1981

MODERN ALGEBRA

SECTION B - LINEAR ALGEBRA

(Lecture 21 Onward)

? omit?

# SECTION B

## LINEAR ALGEBRA

TEXT: Florey, Francis G. *"Elementary Linear Algebra with Applications"* Prentice-Hall, 1979.

The course consists largely of selected reading from Chapters 5, 6, 7 and 8 of the text. It is therefore essential that you obtain a copy as soon as possible. Additional notes and some remarks which supplement material in the text are included in the *course description.*

Linear Algebra is the study of "vector (i.e. linear) spaces". The concept of a vector space is an abstract one. Vector space is not a 'concrete' object, though particular examples are. Anything which can be made to satisfy the axioms <u>is</u> a vector space. The importance of the concept lies in the extremely large number of "mathematical objects" which exhibit the structure of a vector space (some of these objects are listed in the examples given in section 5.1 p. 163 - 169 of the text). By recognising these objects as vector spaces we can, in one sense, think of them as being the same. Since ordinary two or three dimensional euclidean space is one such example, when we attempt to prove a result for some less familar example we can see what the problem amounts to in ordinary space, use our "geometrical intuition" to find a proof in that case, then, provided our proof only relied upon the vector space structure of euclidean space, translate it back to obtain the result sought. Further, once a result is proved for general vector spaces we then know it is true in <u>all</u> the examples. Thus, once we have accumulated a body of results for vector spaces, showing something is a vector space amounts to immediately knowing a lot about it.

The applications of linear algebra are numerous. A few are treated briefly in the text. Unfortunately there is insufficient time to include these sections in the course. I hope that you will none-the-less inspect them and perhaps make the effort to work through one or two that particularly interest you.

As with any body of 'formal' mathematics, the theory of vector spaces develops upon itself. Subsequent work rests heavily upon earlier definitions and results. It is therefore essential that you learn definitions and results as they occur (if necessary write them on the ceiling above your bed and recite them each night). While it is important that you strive for understanding, as opposed to learning parrot fashion, initially you should not waste too much time worrying about what a definition or result means. A feel for this should develop as you progress through the course. Most students find the early part of the course difficult. It is a different kind of mathematics than you are probably used to. However with perserverence most students find that it eventually fits together and really isn't all that hard.

Working problems is probably the best way to develop a feel for the material. Problems form an integral part of the course. Each of the sections prescribed from the text is concluded by a set of exercises. You should attempt as many of these as possible as soon as they are encountered. Assignment questions are drawn largely from among these exercises. Do not let problems accumulate until the end. Also don't be discouraged if you are unable to do some of them. All mathematicians at all levels, experience this. Be on the look out for ways of using previous results to simplify and assist in the proofs of current problems. The proofs of theorems, and lemmas are important here too. Earlier proofs frequently contain the essential techniques for solving a current problem. You should treat the proofs of theorems etc. much the same way as you would a worked example.

Requests for additional explanations or assistance are welcomed and will be dealt with as promptly as possible. May I wish you success and enjoyment from your studies.

Dr. Brailey Sims
(Lecturer)

NOTATION: Some of the symbols available to the printer are not present on typewriters or are difficult to reproduce by hand. It is therefore necessary to adopt a sligthly different notation in written work, typed notes, exam papers etc. than the one employed in the text. As you will see from the following table I have tried to choose as similar a notation as possible to that of the text. I hope it will not cause you any inconvenience.

The handwritten notation is merely a suggestion. Provided it is clear what you mean I don't mind what notations or style you choose to use in assignments, exams, etc, so find one with which you feel comfortable.

As the notation in the right hand column suggests, different authors employ radically different notations. The alternative one given is perhaps the most commonly used notation. (Indeed, it is the one I normally use. While I will endeavour to parallel the text as nearly as possible, please forgive me for any lapses into this alternative notation.)

| EXPLANATION | FLOREY | TYPEWRITTEN EQUIVALENT AS WILL BE USED IN NOTES, EXAMS ETC | SUGGESTED HANDWRITTEN EQUIVALENT | A FREQUENTLY USED ALTERNATIVE NOTATION OF WHICH YOU SHOULD BE AWARE |
|---|---|---|---|---|
| the field of real numbers | $\mathbf{R}$ | $\mathrm{R}$ | $\mathcal{R}$ | $\mathbb{R}$ |
| elements of the scalar field (Real numbers) | a,b,...,x,y,z | a,b,...,x,y,z | a, b, ..., x, y, z | $\alpha, \beta, \ldots, \omega$ |
| the set of ordered n-tuples of real numbers | $\mathbf{R}_n$ | $\mathbf{R}_n$ | $\mathcal{R}_n$ | $\mathbb{R}^n$ |
| a function (mapping, transformation) from the set A (the domain of f) to the set B (the co-domain or target set of f) | $f: A \to B$ | $f: A \to B$ | $f: A \to B$ | $f: A \to B$ |
| the function f for which the image of each point x of the domain is specified by the expression $E(x)$. | $f: E(x)$ or $x \overset{f}{\to} E(x)$ | $f: E(x)$ or $x \overset{f}{\to} E(x)$ or $f: x \mapsto E(x)$ | $f: \mathcal{E}(x)$ or $x \xrightarrow{f} \mathcal{E}(x)$ or $f: x \longmapsto \mathcal{E}(x)$ | $f: x \mapsto E(x)$ |
| a composite of the previous two notations | | $f: A \to B: x \mapsto E(x)$ | $f: A \to B: x \mapsto \mathcal{E}(x)$ | $f: A \to B: x \mapsto E(x)$ |
| vector spaces | **V, U, W** | V, U, W | $\mathcal{V}, \mathcal{U}, \mathcal{W}$, | V, U, W |
| a more complete notation for a vector space reminding us of the presence of the operations of (vector) addition and scalar multiplication | ( **V**, +, sm) | (V, +, sm) | $(\mathcal{V}, +, sm)$ | (V, +, ·) |
| vectors (elements of a vector space) | X, Y, Z | X, Y, Z | X, Y, Z | $\underset{\sim}{x}, \underset{\sim}{y}, \ldots$ or $\mathbf{x}, \mathbf{y}, \ldots$ (in printed works). |

| | | | | |
|---|---|---|---|---|
| the zero vector in any vector space | $\mathbf{0}$ | $0$ | $\mathit{0}$ | $\underset{\sim}{0}$ or $\mathbf{0}$ |
| the additive inverse of the vector X (equal to -1X) | $-X$ | $-X$ | $-X$ | $-\underset{\sim}{x}$ or $-\mathbf{x}$ |
| the subspace generated (spanned) by the vectors $X_1, X_2, \ldots$ (sometimes referred to as the "span" of $X_1, X_2, \ldots$) | $\langle X_1, X_2, \ldots\rangle$ | $\langle X_1, X_2, \ldots\rangle$ | $\langle X_1, X_2, \ldots\rangle$ | $\langle \underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots\rangle$ or span $\{\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots\}$ or linear hull $\{\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots\}$ |

BACKGROUND for the course

The material in Chapters 1 to 4 of the text provides essential background material to that of the course proper.

For example:

The material of Chapter 1, sections 1.1 and 1.3, is important in motivating the fundamental ideas developed in Chapter 5;

Section 2.2 provides basic motivation for the material of Chapter 8;

Chapter 3 (in particular the matrix work - sections 3.4, 3.5, 3.6, 3.7 and 3.8) is basic to the work in Chapters 6 and 7;

For Chapter 7 the material of Chapter 4, sections 4.1 and 4.2, is also essential.

With the exception of some material in sections 4.1 and 4.2 the content of all these early chapters should be familiar to you from your First Semester work in Pure Mathematics 111-22. It is essential that you revise this preliminary material, however, it is not necessary to complete this revision before starting the course proper. Rather I would suggest that you allocate some time each week to the revision of this material, pacing your revision so that it keeps ahead of where it is needed in the course proper - as detailed above. To this end I will include some questions on this early material in the assignments.

You should aim to achieve understanding and computational skill from this revision, a detailed knowledge of proofs for the material of Chapters 1 to 4 is not essential.

## COURSE DETAILS

(Background preparation: Briefly read through Chapter 1 sections 1.1 and 1.3 - the material should be familiar to you from First Year, the only difference in notation is the use of $||A||$ instead of $|A|$ for the "length" (Magnitude) of the vector A.)

| | |
|---|---|
| Lecture* 1, Lecture 2 | section 5.1 |
| Lecture 3 | section 5.2 |
| Lecture 4, Lecture 5 | section 5.3 including the proof of theorem 5.6 |
| Lecture 6, Lecture 7 | section 5.4 |
| Lecture 8 | Section 5.5 |

Before proceeding to the next section of work you should "revise" the material in section 6.1 (pp. 206-210).

| | |
|---|---|
| Lecture 9, Lecture 10 | Section 6.2 - paying particular attention to the definition 6.2 and examples 6.4 and 6.7.<br>Section 6.3 (pp. 216-219) - excluding the applications. |
| Lecture 11 | Section 6.4 |
| Lecture 12 | Section 6.5 - including the proof of Theorem 6.10. |
| Lecture 13, Lecture 14 | Section 6.6. (Note at this point you may find it necessary to revise the work of Chapter 3, Sections 3.4, 3.5 and 3.6, on Matrix Operations.) |

* Throughout the course you should aim to cover approximately two lectures a week.

Interpretation of some results in Section 6.6

Note that in the verification (p. 165) that $\mathsf{F}$, the set of all real valued functions defined on $R$ (Example 5.1(e)), is a vector space under point-wise defined addition and scalar multiplication the fact that the common domain of all the functions is $R$ was not needed, indeed the domain could have been any set $A$ and the arguments would have established that $\mathsf{F}(A,R)$ the set of all real valued functions defined on $A$ is a vector space. Further it was only the vector space properties of $R$ which were needed and so the same arguments in fact prove that $\mathsf{F}(A,\mathsf{W})$ is a vector space where $A$ is any arbitrary set and $\mathsf{W}$ is any vector space. In particular *if* $\mathsf{V}$, $\mathsf{W}$ *are two vector spaces, then* $\mathsf{F}(\mathsf{V},\mathsf{W})$ *the set of all functions (transformations) from* $\mathsf{V}$ *to* $\mathsf{W}$ *is a vector space with respect to the point-wise defined operations.*

$$\left.\begin{array}{ll}\text{addition:} & (f+g)(v) = f(v)+g(v) \\ \text{scalar multiplication:} & (rf)(v) = r(f(v))\end{array}\right\} \begin{array}{l} f,\ g \in \mathsf{F}(\mathsf{V},\mathsf{W}) \\ v \in \mathsf{V} \\ r \in R\end{array}$$

If we denote by $\mathsf{L}(\mathsf{V},\mathsf{W})$ the subset of $\mathsf{F}(\mathsf{V},\mathsf{W})$ consisting of all linear transformations from $\mathsf{V}$ to $\mathsf{W}$, then exercise 6.6, problem 8 establishes that $\mathsf{L}(\mathsf{V},\mathsf{W})$ is a subspace of $\mathsf{F}(\mathsf{V},\mathsf{W})$. Thus, *the set of all linear transformations from one vector space into another, is itself a vector space with respect to point-wise defined addition and scalar-multiplication.*

(qv. exercises 6.6, problem 11)

Indeed, when $\mathsf{V}$ and $\mathsf{W}$ are finite dimensional, theorem 6.11 may be interpreted as establishing an isomorphism between $\mathsf{L}(\mathsf{V},\mathsf{W})$ and the space of $\dim \mathsf{W} \times \dim \mathsf{V}$ - matrices (example 5.1(d) of p.164) - see exercises 6.6, problem 13.

Lecture 15 Section 6.7 - note that in the proof of theorem 6.16 the assumption that $\mathsf{V}$ is finite dimensional is not used, thus the theorem is valid for any vector space, $\mathsf{V}$, finite or infinite dimensional. It is only in the Corollary 6.2 that finite dimensionality is needed [indeed it is essential, as the conclusions of problems 16 and 17 of exercise 6.4 do not, in general, hold true for infinite dimensional spaces. (Optional: Can you give examples which demonstrate this?)].

To the conclusions of Corollary 6.2 you should add

(iv) T is non-singular if and only if there exists $S: V \to V$ such that $T \circ S = I$.

(v) T is non-singular if and only if there exists $S: V \to V$ such that $S \circ T = I$.

That is, for <u>finite dimensional</u> $V$, T is invertible if and only if it has either a right or left inverse. [Show that in case either (iv) or (v) applies, then the right or left inverse, S, is in fact $T^{-1}$ the unique inverse of T.]

To prove (iv), (v). Note that from the appropriate argument in the proof of theorem 6.16, T is onto (1-1) if T has a right (left) inverse and then apply (i), (iii) of Corollary 6.2.

[At this point you could look at section 6.8, however the question of matrix rank will be taken up again after our work on inner-products and a slightly different approach adopted. The study of this section could therefore be deferred till then.

Note: Only definition 6.11 and the first five lines of the proof to theorem 6.20, which establish: *The column rank of* A *equals the rank of* $T: R_n \to R_m$, *where* T *is the linear mapping represented by* A *relative to the standard bases;* are relevant to the course, the remaining results will be established differently.]

