RBOT101: Mathematical Foundations of Robotics

Instructor: Dr. Lekan Molu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

PREAMBLE

Consider this the roadmap for this course. Please read through the syllabus posted on Moodle2 carefully and feel free to share any questions that you may have. Please print a copy of the Syllabus for reference. Some relevant parts of the Syllabus are repeated here but the Moodles reference should serve as your guide throughout the ten weeks of this course.

1.1 Course Description

This course focuses on the algorithmic and mathematical concepts with respect classical and recent methods for solving real-world problems in robotics. While some students may have encountered some of the concepts we will be treating in past courses or avenues of study, we will provide the breadth and depth necessary for equipping students to be world-class roboticists. The topics covered by this course shall include the configuration space, rigid bodies, semi-rigid soft bodies, as well as their motions in $\mathbb{R}^n$, wrenches, homogeneous transformations, optimal algorithms for rigid body rotations, linear systems theory, probability theory, the Kalman filter. The course will begin and end with a self-assessment to allow students to gauge their strengths and weaknesses in these topics. References for further, in-depth study in each topic are provided at the end of this course.

1.2 Course Outcomes

After taking this course, each student will be able to

- Develop mathematical tools for solving fundamental kinematic problems in robot operation;
- Formulate optimal state estimation tools for solving real-time smoothing and filtering operations in robotics;

- Integrate state estimation with rigid and semi-rigid soft bodies to solve real-world automation problems; and 4. Use open-source Python, and C++ tools to solve classical and emerging problems in robotics in our day.

1.3 Prerequisites

An undergraduate-level understanding of linear algebra, analytical mechanics, Python and C++ programming.

1.4 Recommended Texts

- Main Texts
  - Simon, Dan. (2007). Optimal state estimation: Kalman, $H - \infty$, and nonlinear approaches. Choice Reviews Online, Vol. 44, pp. 44-3334-44–3334. https://doi.org/10.5860/choice.44-3334
  - Murray, R. M., Li, Z., and Sastry, S. S. (1994). A Mathematical Introduction to Robotic Manipulation. Book (Vol. 29). Free PDF preprint downloadable from, Murray's website.
  - Theory of Screws: A Study in the Dynamics of a Rigid Body by Robert Stawell Ball, Dublin: Hodges, Foster, and Co., Grafton-Street. a. Textbooks:
- Secondary Text
  - Modern Robotics: Mechanics, Planning, and Control. Free PDF preprint downloadable from Author's Northwestern University Website.
- Auxiliary Text:
  - Theory of Screws: A Study in the Dynamics of a Rigid Body by Robert Stawell Ball, Dublin: Hodges, Foster, and Co., Grafton-Street (Should be downloadable via Interlibrary Loan).

1.5 Recommended Journals

- IEEE Transactions on Robotics.
- The International Journal of Robotics Research.
- The IEEE International Conference on Robotics and Automation (ICRA).
- IEEE/Robotics Society of Japan International Conference on Intelligent Robots and Systems (IROS).
- Robotics and Autonomous Systems, An Elsevier Journal.

1.6 Required Software

- A working knowledge of python and the anaconda environment.
- ROS 1.x Installation Instructions: ros 1.x website.
- ROS 2 installation ros 2.0 website.

1.7 Online Course Content

This course will be conducted completely online using Brandeis' LATTE site. The site contains the course syllabus, assignments, our discussion forums, links/resources to course-related professional organizations and sites, and weekly checklists, objectives, outcomes, topic notes, self-tests, and discussion questions. Access information is emailed to enrolled participants before the start of the course. To begin participating in the course, review the “Welcoming Message" and the “Week 1 Checklist."

1.8 Errata

If in the course of using these notes, you find sentence errors, errata or mistakes in equations, please annotate them and upload it to the discussion forum. Points will awarded, at the discretion of the instructor, for such help.

CHAPTER 2

MATRIX ANALYSIS.

Our goal here is to introduce the student to the study of matrix theory. Matrices are symbolism of the important transformations in everyday life; these transformations lie at the heart of mathematics and robotics. The contents of this topic are thus positioned toward the aspiration of roboticists, engineers of all stripes and scientists. Specifically, we are concerned with the *theory of symmetric matrices*, which is important for all fields, *matrices and differential equations*, necessary for engineering and robotics, as well as *positive matrices*, necessary for probability theory. Most of the texts im this chapter are drawn from Richard Bellman's Matrix Analysis Book given in the Syllabus.

