Midterm 1 Review

Vikas Dhiman

February 16, 2022

Contents

1 Linear algebra review

Definition 1 (Matrix). *A real matrix A with n rows and m columns is defined as a set of real numbers $\{a_{11}, a_{12}, \dots, a_{nm}\}$, arranged in an 2D grid with n rows and m columns :*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \tag{1}$$

The set of all possible real matrices with n rows and m columns is denoted as $\mathbb{R}^{n\times m}$, where $\mathbb{R}$ denotes the set of all real numbers.

Any matrix A with with n rows and m columns is said to lie in the set of $\mathbb{R}^{n\times m}$. $A \in \mathbb{R}^{n\times m}$ is read aloud as "A lies in the set of all n cross m real matrices".

Definition 2 (Vector or Column vector). *A column vector or a vector $\mathbf{x}$ is a matrix with only one column.*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2}$$

The set of all possible real vectors with n rows is denoted as $\mathbb{R}^{n\times 1}$ or more simply $\mathbb{R}^n$.

A vector is denoted by bold-font small letter, for example, $\mathbf{x}, \mathbf{y}, \mathbf{z}$. A matrix is denoted by capital letters, A, B, M, P, K.

A matrix $A \in \mathbb{R}^{n\times m}$ is often denoted a set m column vectors of dimension $n \times 1$,

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \end{bmatrix},$$
$$\text{where } \mathbf{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix}, \qquad \text{for all } i \in \{1, \dots, m\}. \quad (3)$$

A block matrix is a matrix denoted in terms of other matrices,

$$A = \left[\begin{array}{ccc|ccc} b_{11} & \dots & b_{1q} & c_{11} & \dots & c_{1r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & \dots & b_{pq} & c_{1s} & \dots & c_{sr} \\ \hline e_{11} & \dots & e_{1v} & d_{11} & \dots & d_{1x} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{u1} & \dots & e_{uv} & d_{1w} & \dots & d_{wx} \end{array}\right] \quad (4)$$
$$= \begin{bmatrix} B & C \\ E & D \end{bmatrix}, \text{ where } B, C, E, D \text{ are matrices.} \quad (5)$$

Definition 3 (Square matrix)**.** *A matrix is said to be square if its number of columns is same as the number of rows. That is matrix $A \in \mathbb{R}^{n \times m}$ is said to be square matrix if $m = n$.*

Definition 4 (Diagonal of a square matrix)**.** *Let A be a square matrix $A \in \mathbb{R}^{n \times n}$ with entries:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad (6)$$

The diagonal of a square matrix A is defined to be the vector

$$diag(A) = \begin{bmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{nn} \end{bmatrix}$$

Definition 5 (Trace of a square matrix)**.** *Trace of a square matrix A is defined as the sum its diagonal elements,*

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

.

Definition 6 (Identity matrix)**.** *An identity matrix I of size n is a square matrix with all its diagonal entries as 1 and non-diagonal entries as 0.*

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (7)$$

1.1 Matrix operations

1.1.1 Transpose

Definition 7 (Transpose)**.** *The matrix transpose $A^\top$ of a matrix A is defined as a matrix where rows of matrix A are the columns of $A^\top$ and vice-versa.*

$$A^\top = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \quad (8)$$

In the matrix as set of m column vectors notation, the transpose is written as m row vectors $\mathbf{a}_i^\top$,

$$A^\top = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}, \quad \mathbf{a}_i^\top = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix},$$
$$\text{for all } i \in \{1, \dots, n\}. \quad (9)$$

1. If A has $n$rows and m columns, then $A^\top$ has m rows and n columns. If $A \in \mathbb{R}^{n \times m}$, then $A^\top \in \mathbb{R}^{m \times n}$.
2. The transpose of a transpose is matrix itself. $(A^\top)^\top = A$.
3. The transpose of a block matrix is block-wise transpose of each matrix,

$$\begin{bmatrix} B & C \\ E & D \end{bmatrix}^\top = \begin{bmatrix} B^\top & E^\top \\ C^\top & D^\top \end{bmatrix}$$

Definition 8 (Row vector)**.** *A row vector is Y is matrix with only one row*

$$Y = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \quad (10)$$

It is common to denote row vectors as tranpose of a column vector. For example, the matrix Y shown above is typically represented $\mathbf{y}^\top$, where $\mathbf{y}$ is a column vector.

$$Y = \mathbf{y}^\top \qquad \text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{11}$$

1.1.2 Vector dot product

Before we define general matrix multiplication, it is easier to define matrix multiplication between a row vector and a column vector $\mathbf{x}^\top \in \mathbb{R}^{1\times n}$ and $\mathbf{y} \in \mathbb{R}^{n\times 1}$

$$\mathbf{x}^\top \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i \tag{12}$$

where $\mathbf{x}^\top = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$

and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

Note that $\mathbf{x}^\top\mathbf{y}$ is same as the vector dot product or the vector inner-product,

$$\mathbf{x}^\top \mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| cos(\theta) = \mathbf{y}^\top \mathbf{x}, \tag{13}$$

where θ is the angle between vectors $\mathbf{x}$ and $\mathbf{y}$ and the vector norm or euclidean norm $\|.\|$ is defined as

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^\top \mathbf{x}} \tag{14}$$

Definition 9 (Unit vector)**.** *A unit vector, typically denoted with a hat,* $\hat{\mathbf{x}}$ *is a vector with euclidean norm as 1. That is* $\|\hat{\mathbf{x}}\| = 1$ *or equivalently* $\mathbf{x}^\top\mathbf{x} = 1$.