Lecture 16 Section 6.9

For the next two sections it will be necessary for you to have revised the material of Chapter 4, sections 4.1 and 4.2.

Lecture 17 Section 7.1

Lecture 18 Section 7.2 (excluding the applications on pp. 298-301).

[Note: The work on quadratic forms in sections 7.3 and 7.4 will be considered after we have made a study of inner-product spaces.]

Background: Read Chapter 2, Section 2.2.

Lecture 19 Section 8.1

## ASSIGNMENT 2 – Linear Algebra

1) Show that the set of strictly positive real numbers, $\mathbb{R}^+$, with the binary operation, "addition" defined by

$$X \text{ "+" } Y = XY$$

and scalar multiplication defined by

$$rX = X^r$$

is a vector space.

2) Exercise 5.1 problems 12, 13, 14 and 15 (pp. 168, 169).

3) Exercise 5.1 problems 19, 20, 21 and 22 (p. 169)

4) Exercise 5.2 problems 3, 8 and 14. (pp. 175, 176).

5) Exercise 5.2 problems 10 and 11 (p. 176).

6) Exercise 5.2 problems 18, 19 and 20 (p. 176).

7) Exercise 5.3 problems 2 and 3 (p. 183).

8) Exercise 5.3 problem 11 (p. 184).

9) Exercise 5.3 problems 13 and 15 (p. 184).

10) Exercise 5.4 problem 2.

# ASSIGNMENT 3 - Linear Algebra

(due 14th June)

11. Exercise 5.4, problems 14 (note remark immediately preceding this problem), 17 and 18 (pp. 191 and 192).
12. Exercise 5.4, problem 20 (p.192).
13. (i) Exercise 5.5, problem 10 (p.198).
    (ii) Is it possible to find two subspaces $\mathbb{U}$ and $\mathbb{W}$ of $R_3$ such that $\dim \mathbb{U} = \dim \mathbb{W}$ and $R_3 = \mathbb{U} \oplus \mathbb{W}$?
14. Exercise 5.5, problems 20, 17, 19 and 21 (pp. 198 and 199)
15. (a) Exercise 6.2, problems 2, 3, 7 [in this problem also verify that every linear transformation $T: R \to R$ is of this form. Hint: Let $k = T(1)$] and 8 (p.215).
    (b) Exercise 6.2, problems 10, 11 and 19 (pp. 215 and 216).
    Verify each of the following functions (transformations) is a linear transformation.

    (i) $T: \mathbb{V} \to \mathbb{F}([a,b],R): f \mapsto \int_a^x f(t)dt,$

    where $\mathbb{V}$ is the space of all integrable functions on $[a,b]$ and $\mathbb{F}([a,b],R)$ is the space of all real valued functions on $[a,b]$.
    [See examples 5.1(i) and (k).]

    (ii) $T: \mathbb{V} \to \mathbb{F}: \quad f \mapsto a(x)f'' + b(x)f' + c(x)f,$ where $a$, $b$ and $c$ are given functions, $\mathbb{V}$ is the space of all twice differentiable functions on $R$ (see example 5.1(j) for example) and $\mathbb{F}$ is defined in example 5.1(e).
    [Remark: Note that the second order differential equation $ay'' + by' + cy = g(x)$ may be written as $T(y) = g$, it is for this reason that such equations are referred to as <u>linear</u> equations.]

    (iii) $L: \mathbb{V} \to \mathbb{F}([0,\infty), R)$ where $L(f)$ is the Laplace Transform of $f$, $L(f)(s) = \int_0^\infty e^{-sx} f(x)dx$. Here $\mathbb{F}([0,\infty),R)$ is the space of all real valued functions defined on the set of positive reals and $\mathbb{V}$ is the space of all real valued functions $f$ such that $\int_0^\infty e^{-sx} f(x)dx$ exists and is finite for all $s \geqslant 0$.

15. (cont.) (iv) $C: V \to F([-\frac{\pi}{2}, \frac{\pi}{2}], R)$ where $C(f)$ is the convolution,

$$C(f)(x) = (\sin * f)(x) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(x - t)f(t)dt.$$

Here, $V$ is the space of all integrable functions on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

[*Note: The conclusion of (iv) would remain valid if* $C$ *were defined with* $\sin(x - t)$ *replaced by any continuous function of two variables* $K(x,t)$. *Such transformations (operators) are basic to the theory of integral equations and arise frequently in problems of Mathematical Physics.*]

(c) Exercise 6.2, problem 20 (p.216), Exercise 6.3, problem 15, 16 (p.224); 7 (p.223) and 19 (p.224).

## ASSIGNMENT 4 - LINEAR ALGEBRA

16. Exercise 6.4, problems 8, 9, 10, 12, 13, 14, 16 and 17 (p.230).

17. (a) Exercise 6.5, problems 4, 7, 11 and 16 (pp. 239-242).

(b) Let $T: P_3 \to P_4$ be defined by

$$T(p)(x) = \int_0^x p(t)dt;$$

that is, $T$ maps each quadratic to a primitive (indefinite integral). For example, $T(3x^2 + x + 1) = x^3 + \frac{1}{2}x^2 + x$.

Find the matrix $[T]_{B_1B_2}$ which represents $T$ relative to the natural bases

$$B_1 = \{1, x, x^2\} \text{ of } P_3$$

and

$$B_2 = \{1, x, x^2, x^3\} \text{ of } P_4.$$

18. Exercise 6.6, problems 2, 4, 8, 9, 12 and 13 [also observe that $(L, +, \circ)$ is a ring, keeping this in mind, may help in some of the subsequent problems] (pp. 250 and 251).

19. Exercise 6.7, problems 1, 2, 3, 4, 5, 8, 12, 14 and 15 (pp. 257 and 258). [Note: Problems 16, 17 and 18 relate to the additional notes included in the course outline for lecture 15 and so may be of interest to you.]

20. Exercise 6.9, problems 7, 12, 15 and 17 (pp. 275-277).

## ASSIGNMENT 5 (Linear Algebra)

21. Exercise 7.1 problems 6, 11 and 12 ( p.292 )

22. Exercise 7.2 problems 8, 10, 16, 17 (pp.302, 303)

23. Exercise 8.1 problems 7, 8, 12, 13, 15 (p.340)

    Also; show that

    $\|(x_1,x_2)\| = |x_1| + |x_2|$

    defines a norm on $R_2$.

    By considering the vectors $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},-\frac{1}{2})$ show that the above norm does not satisfy the parallelogram rule (see F. Problem 12 p.340) and so conclude that this norm does <u>not</u> arise as $\sqrt{(X,X)}$ for <u>any</u> inner-product on $R_2$. Thus, *the notion of a normed space is more general than that of a Euclidean space.*

24. Exercise 8.2 problems 7, 9, 10, 11, 14 (P.348). Also prove the converse of 7 - that is, if A,B are vectors in the (real) innerproduct space V, ( , ) such that $\|A - B\|^2 = \|A\|^2 + \|B\|^2$ , then A and B are orthogonal.

    *Question 5-7 refer to the supplementary notes and not to Florey.*

25. Prove Theorem 21.1, Exercise 21.1

## ASSIGNMENT 6 (Linear Algebra)

26. Let $T : R_3 \to R_2$ be defined by $T(x_1,x_2,x_3) = (x_1+2x_2+x_3, \frac{1}{2}x_2+\frac{1}{2}x_3)$.

    Find the matrix M which represents T with respect to the standard bases for $R_3$ and $R_2$. Determine the adjoint M* and hence T*. (Verify your results directly from the definition of T*).

    Exercise 22.1

27. The exercises on p.39 and p.41

    Also verify by direct calculation (i.e., without referring to theorem 25.2) that, if $\lambda_1,\lambda_2$ are two distinct eigenvalues of the self-adjoint transformation T, then corresponding eigenvectors $X_1$, $X_2$ are orthogonal.

    [Hint: consider $(\lambda_1 - \lambda_2)(X_1,X_2)$.]

28. Exercise 7.3 problems 3, 7 and 11 (pp.310, 311)

29. Exercise 7.4 problems 11, 15, 19 and 22 (pp.319, 320)

30. Exercise 7.5 problems 2, 8 and 13 (pp.329, 330). Also:

If the n×n matrix A has a complete set of eigenvectors $X_1, X_2, \ldots, X_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (not necessarily distinct), prove that

$$\prod_{i=1}^{n} (A-\lambda_i) \equiv (A-\lambda_1)(A-\lambda_2)\ldots(A-\lambda_n) = 0$$

[Hint: for any i,j show that $(A-\lambda_i)(A-\lambda_j)=(A-\lambda_j)(A-\lambda_i)$ and so show that $\left[\prod_{i=1}^{n} (A-\lambda_i)\right] (X) = 0$ for any $X \in R^n$].

Note that; under the assumptions on A we have, A is similar to the diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 & \ldots \\ 0 & \lambda_2 & \ldots \\ \ldots & & \\ \ldots & & \lambda_n \end{bmatrix}$$

and so the characteristic polynomial of A is

$$\det (A-\lambda I) = \prod_{i=1}^{n} (\lambda_i - \lambda)$$

The above result is therefore a special case of the Hamilton - Cayley Theorem which is sufficiently general to encompass any self-adjoint, and therefore any symmetric, matrix A.

31. (Optional) - *Simultaneous diagonalization of two quadratic forms.*

(this result is of use in the advanced theory of vibrations and certain aspects of algebraic geometry, where you may meet it in third year).

Let $P(X) = X^T AX$ and $Q(X) = X^T BX$ be two quadratic forms on $R_n$ and further assume that $P$ is positive definite, that is the (real) symmetric matrix A has all its eigenvalues strictly positive.

Then by Florey Section 7.4, there exists exists a matrix P s.t. $P^{-1} = P^T$ and

$$P^T AP = \begin{bmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \cdots \\ \cdot\ \cdot & & \\ \cdot\ \cdot\ \cdot & & \cdots \\ \cdots & & \lambda_n \end{bmatrix} \quad \text{is a diagonal matrix}$$

with $\lambda_1, \lambda_2, \ldots, \lambda_n$ n strictly positive real numbers (not necessarily all distinct).

$$\text{Let } R = P \begin{bmatrix} 1/\sqrt{\lambda_1} & 0 & \cdots \\ 0 & 1/\sqrt{\lambda_2} & \cdots \\ \cdots & \cdots & \\ \cdots & & 1/\sqrt{\lambda_n} \end{bmatrix},$$

Then $R^T AR = I_n$, the n×n identity matrix (<u>show this</u>) and Further $R^T BR$ is still a (real) symmetric matrix, since B is (<u>verify this</u>) and so again, by Florey section 7.4, there exists a real symmetric matrix $Q$ such that $Q^{-1} = Q^T$ and $Q^T(R^T BR)Q$ is a diagonal matrix.

Now let $S = RQ$ then

$S^T BS$ is diagonal (why?)

and

$$S^T AS = Q^T R^T ARQ = Q^T I_n Q = I_n \quad \text{(Why?)}$$

That is, both A and B are simultaneously diagonalized by the matrix S.

Relative to the "new coordinates"

$U = S^{-1}X$ (<u>how do we know $S^{-1}$ exists?</u>) we therefore have that both the quadratic forms $P$ and $Q$ are in diagonal (or <u>normal</u>) form.

Note: $U = [X]_{B_2}$ - the coordinates of X relative to the basis $B_2 = \{SX_1, SX_2, \ldots, SX_n\}$ where $B_1 = \{X_1, X_2, \ldots, X_n\}$ is the standard basis for $R_n$.

By following the above steps simultaneously diagonalize the quadratic forms on $R_2$ which corespond to

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \qquad\qquad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

In what follows it is convenient to replace the real numbers $\mathbb{R}$ by the complex numbers $\mathbb{C}$ as the field of scalars over which our vector space is defined. (The eigenvalues of a linear transformation of a finite dimensional space are the roots of a polynomial - the characteristic polynomial - and so may be complex numbers).

The will not alter any of the basic theorems on linear algebras, as only the <u>field</u> properties of $\mathbb{R}$ have been used.

However, in the definition of inner-product we must replace 1P by

$$(Y,X) = \overline{(X,Y)} \quad \text{for all } X \text{ and } Y \text{ in } \mathbb{V}$$

(here, and elsewhere, " $\overline{\phantom{x}}$ " denotes complex conjugate).

Consequently we have

$(\lambda X,Y) = \lambda(X,Y)$ for all X, Y in $\mathbb{V}$ and $\lambda$ in $\mathbb{C}$,

but

$(X,\lambda Y) = \bar{\lambda}(X,Y)$. (Verify this.)

All the other properties of an inner-product remain unchanged.

[If this modification were not made, for any $X \neq 0$ in $\mathbb{V}$ we would have

$$\begin{aligned} 0 &< (iX,iX) \quad \text{(by 3P)} \\ &= i(X,iX) \\ &= (i)^2(X,X) \\ &= -(X,X) \\ &< 0 \quad \text{(as, again by 3P, } (X,X) > 0) \end{aligned}$$

and so our axioms would be inconsistent.]

The following may serve as a heuristic motivation for the definition of an inner-product in the space of all continuous real (or complex valued) functions on the interval $[a,b]$, $V = C[a,b]$:

$$(f,g) \;=\; \int_a^b f(x)\,\overline{g(x)}\,dx$$

(qv. Florey Example 8.1, pp. 333-334.)