2.1 Maximization and Minimization

Of importance to us in this section is to ascertain the range of values of *homogeneous quadratic functions* of two variables and how it is connected to the determination of the maximum or minimum of a general function of two variables.

2.1.1 Maximization of Functions of a Variable

Suppose $f(x)$ is a real function of the real variable x for $x \in [a, b]$, and let us suppose that it is a Taylor series of the form

$$f(x) = f(c) + f'(x-c) + f''\frac{(x-c)^2}{2!} + \dots \tag{2.1.1}$$

around every point in the open interval (a, b). We define a *stationary point* of $f(x)$ to be a point where $f'(x) = 0$ and it is the point that determines if c is a point at which $f(x)$ is a relative maximum, a relative minimum, or a stationary point of a subtle characteristic. If c is a stationary point, we must have

$$f(x) = f(c) + f''\frac{(x-c)^2}{2!} + \dots \tag{2.1.2}$$

If $f''(c) > 0$, then $f(x)$ has a relative minimum at $x = c$. Otherwise, if $f''(c) < 0$, $f(x)$ has a relative maximum at $x = c$. Whereas, if $f''(c) = 0$, we must needs consider further terms in the expansion.

Quiz 1. Suppose that $f''(c) = 0$, what are the sufficient conditions that c must furnish to be a relative minimum?

2.1.2 Maximization of Functions of Two Variables

Now, suppose that we have two variables x, y as arguments of a function f, defined over the rectangle $a_1 \leq x \leq b_1$, $a_2 \leq y \leq b_2$, and possessing a convergent Taylor series around each point (c_1, c_2) within the region. Then, for sufficiently small $|x - c_1$ and $|y - c_2|$, we have

$$\begin{aligned} f(x,y) = f(c_1,c_2) + (x-c_1)\frac{\partial f}{\partial c_1} + (y-c_2)\frac{\partial f}{\partial c_2} + \frac{(x-c_1)^2}{2}\frac{\partial^2 f}{\partial c_1^2} \\ + (x-c_1)(y-c_2)\frac{\partial^2 f}{\partial c_1 \partial c_2} + \frac{(y-c_2)^2}{2}\frac{\partial^2 f}{\partial c_2^2} + \dots \end{aligned} \tag{2.1.3}$$

where

$$\begin{aligned} \frac{\partial f}{\partial c_1} = \frac{\partial f}{\partial x} \text{ at } x = c_1, \qquad y = c_2 \\ \frac{\partial f}{\partial c_2} = \frac{\partial f}{\partial y} \text{ at } x = c_1, \qquad y = c_2 \text{ e.t.c.} \end{aligned} \tag{2.1.4}$$

As before, the stationary point of $f(x,y)$ is defined to be (c_1, c_2) so that $\frac{\partial f}{\partial c_1} = 0$ and $\frac{\partial f}{\partial c_2} = 0$; and the behavior of $f(x,y)$ in the immediate neighborhood of (c_1, c_2) depends on the nature of the quadratic terms in the expansion of (2.1.3),

$$Q_2(x,y) = a(x-c_1)^2 + 2b(x-c_1)(y-c_2) + c(y-c_2)^2 \tag{2.1.5}$$

where $a = \frac{1}{2}\frac{\partial^2 f}{\partial c_1^2}$, $2b = \frac{\partial^2 f}{\partial c_1 \partial c_2}$, and $c = \frac{1}{2}\frac{\partial^2 f}{\partial c_2^2}$.

Suppose we set $x - c_1 = u$ and $y - c_2 = v$, then we can write a quadratic expression in variables u and v $i.e.$

$$Q(u,v) = au^2 + 2buv + cv^2 \tag{2.1.6}$$

whereupon we are interested in the behavior of $Q(u, v)$ in the vicinity of $u = v = 0$ and the fact that $Q(u, v)$ is homogeneous allows us to examine the range of values of $Q(u, v)$ for the set of values on $u^2 + v^2 = 1$.