Definition 10 (Orthogonal vectors)**.** *Two vectors,* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathbb{R}^n$ *are said to be orthogonal if and only if their dot product is zero* $\mathbf{x}^\top\mathbf{y} = 0$.

Definition 11 (Orthonormal vectors)**.** *A set of vectors,* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^n$ *are said to be orthonormal if and only if they are all unit vectors* $\mathbf{x}_i^\top \mathbf{x}_i = 1$ *and they are pair-wise orthogonal,* $\mathbf{x}_i^\top \mathbf{x}_j = 0$ *for all* $i \neq j$.

1.1.3 Matrix multiplication

The matrix multiplication between matrix $A \in \mathbb{R}^{n\times m}$ and matrix $B \in \mathbb{R}^{m\times p}$ (note that A has m columns while B has m rows; the only case when matrix multiplication is defined) is easier defined if matrix A is written in terms of row vectors while matrix B is written in terms of column vectors. Let the matrix A is written in terms of row vectors $\mathbf{a}_i^\top \in \mathbb{R}^{1\times m}$ and the matrix B is written in terms of column vectors $\mathbf{b}_i \in \mathbb{R}^{m\times 1}$. Then the matrix multiplication $AB \in \mathbb{R}^{n\times p}$ is defined as the matrix,

$$AB = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_p \end{bmatrix} \tag{15}$$

$$= \begin{bmatrix} \mathbf{a}_1^\top \mathbf{b}_1 & \mathbf{a}_1^\top \mathbf{b}_2 & \dots & \mathbf{a}_n^\top \mathbf{b}_p \\ \mathbf{a}_2^\top \mathbf{b}_1 & \mathbf{a}_2^\top \mathbf{b}_2 & \dots & \mathbf{a}_n^\top \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_n^\top \mathbf{b}_1 & \mathbf{a}_n^\top \mathbf{b}_2 & \dots & \mathbf{a}_n^\top \mathbf{b}_p \end{bmatrix} \tag{16}$$

Block matrix multiplication Block matrix multiplication works in a similar way as scalar multiplication as long as sub-matrix multiplication is properly defined,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \begin{bmatrix} AP + BR & AQ + BS \\ CP + DR & CQ + DS \end{bmatrix} \tag{17}$$

Definition 12 (Orthogonal matrices)**.** *A square matrix* A *is said to be orthogonal if and only if* $A^\top A = I$

1.1.4 Transpose of matrix multiplication

$$(AB)^\top = B^\top A^\top$$

1.1.5 Properties of trace operator

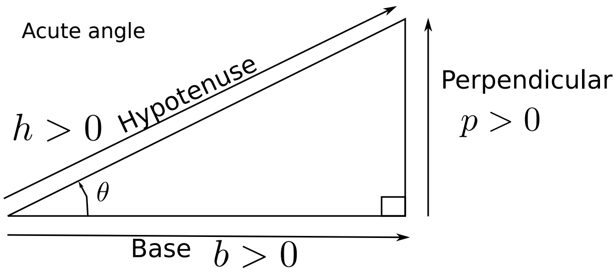
Trace is a linear operator:

$$\mathrm{tr}(\alpha A + \beta B) = \alpha\,\mathrm{tr}(A) + \beta\,\mathrm{tr}(B), \tag{18}$$

for compatible matrices A and B and scalars α and β.

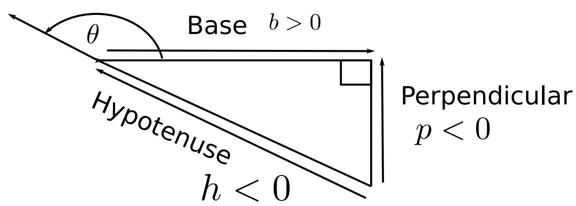
2 Trignometry review

Acute angle

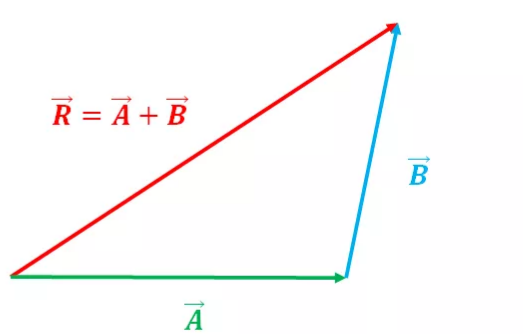


$\tan(\theta) = \frac{p}{b}$ $\quad \sin(\theta) = \frac{p}{h}$ $\quad \cos(\theta) = \frac{b}{h}$

