponding eigenvector

$\mathbf{v}_i$. We can put all n eigen-equations $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ together and obtain the more compact form:

$$\mathbf{A}\mathbf{V} = \mathbf{A}[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [\lambda_1 \mathbf{v}_1, \ldots, \lambda_n \mathbf{v}_n] = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} = \mathbf{V}\mathbf{\Lambda} \tag{9.4}$$

where we have defined

$$\mathbf{\Lambda} = diag[\lambda_1, \ldots, \lambda_n] \text{ and } \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]. \tag{9.5}$$

The eigen-equations can be written in some alternative forms:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \text{ or } \mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda}. \tag{9.6}$$

In the first form, $\mathbf{A} = \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ is expressed as a product of three matrices, called *eigenvalue decomposition* of the matrix; in the second form, $\mathbf{A}$ is diagonalized by its eigenvector matrix $\mathbf{V}$ to become a diagonal matrix, its eigenvalue matrix $\mathbf{\Lambda}$.
Given $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$, we have the following:

- $\mathbf{A}^T$ has the same eigenvalues and eigenvectors as $\mathbf{A}$.

  Proof: As a matrix $\mathbf{A}$ and its transpose $\mathbf{A}^T$ have the same determinant, they have the same characteristic polynomial:

  $$|\mathbf{A} - \lambda \mathbf{I}| = |(\mathbf{A} - \lambda \mathbf{I})^T| = |\mathbf{A}^T - \lambda \mathbf{I}|, \tag{9.7}$$

  therefore they have the same eigenvalues and eigenvectors.

- The eigenvalues and eigenvectors of $\mathbf{A}^*$ are the complex conjugate of the eigenvalues and eigenvectors of $\mathbf{A}$.

- $\mathbf{A}^T\mathbf{A}$ has the same eigenvectors as $\mathbf{A}$, but its eigenvalues are λ^2.

  Proof:

  $$\mathbf{A}^T\mathbf{A}\mathbf{v} = \mathbf{A}^T \lambda \mathbf{v} = \lambda^2 \mathbf{v}. \tag{9.8}$$

- $\mathbf{A}^k$ has the same eigenvectors as $\mathbf{A}$, but its eigenvalues are $\{\lambda_1^k, \ldots, \lambda_n^k\}$, where k is a positive integer.

Proof:

$$\mathbf{A}^2\mathbf{v} = \mathbf{A}\mathbf{A}\mathbf{v} = \mathbf{A}\lambda\mathbf{v} = \lambda^2\mathbf{v}. \tag{9.9}$$

This result can be generalised to

$$\mathbf{A}^k\mathbf{v} = \lambda^k\mathbf{v}. \tag{9.10}$$

- In particular when $k = -1$, i.e., the eigenvalues of $\mathbf{A}^{-1}$ are $\{\lambda_1^{-1}, \dots, \lambda_n^{-1}\}$.

  Proof:
  Let $\mathbf{\Lambda} = diag(\lambda_1, \dots, \lambda_n)$ and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ be the eigenvalues and eigenvector matrices of a square matrix $\mathbf{A}$:

$$\mathbf{A}\mathbf{V} = \mathbf{\Lambda}\mathbf{V} \tag{9.11}$$

  and $\mathbf{\Sigma} = diag(\sigma_1, \dots, \sigma_n)$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$ be the eigenvalue and eigenvector matrices of $\mathbf{B} = \mathbf{R}^*\mathbf{A}\mathbf{R}$, a unitary transform of $\mathbf{A}$:

$$\mathbf{B}\mathbf{U} = (\mathbf{R}^*\mathbf{A}\mathbf{R})\mathbf{U} = \mathbf{U}\mathbf{\Sigma}. \tag{9.12}$$

  Left multiplying $\mathbf{R}$ on both sides we get the eigenequation of $\mathbf{A}$

$$\mathbf{R}\mathbf{R}^*\mathbf{A}\mathbf{R}\mathbf{U} = \mathbf{A}(\mathbf{R}\mathbf{U}) = (\mathbf{R}\mathbf{U})\mathbf{\Sigma}. \tag{9.13}$$

  We see that $\mathbf{A}$ and $\mathbf{B} = \mathbf{R}^*\mathbf{A}\mathbf{R}$ have the same eigenvalues $\mathbf{\Sigma} = \mathbf{\Lambda}$ and their eigenvector matrices are related by $\mathbf{V} = \mathbf{R}\mathbf{U}$ or $\mathbf{U} = \mathbf{R}^*\mathbf{V}$.

- Given all eigenvalues $\lambda_1, \dots, \lambda_n$ of a matrix $\mathbf{A}$, its trace and determinant can be obtained as

$$tr(\mathbf{A}) = \sum_{k=1}^{n} \lambda_k, \ \det(\mathbf{A}) = \prod_{k=1}^{n} \lambda_k. \tag{9.14}$$

- The *spectrum* of an $n \times n$ square matrix $\mathbf{A}$ is the set of its eigenvalues $\{\lambda_1, \dots, \lambda_n\}$. The *spectral radius* of $\mathbf{A}$, denoted by $\rho(\mathbf{A})$, is the maximum of the absolute values of the elements in the spectrum:

$$\rho\mathbf{A} = \max(|\lambda_1|, \dots, |\lambda_n|), \tag{9.15}$$

  where $|z| = \sqrt{x^2 + y^2}$ is the modulus of a complex number $z = x + jy$. If all eigenvalues are sorted such that $|\lambda_1| \geq \cdots \geq |\lambda_n|$ then $\rho()\mathbf{A}) = |\lambda_1| = |\lambda_{max}|$. As the eigenvalues of $\mathbf{A}^{-1}$ are $\{1/\lambda_{max}, \dots, 1/\lambda_{min}\}$, $\rho(\mathbf{A}^{-1}) = 1/|\lambda_{min}|$.