A clue as to how an inner product may be defined in V may be found by examining the definition of the inner-product in the finite case.

For $\underset{\sim}{x} \;=\; (x_1, x_2, x_3, \ldots, x_n)$

$\underset{\sim}{y} \;=\; (y_1, y_2, y_3, \ldots, y_n)$ ($x_j$, $y_j$ complex numbers)

$$(x,y) \overset{\text{defn}}{=\!=} x_1\bar{y}_1 + x_2\bar{y}_2 + x_3\bar{y}_3 + \ldots + x_n\bar{y}_n$$

Now, the vector $\underset{\sim}{x}$ may be regarded as the "baby" function, whose domain $\{1,2,3,\ldots,n\}$ is mapped into the complex numbers according to

$$\begin{aligned} 1 &\longmapsto x_1 \\ 2 &\longmapsto x_2 \\ 3 &\longmapsto x_3 \\ &\;\vdots \\ n &\longmapsto x_n \end{aligned}$$

which we summarize by writing $\underset{\sim}{x}$: $j \to x_j$. Similarly y is equivalent to the function $\underset{\sim}{y}$: $j \to y_j$.

From this point of view the individual terms, $x_1y_1$, $x_2y_2$, ..., of the inner-product $\underset{\sim}{x}.\underset{\sim}{y}$ represent the value of $\underset{\sim}{x}$ at each domain point j ($x_j$) multiplied by the conjugate of the value of $\underset{\sim}{y}$ at the same domain point. The inner-product is obtained by summing these products over all possible domain points.

In the case of interest to us, our vectors are again complex valued functions defined on the domain $[a,b]$. For f, g $\in$ V the corresponding point value product is $f(x)\,\overline{g(x)}$ and the "sum" over all such values corresponds to

$$\int_a^b f(x)\,\overline{g(x)}\,dx.$$

We are therefore led to define an <u>inner product</u> in V by

$$(f,g) \;=\; \int_a^b f(x)\,\overline{g(x)}\,dx \quad \text{for all } f,\ g \in V.$$

These notes further the study of Euclidean (or inner-product) spaces initiated in Chapter 8 of Florey.

## Lecture 21 - ORTHOGONAL COMPLEMENTS.

If X,Y are orthogonal vectors in the Euclidean space $V$, ( , ); that is, $(X,Y) = 0$, then we will write $X \perp Y$. Now let $S$ be any non-empty subset of $V$. The orthogonal complement of $S$, denoted by $S^{\perp}$ is the set of vectors in $V$ which are orthogonal to every element of $S$. That is

$$S^{\perp} = \{X \in V : (X,S) = 0 \text{ for all } S \in S\}$$

THEOREM 21.1: *The orthogonal complement $S^{\perp}$ of any non-empty subset $S$ of the Euclidean space $V$, ( , ) is a subspace of $V$.*

Proof. EXERCISE.

We now turn to the important special case when $S = M$ is itself a subspace of V.

THEOREM 21.2: *Let $M$ be a subspace of the finite dimensional Euclidean space $V$, ( , ), then $V$ is the direct sum of $M$ and $M^{\perp}$; that is*

$$V = M \oplus M^{\perp}$$

Proof. Let $X_1, X_2, \ldots, X_m$ be a basis for $M$ and let $X_1, X_2, \ldots, X_m, X_{m+1}, \ldots, X_n$ be an extension to a basis for $V$.

Using the Gram-Schmidt orthogonalization procedure convert $X_1, X_2, \ldots, X_n$ to an orthogonal basis $Y_1, Y_2, \ldots, Y_n$ for $V$:

$$Y_1 = X_1$$

$$Y_2 = X_2 - \frac{(X_2,Y_1)}{(Y_1,Y_1)} Y_1$$

$$Y_3 = X_3 - \frac{(X_3,Y_1)}{(Y_1,Y_1)} Y_1 - \frac{(X_3,Y_2)}{(Y_2,Y_2)} Y_2$$

etc.

Then, since $Y_k$ is a linear combination of $X_1, X_2, \ldots, X_k$ we have that $Y_1, Y_2, \ldots, Y_m$ is a basis for $M$ and so it suffices to show that $Y_{m+1}, \ldots, Y_n$ is a basis for $M^{\perp}$.

Clearly $< Y_{m+1}, \ldots, Y_n > \subseteq M^{\perp}$. (Why?) On the other hand, if $V = a_1Y_1 + a_2Y_2 + \ldots + a_mY_m + a_{m+1}Y_{m+1} + \ldots + a_nY_n$ is an element of $M^{\perp}$ we must have

$(V, Y_1) = 0$ as $Y_1 \in M$ and so

$$0 = (a_1Y_1 + a_2Y_2 + \ldots + a_nY_n, Y_1) = a_1(Y_1, Y_1) + a_2(Y_2, Y_1) + \ldots + a_n(Y_n, Y_1)$$

$= a_1(Y_1, Y_1)$. Since $(Y_1, Y_1) \neq 0$ we therefore have $a_1 = 0$.

Similarly, $(V, Y_2) = 0$ and so $a_2 = 0$

$(V, Y_3) = 0$ and so $a_3 = 0$

$\vdots$

$(V, Y_m) = 0$ and so $a_m = 0$.

Thus, $V = a_{m+1}Y_{m+1} + \ldots + a_nY_n$ is an element of $< Y_{m+1}, \ldots Y_n >$ and so we have $M^{\perp} \subseteq < Y_{m+1}, \ldots, Y_n >$, completing the proof.

COROLLARY 21.1: *For $M$ a subspace of the finite dimensional Euclidean space $V$ we have*

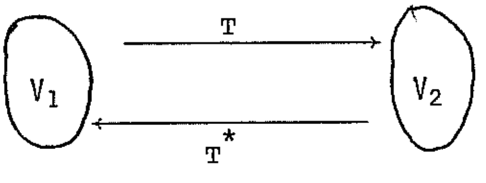
$$\dim V = \dim M + \dim M^{\perp}$$

EXERCISE (optional) 21.1: For any non-empty subset $S$ of the Euclidean space $V$, ( , ) show that the orthogonal complement of the orthogonal complement of $S$ is the span of $S$; that is, $(S^{\perp})^{\perp} = \langle S \rangle$.

ADJOINT TRANSFORMATIONS.

We aim to show that if $V_1, ( \; , \; )_1$ and $V_2, ( \; , \; )_2$ are two Euclidean spaces and T is a linear transformation from $V_1$ to $V_2$, then there exists a linear transformation $T^*$ (the "adjoint" of T) from $V_2$ to $V_1$ such that

$$(TX,Y)_2 = (X,T^*Y)_1 \quad \text{for all } X \in V_1 \text{ and } Y \in V_2$$




EXAMPLE 21.1: Let $V_1 = V_2 = R_2$ with inner-product $(X,Y) = X.Y = x_1y_1 + x_2y_2$ where $X = (x_1,x_2)$ and $Y = (y_1,y_2)$. Then, if $T : V_1 \to V_2$ is defined by

$$T((x_1,x_2)) = (x_1-x_2, \; x_1+x_2)$$

we see that

$$\begin{aligned}(T(x_1,x_2),(y_1,y_2)) &= (x_1-x_2)y_1 + (x_1+x_2)y_2 \\ &= x_1(y_1+y_2) + x_2(y_2-y_1) \\ &= ((x_1,x_2),(y_1+y_2,y_2-y_1))\end{aligned}$$

and so $(TX,Y) = (X,T^*Y)$, where $T^*$ is the linear transformation defined by

$$T^*((y_1,y_2)) = (y_1+y_2, y_2-y_1)$$

EXERCISE 21.1: Determine the matrix representations for T and $T^*$ in the above example with respect to the standard basis for $R_2$, $\{(1,0),(0,1)\}$.

The desired conclusion will be seen to be a direct consequence of the following theorem, which represents a special case of the finite dimensional Riesz-Fréchet representation theorem.

THEOREM 21.3: *Let* $V_1$, $(\ ,\ )_1$ *and* $V_2$ $(\ ,\ )_2$ *be two finite dimensional Euclidean spaces and let* T *be a linear transformation of* $V_1$ *into* $V_2$, *then for each vector* $Y \in V_2$ *there exists a unique vector* $V_Y \in V$ *such that* $(TX,Y)_2 = (X,V_Y)_1$ *for all* X *in* $V_1$.

EXAMPLE 21.2: For $V_1,V_2$ and T as in example 21.1 and $Y = (1,2)$ we see that

$$(T(X),Y) = 3x_1 + x_2 = (X,(3,1))$$

for all $X = (x_1,x_2) \in V_1$. Thus $V_y = (3,1)$.

<u>Lecture 22</u> - proof of theorem 21.3 and construction of the adjoint transformation.

PROOF (of theorem 21.3): Let $M = \{X \in V_1 : (TX,Y)_2 = 0\}$, then M is a subspace of $V_1$ (check this) and so by theorem 21.2.

$$V_1 = M \oplus M^\perp .$$

If $M^\perp = \{0\}$, then $M = V_1$ and for all $X \in V_1$ we have that

$$(TX,Y) = 0 = (X,0)$$

therefore it suffices to take $V_y = 0$. If $M^\perp \neq \{0\}$ we may choose a <u>non-zero</u> vector V from $M^\perp$. We begin by observing that for all $X \in V_1$

$$(T[(TV,Y)_2 X - (TX,y)_2 V],Y)_2$$

$$= (TV,Y)_2(TX,Y)_2 - (TX,Y)_2(TV,Y)_2$$

$$= 0.$$

and so conclude that

$$(TV,Y)_2 X - (TX,Y)_2 V \in M.$$

In particular then for all $X \in V_1$,

$$((TV,Y)_2 X - (TX,Y)_2 V, V)_1 = 0 \qquad (\text{as } V \in M^{\perp})$$

or

$$(TV,Y)_2 (X,V)_1 - (TX,Y)_2 (V,V)_1 = 0.$$

Since $(V,V)_1 \neq 0$ (3P. of the definition of inner-product, Florey p.333), this last equation may be rearranged to give

$$(TX,Y)_2 = \frac{(TV,Y)_2}{(V,V)_1} (X,V)_1$$

or

$$(TX,Y)_2 = (X, \frac{\overline{(TV,Y)_2}}{\overline{(V,V)_1}} V)$$

for all $X \in V_1$.

Thus the theorem is satisfied by taking

$$V_Y = \frac{\overline{(TV,Y)_2}}{\overline{(V,V)_1}} V$$

(which we note depends only on Y and V <u>not</u> on X).

It only remains to prove that the element $V_Y$, whose existence has been established above, is indeed unique. Thus, assume $V_Y'$ is another such vector, then we have

$$(X,V_Y)_1 = (TX,Y) = (X,V_Y')$$

or $(X,V_Y - V_Y') = 0$ for all $X \in V_1$. In particular, taking $X = V_Y - V_Y'$ we have

$$(V_Y - V_Y', V_Y - V_Y') = 0$$

and so, by P3. of the definition for inner-product we have $V_Y - V_Y' = 0$ or $V_Y' = V_Y$.

As a result of Theorem 21.3, for $T \in L(V_1,V_2)$ we may define the <u>adjoint transformation</u>

$T^*: V_2 \to V_1$ by

$T^*(Y) = V_Y$

(here $V_Y$ is the uniques element of $V_1$ such that $(TX,Y)_2 = (X,V_Y)_1$ for all $X \in V_1$.)

That is

| $(TX,Y)_2 = (X,T^*Y)_1$ for all $X \in V_1$, $Y \in V_2$ |
|---|

Theorem 22.1: *For $T \in L(V_1,V_2)$ the adjoint transformation $T^*$ is a linear mapping from $V_2$ to $V_1$, that is $T^* \in L(V_2,V_1)$*

PROOF: For $Y,\ Z \in V_2$, scalar $\lambda$ and each $X \in V_1$ we have

$$\begin{aligned}(X,T^*(Y+\lambda Z))_1 &= (TX,Y+\lambda Z)_2\\ &= (TX,Y)_2 + \bar{\lambda}(TX,Z)_2\\ &= (X,T^*Y)_1 + \bar{\lambda}(X,T^*Z)_1\\ &= (X,T^*Y)_1 + (X,\lambda T^*Z)_1\\ &= (X,T^*Y+\lambda T^*Z)_1.\end{aligned}$$

Thus, $(X,T^*(Y+\lambda Z))_1 - (T^*Y+\lambda[T^*Z])_1 = 0$ for all $X \in V_1$. In particular, $(T^*(Y+\lambda Z) - [T^*Y+\lambda T^*Z],\ T^*(Y+\lambda Z) + [T^*Y - \lambda T^*Z])_1 = 0$ and so, by 3P. of the definition for inner-product, we have

$$T^*(Y+\lambda Z) - [T^*Y+\lambda T^*Z] = 0$$

or

$$T^*(Y+\lambda Z) = T^*Y + \lambda T^*Z.$$

For all $Y,Z \in V_2$ and scalars $\lambda$.