Obtuse angle



3 Triangle law of vector addition



4 2D Rotation matrix

Definition 13 (2D Cartesian Coordinate frame)**.** *A 2D cartesian coordinate frame is defined as a set of mutually orthogonal unit vectors* $\hat{\mathbf{x}} \in \mathbb{R}^2$ *and* $\hat{\mathbf{y}} \in \mathbb{R}^2$ *called the basis vectors* $B = [\hat{\mathbf{x}}, \hat{\mathbf{y}}]$ *along with an origin* $\mathbf{o} \in \mathbb{R}^2$*. Thus the tuple* $(B, \mathbf{o})$ *form a coordinate frame. A coordinate frame is denoted by curly braces around it, for example,* $\{C\}$ *or* $\{W\}$*.*

Example 1 (2D Coordinate frame)**.** *The figure 1 contains two coordinate frames the one shown in red and the one shown in green. Both have the same origin, but different basis vectors. The* $\{W\}$ *coordinate frame shown in green has basis vectors* $B_w = [\hat{\mathbf{x}}_w, \hat{\mathbf{y}}_w]$*. The same notation is used for the* $\{C\}$ *coordinate frame* $B_c = [\hat{\mathbf{x}}_c, \hat{\mathbf{y}}_c]$*. Note that the basis vectors of* $\{C\}$ *coordinate frame can be expressed in terms of* $\{W\}$ *coordinate frame by triangle law of vector addition,*

$$\begin{aligned} \hat{\mathbf{x}}_c &= |\overrightarrow{OA}|\hat{\mathbf{x}}_w + |\overrightarrow{AB}|\hat{\mathbf{y}}_w \\ \hat{\mathbf{y}}_c &= -|\overrightarrow{PQ}|\hat{\mathbf{x}}_w + |\overrightarrow{OP}|\hat{\mathbf{y}}_w \end{aligned} \tag{19}$$

In the triangle ΔOAB *(Fig 1),*

$$\cos(\theta) = \frac{|\overrightarrow{OA}|}{|\overrightarrow{OB}|} = \frac{|\overrightarrow{OA}|}{\|\hat{\mathbf{x}}_c\|} = |\overrightarrow{OA}| \tag{20}$$

$$\sin(\theta) = \frac{|\overrightarrow{AB}|}{|\overrightarrow{OB}|} = \frac{|\overrightarrow{AB}|}{\|\hat{\mathbf{x}}_c\|} = |\overrightarrow{AB}| \tag{21}$$

Similarly in the right triangle ΔOPQ *(Fig 1),*

$$\cos(\theta) = \frac{|\overrightarrow{OP}|}{|\overrightarrow{OQ}|} = \frac{|\overrightarrow{OP}|}{\|\hat{\mathbf{y}}_c\|} = |\overrightarrow{OP}| \tag{22}$$

$$\sin(\theta) = \frac{|\overrightarrow{PQ}|}{|\overrightarrow{OQ}|} = \frac{|\overrightarrow{PQ}|}{\|\hat{\mathbf{y}}_c\|} = |\overrightarrow{PQ}| \tag{23}$$

Putting these values back in (19)*, we get,*

$$\begin{aligned} \hat{\mathbf{x}}_c &= cos(\theta)\hat{\mathbf{x}}_w + sin(\theta)\hat{\mathbf{y}}_w \\ \hat{\mathbf{y}}_c &= -sin(\theta)\hat{\mathbf{x}}_w + cos(\theta)\hat{\mathbf{y}}_w \end{aligned} \tag{24}$$

These equations can be written in matrix notation as,

$$\begin{aligned} \hat{\mathbf{x}}_c &= \begin{bmatrix} \hat{\mathbf{x}}_w & \hat{\mathbf{y}}_w \end{bmatrix} \begin{bmatrix} cos(\theta) \\ sin(\theta) \end{bmatrix} = B_w \begin{bmatrix} cos(\theta) \\ sin(\theta) \end{bmatrix} \\ \hat{\mathbf{y}}_c &= \begin{bmatrix} \hat{\mathbf{x}}_w & \hat{\mathbf{y}}_w \end{bmatrix} \begin{bmatrix} -sin(\theta) \\ cos(\theta) \end{bmatrix} = B_w \begin{bmatrix} -sin(\theta) \\ cos(\theta) \end{bmatrix} \end{aligned} \tag{25}$$

The full basis matrix of coordinate frame $\{C\}$ can be written as

$$\begin{aligned} B_c &= \begin{bmatrix}\hat{\mathbf{x}}_c & \hat{\mathbf{y}}_c\end{bmatrix} \\ &= \begin{bmatrix} B_w \begin{bmatrix} cos(\theta) \\ sin(\theta)\end{bmatrix} & B_w \begin{bmatrix} -sin(\theta) \\ cos(\theta)\end{bmatrix}\end{bmatrix} \\ &= B_w \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta)\end{bmatrix} \end{aligned} \tag{26}$$