If $Q(u, v) > 0$ for all u and v distinct from $u = v = 0$, $f(x, y)$ will have a relative minimum at $x = c_1$, $y = c_2$; and if $Q(u, v) < 0$ for all u and v distinct from $u = v = 0$, $f(x, y)$ will have a relative maximum at $x = c_1$, $y = c_2$; The stationary point is a *saddle point* if $Q(u, v)$ can take on both positive and negative values.

2.1.3 Algebraic Approach

How do we determine which of the three situations described in the foregoing occur for any given quadratic form, $au^2 + 2buv + cv^2$, with real coefficients. To determine the sign of $Q(u, v)$, we complete the square in $au^2 + 2buv$ and write $Q(u, v)$ as

$$Q(u, v) = a\left(u + \frac{bv}{a}\right)^2 + \left(c - \frac{b^2}{a}\right)v^2 \tag{2.1.7}$$

provided that $a \neq 0$.

If $a = c = -$, then $Q(u, v) \equiv 2buv$. If $b \neq 0$, then $Q(u, v)$ can be positive or negative. If however, $b = 0$, the quadratic form is eliminated.

If $a \neq 0$, from (2.1.7), we must have a $Q(u, v) > 0$ for all unique u and v different from the pair $(0, 0)$ provided that $a > 0$ and $c - \frac{b^2}{a} > 0$.

In the same vein, $Q(u, v) < 0$ for all nontrivial u and v, provided that we have the inequalities, $a < 0$ and $c - \frac{b^2}{a} < 0$.

Positivity Requirement

A set of *necessary and sufficient* conditions that $Q(u, v)$ be positive for all nontrivial u and v is that

$$a > 0, \qquad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0. \tag{2.1.8}$$

2.1.4 Analytic Approach

To find the range of values of $Q(u,v)$, we can examine the set of values that $Q(u,v)$ occupies on the circle $u^2+v^2=1$. If Q is to be positive for all nontrivial values of u and v, we must have

$$\min_{u^2+v^2=1} Q(u,v) > 0 \tag{2.1.9}$$

and to have $Q(u,v)$ negative for all u and v on the unit circle, we must have

$$\max_{u^2+v^2=1} Q(u,v) < 0. \tag{2.1.10}$$

Introducing a Lagrange multiplier, λ, we can rewrite the problem as

$$R(u,v) = aU^2 + 2buv + cv^2 - \lambda(u^2+v^2). \tag{2.1.11}$$

At the stationary points, we must have $\frac{\partial R}{\partial u} = \frac{\partial R}{\partial v} = 0$ so that

$$\begin{aligned} au + bv - \lambda u &= 0 \\ bu + cv - \lambda v &= 0 \end{aligned} \tag{2.1.12}$$

whereupon, we see that λ satisfies

$$\begin{vmatrix} a-\lambda & b \\ b & c-\lambda \end{vmatrix} = 0 \tag{2.1.13}$$

$$\lambda^2 - (a+c)\lambda + ac - b^2 = 0. \tag{2.1.14}$$

The roots of (2.1.14) are real seeing that the discriminant is non-negative $i.e.$

$$(a+c)^2 - 4\left(ac-b^2\right) = (a-c)^2 + 4b^2, \tag{2.1.15}$$

and as long as $a \neq 0$ and $b \neq 0$, the roots are distinct.

If $b=0$, the roots of the quadratic in (2.1.14) becomes $\lambda_1 = a$, $\lambda_2 = c$. For $\lambda_1 = a$, the linear set of equations from (2.1.12) becomes

$$(a-\lambda_1)\,u = 0 \qquad (c-\lambda_1)\,v = 0 \tag{2.1.16}$$

which leaves u arbitrary and $v = 0$, if $a \neq c$.

Whereas if $b \neq 0$, we obtain the nontrivial solutions of (2.1.12) by using one equation and discarding the other. Therefore, u and v are connected by the relation

$$(a - \lambda_1)\, u = -bv. \tag{2.1.17}$$

For the exact solution, we can add the normalization requirement that $u^2 + v^2 = 1$ so that the values of u and v are

$$\begin{aligned} u_1 &= -b/\left(b^2 + (a-\lambda_1)^2\right)^{1/2} \\ v_1 &= (a-\lambda_1)/\left(b^2 + (a-\lambda_1)^2\right)^{1/2} \end{aligned} \tag{2.1.18}$$

with another set (u_2, v_2) determined in a similar fashion when λ_2 is used in place of λ_1.