The mapping $T \longrightarrow T^*$ is of considerable importance with many properties analogous to the operation of complex conjugation. For example:

(i) $(T+S)^* = T^*+S^*$, $T,S \in L(V_1,V_2)$;

(ii) $(\lambda T)^* = \bar{\lambda}\, T^*$, $T \in L(V_1,V_2)$, scalar $\lambda$;

(iii) $(SoT)^* = T^*oS^*$, $T \in L(V_1,V_2)$, $S \in L(V_2,V_3)$;

(iv) If $T$ is invertible, then so is $T^*$ with

$(T^*)^{-1} = (T^{-1})^*$, $T \in L(V_1,V_2)$.

Proof of (iii): For all $X \in V_1$, $Z \in V_3$ we have

$$\begin{aligned}(X,(SoT)^*Z)_1 = (SoT(X),Z)_3 &= (S(T(X)),Z)_3\\ &= (T(X),S^*Z)_2\\ &= (X,T^*(S^*Z))_1\\ &= (X,T^*oS^*(Z))_1.\end{aligned}$$

EXERCISE 22.1. Prove (i),(ii) and (iv)

LECTURE 23 - Matrix representation of $T^*$ with respect to a pair of orthonormal bases.

Let $V_1$, $(\ ,\ )_1$ and $V_2$, $(\ ,\ )_2$ be two Euclidean spaces with orthonormal bases

$\mathcal{B}_1 = \{X_1,X_2,\ldots,X_n\}$ and $\mathcal{B}_2 = \{Y_1,Y_2,\ldots,Y_m\}$

respectively, and let $T \in L(V_1,V_2)$.

Recalling that the entry in the j'th row and i'th column, $t^*_{ji}$, of the n×m matrix $[T^*]_{\mathcal{B}_2\mathcal{B}_1}$ representing the adjoint $T^*$ is the coefficient of $X_i$ in the expansion of $T^*Y_j$ with respect to the basis $\mathcal{B}_1$ we have by Florey, Corollary 8.2 p.347,

$$\begin{aligned} t^*_{ji} &= (T^*Y_j,X_i)_1 \\ &= \overline{(X_i,\ T^*Y_j)_1} \\ &= \overline{(TX_i,Y_j)} \end{aligned}$$

But, again by Florey Corollary 8.2, this is the conjugate of the coefficient of $Y_j$ in the expansion of $TX_i$ with respect to the basis $\mathcal{B}_1$.

Thus we see that $t^*_{ji}$, the entry in the j'th row and i'th column of $[T^*]_{\mathcal{B}_2\mathcal{B}_1}$ is the conjugate of the entry in the i'th row and j'th column of $[T]_{\mathcal{B}_1\mathcal{B}_2}$.

That is,

$$\boxed{t^*_{ji} = \bar{t}_{ij}}$$

DEFINITION: The adjoint matrix of the m×n matrix M, is the n×m matrix M* whose entry in the j-i'th place is the conjugate of the i-j'th entry in M.

That is, the i'th row of M* is the "conjugate" of the i'th column of M.

[Note: If the entries of M are all real, then M* is the transpose of M, $M^T$.]

In terms of this definition we have

$$[T^*]_{\mathcal{B}_2\mathcal{B}_1} = [T]^*_{\mathcal{B}_1\mathcal{B}_2}$$

EXAMPLES:

1) Let $M = \begin{bmatrix} 1 & 2 \\ 3 & 1+i \\ 2-i & 4 \end{bmatrix}$

then

$$M^* = \begin{bmatrix} 1 & 3 & 2+i \\ 2 & 1-i & 4 \end{bmatrix}$$

2) Let $V_1 = R_3$ and $V_2 = R_2$ with "dot" product for inner-product in both spaces.

Let $\mathcal{B}_1$ denote the standard orthonormal basis for $V_1$, that is $\mathcal{B}_1 = \{(1,0,0),(0,1,0),(0,0,1)\}$. Similarly let $\mathcal{B}_2 = \{(1,0),(0,1)\}$.

If $T:V_1 \to V_2$ is given by

$$T(x_1,x_2,x_3) = (x_1+2x_2,\ x_1- x_3)$$

then

$$[T]_{B_1B_2} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

Thus

$$[T^*]_{B_2B_1} = [T]^*_{B_1B_2} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 0 & -1 \end{bmatrix}$$

and so $T^*(y_1,y_2) = (y_1+y_2, 2y_1, -y_2)$

Verify this by direct substitution into

$$(TX,Y)_2 = (X,T^*Y)_1 :$$

$$\text{L.H.S.} = (T(x_1,x_2,x_3),(y_1,y_2))_2$$

$$= (x_1 + 2x_2, x_1-x_3).(y_1y_2)$$

$$= (x_1y_1 + 2x_2y_1 + x_1y_2 - x_3y_2$$

$$\text{R.H.S.} = (x_1,x_2,x_3), T^*(y_1,y_2))_1$$

$$= (x_1,x_2,x_3).(y_1+y_2,2y_1,-y_2)$$

$$= (x_1y_1+x_1y_2 + 2x_2y_1 - x_3y_2$$

$$= \text{L.H.S.}$$

<u>LECTURE 24</u>

Theorem 24.1: *For any* $T \in L(V_1,V_2)$ *where* $V_1,(\ ,\ )_1$ *and* $V_2,(\ ,\ )_2$ *are Euclidean spaces we have* $\text{Ker } T \subseteq (T^*(V_2))^{\perp}$ *the orthogonal complement of the range of the adjoint.*

Proof. If $X \in \text{Ker } T$ we have

$TX = 0$, so $(TX,Y) = 0$ for all $Y \in V_2$, thus $(X,T^*Y) = 0$ for all $Y \in V_2$. That is, $X$ is orthogonal to every element of $T^*(V_2)$ and the result is established.

Corollary 24.1: *The rank of* $T$ *equals the rank of* $T^*$; *that is* $\dim T(V_1) = \dim T^*(V_2)$.

Proof. By Florey Theorem 5.12 p.195,

$$\begin{aligned}\dim T(V_1) &= \dim V_1 - \dim \operatorname{Ker} T \\ &\geq \dim V_1 - \dim (T^*(V_2))^{\perp} \text{ by Theorem 24.1} \\ &= \dim V_1 - \dim V_1 - \dim T^* (V_2)) \text{ by Corollary 21.1} \\ &= \dim T^*(V_2).\end{aligned}$$

Similarly $\dim T^*(V_2) \geq \dim(T^*)^*(V_1)$

$$= \dim T(V_1), \text{ as } (T^*)^* = T$$

Thus $\dim T(V_1) = \dim T^*(V_2)$ as required.

Corollary 24.2: *If* M *is any* m×n *matrix, the row rank of* M *equals the column rank of* M.

Proof. For definitions of row and column ranks see Florey p.259. Now, let T denote the linear transformations from $R_n$ to $R_m$ whose matrix representation relative to the standard bases is M, then the column rank of M equals dim T $(R_n)$, see the first five lines to the proof of Theorem 6.20 in Florey p.262. By Corollary 24.1 we therefore have that the column rank of M equals dim $T^*(R_m)$ which by the same argument as above equals the column rank of M*. Since the columns of M* are the rows of M the result is established.

## LECTURE 25 - Self-adjoint transformations

DEFINITION: Let $V,(\ ,\ )$ be a Euclidean space, $T \in L(V,V)$ is a self-adjoint transformation if $T^* = T$.

Thus T is self-adjoint if and only if $(TX,Y) = (X,TY)$ all $X,Y \in V$.

By the results of lecture 23 we see that T is self-adjoint if and only if

$$[T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}} = [T]_{\mathcal{B}} \text{ for any orthonormal basis } \mathcal{B} \text{ of } V.$$

EXAMPLES:

1) The identity map $I_V$ is self-adjoint on $V$ . For any basis $\mathcal{B}$

we have

$$[I_V]_{\mathcal{B}} = I_n = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ . & . & . & \dots \end{bmatrix} = I_n^* = [I_V^*]_B$$

2) $T : R_2 \to R_2 : (x_1,x_2) \longrightarrow (x_1+x_2, x_1-x_2)$

is self-adjoint when inner-product equals "dot" product.

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^* = [T^*]_{\mathcal{B}}$$

where $\mathcal{B} = \{(1,0),(0,1)\}$.

[Also verify this by direct substitution into

$(TX,Y) = (X,TY)$.]

3) For $V_1$, $(\ ,\ )_1$ and $V_2$, $(\ ,\ )_2$ any two Euclidean spaces and $T \in L(V_1,V_2)$,

$T^*T$ is self-adjoint on $V_1$

Proof - EXERCISE.

Theorem 25.1: *If* $V$, $(\ ,\ )$ *is a Euclidean space and* $T \in L(V,V)$ *is self-adjoint, then the eigenvalues of* $T$ *are all real.*

Proof. Let $\lambda$ be an eigenvalue of $T$ corresponding to the (non-zero) eigenvector $X$, then

$$\lambda(X,X) = (\lambda X,X) = (TX,X) = (X,TX) = (X,\lambda X) = \bar{\lambda}(X,X).$$

Thus $(\lambda-\bar{\lambda})(X,X)=0$ and so since, by P3. of the definition for inner-product, $(X,X) \neq 0$ we have $\lambda-\bar{\lambda} = 0$ or $\lambda=\bar{\lambda}$, that is $\lambda$ is real.

EXERCISE: If $\lambda$ is an eigenvalue of $T^*T$ where $T \in L(V_1,V_2)$; $V_1,(\ ,\ )_1$; $V_2$, $(\ ,\ )_2$ are Euclidean spaces, show that $\lambda \geq 0$.

THEOREM 25.2: *If* $T$ *is a self-adjoint transformation of the Euclidean space* $V$, $(\ ,\ )$ *into itself, then* $T$ *has a complete set of eigenvectors which form an* <u>*orthogonal*</u> *basis for* $V$.

The proof is broken into several steps (lemmas).

Step I) - *Invariant Subspaces*

Basic to the proof is the notion of an invariant subspace: The subspace $U$ of the vector space $V$ is invariant under the linear mapping T if $T(U) \subseteq U$; that is T maps each element of $U$ to an element of $U$. By restricting the domain of T to $U$ we may think of T as defining a linear mapping from $U$ to $U$. If $V$ is a Euclidean space with innerproduct ( , ) and T is self-adjoint; that is $(TX,Y) = (X,TY)$ for all $X, Y \in V$, then it is certainly true that $(TX,Y)=(X,TY)$ for all $X,Y \in U$ and so T restricted to U defines a self-adjoint transformation on $U$.

Examples of invariant subspaces:

(1) Ker T is invariant under T - indeed $T(\text{Ker } T) = \{0\} \leq \text{Ker } T$.

(2) (Important to our proof)

Let $X_1,X_2,\ldots,X_m$ be eigenvectors of T with corresponding eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_m$, then $U = \langle X_1,X_2,\ldots,X_m \rangle$ the subspace spanned by $X_1,X_2,\ldots,X_m$ is invariant under T.

- To see this, let $X \in U$, then

$X = a_1X_1 + a_2X_2 + \ldots + a_mX_m$ for some scalars $a_1,a_2,\ldots,a_m$ and so

$$T(X) = a_1T(X_1)+\ldots+a_mT(X_m)$$

$$= a_1\lambda_1X_1+\ldots+a_m\lambda_mX_m$$

which is a linear combination of $X_1,\ldots,X_m$ and so in $U$.

(3) (Important to our proof) If T is a self-adjoint transformation of $V$ and $U$ is an invariant subspace under T, then $U^{\perp}$ is also invariant under T.

- Let $X \in U^{\perp}$; that is $(U,X)=0$ for all $U \in U$, we must show $(U,TX)=0$ for all $U \in U$. But, $(U,TX) = (TU,X)$ and $TU \in U$, since $U$ is invariant under T, we therefore have $(TU,X)=0$ and so $(U,TX)=0$ as required.

EXERCISE: Show that $U$ is invariant under T if and only if $U^{\perp}$ is invariant under T* (Note: (3) above is a particular consequence of this.)

Step II) Let dim V = n. Since the eigenvalues of T are the roots of the n'th degree characteristic polynomial, the fundamental theorem of algebra guarantees that T has at least one eigenvalue $\lambda_1$ (possibly repeated n-times). Let $X_1$ be an eigenvector which corresponds to $\lambda_1$; that is $TX_1 = \lambda_1 X_1$, $X_1 \neq 0$. Now let $U_1$ be the one dimensional subspace spanned by $X_1$, $U_1 = \langle X_1 \rangle$. Then, by (2) of Step I, $U_1$ is invariant under T and so by (3) of Step I, $U_1^{\perp}$ is also invariant under T. By the remarks at the start of Step I we may therefore regard T as defining a self-adjoint transformation from $U_1^{\perp}$ to $U_1^{\perp}$.

By the same reasoning as above the self-adjoint operator obtained by restricting T to $U_1^{\perp}$ has at least one eigenvalue $\lambda_2$ (possibly equal to $\lambda_1$) with corresponding eigenvalue $X_2 \in U_1^{\perp}$. That is; $TX_2 = \lambda_2 X_2$, $X_2 \neq 0$ with $X_2$ orthogonal to $X_1 \in U_1$ $(X_2 \in U_1^{\perp})$.

Since $X_1, X_2$ are linearly independent we may form the 2-dimensional subspace $U_2 = \langle X_1, X_2 \rangle$. Again by Step I, $U_2$ and therefore $U_2^{\perp}$ are invariant under T and so by restricting T to $U_2^{\perp}$ we obtain an eigenvalue and corresponding eigenvector $X_3$. That is;

$TX_3 = \lambda_3 X_3$, $X_3 \neq 0$ and $X_3$ is orthogonal to both $X_1$ and $X_2$.