- If $\mathbf{A}$ and $\mathbf{B}$ are similar, i.e.,

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \tag{9.16}$$

then they have the same eigenvalues.

Proof:

Let $\mathbf{\Lambda}$ and $\mathbf{V}$ be the eigenvalues and eigenvector matrices of $\mathbf{A}$:

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}, \; \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \tag{9.17}$$

then we have

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{P} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}, \tag{9.18}$$

i.e., $\mathbf{\Lambda}$ and $\mathbf{U} = \mathbf{P}^{-1}\mathbf{V}$ are the eigenvalue and eigenvector matrices of $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

- if $\mathbf{A}$ is Hermitian $\mathbf{A}^* = \mathbf{A}^T = \mathbf{A}$ (symmetric $\mathbf{A}^T = \mathbf{A}$ if real) (e.g., the covariance matrix of a random vector), then all of its eigenvalues $\bar{\lambda}_i = \lambda_i$ are real, and all of its eigenvectors are orthogonal, $\mathbf{v}_i^*\mathbf{v}_j = 0$ $(i \neq j)$ i.e., $\mathbf{V}^{-1} = \mathbf{V}^*$. And we further have

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}^T = \mathbf{U}\mathbf{U}^T \tag{9.19}$$

where $\mathbf{U} = \mathbf{V}\mathbf{\Lambda}^{1/2}$.

Proof:

Let λ and $\mathbf{v}$ be an eigenvalue and the corresponding eigenvector of a Hermitian matrix $\mathbf{A} = \mathbf{A}^*$, i.e., $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, then we have

$$\begin{aligned} (\mathbf{A}\mathbf{v})^*\mathbf{v} &= (\lambda\mathbf{v})^*\mathbf{v} = \bar{\lambda}\mathbf{v}^*\mathbf{v} = \bar{\lambda}||\mathbf{v}||^2 \\ &= \mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{v}^*\mathbf{A}^*\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} = \lambda||\mathbf{v}||^2 \end{aligned} \tag{9.20}$$

i.e., $\bar{\lambda} = \lambda$ is real. We also have

$$\begin{aligned} \mathbf{v}_i^*\mathbf{A}\mathbf{v}_j &= \mathbf{v}_i^*\lambda_j\mathbf{v}_j = \lambda_j\mathbf{v}_i^*\mathbf{v}_j \\ &= (\mathbf{v}_j^*\mathbf{A}\mathbf{v}_i)^* = (\mathbf{v}_j^*\lambda_i\mathbf{v}_i)^* = \bar{\lambda}_i\mathbf{v}_i^*\mathbf{v}_j = \lambda_i\mathbf{v}_i^*\mathbf{v}_j \end{aligned} \tag{9.21}$$

i.e.,

$$\lambda_j\mathbf{v}_i^*\mathbf{v}_j = \lambda_i\mathbf{v}_i^*\mathbf{v}_j, \text{ or } (\lambda_i - \lambda_j)\mathbf{v}_i^*\mathbf{v}_j = 0. \tag{9.22}$$

As $\lambda_i \neq \lambda_j$, we get $\mathbf{v}_i^* \mathbf{v}_j = 0$, i.e., the eigenvectors corresponding to different eigenvalues are orthogonal.

When all eigenvectors are normalised $\mathbf{v}_i^* \mathbf{v}_i = ||\mathbf{v}_i||^2 = 1$, they become orthonormal

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \mathbf{v}_i^* \mathbf{v}_j = \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases} \tag{9.23}$$

i.e., the eigenvector matrix $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]^T$ is unitary (orthogonal if $\mathbf{A}$ is real):

$$\mathbf{V}^{-1} = \mathbf{V}^* \text{ i.e. } \mathbf{V}^* \mathbf{V} = \mathbf{I} \tag{9.24}$$

and we have

$$\mathbf{V}^{-1} \mathbf{A} \mathbf{V} = \mathbf{V}^* \mathbf{A} \mathbf{V} = \mathbf{\Lambda}. \tag{9.25}$$

Left and right multiplying by $\mathbf{V}$ and $\mathbf{V}^* = \mathbf{V}^{-1}$ respectively on the two sides, we get

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^* = [\mathbf{v}_1, \dots, \mathbf{v}_n] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_n^* \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^*. \tag{9.26}$$

This is the *spectral theorem* indicating that $\mathbf{A}$ can be written as a linear combination of n matrices $\mathbf{v})i\mathbf{v}_i^*$ weighted by λ_i $(i = 1, \dots, n)$.

The significance of this property is that a linear operation $\mathbf{y} = \mathbf{A}\mathbf{x}$ applied to vector $\mathbf{x}$ can be mapped to a new vector space in which the operations of the components are independent of each other. Consider

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^* \mathbf{x}. \tag{9.27}$$

Pre-multiplying $\mathbf{V}^*$ on both sides we get

$$\mathbf{y}' = \mathbf{V}^* \mathbf{y} = \mathbf{\Lambda} \mathbf{V}^* \mathbf{x} = \mathbf{\Lambda} \mathbf{x}', \tag{9.28}$$

where we have defined $\mathbf{x}' = \mathbf{V}^*\mathbf{x}$ and $\mathbf{y}' = \mathbf{V}^*\mathbf{y}$, the unitary transform of $\mathbf{x}$ and $\mathbf{y}$, respectively. We see that for the ith component we have the following, independent of all other components:

$$y_i' = \lambda_i x_i' \tag{9.29}$$

- The entries on the diagonal of an upper (or lower) triangular matrix are its eigenvalues.