Definition 14 (2D Coordinates of a point)**.** *The coordinate of a point $\mathbf{p}$ in a given coordinate frame $\{W\}$ with basis vectors $B_w = [\hat{\mathbf{x}}_w, \hat{\mathbf{y}}_w]$ and origin $\mathbf{o}_w = \begin{bmatrix} o_x \\ o_y \end{bmatrix}$ is defined as the vector $\mathbf{p}_w = \begin{bmatrix} p_{wx} \\ p_{wy} \end{bmatrix}$ such that,*

$$\begin{aligned} \mathbf{p} &= (p_{wx} + o_x)\hat{\mathbf{x}}_w + (p_{wy} + o_y)\hat{\mathbf{y}}_w \\ &= \begin{bmatrix}\hat{\mathbf{x}}_w & \hat{\mathbf{y}}_w\end{bmatrix} \left(\begin{bmatrix} p_{wx} \\ p_{wy} \end{bmatrix} + \begin{bmatrix} o_x \\ o_y \end{bmatrix} \right) \\ &= B_w(\mathbf{p}_w + \mathbf{o}_w) \end{aligned} \tag{27}$$

Example 2 (Fig 1)**.** *The point $\mathbf{p}$ can be represented in coordinate frames $\{W\}$ and $\{C\}$. Let the projection on the basis $B_c = [\hat{\mathbf{x}}_c, \hat{\mathbf{y}}_c]$ be $\mathbf{p}_c$, while that on $B_w = [\hat{\mathbf{x}}_w, \hat{\mathbf{y}}_w]$ be $\mathbf{p}_w$. Since both the coordinate frames have same origin, we assume $\mathbf{o}_w = \mathbf{o}_c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. We have*

$$\mathbf{p} = B_w \mathbf{p}_w = B_c \mathbf{p}_c \tag{28}$$

Theorem 1 (2D Rotation matrix)**.** *In a coordinate transformation as given in Fig 1, the coordinates in frame $\{C\}$, $\mathbf{p}_c$ can be converted into coordinates in frame $\{W\}$, $\mathbf{p}_w$ with the same origin by using a rotation matrix ${}^W R_C(\theta)$,*

$$\begin{aligned} \mathbf{p}_w &= {}^W R_C(\theta) \mathbf{p}_c \\ \text{where } {}^W R_C(\theta) &= \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta)\end{bmatrix} \end{aligned} \tag{29}$$

Proof. First note that the basis matrix of any coordinate frame $\{W\}$ is orthogonal,

$$\begin{aligned} B_w^\top B_w &= \begin{bmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}}\end{bmatrix}^\top \begin{bmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}}\end{bmatrix} \\ &= \begin{bmatrix}\hat{\mathbf{x}}^\top \\ \hat{\mathbf{y}}^\top\end{bmatrix} \begin{bmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}}\end{bmatrix} \\ &= \begin{bmatrix}\hat{\mathbf{x}}^\top\hat{\mathbf{x}} & \hat{\mathbf{x}}^\top\hat{\mathbf{y}} \\ \hat{\mathbf{y}}^\top\hat{\mathbf{x}} & \hat{\mathbf{y}}^\top\hat{\mathbf{y}}\end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix} = I \end{aligned} \tag{30}$$

Left-multiply $B_w^\top$ to both sides of (28)

$$B_w^\top B_w \mathbf{p}_w = B_w^\top B_c \mathbf{p}_c \tag{31}$$

Replace $B_w^\top B_w = I$.

$$I\mathbf{p}_w = B_w^\top B_c \mathbf{p}_c \text{ or } \mathbf{p}_w = B_w^\top B_c \mathbf{p}_c \tag{32}$$

Substitute value of B_c from (26), to get

$$\mathbf{p}_w = B_w^\top \left(B_w \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta)\end{bmatrix} \right) \mathbf{p}_c. \tag{33}$$

Again use $B_w^\top B_w = I$ to get,

$$\mathbf{p}_w = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta)\end{bmatrix} \mathbf{p}_c. \tag{34}$$

Defining ${}^W R_C(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta)\end{bmatrix}$, we get the desired result.