CHAPTER 3

VECTORS AND MATRICES

In the previous chapter, we looked into the problem of the minima and maxima (locally) of a function of a single and two variables. Suppose that we have N variables, and proceed in a similar manner as before, we see that finding basic necessary and sufficient conditions that ensure the positivity of a quadratic form of N variables are of the form

$$Q(x_1, x_2, \ldots, x_N) = \sum_{i,j=1}^{N} a_{ij} x_i x_j \tag{3.0.1}$$

We will thus develop a notation that allows us to solve the problem *analytically* using a minimum of arithmetic or analytical calculation. In this light, we will develop a notation that allows us to study linear transformations such as

$$y_i = \sum_{j=1}^{N} a_{ij} x_j \qquad i = 1, 2, \ldots, N \tag{3.0.2}$$

3.1 Vectors

We shall define a set of N complex-valued numbers as a *vector*, written as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{3.1.1}$$

The vector $\boldsymbol{x}$ in (3.1.1) shall be called a *column vector*. If the elements of the vector are stacked horizontally, $i.e.$

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \vdots & x_N, \end{bmatrix} \tag{3.1.2}$$

then we shall call it a *row vector*.

Going forward, we shall use the notation of (3.1.1) to represent all forms of vectors we shall be using. When we mean a row vector, we shall use the notation of a transpose of (3.1.1), $i.e.$ $\boldsymbol{x}^T$. Bold font letters such as $\boldsymbol{x}$, or $\boldsymbol{y}$ shall denote vectors and lower-case letters with subscripts i such as x_i, y_i, z_i or p_i, q_i, r_i shall denote the components of a vector. When discussing a particular set of vectors, we shall use the superscripts $\boldsymbol{x}^1, \boldsymbol{x}^2$ e.t.c. N shall denote the dimension of a vector $\boldsymbol{x}$.

One-dimensional vectors are called *scalars* and shall be our quantities of analysis. When we write $\bar{\boldsymbol{x}}$, we shall mean the vector whose components are the complex conjugates of the elements of $\boldsymbol{x}$.

3.2 Addition of Vectors

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be equal if all of their components, (x_i, y_i) are equal for $i = 1, 2, \ldots, N$. Addition is the simplest of the arithmetic operations on vectors. We shall write the sum of two vectors as $\boldsymbol{x} + \boldsymbol{y}$ so that

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots x_N + y_N, \end{bmatrix} \tag{3.2.1}$$

whereupon we note that the " + " sign connecting $\boldsymbol{x}$ and $\boldsymbol{y}$ is different from the one connecting x_i and y_i.

Homework 1. Prove that we have the *commutativity*, $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$, and the associativity $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$

Homework 2. Just as we showed the addition property of two vectors above, show the subtraction property of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

3.3 Scalar Multiplication

When a vector is multiplied by a scalar, we shall write it out as follows

$$c_1 \boldsymbol{x} = \boldsymbol{x} c_1 = \begin{bmatrix} c_1 x_1 \\ c_1 x_2 \\ c_1 x_N \end{bmatrix} \tag{3.3.1}$$

3.4 The Inner Product of Two Vectors

This is a scalar function of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{N} x_i y_i. \tag{3.4.1}$$

Further to the above, we define the following properties for inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \tag{3.4.2a}$$

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{u} + \boldsymbol{v} \rangle = \langle \boldsymbol{x}, \boldsymbol{u} \rangle + \langle \boldsymbol{x}, \boldsymbol{v} \rangle + \langle \boldsymbol{y}, \boldsymbol{u} \rangle + \langle \boldsymbol{y}, \boldsymbol{v} \rangle \tag{3.4.2b}$$

$$\langle c_1 \boldsymbol{x}, \boldsymbol{y} \rangle = c_1 \langle \boldsymbol{x}, \boldsymbol{y} \rangle \tag{3.4.2c}$$

The above is an easy way to *multiply* two vectors. The inner product is important because $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$ can be considered as the square of the “length" of the real vector $\boldsymbol{x}$.