  Proof:)

  Let $\mathbf{A}$ be an upper triangular matrix with $a_{i,j} = 0$ for all $i > j$. The eigenvalues of $\mathbf{A}$ are the roots fo the following homogeneous characteristic equations:

  $$\det(\mathbf{A} - \lambda\mathbf{I}) = \prod_{i=1}^{n} (a_{ii} - \lambda) = 0. \tag{9.30}$$

  The first equal sign is due to the fact that $\mathbf{A} - \lambda\mathbf{I}$ is also an upper-triangular matrix, and the determinant of an upper-triangular matrix is the product of all its diagonal entries. We therefore see that each diagonal entry a_{ii}, as a root of the characteristic equation, is also an eigenvalue of $\mathbf{A}$.

- Similar matrices have the same eigenvalues.

  Proof:

  Let λ and $\mathbf{v}$ be an eigenvalue and the corresponding eigenvector of $\mathbf{A}$ satisfying $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, and $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ be a similar matrix of$\mathbf{A}$. We have $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$ and

  $$\mathbf{A}\mathbf{v} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}\mathbf{v} = \lambda\mathbf{v}, \text{ i.e., } \mathbf{B}\mathbf{P}^{-1}\mathbf{v} = \lambda\mathbf{P}^{-1}\mathbf{v}. \tag{9.31}$$

  In other words, λ is also the eigenvalue of $\mathbf{B}$ with the corresponding eigenvector $\mathbf{P}^{-1}\mathbf{v}$.

- All eigenvalues of a *stochastic matrix* $\mathbf{P}$ are no greater than 1. A stochastic matrix of which component P_{ij} is the probability for a state transition of a Markov process (chain) from s_i to s_j, i.e. $0 \le P_{ij} \le 1$ and $\sum_j P_{ij} = 1$.

  Proof:

  First, as $\mathbf{P}\mathbf{1} = 1$, $\lambda = 1$ is one of the eigenvalues of $\mathbf{P}$. Next let λ and $\mathbf{v}$ be an eigenvalue and the corresponding eigenvector of $\mathbf{P}$, i.e., $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$, and let $|v_k| \geq \max\{|v_1|, \ldots, |v_n|\}$. The kth row of the eigenequation is

$$|\lambda v_k| \leq |\lambda||v_k| = \left|\sum_{j=1}^{n} P_{kj} v_j\right| \leq \sum_{j=1}^{n} P_{kj}|v_j| \leq |v_k| \sum_{j=1}^{n} P_{kj} = |v_k| \tag{9.32}$$

i.e. $|\lambda| \leq 1$.

Chapter 10

Generalized Eigenvalue Problem

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node10.html`
The *generalized eigenvalue problem* of two symmetric matrices $\mathbf{A} = \mathbf{A}^T$ and $\mathbf{B} = \mathbf{B}^T$ is to find a scalar λ_i and the corresponding vector $\mathbf{v}_i$ for the following generalized eigen equation to hold:

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{B}\mathbf{v}_i, \qquad (i = 1, \ldots, n). \tag{10.1}$$

or in matrix form:

$$\mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\mathbf{\Lambda}. \tag{10.2}$$

The eigenvalue and eigenvector matrices $\mathbf{\Lambda}$ and $\mathbf{V}$ can be found in the following steps:

- Solve the eigenvalue problem of $\mathbf{B}$ to find its diagonal eigenvalue matrix $\mathbf{\Lambda}_B$ and orthogonal eigenvectors matrix $\mathbf{V}_B = (\mathbf{V}_B^T)^{-1}$ so that

$$\mathbf{B}\mathbf{V}_B = \mathbf{V}_B\mathbf{\Lambda}_B, \qquad \text{or} \qquad \mathbf{V}_B^{-1}\mathbf{B}\mathbf{V}_B = \mathbf{V}_B^T\mathbf{B}\mathbf{V}_B = \mathbf{\Lambda}_B. \tag{10.3}$$

- Left and right multiplying both sides of the second equation above by $\mathbf{\Lambda}^{-1/2}$ (whitening) we get

$$\mathbf{\Lambda}_B^{-1/2}(\mathbf{V}_B^T\mathbf{B}\mathbf{V}_B)\mathbf{\Lambda}_B^{-1/2} = \mathbf{\Lambda}_B^{-1/2}\mathbf{\Lambda}_B\mathbf{\Lambda}_B^{-1/2} = \mathbf{I}. \tag{10.4}$$

We define

$$\mathbf{V}_B^{'} = \mathbf{V}_B \mathbf{\Lambda}_B^{-1/2}, \tag{10.5}$$

and the above becomes

$$(\mathbf{V}_B^{'})^T \mathbf{B} \mathbf{V}_B^{'} = \mathbf{I}. \tag{10.6}$$

Note that $\mathbf{V}_B^{'}$ is not orthogonal

$$(\mathbf{V}^{'})_B^{-1} = (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2})^{-1} = \mathbf{\Lambda}_B^{1/2} \mathbf{V}_B^{-1} = \mathbf{\Lambda}_B^{1/2} \mathbf{V}_B^T \neq \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_B^T = \mathbf{V}_B^T. \tag{10.7}$$

- Apply the same transformation to $\mathbf{A}$:

$$(\mathbf{V}_B^{'})^T \mathbf{A} \mathbf{V}_B^{'} = (\mathbf{\Lambda}_B^{-1/2} \mathbf{V}_B^T) \mathbf{A} (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2}) = \mathbf{A}^{'}. \tag{10.8}$$