Forming the 3-dimensional subspace $U_3 = \langle X_1, X_2, X_3 \rangle$ and continuing in this way we obtain mutually orthogonal eigenvectors $X_1, X_2, X_3, X_4, X_5, \ldots$. Clearly this proceedure can continue until we obtain n mutually orthogonal eigenvectors $X_1, X_2, \ldots, X_n$, for which $U_n = \langle X_1, X_2, \ldots, X_n \rangle$ has dim n and so equals $V$, proving the theorem.

The material from here on is not included in Florey "Elementary Linear Algebra with Applications". The following might be useful references.

Strang, Gilbert : "*Linear Algebra and its Applications*" 2nd edition 1980, Academic Press.

(See Section 6.4 for relevant material on Rayleigh's Principle for determining the smallest and largest eigenvalue of a real symmetric matrix.)

Isaacson and Keller : "*Analysis of Numerical Methods*", John Wiley.

(See Chapter 4 Section 2 for a discussion of the Power Method of approximating the "largest" eigenvalue of a matrix.)

Marsden, Jerold E. : "*Elementary Classical Analysis*", Freeman.

(Chapter 2 is concerned with topology in euclidean n-space.)

NUMERICAL DETERMINATION OF EIGENVALUES

Lecture 29:

In all our problems so far we have considered relatively small matrices (4×4 or less). To determine the eigenvalues of such a matrix we have first determined the characteristic polynomial $p(\lambda) = \det(M - \lambda I)$ and then extracted its roots (if necessary this last step could have been done numerically using "half-interval" or Newton's method, for example). Let us consider the situation when M is an n×n matrix with n large. If we compute the n×n determinant of the first step by cofactor expansion, say $\det A = a_{11}A_{11} + a_{21}A_{21} + \dots + a_{n1}A_{n1}$ , then if each cofactor is available we require n multiplications and a similar number of additions. Now each cofactor is an $(n-1)\times(n-1)$ determinant and so could itself be expanded into cofactors, requiring $(n-1)$ multiplications. A simple inductive argument shows that we will require the order of $n\times(n-1)\times(n-2)\times\dots\times 2\times 1 = n!$ operations to determine $p(\lambda)$ in this way. Even if we could perform the operations on a high speed computer it would require *billions* of years to determine a 25×25 determinant in this way. Problems of this magnitude occur quite frequently in Economic, Ecological and other applications of linear algebra. Clearly, alternative procedures are necessary. Fortunately in many applications it is not necessary to determine all the eigenvalues of M, frequently a knowledge of the "largest" or "smallest" eigenvalue and corresponding eigenvectors is sufficient. It is to the problem of finding these that we address ourselves.

*Rayleigh's Principle.*

Let M be a real symmetric nxn-matrix with eigenvalues (necessarily real, though not necessarily all distinct) $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ . Define the Rayleigh quotient of M to be the function

$$R : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$$

defined by

$$R(X) = \frac{X^T M X}{X^T X} \qquad \text{where } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

For example, if $M = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ we have

$$R(X) = \frac{[x_1, x_2] \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{[x_1, x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$= \frac{3x_1^2 + 2x_1x_2 + 3x_2^2}{x_1^2 + x_2^2} .$$

Our main result is the following.

Theorem 29.1 : *The maximum value of the Rayleigh quotient of* M,

$$R(X) = \frac{X^T M X}{X^T X}$$ , *is* $\lambda_n$ *(the largest eigenvalue of* M*).*

*Further, if* $X_n$ *is an eigenvector of* M *corresponding to the eigenvalue* $\lambda_n$*, then* R(X) *achieves its maximum at* $X = X_n$.

Proof. Let $\mathcal{B} = \{X_1, X_2, \ldots, X_n\}$ be an ordered (orthogonal) basis of eigenvectors corresponding to the eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ of M. (Such a basis exists by lecture 25, theorem 25.2)

Since the numerator in Rayleigh's quotient is a quadratic form we have by lecture 27 that

$$R(X) = \frac{X^T P D P^T X}{X^T X}$$

where D is the diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 & \cdot & \cdot & \cdot \\ 0 & \lambda_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \lambda_n \end{bmatrix}$$

and P is the matrix representing the change of basis from $\mathcal{B}$ to the natural basis of $\mathbb{R}^n$. Recall $P^{-1} = P^T$ as M is symmetric.

Further, $X^TX = X^TIX = X^TPP^{-1}X = X^TP\,P^TX = (P^TX)^T(P^TX)$ and so

$$R(X) = \frac{(P^TX)^T D (P^TX)}{(P^TX)^T(P^TX)}$$

$$= \frac{Y^TDY}{Y^TY}$$

$$= \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2}$$

where $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = P^TX$ is the coordinate vector of X relative to the basis $\mathcal{B}$.

The result now follows from the observation that

$$P^TX_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$ and so for any $X \neq 0$ if

$$P^TX = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$ we have

$$R(X_n) = [0,0,\ldots,0,1]\, D \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$= \lambda_n$$

$$= \frac{\lambda_n y_1^2 + \lambda_n y_2^2 + \ldots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2}$$

$$\geq \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2} \quad (\text{as } \lambda_1, \lambda_2, \ldots \leq \lambda_n)$$

$$= R(X)$$

EXAMPLE: For $M = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ we had

$$R(X) = \frac{3x_1^2 + 2x_1x_2 + 3x_2^2}{x_1^2 + x_2^2}$$

$$= 3 + \frac{2x_1x_2}{x_1^2 + x_2^2}$$

$$= 3 + \frac{2}{u + \frac{1}{u}} \quad \text{where } u = x_1/x_2$$

clearly this is a maximum when $u + \frac{1}{u}$ has a positive minimum that is when $u = 1$ (or $x_1 = x_2$) and so we conclude that 4 is the largest eigenvalue of M and that (1,1) is a corresponding eigenvector. [Indeed the eigenvalues of M are 2 and 4]

REMARKS: I) For any vector $X_0 \in \mathbb{R}^n$ we see that the value $R(X_0)$ is a lower bound on the largest eigenvalue of M.

II) Since $R(X) = R(cX)$ for any scalar $c \neq 0$ (why?) we may, without loss of generality, seek the maximum of R(X) for vectors X with $X^TX (=\|X\|^2) = 1$.

In this form the problem becomes; *maximise*

$$X^TMX$$

*subject to the constraint*

$$X^TX = 1$$

and so is conveniently handled by the method of Lagrange multipliers.

III) By replacing M by $-M$ we readily conclude that the minimum value of the Rayleigh quotient is the smallest eigenvalue of M and that this value is achieved at any corresponding eigenvector.

EXCERCISES:

1) Use Rayleigh's principle to find the largest eigenvalue, and a corresponding eigenvector of

a) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and b) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

2) For any symmetric matrix M show that the largest eigenvalue is at least as large as the largest value on the diagonal of M. [Hint: Consider the value of the Rayleigh quotient at each of the natural basis elements of $\mathbb{R}^n$.]

3) Let M and N be two symmetric matrices, show that the largest eigenvalue of M+N is no larger than the sum of the largest eigenvalues of M and N.

Lecture 30: *The Power Method*

For non-symmetric matrices the problem of determining the largest eigenvalue is not so simple as it was for symmetric ones. In general we must resort to approximate methods. One such procedure, applicable in certain circumstances, is introduced below.

Suppose that M is an n×n matrix which has a complete set of eigenvectors $X_1, X_2, \ldots, X_n$, corresponding to the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ (not necessarily all distinct). Further suppose that M has one eigenvalue of larger modulus than all the others and that the eigenspace corresponding to this eigenvalue is of dimension one. Without loss of generality we may take this eigenvalue to be $\lambda_1$.

Any vector $X \in \mathbb{R}^n$ may be expanded uniquely as

$$X = x_1X_1 + x_2X_2 + \ldots + x_nX_n$$

and so we have

$$MX = x_1\lambda_1 X_1 + x_2\lambda_2 X_2 + \dots + X_n\lambda_n X_n$$

$$M^2X = x_1\lambda_1{}^2 X_1 + x_2\lambda_2{}^2 X_2 + \dots + x_n\lambda_n^2 X_n$$

and in general

$$M^kX = x_1\lambda_1^k X_1 + x_2\lambda_2^k X_2 + \dots + x_n\lambda_n^k X_n$$

$$= \lambda_1^k (x_1 X_1 + x_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k X_2 + \dots + x_n \left(\frac{\lambda_n}{\lambda_1}\right)^k X_n)$$

Thus, provided we start with an X for which $x_1 \neq 0$, we see that the "direction" of $M^kX$ approaches that of $X_1$ as $k\to\infty$ (for j = 2,3,...,n $\left|\frac{\lambda_j}{\lambda_1}\right|^k \to 0$ as $k\to\infty$). Of course, without normalization these successive iterates need not themselves converge, however it does follow that as $k\to\infty$

$$\frac{(M(M^kX), M^kX)}{(M^kX, M^kX)} \to \lambda_1$$

EXAMPLE: Let $M = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$ [From the onset let us note that the eigenvalues of M are 3 and 2 with corresponding eigenvectors (1,2) and (1,1).]

Starting with $X = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and computing the successive iterates we obtain

$MX = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ (4), $M^2X = \begin{bmatrix} 5 \\ 14 \end{bmatrix}$ (2.8), $M^3X = \begin{bmatrix} 19 \\ 46 \end{bmatrix}$ (2.42), $M^4X = \begin{bmatrix} 65 \\ 146 \end{bmatrix}$ (2.25)

$M^5X = \begin{bmatrix} 210 \\ 454 \end{bmatrix}$ (2.16) $M^6X = \begin{bmatrix} 664 \\ 1396 \end{bmatrix}$ (2.10) $M^7X = \begin{bmatrix} 2060 \\ 4256 \end{bmatrix}$ (2.07)

The figure in brackets beneath each iterate is the ratio of the second to first component which should be approaching 2 if these iterates are to successively more closely approximate multiples of the eigenvector (1,2).

The corresponding approximations to the largest eigenvector (3) are

$$\frac{((5,14),(1,4))}{((1,4),\ (1,4))} = 3.59$$

$$\frac{((19,46),(5,14))}{((5,14),(5,14))} = 3.34$$

$$\frac{((65,146),(19,46))}{((19,46),(19,46))} = 3.21$$

$$\frac{((210,454),(65,146))}{((65,146),(65,146))} = 3.15$$

$$\frac{((664,1396),(210,454))}{((210,454),(210,454))} = 3.09$$

$$\frac{((2060,4256),(664,1396))}{((664,1396),(664,1396))} = 3.06$$

Several refinements to improve the rate of convergence are available, however the above method serves to illustrate the basic idea. To be "certain" of the approximation obtained it is strictly necessary to repeat the procedure starting in turn with each element of a basis for $\mathbb{R}^n$ (the sequence of iterates which converge to the largest estimate for $\lambda_1$ then being chosen). Clearly this procedure is best carried out on a computing device. If possible you should as an EXERCISE prepare a programme for the above mehtod and run it for several different matrices and starting vectors.

EXERCISES 1) Use the power method to estimate the largest eigenvalue (and corresponding eigenvector) for each of the following matrices (in each case determine at least 5 successive iterates).

a) $\begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$

**2) Prove the convergence result claimed for the above procedure.

[HINT: Let $X_k = M^k X = \lambda_1^k (x_1 X_1 + Z_k)$

where $$Z_k = \left(\frac{\lambda_2}{\lambda_1}\right)^k x_2 X_2 + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^k x_n X_n .$$

Note that $\|Z_k\| \to 0$ and so deduce that if $Y_k = \frac{X_k}{\|X_k\|} - \frac{x_1 X_1}{|x_1| \|X_1\|}$,

then $Y_k \to 0$ as $k \to \infty$ .

Hence conclude that

$$\frac{(MX_k, X_k)}{(X_k, X_k)} = \left(M\left[Y_k + \frac{x_1 X_1}{|x_1| \|X_1\|}\right], Y_k + \frac{x_1 X_1}{|x_1| \|X_1\|}\right) \to \lambda_1 ,$$

(Note, at this step you may need to use the "obvious" estimate that

$\|MY_k\| \le n(\max\{|m_{ij}|\}) \|Y_k\|$.)

---

TOPOLOGY IN EUCLIDEAN SPACES

Lecture 31.

In Florey chapter 8 the "distance between" two vectors X,Y of the euclidean space $\mathbb{V}$, ( , ) was defined as

$$d(X,Y) = \|X-Y\|$$

$$= \sqrt{(X-Y, X-Y)} .$$

The function d(X,Y) is referred to as a *metric* on $\mathbb{V}$ and was seen to satisfy the following.

i) $d(X,Y) \ge 0$ for all X,Y in $\mathbb{V}$

ii) $d(X,Y) = 0$ if and only if X=Y

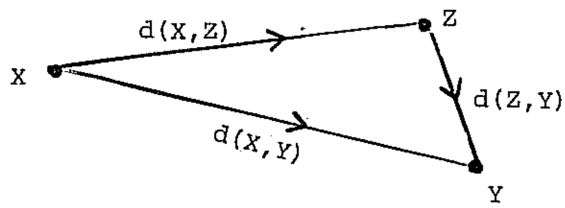
iii) $d(X,Y) = d(Y,X)$ for all X,Y in $\mathbb{V}$

[that is, the metric is symmetric in its two arguments X and Y]

iv) For any three vectors X,Y and Z in $V$ we have

$d(X,Y) \leq d(X,Z) + d(Z,Y)$

[this last inequality is often referred to as the triangle inequality - effectively it states that the length of one side of a triangle cannot exceed the sum of the lengths of the other two sides.