Homework 3. Prove that $\langle a\boldsymbol{x} + b\boldsymbol{y}, a\boldsymbol{x} + b\boldsymbol{y} \rangle = a^2 \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2ab \langle \boldsymbol{x}, \boldsymbol{x} \rangle + b^2 \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ is a non-negative quadratic form in the scalar variables a and b if $\boldsymbol{x}$ and $\boldsymbol{y}$ are real.

Homework 4. Hence, show that for real-valued vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, that the Cauchy-Schwarz Inequality $\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ holds.

Homework 5. Using the above result, show that for any two complex vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leq \langle \boldsymbol{x}, \bar{\boldsymbol{x}} \rangle \langle \boldsymbol{y}, \bar{\boldsymbol{y}} \rangle$

Homework 6. Show that the *triangle inequality*

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle^{\frac{1}{2}} \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{\frac{1}{2}} + \langle \boldsymbol{y}, \boldsymbol{y} \rangle^{\frac{1}{2}}$$

holds for any two real-valued variables.

3.5 Orthogonality

Two vectors are said to be orthogonal if their inner product is 0 $i.e.$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \tag{3.5.1}$$

When the set of real vectors $\{\boldsymbol{x}^i\}$ possess the property that $\langle \boldsymbol{x}^i, \boldsymbol{y}^i \rangle = 1$, then we say they are *orthonormal.*

Homework 7. show that $\boldsymbol{x}^i$ are mutually orthogonal and normalized $i.e.$ orthonormal for the following N-dimensional Euclidean basis coordinate vectors

$$\boldsymbol{x}^1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \boldsymbol{x}^2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \qquad \boldsymbol{x}^N = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{3.5.2}$$

3.6 Matrices

We can write an array of complex numbers in the form

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NN} \end{bmatrix} \tag{3.6.1}$$

The matrix of (3.6.1) shall be called a *square matrix*. The quantities x_{ij} are the *elements* of the matrix X; the quantities $x_{i1}, x_{i2}, \dots, x_{iN}$ are the ith *rows* of the matrix X and the quantities $x_{1j}, x_{2j}, \dots, x_{Nj}$ are the jth columns of X. We denote matrices with upper case letters or the lower-case subscript notations

$$X = (x_{ij}) \tag{3.6.2}$$

while the *determinant* of the array associated with (3.6.1) shall be denoted $|X|$ or $|x_{ij}$.

Similar to the equality definition between vectors, two matrices are said to be equal if and only if their elements are equal $i.e.$

$$A + B = (a_{ij} + b_{ij}) \tag{3.6.3}$$

Scalar multiplication of a matrix can be expressed as

$$c_1 X = X c_1 = (c_1 x_{ij}) \tag{3.6.4}$$

Lastly, by $\bar{X}$ we shall mean the matrix whose elements are the complex conjugates of X. X is a real matrix if the elements of X are real.

3.7 Vector by Matrix Multiplication

Recall the linear transformation

$$y_i = \sum_{j=1}^{N} a_{ij} x_j \qquad i = 1, 2, \ldots, N \tag{3.7.1}$$

where a_{ij} are complex quantities. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ related as above, we have

$$\boldsymbol{y} = A\boldsymbol{x} \tag{3.7.2}$$

to describe the multiplication of a vector $\boldsymbol{x}$ by a matrix X.

Homework 8. Consider the identity matrix I, so defined

$$I = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} \tag{3.7.3}$$

$i.e.\ I = (\delta_{ij})$, where δ_{ij} is the Kronecker delta symbol, defined as

$$\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \tag{3.7.4}$$

Show that

$$\delta_{ij} = \sum_{k=1}^{N} \delta_{ik}\delta_{kj} \tag{3.7.5}$$

Homework 9. Show that

$$\langle Ax, Ax \rangle = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} a_{ij} x_j \right)^2 \tag{3.7.6}$$