□

Theorem 2 (Orthogonality of 2D Rotation matrices)**.** *All 2D rotation matrices are orthogonal $R^\top R = I$ have determinant as one $\det(R) = 1$. If any square matrix $A \in \mathbb{R}^{2\times 2}$ is orthogonal $A^\top A = I$ and has determinant 1, $\det(A) = 1$, then it is a valid rotation matrix.*

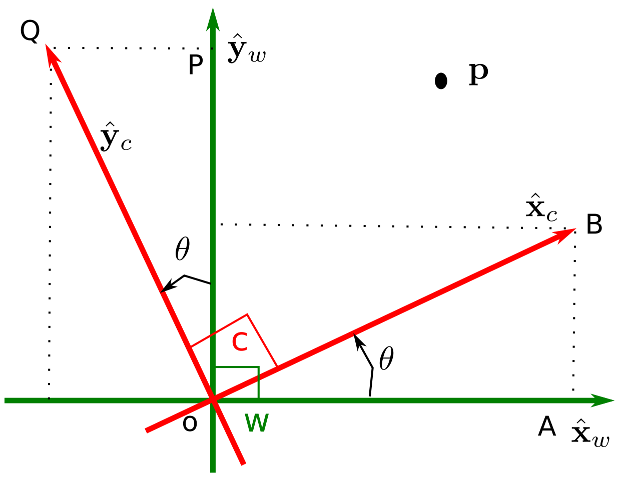


Figure 1: The coordinate frame $\{C\}$ is rotated around origin by an θ from coordinate frame $\{W\}$.

Proof.

$$
\begin{aligned}
R^\top R &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}^\top \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\
&= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\
&= \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & -\cos(\theta)\sin(\theta) + \sin(\theta)\cos(\theta) \\ -\sin(\theta)\cos(\theta) + \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (35)
\end{aligned}
$$

$$
\begin{aligned}
\det(R) &= \det \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\
&= \cos^2(\theta) + \sin^2(\theta) = 1 \qquad (36)
\end{aligned}
$$

Denote the columns of square matrix A which is orthogonal with determinant 1 as $A = [\mathbf{a}_1, \mathbf{a}_2]$. Since A is orthogonal, we have

$$
\begin{aligned}
A^\top A &= \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{a}_1 & \mathbf{a}_1^\top \mathbf{a}_2 \\ \mathbf{a}_2^\top \mathbf{a}_1 & \mathbf{a}_2^\top \mathbf{a}_2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (37)
\end{aligned}
$$

This implies that $\mathbf{a}_1$ and $\mathbf{a}_2$ are mutually orthogonal unit vectors. Let $\mathbf{a}_1 = [\cos(\theta), \sin(\theta)]$ because any 2D unit vector can be written in cos,sin form, where $\theta = \operatorname{arctan2}(a_{12}, a_{11})$. Next we know that $\mathbf{a}_1^\top \mathbf{a}_2 = 0$ and that $\mathbf{a}_2$ is unit vector. For any unit 2D vector $[u, v]^\top$, there are only two unit vectors perpendicular to it $[-v, u]^\top$ and $[v, -u]^\top$. Then we have only two options for $\mathbf{a}_2$ are either $[-\sin(\theta), \cos(\theta)]$ or $[\sin(\theta), -\cos(\theta)]$. But we also know that the determinant of A is 1. The second option for $\mathbf{a}_2$ leads to determinant of -1.

$$
\det \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \det \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} = -1 \qquad (38)
$$

Hence, we have

$$
A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = R(\theta)
$$

□

5 2D Transformation matrix

To consider the rotation and translation case, we consider the case shown in Fig 2. We have an intermediate frame $\{I\}$ which has only rotation from $\{C\}$ frame. We assume that basis vectors $\{I\}$ are parallel to $\{W\}$ which make it translation only conversion. We can convert from $\mathbf{p}_c$ to $\mathbf{p}_I$ using the rotation matrix derived in the previous section,

$$
\mathbf{p}_I = B_I^{-1} B_c \mathbf{p}_c = R(\theta)\mathbf{p}_c. \qquad (39)
$$

We can account for the translation of the frame $\mathbf{p}_I$ by noticing that the coordinate frames only differ in origin, such that $B_c\mathbf{o}_c = B_w(\mathbf{o}_w + {}^w\mathbf{t}_c)$, where the translation ${}^w\mathbf{t}_c$ is measured in world coordinate frame.

$$
\begin{aligned}
\mathbf{p} &= B_c(\mathbf{p}_c + \mathbf{o}_c) = B_w(\mathbf{p}_w + \mathbf{o}_w) \\
\implies & B_c\mathbf{p}_c + B_c\mathbf{o}_c = B_w\mathbf{p}_w + B_w\mathbf{o}_w \\
\implies & B_c\mathbf{p}_c + (B_c\mathbf{o}_c - B_w\mathbf{o}_w) = B_w\mathbf{p}_w \\
\implies & B_c\mathbf{p}_c + B_w{}^w\mathbf{t}_c = B_w\mathbf{p}_w \\
\implies & B_w^\top B_c\mathbf{p}_c + {}^w\mathbf{t}_c = \mathbf{p}_w \\
\implies & \mathbf{p}_w = R(\theta)\mathbf{p}_c + {}^w\mathbf{t}_c \qquad (40)
\end{aligned}
$$

This relation is often written in terms of homogeneous coordinates which are obtained by appending 1 to euclidean coordinates $\underline{\mathbf{p}}_w = \begin{bmatrix} \mathbf{p}_w \\ 1 \end{bmatrix}$ and $\underline{\mathbf{p}}_c = \begin{bmatrix} \mathbf{p}_c \\ 1 \end{bmatrix}$.