Note that $\mathbf{A}^{'}$ is symmetric as well as $\mathbf{A}$:

$$\mathbf{A}^{'T} = (\mathbf{V}_B^{'T} \mathbf{A} \mathbf{V}_B^{'})^T = \mathbf{V}_B^{'T} \mathbf{A} \mathbf{V}_B^{'} = \mathbf{A}^{'}. \tag{10.9}$$

- Diagonalize $\mathbf{A}^{'}$

As $\mathbf{A}$ is symmetric, it can be diagonalized by its orthogonal eigenvector marix $\mathbf{V}_A$:

$$\mathbf{V}_A^T \mathbf{A}^{'} \mathbf{V}_A = \mathbf{\Lambda}. \tag{10.10}$$

i.e.,

$$\begin{array}{ll} \mathbf{V}_A^T (\mathbf{\Lambda}_B^{-1/2} \mathbf{V}_B^T \mathbf{A} \mathbf{V}_B \mathbf{\Lambda}_B^{-1/2}) \mathbf{V}_A & = (\mathbf{V}_A^T \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_B^T) \mathbf{A} (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A \\ & = \mathbf{V}^T \mathbf{A} \mathbf{V} = \mathbf{A} \end{array}. \tag{10.11}$$

where we have defined

$$\mathbf{V} = \mathbf{V}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A, \tag{10.12}$$

which is not orthgonal:

$$\mathbf{V}^{-1} = (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A)^{-1} = \mathbf{V}_A^{-1} \mathbf{\Lambda}_B^{1/2} \mathbf{V}_B^{-1} = \mathbf{V}_A^T \mathbf{\Lambda}_B^{1/2} \mathbf{V}_B^T \neq \mathbf{V}_A^T \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_B^T = \mathbf{V}^T. \tag{10.13}$$

- This $\mathbf{V}$ also diagonalizes $\mathbf{B}$:

$$\begin{aligned} \mathbf{V}^T \mathbf{B} \mathbf{V} &= (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A)^T \mathbf{B} (\mathbf{V}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A) = \mathbf{V}_A^T \mathbf{\Lambda}_B^{-1/2} (\mathbf{V}_B^T \mathbf{B} \mathbf{V}_B) \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A \\ &= \mathbf{V}_A^T \mathbf{\Lambda}_B^{-1/2} \mathbf{\Lambda}_B \mathbf{\Lambda}_B^{-1/2} \mathbf{V}_A = \mathbf{V}_A^T \mathbf{V}_A = \mathbf{I} \end{aligned} . \tag{10.14}$$

- Now we have:

$$\left\{ \begin{array}{l} \mathbf{V}^T \mathbf{A} \mathbf{V} = \mathbf{\Lambda} \\ \mathbf{V}^T \mathbf{B} \mathbf{V} = \mathbf{I} \end{array} \right. . \tag{10.15}$$

Right multiplying both sides of the second equation by $\mathbf{\Lambda}$ we get $\mathbf{V}^T \mathbf{B} \mathbf{V} \mathbf{\Lambda} = \mathbf{\Lambda}$, then equating the left-hand side to that of the first equation, we get:

$$\mathbf{V}^T \mathbf{A} \mathbf{V} = \mathbf{V}^T \mathbf{B} \mathbf{V} \mathbf{\Lambda}, \qquad \text{i.e.} \qquad \mathbf{A} \mathbf{V} = \mathbf{B} \mathbf{V} \mathbf{\Lambda}, \tag{10.16}$$

i.e, $\mathbf{\Lambda}$ and $\mathbf{V}$ are the eigenvalue and eigenvector matrices of the generalized eigenvalue problem.

Note, however, as shown above, $\mathbf{V}$ is not orthogonal.

Chapter 11

Rayleigh Quotient

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node11.html`
The *Rayleigh quotient* $R(\mathbf{A}$ of a hermitian matrix $\mathbf{A} = \mathbf{A}^*$ (symmetric if $\mathbf{A}$ is real) is defined as:

$$R(\mathbf{A}) = \frac{\mathbf{w}^* \mathbf{A} \mathbf{w}}{\mathbf{w}^* \mathbf{w}}, \tag{11.1}$$

where $\mathbf{w}$ is any non-zero vector. In particular, if $\mathbf{w} = \mathbf{v}_i$ is the ith eigenvector of $\mathbf{A}$, then the Rayleigh quotient is the corresponding to eigenvalue λ_i:

$$R(\mathbf{A}) = \frac{\mathbf{v}_i^* \mathbf{A} \mathbf{v}_i}{\mathbf{v}_i^* \mathbf{v}_i} = \frac{\lambda_i \mathbf{v}_i^* \mathbf{v}_i}{\mathbf{v}_i^* \mathbf{v}_i} = \lambda_i. \tag{11.2}$$

The Rayleigh quotient of two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ is a function of a vector $\mathbf{w}$ defined as:

$$R(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{B} \mathbf{w}}. \tag{11.3}$$

To find the optimal $\mathbf{w}$ corresponding to the extremum (maximum or minimum) of $R(\mathbf{w})$, we find its derivative with respect to $\mathbf{w}$:

$$\frac{d}{d\mathbf{w}} R(\mathbf{w}) = \frac{2\mathbf{A}\mathbf{w}(\mathbf{w}^T \mathbf{B} \mathbf{w}) - 2\mathbf{B}\mathbf{w}(\mathbf{w}^T \mathbf{A} \mathbf{w})}{(\mathbf{w}^T \mathbf{B} \mathbf{w})^2}. \tag{11.4}$$