These are properties which, as the German mathematician Herman Minkowski (1864-1909) remarked in 1906, "any notion of distance ought to possess". Any set (not necessarily a vector or euclidean space) on which a function d satisfying the above four properties is defined is referred to as a metric space (see Exercise 1 at the conclusion of this lecture.) Such spaces were introduced and extensively studied by the French mathematician Maurice Fréchet (1878-1973) in 1906. They have proved to be of considerable importance in modern mathematics and what follows serves as an introduction to their study. We will however confine ourselves to examples of metrics which arise from an inner-product according to the formula

$$d(X,Y) = \sqrt{(X-Y,X-Y)}.$$

When $V = R^n$ and the inner-product is 'dot product' this gives the usual euclidean formula for the distance between two points $X = (x_1,x_2,\ldots,x_n)$ and $Y = (y_1,y_2,\ldots,y_n)$: $d(X,Y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2}$.

Of course, a different inner-product on $R^n$ will lead to an alternative notion of distance. For example, in the "plane" $R^2$

$$(X,Y) = x_1y_1 + \tfrac{1}{2}(x_1y_2 + x_2y_1) + \frac{1}{3}x_2y_2$$

defines an inner-product * for which the distance between two points is different to that of ordinary Euclidean geometry: $d((0,0),(1,1)) = \sqrt{7/3}$ for instance. (See Exercise 2 at the end of the lecture.)

For euclidean spaces other than $\mathbb{R}^n$ this provides us with a "distance" between objects with which we would not normally associate such a measure. For example, if $V = C[0,1]$ the space of continuous functions on [0,1] and for $f,g \in C[0,1]$ we take $(f,g) = \int_0^1 f(x)g(x)dx$, then we have as the "distance" between f and g;

$$d(f,g) = \sqrt{\int_0^1 (f(x)-g(x))^2 dx}.$$

Frequently we speak figuratively as though some notion of distance were present, for instance: 'Orange is a colour nearer red than violet'; 'For small x, $x - \frac{x^3}{6}$ is a closer approximation to sin x, than x'.

Such notions are made precise by the presence of a metric. Thus in terms of the metric given above we have

$$d(x, \sin x) \doteq 0.061$$

while $$d\left(x - \frac{x^3}{6}, \sin x\right) \doteq 0.002.$$

---

* Indeed, identifying $X = (x_1, x_2)$ with the polynomial $x(t) = x_2 t + x_1$ and $Y$ with $y(t) = y_2 t + y_1$ we have $(X,Y) = \int_0^1 x(t)y(t)dt$, the usual inner-product for the space of polynomials on [0,1].

It is also worth remarking that in terms of the last metric the "root mean square voltage" of electronics is the distance of the voltage function from zero. A similar interpretation of the "variance" in statistical theory is also possible.

Using this generalized notion of distance it is possible to define for arbitrary euclidean spaces (indeed metric spaces) topological concepts similar to those for the real line or the plane, these include: the convergence of sequences; the interior and boundary of a set and the continuity of functions. It is to the development of these ideas that the remainder of the course is devoted.

Exercises: 1) (i) For points $X = (x_1,x_2)$ and $Y = (y_1,y_2)$ of the "plane" $\mathbb{R}^2$ show that each of the following define metrics.

(a) $d(X,Y) = |x_1-y_1| + |x_2-y_2|$

(b) $d(X,Y) = \begin{cases} 0 & \text{if } X = Y \\ \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2} & \text{otherwise} \end{cases}$

(This metric is sometimes termed the "post-office" metric, can you see why?)

(c) $d(X,Y) = \begin{cases} 0 \text{ if } X=Y \\ 1 \text{ if } X\neq Y \end{cases}$

(This metric is referred to as the "discrete metric", and may be defined on any set.)

(ii) In conventional geometry a circle is defined to be a set of points equidistant from a given point - the centre. When 'distance' is measured according to each of the three formulas of part i), sketch the "circles" of centre (1,1) and radius $1\frac{1}{2}$. That is, sketch

$$\{X \in \mathbb{R}^2 : d(X,(1,1)) = 1\tfrac{1}{2}\}$$

2) Given that

$$(X,Y) = x_1y_1 + \tfrac{1}{2}(x_1y_2 + x_2y_1) + \frac{1}{3}x_2y_2.$$

defines an inner-product on $\mathbb{R}^2$ determine a formula for the associated distance between two points

$$d(X,Y) = \sqrt{(X-Y,X-Y)}.$$

Verify that $d((0,0),(1,1)) = \sqrt{7/3}.$

Sketch the "circle"; centre (0,0) and radius 1 which results when distances are measured according to the formula obtained.

3) For any metric prove the important inequality

$$|d(X,Z) - d(Z,Y)| \leq d(X,Y)$$

(This corresponds to: the difference in lengths of two sides a triangle cannot exceed the length of the third side.)

Lecture 32.

Let us begin by observing that on the 'line' $\mathbb{R}$, the distance between two points x and y is

$$d(x,y) = |x-y| \quad ( = \sqrt{(x-y)^2}).$$

Recalling that a sequence of real numbers

$$x_1, x_2, \ldots, x_n, \ldots,$$

[which we abbreviate to $(x_n)_{n=1}^{\infty}$ or simply $(x_n)$] converges to x if, "given any $\epsilon > 0$, there exists an N such that $|x_n - x| < \epsilon$ whenever $n > N$", we see that the notion of convergence may be extended to sequences of vectors in an inner-product space as follows.

DEFINITION: Let $V$, ( , ) be an euclidean space and $d(X,Y) = \sqrt{(X-Y,X-Y)}$ be the associated metric. We say the sequence of vectors $(X_n) \subset V$ *converges* to $X \in V$ if; given an $\epsilon > 0$, there exists an N such that $d(X_n,X) < \epsilon$ whenever $n > N$.

We often abreviate this by writting $d(X_n,X) \to 0$, or simply $X_n \to X$.

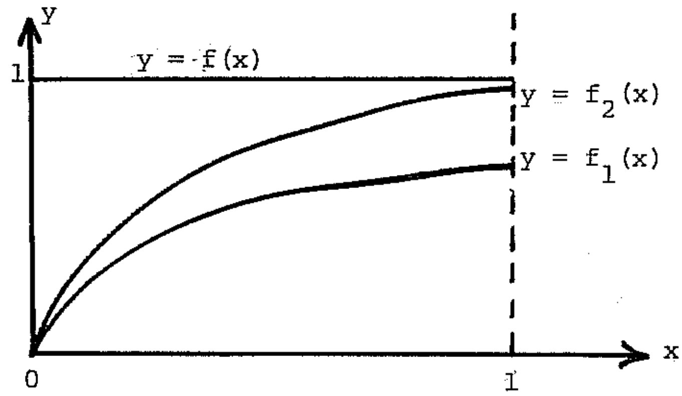
EXAMPLES: 1) Let $V = C[0,1]$ with inner-product defined by

$$(f,g) = \int_0^1 f(x)g(x)dx.$$

Then the sequence of functions

$$f_n(x) = 1 - e^{-nx} \qquad (n=1,2,3,\ldots)$$

converges to the constant function $f(x) \equiv 1$.




To see this, observe that

$$\begin{aligned} d(f_n,d) &= \sqrt{(f_n-f,f_n-f)} \\ &= \left(\int_0^1 (1 - e^{-nx} - 1)^2 dx\right)^{\frac{1}{2}} \\ &= \left(\int_0^1 e^{-2nx} dx\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{2n}[1 - e^{-2n}]\right)^{\frac{1}{2}} \\ &\le \frac{1}{\sqrt{2n}} \end{aligned}$$

and that $\frac{1}{\sqrt{2n}} \to 0$ as $n \to \infty$.

2) Let $V$, $(.,.)$ denote $R^m$ with inner-product 'dot-product' and let

$$\begin{aligned} X_1 &= (x_1^1, x_2^1, x_3^1, \ldots, x_m^1) \\ X_2 &= (x_1^2, x_2^2, x_3^2, \ldots, x_m^2) \\ &\ldots \\ X_n &= (x_1^n, x_2^n, x_3^n, \ldots, x_m^n) \\ &\ldots \end{aligned}$$

be a sequence of vectors in $V$. *We will show that*

$X_n \to X = (x_1, x_2, \ldots, x_m)$

*if and only if* $x_i^n \to x_i$ *as* $n \to \infty$ *for each* $i = 1,2,\ldots,m$.

That is if and only if each of the component sequences converges to the corresponding component of $X$.

$$
\begin{array}{lcl}
X_1 & = & (\boxed{x_1^1}, x_2^1, x_3^1, \ldots, x_m^1) \\
X_2 & = & (\boxed{x_1^2}, x_2^2, x_3^2, \ldots, x_m^2) \\
\cdots & & \\
X_n & = & (\boxed{x_1^n}, x_2^n, x_3^n, \ldots, x_m^n) \\
\cdots & & \\
\downarrow & \Leftrightarrow & \quad \downarrow \quad\; \downarrow \quad \downarrow \qquad\quad\; \downarrow \\
X & = & (\boxed{x_1}, x_2, x_3, \ldots, x_m)
\end{array}
$$

($\Rightarrow$) Let $X_n \to X$; that is,

$$
\begin{aligned}
d(X_n, X) &= \sqrt{(X_n - X, X_n - X)} \\
&= \sqrt{\sum_{i=1}^{m} (x_i^n - x_i)^2} \\
&\to 0
\end{aligned}
$$

Thus, given $\varepsilon > 0$ there exists an $N$ such that

$d(X_n, X) < \varepsilon$ for $n > N$, but then, for $n > N$ we have for each $i_0 = 1,2,\ldots,m$ that

$$
\begin{aligned}
|x_{i_0}^n - x_{i_0}| &= \sqrt{(x_{i_0}^n - x_{i_0})^2} \\
&\le \sqrt{\sum_{i=1}^{m} (x_i^n - x_i)^2} \\
&= d(X_n, X) \\
&< \varepsilon
\end{aligned}
$$

and so $x_{i_0}^n \to x_{i_0}$ as required.

($\Leftarrow$) Assume for each $i=1,2,\ldots,m$ we have

$x_i^n \to x_i$ as $n \to \infty$; that is, given $\varepsilon > 0$ for each $i$ there exists $N_i$ such that

$|x_i^n - x_i| < \varepsilon/\sqrt{m}$ for $n > N_i$ (Use $\varepsilon/\sqrt{m}$ in place of $\varepsilon$ in the definition of convergence.)

Now, let $N$ be the maximum of $N_1, N_2, \ldots, N_m$, then for $n > N$ we have

$|x_i^n - x_i| < \varepsilon/\sqrt{m}$ for all $i=1,2,\ldots,m$

and so

$$d(X_n, X) = \sqrt{\sum_{i=1}^{m} (x_i^n - x_i)^2}$$
$$\leq \sqrt{\sum_{i=1}^{m} \varepsilon^2/m}$$
$$= \varepsilon.$$

Thus $X_n \to X$, establishing the result.

DEFINITION: Let $\mathcal{A}$ be a subset of the inner-product space $\mathcal{V}$, ( , ). We say that $X \in \mathcal{V}$ is a *limit point of* $\mathcal{A}$ if there exists a sequence $(A_n)$ of elements of $\mathcal{A}$ which converges to $X$.

NOTE: Every element $A$ of $\mathcal{A}$ is a limit point of $\mathcal{A}$. To see this note that the "constant sequence" $A, A, A, \ldots, A, \ldots$ is a sequence of elements of $\mathcal{A}$ converging to $A$.

EXAMPLES: 1) Let $\mathcal{V}, \langle\ ,\ \rangle$ be $\mathcal{R}$ with inner-product ordinary multiplication and corresponding norm $\|x\| = |x|$ the absolute value function. If $\mathcal{A}$ is the open interval $(0,1)$ then:

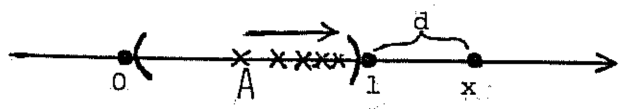
(a) 1 is a limit point of $\mathcal{A}$. The sequence $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, 1-\frac{1}{n}, \ldots$ is a sequence of elements of $\mathcal{A}$ and

$|(1 - \frac{1}{n}) - 1| = \frac{1}{n} \to 0$ as $n \to \infty$.