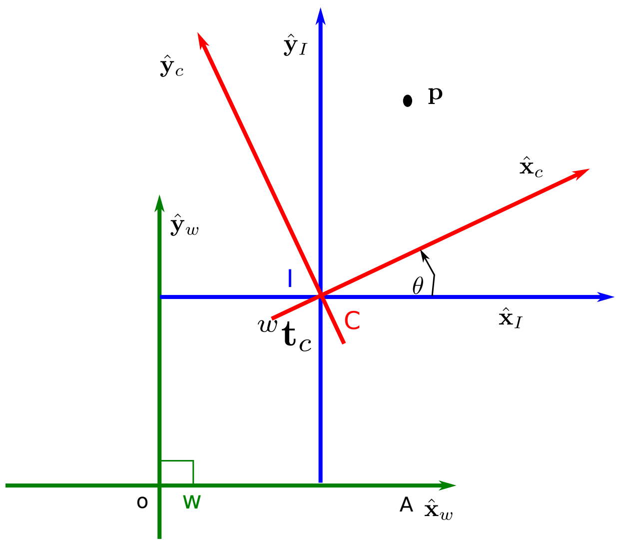
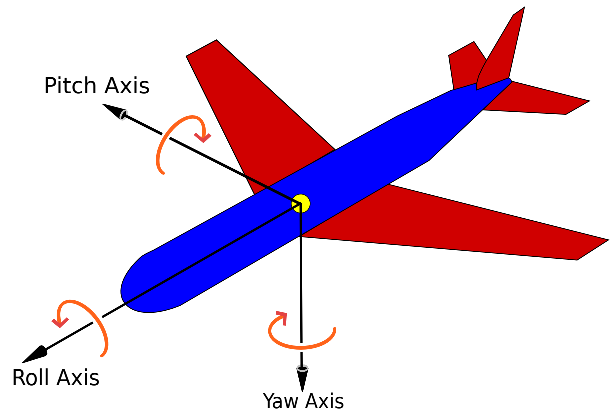


Figure 2: The coordinate frame $\{C\}$ is rotated around origin by an θ from coordinate frame $\{W\}$ and then shifted by translation ${}^w\mathbf{t}_c$.

The matrix that transforms homogeneous coordinates in one coordinate frame to another is called the transformation matrix. For 2D systems it is 3×3 matrix denoted by ${}^wT_c$,

$$\underline{\mathbf{p}}_w = \begin{bmatrix} R(\theta) & {}^w\mathbf{t}_c \\ \mathbf{0}^\top & 1 \end{bmatrix} \underline{\mathbf{p}}_c = {}^wT_c\underline{\mathbf{p}}_c \tag{41}$$

6 Principal 3D Rotations



2D Rotation can be easily extended to rotation around an axis in 3D. Rotation around X-axis, Y-axis, Z-axis is respectively given by,

$$R_x(\phi) = \text{Roll}(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix}$$

$$R_y(\theta) = \text{Pitch}(\theta) = \begin{bmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & & \cos(\phi) \end{bmatrix}$$

$$R_z(\psi) = \text{Yaw}(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{42}$$

7 3D Rotation matrix from Euler angles

Euler angles can be applied sequentially in one of the two ways:

1. Proper Euler angles (z-x-z, x-y-x, y-z-y, z-y-z, x-z-x, y-x-y)
2. Tait-Bryan angles (x-y-z, y-z-x, z-x-y, x-z-y, z-y-x, y-x-z).

One of the most common application of Euler angles is X-Y-Z:

$$R(\phi, \theta, \psi) = R_z(\psi)R_y(\theta)R_x(\phi) = \text{Yaw}(\psi)\text{Pitch}(\theta)\text{Roll}(\phi). \tag{43}$$

Note that the rotation matrix application is read from right to left.

$$R(\phi,\theta,\psi) = \begin{bmatrix} c_\psi & -s_\psi & 0 \\ s_\psi & c_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_\theta & 0 & s_\theta \\ 0 & 1 & 0 \\ -s_\theta & 0 & c_\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\phi & -s_\phi \\ 0 & s_\phi & c_\phi \end{bmatrix}$$
$$= \begin{bmatrix} c_\psi & -s_\psi & 0 \\ s_\psi & c_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_\theta & s_\theta s_\phi & s_\theta c_\phi \\ 0 & c_\phi & -s_\phi \\ -s_\theta & c_\theta s_\phi & c_\theta c_\phi \end{bmatrix}$$
$$= \begin{bmatrix} c_\psi c_\theta & c_\psi s_\theta s_\phi - s_\psi c_\phi & c_\psi s_\theta c_\phi + s_\psi s_\phi \\ s_\psi c_\theta & s_\psi s_\theta s_\phi + c_\psi c_\phi & s_\psi s_\theta c_\phi - c_\psi s_\phi \\ -s_\theta & c_\theta s_\phi & c_\theta c_\phi \end{bmatrix} \quad (44)$$