Setting it to zero we get:

$$\mathbf{A}\mathbf{w}(\mathbf{w}^T\mathbf{B}\mathbf{w}) = \mathbf{B}\mathbf{w}(\mathbf{w}^T\mathbf{A}\mathbf{w}), \qquad \text{i.e.} \qquad \mathbf{A}\mathbf{w} = \frac{\mathbf{w}^T\mathbf{A}\mathbf{w}}{\mathbf{w}^T\mathbf{B}\mathbf{w}}\mathbf{B}\mathbf{w} = R(\mathbf{w}) = \lambda\mathbf{B}\mathbf{w}. \tag{11.5}$$

The second equation can be recognised as a generalised eigenvalue problem with $\lambda = R(\mathbf{w})$ being the eigenvalue and $\mathbf{w}$ the corresponding eigenvector. Solving this we get the vector $\mathbf{w}$ corresponding to the maximum/minimum eigenvalue $\lambda = R(\mathbf{w})$, which maximizes/minimizes the Rayleigh quotient.

Chapter 12

Normal Matrices and Diagonalizability

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node12.html`
Theorem: The product of two unitary matrices is unitary.
Proof: Let $\mathbf{U}$ and $\mathbf{V}$ be unitary, i.e. $\mathbf{U}^* = \mathbf{U}^{-1}$ and $\mathbf{V}^* = \mathbf{V}^{-1}$, thrn $\mathbf{UV}$ is unitary:

$$(\mathbf{UV})^* = \mathbf{V}^*\mathbf{U}^* = \mathbf{V}^{-1}\mathbf{U}^{-1} = (\mathbf{UV})^{-1}. \tag{12.1}$$

Theorem: Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are simultaneously diagonalizable if and only if they commute:
Proof: (reference)

- Let $\mathbf{A}$ and $\mathbf{B}$ be simultaneously diagonalizable by $\mathbf{R}$

  $$\mathbf{R}^{-1}\mathbf{AR} = \mathbf{\Lambda}_A, \qquad \mathbf{R}^{-1}\mathbf{BR} = \mathbf{\Lambda}_B, \tag{12.2}$$

  then

  $$\begin{aligned} \mathbf{AB} &= (\mathbf{R\Lambda}_A\mathbf{R}^{-1})(\mathbf{R\Lambda}_B\mathbf{R}^{-1}) = (\mathbf{R\Lambda}_A\mathbf{\Lambda}_B\mathbf{R}^{-1}) = (\mathbf{R\Lambda}_B\mathbf{\Lambda}_A\mathbf{R}^{-1}) \\ &= (\mathbf{R\Lambda}_B\mathbf{R}^{-1})(\mathbf{R\Lambda}_A\mathbf{R}^{-1}) = \mathbf{BA} \end{aligned} . \tag{12.3}$$

- Let $\mathbf{A}$ and $\mathbf{B}$ commute, i.e., $\mathbf{AB} = \mathbf{BA}$. Assuming $\mathbf{u}$ is an eigenvector of $\mathbf{A}$ corresponding to eigenvalue λ, i.e., $\mathbf{Au} = \lambda\mathbf{u}$, then:

$$\mathbf{ABu} = \mathbf{B}(\mathbf{Au}) = \lambda\mathbf{Bu}, \tag{12.4}$$

  we see that $\mathbf{Bu}$ is also an eigenvector of $\mathbf{A}$ corresponding to the same eigenvalue λ, i.e., $\mathbf{Bu}$ must be scaled version of $\mathbf{u}$ (in the same 1-D space). $\mathbf{Bu} = \gamma\mathbf{u}$, i.e., $\mathbf{u}$ is also an eigenvector of $\mathbf{B}$.

Theorem: A matrix is normal if and only if it is unitarily diagonalizable.
Proof: (reference)

- If $\mathbf{A}$ is unitarily diagonalizable:

$$\mathbf{AU} = \mathbf{U\Lambda}, \qquad \mathbf{U}^{-1}\mathbf{AU} = \mathbf{U}^*\mathbf{AU} = \mathbf{\Lambda}, \qquad \mathbf{A} = \mathbf{U\Lambda U}^*, \tag{12.5}$$

  where $\mathbf{U}^* = \mathbf{U}^{-1}$ is unitary and $\mathbf{\Lambda}$ is a diagonal matrix satisfying $\mathbf{\Lambda}^*\mathbf{\Lambda} = \mathbf{\Lambda\Lambda}^*$, then $\mathbf{A}$ is normal:

$$\begin{aligned} \mathbf{AA}^* &= (\mathbf{U\Lambda U}^*)(\mathbf{U\Lambda U}^*)^* = (\mathbf{U\Lambda U}^*)(\mathbf{U\Lambda}^*\mathbf{U}^*) = \mathbf{U\Lambda\Lambda}^*\mathbf{U}^* \\ &= \mathbf{U\Lambda}^*\mathbf{\Lambda U}^* = (\mathbf{U\Lambda}^*\mathbf{U}^*)(\mathbf{U\Lambda U}^*) = \mathbf{A}^*\mathbf{A} \end{aligned}. \tag{12.6}$$

- If $\mathbf{A}$ is normal, then it is diagonalizable by unitary matrices. First we show any matrix $\mathbf{A}$ can be written as:

$$\mathbf{A} = \mathbf{B} + i\mathbf{C}, \tag{12.7}$$

  where

$$\mathbf{B} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^*) = \mathbf{B}^*, \qquad \mathbf{C} = -\frac{1}{2}(\mathbf{A} - \mathbf{A}^*) = \mathbf{C}^*, \tag{12.8}$$

  are both Hermitian, and diagonalizable by unitary matrix. As $\mathbf{A}$ is normal, we have:

$$0 = \mathbf{AA}^* - \mathbf{A}^*\mathbf{A} = (\mathbf{B}+i\mathbf{C})(\mathbf{B}-i\mathbf{C}) - (\mathbf{B}-i\mathbf{C})(\mathbf{B}+i\mathbf{C}) = 2i(\mathbf{CB}-\mathbf{BC}). \tag{12.9}$$