3.8 Matrix by Matrix Multiplication

Consider (3.7.2). Now, suppose our goal is to generate a second-order linear transformation so defined

$$\boldsymbol{z} = B\boldsymbol{y} \tag{3.8.1}$$

which converts the components of $\boldsymbol{y}$ into components of $\boldsymbol{z}$. To express the components of $\boldsymbol{z}$ in terms of the components of $\boldsymbol{x}$ this, we write

$$z_i = \sum_{k=1}^{N} b_{ik} y_k = \sum_{k=1}^{N} b_{ik} \left(\sum_{j=1}^{N} a_{kj} x_j \right) \tag{3.8.2}$$

$$= \sum_{j=1}^{N} \left(\sum_{k=1}^{N} b_{ik} a_{kj} \right) \boldsymbol{x}_j \tag{3.8.3}$$

Introducing $C = (c_{ij})$ defined as

$$c_{ij} = \sum_{k=1}^{N} b_{ik} a_{kj} \qquad i, j = 1, 2, \ldots, N \tag{3.8.4}$$

we may write

$$\boldsymbol{z} = C\boldsymbol{x} \tag{3.8.5}$$

Since, formally

$$\boldsymbol{z} = B\boldsymbol{y} = B(A\boldsymbol{x}) = B(A\boldsymbol{x}) = (BA)\boldsymbol{x} \tag{3.8.6}$$

so that

$$C = BA \tag{3.8.7}$$

Note the ordering of the matrix product above.

Homework 10. Show that

$$f(\theta_1)f(\theta_2) = f(\theta_2)f(\theta_1) = f(\theta_1 + \theta_2) \tag{3.8.8}$$

where

$$f(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.8.9}$$

Homework 11. Show that

$$\left(a_1^2 + a_2^2\right)\left(b_1^2 + b_2^2\right) = \left(a_2 b_1 + a_1 b_2\right)^2 + \left(a_1 b_1 + a_2 b_2\right)^2 \tag{3.8.10}$$

Hint: $|AB| = |A||B|$,

3.9 Non-Commutativity

Matrix multiplication is not commutative, *i.e.* $AB \neq BA$. For an example, consider the following 3×3 matrices

$$A = \begin{bmatrix} 5 & 6 & 9 \\ 2 & 1 & 6 \\ 3 & 6 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 4 & 13 \\ 23 & 6 & 24 \\ 8 & 3 & 9 \end{bmatrix} \tag{3.9.1}$$

where

$$AB = \begin{bmatrix} 215 & 83 & 290 \\ 73 & 32 & 104 \\ 213 & 75 & 264 \end{bmatrix} \qquad \text{and } BA = \begin{bmatrix} 52 & 88 & 150 \\ 199 & 288 & 459 \\ 73 & 105 & 171 \end{bmatrix} \tag{3.9.2}$$

so that $AB \neq BA$. If, however, $AB = BA$, we say A and B *commute*.

3.10 Associativity

Associativity of matrix multiplication gets preserved unlike the commutativity. So for matrices A, B, and C, we have

$$(AB)\,C = A\,(BC) \tag{3.10.1}$$

that is, the product ABC is unambiguously defined without the parentheses. To prove this, we write the ijth element of AB as

$$a_{ik}b_{kj} \tag{3.10.2}$$

so that the definition of multiplication implies that

$$(AB)\,C = [(a_{ik}b_{kl})\,c_{lj}] \tag{3.10.3}$$

$$A\,(BC) = [a_{ik}\,(b_{kl}c_{lj})] \tag{3.10.4}$$

which establishes the equality $(AB)\,C$ and $A\,(BC)$.

3.11 Invariant Vectors

The problem of finding the minimum or maximum of $Q = \sum_{i,j=1}^{N} a_{ij}\boldsymbol{x}_i\boldsymbol{x}_j$ for $\boldsymbol{x}_i$ satisfying the relation $\sum_{i=1}^{N} \boldsymbol{x}_i^2 = 1$ can be reduced to the problem of finding the values of the scalar λ that satisfies the set of linear homogeneous equations

$$\sum_{j=1}^{N} a_{ij}\boldsymbol{x}_j = \lambda\boldsymbol{x}_i, \qquad i = 1, 2, \ldots, N \tag{3.11.1}$$

which possesses nontrivial solutions. Vectorizing, we have

$$A\boldsymbol{x} = \lambda\boldsymbol{x} \tag{3.11.2}$$

Here, $\boldsymbol{x}$ signifies the direction indicated by the N direction numbers $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, and we are searching for the directions that are invariant.