For a given 3D rotation matrix R, whose elements are r_{ij} as follows

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad (45)$$

the roll, pitch, yaw angles can be read as,

$$\phi = \operatorname{arctan2}(r_{32}, r_{33}) \quad (46)$$
$$\theta = -\arcsin(r_{31}) \quad (47)$$
$$\psi = \operatorname{arctan2}(r_{22}, r_{21}) \quad (48)$$

7.1 Gimbal lock

When pitch $\theta = \frac{\pi}{2}$, then yaw-axis (Z-axis) coincides with roll-axis (X-axis). In such a case, inversion from a rotation matrix leads to infinitely possible solutions, because $c_\theta = 0$ and that leads to $r_{32} = r_{33} = r_{22} = r_{21} = 0$.

7.2 Orthogonality and determinant

Let 3D rotation be represented by a block matrix of 2D rotation $R_2(\phi)$.

$$R_x(\phi) = \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & R_2(\phi) \end{bmatrix}$$

$$R_x^\top(\phi) R_x(\phi) = \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & R_2(\phi) \end{bmatrix}^\top \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & R_2(\phi) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & R_2^\top(\phi) R_2(\phi) \end{bmatrix} = I \quad (49)$$

Same check can be applied to $R_y(\theta)$ and $R_z(\psi)$ as well. If two matrices A and B are orthogonal, then AB is also orthogonal:

$$(AB)^\top (AB) = B^\top A^\top AB = B^\top IB = B^\top B = I. \quad (50)$$

Hence any combination of principal rotations is also orthogonal.

Similar procedure can be followed to establish that $\det(R) = 1$.

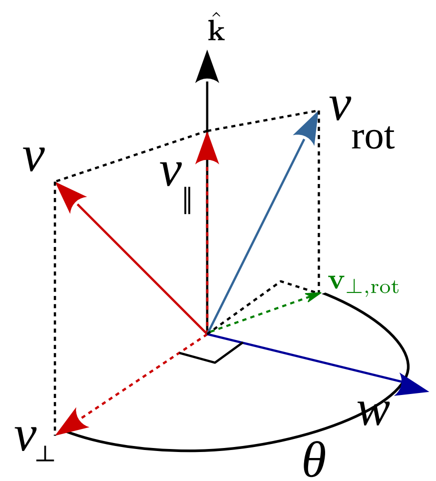
8 3D Transformation matrix

For 3D systems transformation matrix is 4×4 matrix denoted by ${}^wT_c$,

$$\underline{\mathbf{p}}_w = \begin{bmatrix} R(\phi,\theta,\psi) & {}^w\mathbf{t}_c \\ \mathbf{0}^\top & 1 \end{bmatrix} \underline{\mathbf{p}}_c = {}^wT_c \underline{\mathbf{p}}_c, \quad (51)$$

where ${}^w\mathbf{t}_c \in \mathbb{R}^3$, $\underline{\mathbf{p}}_c = \begin{bmatrix} \mathbf{p}_c \\ 1 \end{bmatrix}$ and $\mathbf{p}_c \in \mathbb{R}^3$.

9 Axis-angle representation



$$\mathbf{w} = \hat{\mathbf{k}} \times \mathbf{v} \tag{52}$$

$$\mathbf{v}_\perp = -\hat{\mathbf{k}} \times \mathbf{w} = -\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \mathbf{v}) \tag{53}$$

$$\mathbf{v}_{\perp,\text{rot}} = \mathbf{v}_\perp \cos(\theta) + \mathbf{w}\sin(\theta) \tag{54}$$

$$\begin{aligned}\mathbf{v}_\text{rot} &= \mathbf{v}_\parallel + \mathbf{v}_{\perp,\text{rot}} \\ &= \mathbf{v} - \mathbf{v}_\perp + \mathbf{v}_\perp\cos(\theta) + \mathbf{w}\sin(\theta) \\ &= \mathbf{v} - (1-\cos(\theta))\mathbf{v}_\perp + \mathbf{w}\sin(\theta) \\ &= \mathbf{v} + (1-\cos(\theta))\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \mathbf{v}) + \sin(\theta)\hat{\mathbf{k}} \times \mathbf{v}\end{aligned} \tag{55}$$

Define cross product matrix K of $\hat{\mathbf{k}} = \begin{bmatrix} k_x \\ k_y \\ k_z \end{bmatrix}$ as,

$$K = \begin{bmatrix} 0 & -k_z & k_y \\ k_z & 0 & -k_x \\ -k_y & k_x & 0 \end{bmatrix}. \tag{56}$$

$$\begin{aligned}\mathbf{v}_\text{rot} &= \mathbf{v} + (1-\cos(\theta))\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \mathbf{v}) + \sin(\theta)\hat{\mathbf{k}} \times \mathbf{v} \\ &= (I + (1-\cos(\theta))K^2 + \sin(\theta)K)\mathbf{v}\end{aligned} \tag{57}$$