We see that $\mathbf{CB} = \mathbf{BC}$, i.e., $\mathbf{B}$ and $\mathbf{C}$ commute, and they can be simultaneously diagonalized by some unitary matrix $\mathbf{U}$:

$$\mathbf{U}^*\mathbf{B}\mathbf{U} = \mathbf{\Lambda}_B, \qquad \mathbf{U}^*\mathbf{C}\mathbf{U} = \mathbf{\Lambda}_C, \tag{12.10}$$

and so can $\mathbf{A} = \mathbf{B} + i\mathbf{C}$:

$$\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{U}^*(\mathbf{B} + i\mathbf{C})\mathbf{U} = \mathbf{\Lambda}_B + i\mathbf{\Lambda}_C = \mathbf{\Lambda}_A. \tag{12.11}$$

Q.E.D.
If $\mathbf{A}$ is normal, i.e., it commutes with its conjugate transpose $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$, then $\mathbf{A}$ and $\mathbf{A}^*$ can be simultaneously diagonalized by their unitary eigenvector matrix $\mathbf{U}$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*, \qquad \mathbf{A}^* = \mathbf{U}\overline{\mathbf{\Lambda}}\mathbf{U}^*. \tag{12.12}$$

Chapter 13

Centering Matrix

From: `http://fourier.eng.hmc.edu/e176/lectures/algebra/node13.html`

The centering matrix is defined as:

$$\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T, \tag{13.1}$$

where $\mathbf{1} = [1, \ldots, 1]^T$ and $\mathbf{1}\mathbf{1}^T$ is an $n \times n$ matrix of all 1's. For example,

$$\mathbf{C}_1 = 0; \tag{13.2}$$

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}; \tag{13.3}$$

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}. \tag{13.4}$$

When $\mathbf{C}_n$ is applied to an n-D vector $\mathbf{v} = [v_1, \ldots, v_n]^T$, we get:

$$\mathbf{C}_n\mathbf{v} = \mathbf{v} - \frac{1}{n}(\mathbf{1}\mathbf{1}^T)\mathbf{v} = \mathbf{v} - \frac{1}{n}\sum_{i=1}^{n} v_i\mathbf{1} = \mathbf{v} - \overline{v}\mathbf{1}, \tag{13.5}$$

where $\overline{v}$ is the mean of all components of $\mathbf{v}$:

$$\bar{v} = \frac{1}{n}\mathbf{1}^T\mathbf{v} = \frac{1}{n}\sum_{i=1}^{n} v_i. \tag{13.6}$$

We see that the mean of $\mathbf{v}$ is removed from the resulting vector $\mathbf{C}_n\mathbf{v}$.

$$\mathbf{C}_n\mathbf{v} = \mathbf{v} - \frac{1}{n}\left(\mathbf{1}\mathbf{1}^T\right)\mathbf{v} = \mathbf{v} - \frac{1}{n}\sum_{i=1}^{n} v_i\mathbf{1} = \mathbf{v} - \bar{v}\mathbf{1} \tag{13.7}$$

where $\bar{v}$ is the mean of all components of $\mathbf{v}$:

$$\bar{v} = \frac{1}{n}\mathbf{1}^T\mathbf{v} = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{13.8}$$

We see that the mean of $\mathbf{v}$ is removed from the resulting vector $\mathbf{C}_n\mathbf{v}$.

Example: $\mathbf{v} = [3,\ -2,\ 5]^T$, with mean $\bar{v} = (3-2+5)/3 = 2$,

$$\mathbf{u} = \mathbf{C}_3\mathbf{v} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix} \tag{13.9}$$

The mean of the resulting vector $\mathbf{u}$ is $\bar{u} = 0$. In particular, if $\mathbf{v} = \mathbf{1}$, then

$$\mathbf{C}_n\mathbf{v} = \mathbf{C}_n\mathbf{1} = \left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}n = \mathbf{0} \tag{13.10}$$

i.e., $\mathbf{C}_n\mathbf{1} = 0\,\mathbf{1}$. We therefore see that $\lambda = 0$ and $\mathbf{v} = \mathbf{1}$ are the eigenvalue and corresponding eigenvector of $\mathbf{C}_n$. Also, consider any zero-mean vector $\mathbf{v} = [v_1, \cdots, v_n]$ satisfying $\sum_{i=1}^{n} v_i = \mathbf{1}^T\mathbf{v} = 0$ with $n-1$ degrees of freedom, we have

$$\mathbf{C}_n\mathbf{v} = \left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{v} = \mathbf{v} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{v} = \mathbf{v} \tag{13.11}$$

i.e., $\mathbf{C}_n\mathbf{v} = 1\,\mathbf{v}$. We therefore see that $\lambda = 1$ and any $\mathbf{v}$ satisfying $\mathbf{1}^T\mathbf{v} = 0$ are the eigenvalue of multiplicity $n-1$ and the corresponding eigenvector of $\mathbf{C}_n$. The certering matrix $\mathbf{C}_n$ has the following properties:

Idempotence: $\mathbf{C}_n^k = \mathbf{C}_n$

$$\mathbf{C}_n^2 = \mathbf{C}_n\mathbf{C}_n = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) = \mathbf{I}_n - \frac{2}{n}\mathbf{1}\mathbf{1}^T + \frac{n}{n^2}\mathbf{1}\mathbf{1}^T = \mathbf{C}_n \tag{13.12}$$

Once the mean of vector $\mathbf{v}$ is removed, its mean is zero and subsequent removal of mean has not effect. $\mathbf{C}_n$ has eigenvalue $\lambda = 1$ of multiplicity $n - 1$ and eigenvalue $\lambda = 0$ of multiplicity 1. $\mathbf{C}_n$ is singular, its inverse does not exist. Once the mean of vector $\mathbf{v}$ is removed, it cannot be reconstructed by an inverse process.