3.12 The Matrix Transpose

We define the transpose of the matrix $A = (a_{ij})$ as $A^T = (a_{ji})$ $i.e.$ the rows of A^T are the columns of A and vice versa. An important consequence of this is that the transformation A on the set of a vector $\boldsymbol{x}$ is same as the transformation of the matrix A^T on the set $\boldsymbol{y}$. This is shown in the following

$$\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = y_1 \sum_{j=1}^{N} a_{1j}\boldsymbol{x}_j + y_2 \sum_{j=1}^{N} a_{2j}\boldsymbol{x}_j + \ldots + y_N \sum_{j=1}^{N} a_{Nj}\boldsymbol{x}_j \tag{3.12.1}$$

which becomes upon rearrangement,

$$\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = x_1 \sum_{i=1}^{N} a_{i1}\boldsymbol{y}_i + x_2 \sum_{i=1}^{N} a_{i2}\boldsymbol{y}_i + \ldots + x_N \sum_{i=1}^{N} a_{iN}\boldsymbol{y}_i \tag{3.12.2}$$

$$= \langle \boldsymbol{x}, A^T\boldsymbol{y}\rangle \tag{3.12.3}$$

We can then regard A^T as the *induced or adjoint transformation* of A.

3.13 Symmetric Matrices

Matrices that satisfy the relation

$$A = A^T \tag{3.13.1}$$

play a crucial role in the study of quadratic forms and such matrices are said to be *symmetric*, with the property that

$$a_{ij} = a_{ji} \tag{3.13.2}$$

Homework 12. Prove that $(A^T)^T = A$

Homework 13. Prove that $\langle A\boldsymbol{x}, B\boldsymbol{y}\rangle = \langle \boldsymbol{x}, A^T B\boldsymbol{y}\rangle$

3.14 Hermitian Matrices

The scalar function for complex vectors is the expression $\langle \boldsymbol{x}, \bar{\boldsymbol{y}} \rangle$. Suppose we define $\boldsymbol{z} = \bar{A}^T \boldsymbol{y}$, then

$$\langle A\boldsymbol{x}, \bar{y} \rangle = \langle \boldsymbol{x}, \bar{\boldsymbol{z}} \rangle \tag{3.14.1}$$

$i.e.$ the induced transformation is now $\bar{A}^T$, the complex conjugate of A. Matrices for which

$$A = \bar{A}^T \tag{3.14.2}$$

are called Hermitian. Note that in some literature, the Hermitian matrix is often written as $A^\star$.

3.15 Orthogonal Matrices

This section has to do with the invariance of distance between matrices, that is, taking the Euclidean measure of distance as the measure of the magnitude of the real-valued vector $\boldsymbol{x}$. The prodding question of interest is to figure out the linear transformation $\boldsymbol{y} = H\boldsymbol{x}$ that leaves the inner product $\langle \boldsymbol{x}, \boldsymbol{z} \rangle$. Mathematically, we express this problem such that

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle H\boldsymbol{x}, H\boldsymbol{x} \rangle \tag{3.15.1}$$

is satisfied for *all* $\boldsymbol{x}$. We know that

$$\langle H\boldsymbol{x}, H\boldsymbol{x} \rangle = \langle \boldsymbol{x}, H^T H\boldsymbol{x} \rangle \tag{3.15.2}$$

and that $H^T H$ is symmetric so that (3.15.1), gives

$$H^T H = I. \tag{3.15.3}$$

Orthogonal Matrix

A real matrix H for which $H^T H = I$ is called *orthogonal*.

3.16 Unitary Matrices

This is the measure of the distance of a complex vector, akin to the invariance condition of real-valued matrices (3.15.3). We define the unitary property as follows:

$$H^{\star}H = I. \tag{3.16.1}$$

Matrices defined as in the foregoing play a crucial role in the treatment of Hernitian matrices, such as the role that orthogonal matrices play in symmetric matrices theory.

REFERENCES