(b) No point $x > 1$ is a limit point of $\mathcal{A}$. To see this, let $d = x - 1$ then $d > 0$ and if $(a_n)$ is any sequence of elements of $A$ then $a_n < 1$ for all n and so

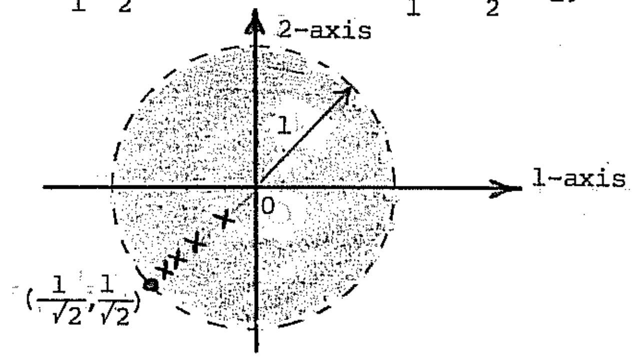
$|x - a_n| > d$

whence $(a_n)$ cannot converge to x.




2) The point $(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})$ is a limit point of the "open disk"

$A = \{x = (x_1,x_2) \in R^2 : \|x\|^2 = x_1^2 + x_2^2 < 1\}$




in $R^2$ with "dot-product" as inner-product.

To see this note that the sequence $(x_n)$ where

$$x_n = \left( -\frac{1}{\sqrt{2}} + \frac{1}{n}, \frac{1}{\sqrt{2}} - \frac{1}{n} \right) \qquad (n=2,3,4,\ldots)$$

is a sequence of points in $A$ [$\|x_n\| = 1 - \frac{2}{n}(\sqrt{2} - \frac{1}{n})$ and $2 - \frac{1}{n} > 0$ for all n, so $\|x_n\|^2 < 1$] which converges to $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, as

$$\|x_n - (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})\| = \|(\frac{1}{n}, -\frac{1}{n})\|$$
$$= \sqrt{2}/n \to 0 \text{ as } n \to \infty.$$

3) In C[0,1] with $\int_0^1 fg$ as inner-product, the function sin x is a limit point of the set $A$ of all polynomials.

Let $p_n(x) = x - \frac{x^3}{6} + \ldots + (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!}$, then from Taylor's Theorem with remainder, for all $x \in [0,1]$ we have

$$|p_n(x) - \sin x| \le \frac{1}{(2n)!}$$

and so

$$\|p_n - \sin\| = \sqrt{\int_0^1 |p_n(x) - \sin x|^2 dx} \le \frac{1}{(2n)!} \to 0$$

as $n \to \infty$.

Intuitively, x is a limit point of $A$ if and only if there are points of $A$ arbitrarily "near" to it.

DEFINITION: Let $A$ be a given subset of the inner-product space $V, \langle , \rangle$. The set of all limit points of $A$ is termed the *closure of $A$* and will be denoted by $\bar{A}$.

NOTE: From the previous "note" we see that always $A \subseteq \bar{A}$.

We will call $A$ a *closed set* if it is its own closure that is if $\bar{A} = A$.

EXAMPLE: From example 1 (a) and (b) above, if $A = (0,1)$ in $R$ with absolute value as norm, then $1 \in \bar{A}$ and no $x > 1$ is in the closure of $A$. Similar arguments (give them) show that $0 \in \bar{A}$ and that no $x < 0$ is in the closure of $A$. Thus we conclude that

$$\bar{A} = \overline{(0,1)} = [0,1] \text{ the "closed" interval from 0 to 1.}$$

Identical arguments also establish that [0,1] is a closed set in the above sense, that is $\overline{[0,1]} = [0,1]$.

EXERCISES: 1) In $R$ with absolute value as norm prove that $\overline{(0,1)} = [0,1]$.

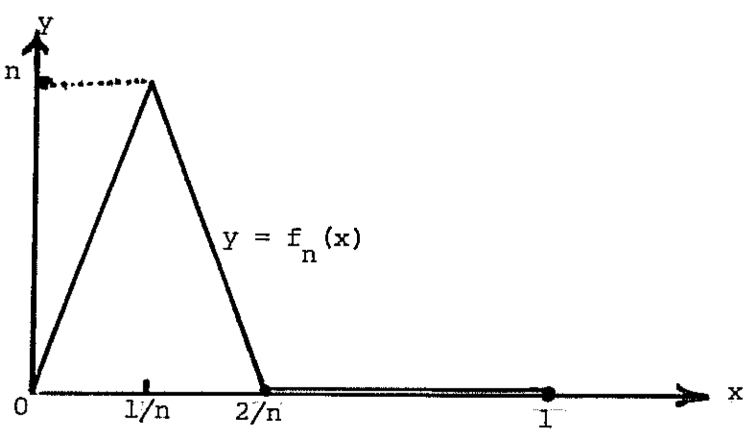
2) (a) In $R^2$ with "dot-product" as inner-product show that the closure of the "open-disk" $\{\underset{\sim}{x} = (x_1,x_2) : \|\underset{\sim}{x}\|^2 = x_1{}^2+x_2{}^2 < 1\}$ is the "closed-disk" $\{\underset{\sim}{x} = (x_1,x_2) : \|\underset{\sim}{x}\|^2 = x_1{}^2+x_2{}^2 \leq 1\}$.

*(b) In $R^n$ with "dot-product" as inner-product show that the "unit-ball" $B = \{\underset{\sim}{x} \in R^n : \|\underset{\sim}{x}\| \leq 1\}$ is a closed set.

3) In any inner-product space $V, ( , )$ show that the intersection of two closed sets is a closed set.

4) Let $f_n$ be defined by

$$f_n(x) = \begin{cases} n^2 x & , 0 \le x \le \frac{1}{n} \\ 2n - n^2 x & , \frac{1}{n} < x \le \frac{2}{n} \\ 0 & , \frac{2}{n} < x \le 1 \end{cases} \qquad (n = 3,4,\ldots)$$

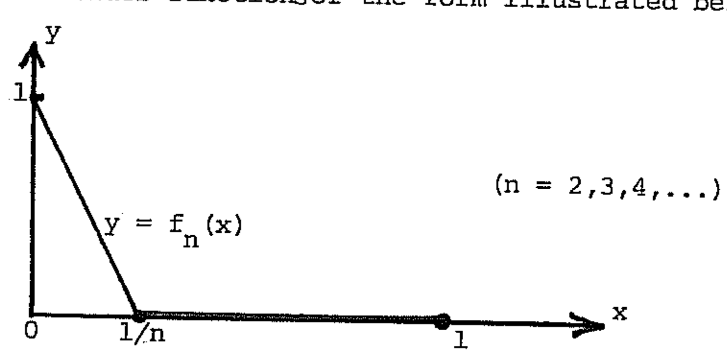



Show that $f_n(x) \to 0$ for each $x \in [0,1]$ and hence deduce that $f_n(x) \to f(x)$ for each $x \in [0,1]$ as $n \to \infty$ is not a sufficient condition to ensure that $\|f_n - f\| \to 0$ as $n \to \infty$ in the space $C[0,1]$ with inner-product defined by

$$(f,g) = \int_0^1 fg.$$

Also give an example to show that the above condition is not necessary; that is, $\|f_n - f\| \to 0$ as $n \to \infty$ need not imply that $f_n(x) \to f(x)$ for all $x \in [0,1]$.

[Hint: consider functionsof the form illustrated below.




(This shows that the analogous results to those of Example 2 on page 13 for sequences in $R^n$ do not apply in $C[a,b]$.)

Lecture 33

The idea, that the closure of a subset A of the inner-product space V, ( , ) consists of the points of A together with those points for which there are points of A arbitrarily near, is made precise in the next lemma.

LEMMA 33.1:

*Let* $V$, ( , ) *be an inner-product space and* $A$ *a given subset of* $V$. *Then* X *is in the closure of* $A$ *if and only if for each* $\varepsilon > 0$ *there exists a point* A *of* $A$ *with* $\|X - A\| < \varepsilon$.

Proof.

($\Rightarrow$) If X is in $\bar{A}$ , then X is a limit point of $A$ and so there exists a sequence $(A_n)$ of points of $A$ with $A_n \to X$, but then, by the definition of convergence, given $\varepsilon > 0$, for some N we have

$\|A_n - X\| < \varepsilon$ for all $n > N$.

Taking $A = A_n$ for any $n > N$ gives the desired result.

($\Leftarrow$) If for each $\varepsilon > 0$ there exists a point A of $A$ with $\|X - A\| < \varepsilon$, then in particular for each n = 1,2,3,.... there is a point of $A$ , call it $A_n$, with $\|X - A_n\| < \frac{1}{n}$ (take $\varepsilon = \frac{1}{n}$ ).

The sequence $(A_n)$ so constructed is such that

$\|A_n - X\| \to 0$ as $n \to \infty$ and so X is a limit point of $A$ and hence in the closure of $A$ as required.

From the notion of closure we can derive a number of others which are of importance in analysis and particularly in advanced calculus.

DEFINITION:

Let $V$, ( , ) be an inner-product space and $A$ a subset of $V$. The *boundary* of $A$ is defined to be the set

$$\text{bdry}(A) = \bar{A} \cap \overline{(V \setminus A)}.$$

Thus X is a *boundary point* of $A$ if and only if it is a limit point of both $A$ and the complement of $A$, $(V \setminus A)$.

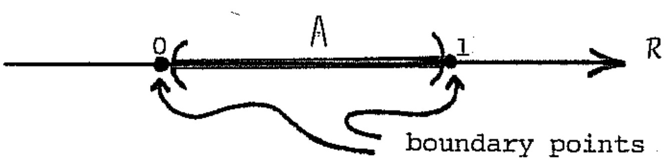
Intuitively, the boundary of $A$ consists of those points for which there are points both "inside" and "outside" of $A$ arbitrarily near to them.

EXAMPLES.

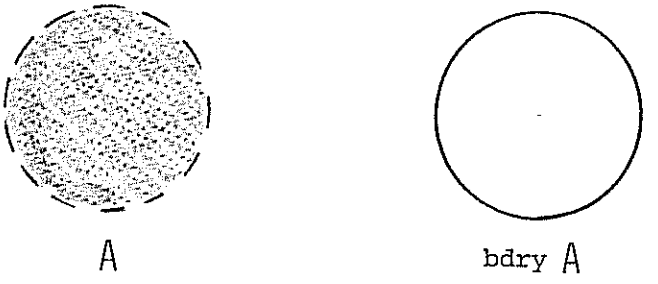
1) In $R$ with absolute value as norm we have for the open interval (0,1) that

$$\begin{aligned} \text{bdry } (0,1) &= \overline{(0,1)} \cap \overline{(R \setminus (0,1))} \\ &= [0,1] \cap \overline{((-\infty,0] \cup [1,\infty))} \\ &= [0,1] \cap ((-\infty,0] \cup [1,\infty)) \\ &= \{0,1\} \end{aligned}$$

(each of the above assertions are readily checked by arguments similar to those of Example 1 on page 15.)




2) In $R^2$ with "dot-product" as inner-product the boundary of both the open disk $\{\underset{\sim}{x} = (x_1,x_2) : \|\underset{\sim}{x}\|^2 = x_1^2 + x_2^2 < 1\}$ and the closed disk $\{\underset{\sim}{x} : \|\underset{\sim}{x}\| \leq 1\}$ is the unit circle $\{\underset{\sim}{x} = (x_1,x_2) : \|\underset{\sim}{x}\|^2 = x_1^2 + x_2^2 = 1\}$




The proof of this is left as an exercise - it should be somewhat easier after the next theorem has been established.

DEFINITION: Let $V$, ( , ) be an inner-product space and $A$ a subset of $V$. The *interior of A* is defined to be

$$\text{int } (A) = A \setminus \text{bdry } (A)$$

and so consists of those points of $A$ which are not boundary points.

<u>EXAMPLES</u>: From the above examples we see that:

(a) The interior of the closed interval [0,1] in $R$ with absolute value as norm is the open interval (0,1); that is

$$\text{int}\ ([0,1]) = (0,1)$$

Similarly

$$\text{int}\ ((0,1)) = (0,1)$$

(b) In $R^2$ with "dot-product" as inner-product the interior of both the "open disk"

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| < 1\}$$

and of the "closed disk"

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| \leq 1\}$$

is the "open disk"

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| < 1\}$$

A set which is its own interior is termed an *<u>open set</u>*; that is $A$ is open if and only if int $A = A$. Thus the "open disk" $\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| < 1\}$ in $R^2$ with 'dot-product' as inner-product is an open set in this sense. Similarly it may be seen that the "open interval" (0,1) is an open subset of $R$ in this sense.

From the definitions we have that

$$\text{int}\ A = A \setminus (\overline{A} \cap \overline{(V \setminus A)})$$

Using some basic set theory we arrive at the following result.

LEMMA 33.2: For any subset $A$ of an inner-product space $V$, ( , ) we have

$$\text{int}\ (A) = V \setminus \overline{(V \setminus A)}$$

PROOF: We begin with the observations that for any sets $A, B, C$ we have

i) $A \setminus B \subseteq C \setminus B$ if $A \subseteq C$

ii) $A \setminus B = A \setminus (A \cap B)$

iii) $A \setminus B \subseteq A \setminus C$ $\qquad C \subseteq B$

iv) If $B \setminus C \subseteq A$, then $B \setminus C \subseteq A \setminus C$.