Chapter 14

Matrix Usage Examples

14.1 Polynomial equations

If we have a $n+1$ number of polynomial equations

$$\mathbf{f}(x) = \sum_{i=0}^{n} \mathbf{a}_i \mathbf{x}^n, \tag{14.1}$$

where

$$\mathbf{a}_i = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} \tag{14.2}$$

is a coefficient vector, and

$$\mathbf{f}(x) = \begin{bmatrix} f(x_0) \\ \vdots \\ f(x_n) \end{bmatrix} \tag{14.3}$$

is a vector of function values at $n+1$ points, then to obtain coefficients we can write in matrix form:

$$\begin{bmatrix} x_0^n & \dots & x_0^0 \\ \vdots & \ddots & \vdots \\ x_n^n & \dots & x_n^0 \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ \vdots \\ f(x_n) \end{bmatrix}, \tag{14.4}$$

$$\mathbf{X}\mathbf{a} = \mathbf{f}(x), \tag{14.5}$$

therefore:

$$\mathbf{a} = \mathbf{X}^{-1}\mathbf{f}(x). \tag{14.6}$$

14.2 Lagrange interpolation polynomials

Example:
Let's have 3 2nd order Lagrange polynomials $f_n(x) \in \mathbb{R}$; $i, j \in \mathbb{N}$; $i, j \in \langle 1, k+1 \rangle$, where k is the polynomial order, defined so that:

$$f_i(x_j) = \sum_{n=0}^{k} a_n x^n = \begin{cases} 1, \; i = j \\ 0, \; i \neq j \end{cases} \quad \forall \quad x_j \in \left\{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\right\}. \tag{14.7}$$

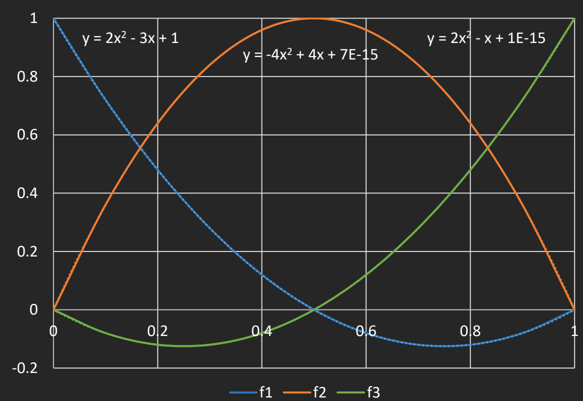

Figure 14.1: 2^{nd} order Lagrange polynomial shape functions.

Then for 1^{st} we have:

$$\begin{array}{ll} f_1(0) & = 1 \\ f_1(1/2) & = 0 \\ f_1(1) & = 0 \end{array}, \tag{14.8}$$

for 2^{nd} we have:

$$\begin{array}{ll} f_2(0) & = 0 \\ f_2(1/2) & = 1 \\ f_2(1) & = 0 \end{array}, \tag{14.9}$$

and for 3^{rd}:

$$\begin{array}{ll} f_2(0) & = 0 \\ f_2(1/2) & = 0 \\ f_2(1) & = 1 \end{array}. \tag{14.10}$$

That gives us following matrix:

$$\begin{array}{r} \\ f_1(0): \\ f_2(1/2): \\ f_3(1): \end{array} \begin{array}{c} \begin{array}{ccc} x^2 & x^1 & x^0 \end{array} \\ \hline \begin{bmatrix} 0 & 0 & 1 \\ 1/4 & 1/2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{ccc} 1. & 2. & 3. \end{array} \\ \hline \begin{bmatrix} a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,0} & a_{2,0} & a_{3,0} \end{bmatrix} \end{array} = \begin{array}{c} \begin{array}{ccc} 1. & 2. & 3. \end{array} \\ \hline \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}. \tag{14.11}$$

$$\begin{bmatrix} a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,0} & a_{2,0} & a_{3,0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/4 & 1/2 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14.12}$$

$$\begin{bmatrix} a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,0} & a_{2,0} & a_{3,0} \end{bmatrix} = \begin{bmatrix} 2 & -4 & 2 \\ -3 & 4 & -1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14.13}$$

$$\begin{bmatrix} a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,0} & a_{2,0} & a_{3,0} \end{bmatrix} = \begin{bmatrix} 2 & -4 & 2 \\ -3 & 4 & -1 \\ 1 & 0 & 0 \end{bmatrix}. \tag{14.14}$$

This means:

$$\begin{array}{ll} f_1(x) & = 2x^2 - 3x + 1 \\ f_2(x) & = -4x^2 + 4x \\ f_3(x) & = 2x^2 - x \end{array}. \tag{14.15}$$