Thus the rotation matrix corresponding to axis-angle $\theta, \hat{\mathbf{k}}$ is given by,

$$R(\theta, \hat{\mathbf{k}}) = I + \sin(\theta)K + (1-\cos(\theta))K^2 \tag{58}$$

To get back θ and $\hat{\mathbf{k}}$ from R, first note that,

$$K^2 = \begin{bmatrix} -k_z^2 - k_y^2 & k_x k_y & k_z k_x \\ k_x k_y & -k_x^2 - k_z^2 & k_z k_y \\ k_x k_z & k_y k_z & -k_x^2 - k_y^2 \end{bmatrix} \tag{59}$$

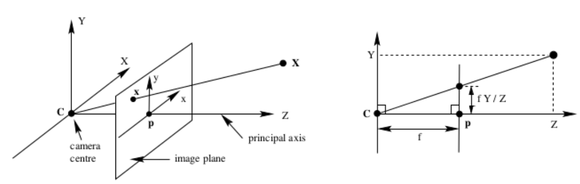
Also we can use trace to separate θ from axis,

$$\begin{aligned}\text{tr}(R) &= \text{tr}(I) + \sin(\theta)\,\text{tr}(K) + (1-\cos(\theta))\,\text{tr}(K^2) \\ &= 3 + 0 + (1-\cos(\theta))(-2(k_x^2 + k_y^2 + k_z^2)). \\ &= 3 - 2 + 2\cos(\theta)\end{aligned} \tag{60}$$

Thus we get $\theta = \arccos(\frac{\text{tr}(R)-1}{2})$. We can estimate axis of rotation as the eigenvector corresponding eigenvalue 1, because $R\hat{\mathbf{k}} = \hat{\mathbf{k}}$.

10 Denavit-Hartenberg transformations

11 Camera projection model



$$\mathbf{K} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{61}$$

$$\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix} \text{ image coordinates in pixels} \tag{62}$$

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \text{ 3D coordinates in world units} \tag{63}$$

$$\underline{\mathbf{u}} = \begin{bmatrix} \mathbf{u} \\ 1 \end{bmatrix} \tag{64}$$

$$\lambda\underline{\mathbf{u}} = K\mathbf{X}, \text{ where } \lambda \neq 0 \tag{65}$$

12 Linear least squares or Pseudo-inverse

Pseudo-inverse of a matrix $\mathbf{A}$ is defined as a matrix $\mathbf{A}^\dagger$, such that$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.

$$\text{if } \mathbf{A} \text{ is tall and full-col rank, then } \mathbf{A}^\dagger = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top \tag{66}$$

$$\text{if } \mathbf{A} \text{ is fat and full-row rank, then } \mathbf{A}^\dagger = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1} \tag{67}$$

[1]

$$\begin{aligned}
&\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 && (68)\\
&= \min_{\mathbf{x}} (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}) && (69)\\
&= \min_{\mathbf{x}} (\mathbf{x}^\top A^\top - \mathbf{b}^\top)(A\mathbf{x} - \mathbf{b}) && (70)\\
&= \min_{\mathbf{x}} (\mathbf{x}^\top A^\top - \mathbf{b}^\top)(A\mathbf{x} - \mathbf{b}) && (71)\\
&= \min_{\mathbf{x}} \mathbf{x}^\top A^\top A\mathbf{x} - \mathbf{b}^\top A\mathbf{x} - \mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b} && (72)
\end{aligned}$$

Recall that a minimum (or maximum) point of a differentiable function $f(\mathbf{x})$, $f'(\mathbf{x}) = 0$. Let us define vector derivative as

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \tag{73}$$

You can verfiy that

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^\top Q\mathbf{x} = 2Q\mathbf{x} \tag{74}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{b}^\top \mathbf{x} = \mathbf{b} \tag{75}$$

At a minimum point $\mathbf{x}$,

$$\frac{\partial}{\partial \mathbf{x}} \left(\mathbf{x}^\top A^\top A\mathbf{x} - \mathbf{b}^\top A\mathbf{x} - \mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}\right) = 0 \tag{76}$$

[1]See Appendix A of Gilbert Strang (1988): Linear Algebra and Its Applications

Note that $\mathbf{b}^\top A\mathbf{x}$ is a scalar, and hence $\mathbf{b}^\top A\mathbf{x} = (\mathbf{b}^\top A\mathbf{x})^\top = \mathbf{x}^\top A^\top \mathbf{b}$.

$$\implies 2A^\top A\mathbf{x} - 2A^\top \mathbf{b} = 0 \tag{77}$$

$$\implies \mathbf{x} = \underbrace{(A^\top A)^{-1} A^\top}_{A^\dagger} \mathbf{b} \tag{78}$$

List of Theorems