Note:
More general Lagrange polynomial definition is:
The *Lagrange interpolation polynomial* is the **polynomial** $P(x)$ of degree $\leq (n-1)$ that passes through n points $\{x_1, y_1 = f(x_1)\}, \{x_2, y_2 = f(x_2)\}, \ldots, \{x_n, y_n = f(x_n)\}$ and is given by:

$$P(x) = \sum_{j=1}^{n} P_j(x). \tag{14.16}$$

where

$$P_j(x) = y_j \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}. \tag{14.17}$$

Written explicitely:

$$\begin{aligned} P(x) = & \frac{(x-x_2)(x-x_3)\ldots(x-x_n)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_n)} y_1 \\ & + \frac{(x-x_1)(x-x_3)\ldots(x-x_n)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_n)} y_2 \\ & + \ldots \\ & + \frac{(x-x_1)(x-x_3)\ldots(x-x_{n-1})}{(x_n-x_1)(x_n-x_4)\ldots(x_n-x_{n-1})} y_n \end{aligned} . \tag{14.18}$$

For our example the polynomial equation can be reduced as the first equation goes through points $y_j = \{1, 0, 0\}$, therefore terms of $P_j(x)$ with y_2 and y_3 are zero and $P_j(x)$ gets reduced to (for curve 2, only second term applies, for curve 3 only the third one):

$$
\begin{aligned}
P_1(x) &= \frac{(x-1/2)(x-1)}{(0-1/2)(0-1)} \times 1 &&= \frac{x^2 - 3/2x + 1/2}{1/2} &&= 2x^2 - 3x + 1 \\
P_2(x) &= \frac{(x-0)(x-1)}{(1/2-0)(1/2-1)} \times 1 &&= \frac{x^2 - x}{-1/4} &&= -4x^2 + 4x \ , \\
P_3(x) &= \frac{(x-0)(x-1/2)}{(1-0)(1-1/2)} \times 1 &&= \frac{x^2 - 1/2x}{1/2} &&= 2x^2 - x
\end{aligned}
\tag{14.19}
$$

where the index of $P(x)$ denotes the curve number $f(x)$, not the *Lagrange polynomial* member.
In other form:

$$
\begin{aligned}
f_1(x) &= (2x-1)(x-1) = 2x^2 - 3x + 1 \\
f_2(x) &= -4x(x-1) = -4x^2 + 4x. \\
f_3(x) &= x(2x-1) = 2x^2 - x
\end{aligned}
\tag{14.20}
$$

14.3 3D and 2D

Rotation Matrices around axis:

$$\mathbf{T}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{14.21}$$

$$\mathbf{T}_y = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{14.22}$$

$$\mathbf{T}_z = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{14.23}$$

14.3.1 Projection from 3D to 2D:

If $\mathbf{X} = \{x, y, z\}^T$ and $\mathbf{X_p}$ is the projected 2D coordinate vector, then:
Orthogonal projection:

$$\mathbf{X}_p = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \end{bmatrix} \mathbf{X} \quad = \begin{bmatrix} S_x * x & S_y * y \end{bmatrix} \tag{14.24}$$

where:
S_x and S_y are scale factors.
Perspective projection:

$$\mathbf{X}_p = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \end{bmatrix} \mathbf{X} * \frac{d}{z+d} \tag{14.25}$$

where:
d is distance from the eye.
To project more coordinates at the same time:
With:

$$\mathbf{X}_{orig} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \tag{14.26}$$

Then **orthogonal** projection is:

$$
\begin{gathered}
\mathbf{X}_p = \mathbf{X}_{orig} \begin{bmatrix} S_x & 0 \\ 0 & S_y \\ 0 & 0 \end{bmatrix} \\
\begin{bmatrix} x_{p1} & y_{p1} \\ x_{p2} & y_{p2} \\ \vdots & \vdots \\ x_{pn} & y_{pn} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \\ 0 & 0 \end{bmatrix}
\end{gathered}
\tag{14.27}
$$

And **perspective** projection is:

$$
\begin{gathered}
\mathbf{X}_p = \mathbf{X}_{orig} \begin{bmatrix} S_x & 0 \\ 0 & S_y \\ 0 & 0 \end{bmatrix} \times \frac{d}{\mathbf{X}_{orig}[3] + d} \\
\begin{bmatrix} x_{p1} & y_{p1} \\ x_{p2} & y_{p2} \\ \vdots & \vdots \\ x_{pn} & y_{pn} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} \frac{d}{z_1+d} \\ \frac{d}{z_2+d} \\ \vdots \\ \frac{d}{z_n+d} \end{bmatrix} \\
x_{p1} = (x_1 * S_x + y_1 * 0 + z_1 * 0) * \frac{d}{z_1 + d} \\
y_{p1} = (x_1 * 0 + y_1 * S_y + z_1 * 0) * \frac{d}{z_1 + d} \\
\vdots \\
x_{pn} = (x_n * S_x + y_n * 0 + z_n * 0) * \frac{d}{z_n + d}
\end{gathered}
\tag{14.28}
$$

14.3.2 Model movement

The rotated model coordinates are defined as:

$$
\mathbf{X}_r = \mathbf{T}_r(\mathbf{X} + \mathbf{T}_t) \tag{14.29}
$$

where:
$\mathbf{X}_r$ are the new rotated coordinates,

$\mathbf{T}_r$ is the rotation matrix
$\mathbf{X}$ are the original coordinates,
$\mathbf{T}_t$ is a translation matrix.

- Then when rotating the **View** of the model, the new rotation matrix is:

$$\mathbf{T}_{r,new} = \mathbf{T}_r \mathbf{T}_{r,prev} \tag{14.30}$$

- When rotating the **model**, the new rotation matrix is:

$$\mathbf{T}_{r,new} = \mathbf{T}_{r,prev} \mathbf{T}_r \tag{14.31}$$