We first show that int $A \subseteq V \setminus \overline{(V \setminus A)}$:

$$\begin{aligned} \text{int } A &= A \setminus (\bar{A} \cap \overline{(V \setminus A)}) \\ &\subseteq \bar{A} \setminus (\bar{A} \cap \overline{(V \setminus A)}) , \quad \text{by (i)} \\ &= \bar{A} \setminus \overline{(V \setminus A)} , \quad \text{by (ii)} \\ &\subseteq V \setminus \overline{(V \setminus A)} , \quad \text{by (i) again.} \end{aligned}$$

To complete the proof we secondly establish that

$$V \setminus \overline{(V \setminus A)} \subseteq \text{int } A .$$

First however, we observe that

$$\begin{aligned} V \setminus \overline{(V \setminus A)} &\subseteq V \setminus (V \setminus A), \quad \text{by (iii)} \\ &= A \end{aligned}$$

and so $\quad V \setminus \overline{(V \setminus A)} = A \setminus \overline{(V \setminus A)} \quad$ BY (iv)

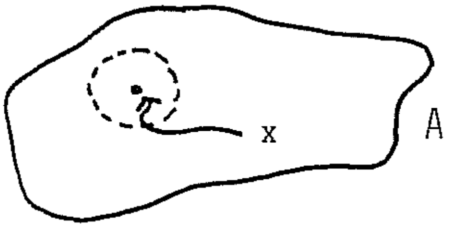
but, then $\quad$

$$\begin{aligned} V \setminus \overline{(V \setminus A)} &= A \setminus \overline{(V \setminus A)} \\ &\subseteq A \setminus (\bar{A} \cap \overline{(V \setminus A)}) , \quad \text{by (iii)} \\ &= \text{int } (A) \end{aligned}$$

as required.

A point of int $(A)$ is termed an *interior point* of $A$ . Combining the previous lemma with lemma 33.1 we arrive at a powerful and useful characterization of interior points.

THEOREM 33.2: Let $V$, ( , ) be an inner-product space and $A$ a subset of $V$ . Then X is an interior point of $A$ if and only if for some $\varepsilon > 0$ we have $\{Y \in V : \|Y-X\| < \varepsilon\}$ is contained in $A$; that is, if and only if there is an "open ball" (of radius $\varepsilon$) centred on X which lies entirely in $A$.




Note: The above characterization is often used as a definition of interior point - see your Advanced Calculus notes for example.

PROOF: First note that from lemma 33.1 with $V \setminus A$ replacing $A$ we have that $X$ is in $\overline{V \setminus A}$ if and only if for each $\varepsilon > 0$ there exists a point $Y$ of $V \setminus A$ with $\|Y - X\| < \varepsilon$. Negating this we see that $X$ is not in $\overline{V \setminus A}$, and so by lemma 33.2 $X$ is in int $A$, if and only if for some $\varepsilon > 0$ there is no point of $V \setminus A$ closer to $X$ than $\varepsilon$. That is for this $\varepsilon$ the set $\{Y \in V : \|Y - X\| < \varepsilon\}$ is disjoint from $V \setminus A$ and so must lie entirely in $A$ as required.

EXERCISES:

1) Prove that the boundary of the "open-disk"

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| < 1\}$$

and of the "closed-disk"

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| \le 1\}$$

is the unit circle

$$\{\underset{\sim}{x} \in R^2 : \|\underset{\sim}{x}\| = 1\}$$

in the space $R^2$ with dot-product as inner-product.

2) Prove that the "open interval" $(0,1)$ in $R$ with absolute value as norm and the "open-disk" of exercise 1) are open sets.

3) For any two sets $A$ and $B$ in an inner-product space $V, (\ ,\ )$, show that

$$\text{int }(A) \cup \text{int }(B) \subseteq \text{int }(A \cup B).$$

Also, give an example to show that equality need not hold; that is, in general $\text{int}(A \cup B) \subsetneq \text{int }(A) \cup \text{int }(B)$.

*4) For any subset $A$ of the inner-product space $V, (\ ,\ )$ show that

i) $\text{int }(A)$ is an open set

ii) $\overline{A} = \text{int }(A) \cup \text{bdry }(A)$

## Lecture 34. Continuity

We conclude our brief introduction to the study of metric spaces by extending the notion of continuity to mappings between inner-product spaces.

A real valued function $f$ of a real variable is continuous at $x_o$

if, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon .$$

This may be generalized to a function

$$f : \mathbb{V} \to \mathbb{W}, \text{ where } \mathbb{V}, ( \, , \, )_V \text{ and } \mathbb{W}, ( \, , \, )_W$$

are two inner-product spaces, as follows.

DEFINITION: Let $\mathbb{V}$, $( \, , \, )_V$ and $\mathbb{W}$, $( \, , \, )_W$ be two inner-product spaces. The function $f : \mathbb{V} \to \mathbb{W}$ is continuous at $X_0 \in \mathbb{V}$ if, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|X - X_0\|_V < \delta \Rightarrow \|f(X) - f(X_0)\|_W < \varepsilon .$$

Here the two norms

$$\|X\|_V = \sqrt{(X,X)_V}$$

$$\|Y\|_W = \sqrt{(Y,Y)_W}$$

are used to replace the absolute value function on $\mathbb{R}$.

EXAMPLES:

1) Let $\mathbb{V} = \mathbb{R}^2$ and $( \, , \, )_V$ be "dot-product". Also let $\mathbb{W} = \mathbb{R}$ with ordinary multiplication as inner-product; that is, absolute value as norm. Then a function $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at the point $(x_0, y_0) \in \mathbb{R}^2$ if, given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x,y) - f(x_0,y_0)| < \varepsilon \text{ whenever } \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta.$$

This should be familiar to you as the definition of continuity used in the multivariable calculus.

2) Let $\mathbb{V} = \mathbb{R}^n$ and $\mathbb{W} = \mathbb{R}^k$ both equipped with dot product as inner-product and let $T$ be the linear mapping from $\mathbb{V}$ to $\mathbb{W}$ defined by

$$T(\underset{\sim}{x}) = M\,\underset{\sim}{x}$$

where $\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$ and M is given $k \times n$ - matrix, $M = [m_{ij}]$

We show that *T is continuous at each point of* $\mathbb{R}^n$.

We first observe that if

$$M_o = \max_{i,j} |m_{ij}| \quad ,$$

then for any $\underset{\sim}{x} \in \mathbb{R}^n$ we have

$$\|T(\underset{\sim}{x})\| = \|(\sum_{j=1}^{n} m_{1j}\, x_j \,, \sum_{j=i}^{n} m_{2j}\, x_j \,, \ldots , \sum_{j=1}^{n} m_{kj}\, x_j)\|$$

$$= (\sum_{i=1}^{k} (\sum_{j=1}^{n} m_{ij}\, x_j)^2)^{\frac{1}{2}}$$

$$\leqslant (\sum_{i=1}^{k} (\sum_{j=1}^{n} m_o\, |x_j|)^2)^{\frac{1}{2}}$$

$$= (\sum_{i=1}^{k} ((m_o, m_o, \ldots, m_o) \cdot (|x_1|, |x_2|, \ldots, |x_n|))^2)^{\frac{1}{2}}$$

and so by the Cauchy Schwarz inequality

$$\|T(\underset{\sim}{x})\| \leqslant (\sum_{i=1}^{k} (\|(m_o, m_o, \ldots, m_o)\| \times \|(|x_1|, |x_2|, \ldots, |x_n|)\|)^2)^{\frac{1}{2}}$$

$$= (\sum_{i=1}^{k} (\sqrt{\sum_{j=1}^{n} m_o^2} \times \sqrt{\sum_{j=1}^{n} x_j^2}\,)^2)^{\frac{1}{2}}$$

$$= (\sum_{i=1}^{k} (\sqrt{n}\, m_o\, \|\underset{\sim}{x}\|)^2)^{\frac{1}{2}}$$

$$= (k\, n\, m_o^2\, \|\underset{\sim}{x}\|^2)^{\frac{1}{2}}$$

$$= \sqrt{kn}\, m_o\, \|\underset{\sim}{x}\| \;.$$

That is $\|T(\underset{\sim}{x})\| \leqslant K\, \|\underset{\sim}{x}\|$ for some positive constant K.

The desired conclusion now follows from the linearity of T.

Thus, given any $\varepsilon > 0$ and $\underset{\sim}{x}_0 \in \mathbb{R}^n$, taking $\delta = \varepsilon/_K$ we have

$$\begin{aligned} \|T(\underset{\sim}{x}) - T(\underset{\sim}{x}_0)\| &= \|T(\underset{\sim}{x} - \underset{\sim}{x}_0)\| \\ &\leqslant K \|\underset{\sim}{x} - \underset{\sim}{x}_0\| \\ &< \varepsilon \end{aligned}$$

whenever $\|\underset{\sim}{x} - \underset{\sim}{x}_0\| < \delta$ and the continuity of T is proved.

[Because of this "automatic" continuity of all linear mappings on <u>finite dimensional</u> spaces analytic notions such as continuity are relatively unimportant in the study of elementary vector spaces and so the theory is largely algebraic.]

3) Let $\mathbb{V} = \mathbb{C}[0,1]$ with inner-product defined by

$$(f,g) = \int_0^1 f(x)\, g(x)\, dx.$$

Then the linear mapping defined by

$$T : \mathbb{V} \to \mathbb{R} : f \mapsto \int_0^1 f(x)\, dx$$

is continuous at each $f \in \mathbb{V}$. That is integration is a continuous operation. To see this note that

$$\begin{aligned} |T(f) - T(g)| &= \left| \int_0^1 \big(f(x) - g(x)\big)\, dx \right| \\ &= \left| \int_0^1 1 \times \big(f(x) - g(x)\big)\, dx \right| \\ &= |(1, f - g)| \quad \text{(here 1 denotes the constant function } 1(x) = 1 \text{ for all } x \in [0,1].) \\ &\leqslant \|1\|\, \|f - g\| \quad \text{(by the Cauchy Schwartz inequality)} \\ &= \|f - g\| \end{aligned}$$

[and so $T(g) \to T(f)$ as $g \to f$.]

4) Let $\mathbb{V}$ be the space of all real valued functions on $[0,1]$ with continuous first derivatives, and inner-product defined by

$$(f,g) = \int_0^1 f(x)\ g(x)\ dx.$$

Then the linear operator of differentiation

$$D : \mathbb{V} \to \mathbb{C}[a,b] : f \mapsto f'$$

is <u>not</u> continuous.

This follows directly from the next theorem if we observe that the sequence of functions $(f_n)$, where $f_n(x) = x^n$, is such that

$$\| f_n \|^2 = \int_0^1 f_n^2 = \int_0^1 x^{2n}\ dx = \frac{1}{2n+1} \to 0$$

so $(f_n)$ converges to the zero function, but

$$\| D(f_n) \|^2 = \int_0^1 n^2\ x^{2(n-1)}\ dx = \frac{n^2}{2(n-1)} \to \infty \qquad \text{as } n \to \infty,$$

so $\big(D(f_n)\big)$ does not converge to $D(0) = 0$.

[This example shows that in infinite dimensional spaces linear mappings need not in general be continuous, in contrast with the conclusion of example 2.]

A well known equivalent to the continuity of the real valued function f of a real variable at the point $x_0$ is that

$$\lim_{n \to \infty} f(x_n) = f(x_0) \quad \text{whenever } (x_n) \text{ is a}$$

sequence of real numbers which converges to $x_0$.

We conclude by establishing the same "sequential characterization" for the continuity of a function between inner-product spaces.

THEOREM 34.1: Let $\mathbb{V}$, $(\ ,\ )_V$ and $\mathbb{W}$, $(\ ,\ )_W$ be two inner-product spaces and let f be a function from $\mathbb{V}$ to $\mathbb{W}$, then f is continuous at $x_0 \in \mathbb{V}$ if and only if whenever the sequence $(x_n)$ in $\mathbb{V}$ is such that $x_n \to x_0$

we have $f(X_n) \to f(X_0)$.

<u>Proof</u>. ($\Rightarrow$) Given any $\varepsilon > 0$, since f is continuous at $X_0$ there exists $\delta > 0$ such that $\|f(X) - f(X_0)\|_W < \varepsilon$ whenever $\|X - X_0\| < \delta$. Now assume $X_n \to X_0$, then from the definition of convergence with $\varepsilon$ replaced by the above $\delta$, we have; there exists N such that

$$\|X_n - X_0\| < \delta \quad \text{for all } n > N.$$

It therefore follows for $n > N$ that $\|f(X_n) - f(X_0)\| < \varepsilon$, and so by definition $f(X_n) \to f(X_0)$ as required.

($\Leftarrow$) Here we establish the contrapositive. That is we show that if f is not continuous there exists at least one sequence $(X_n)$ converging to $X_0$ for which $f(X_n)$ does not converge to $f(X_0)$. Thus assume f <u>is not</u> continuous at $X_0$; that is, there exists an $\varepsilon > 0$ such that for each $\delta > 0$ there is at least one point X with $\|X - X_0\|_V < \delta$ but $\|f(X) - f(X_0)\|_W \geq \varepsilon$ (otherwise, for every $\varepsilon$ there would be a $\delta$ such that

$$\|X - X_0\|_V < \delta \Rightarrow \|f(X) - f(X_0)\|_W < \varepsilon$$

and so f would be continuous at $X_0$).

In particular then for each natural number n, taking $\delta = \frac{1}{n}$, we have that there exists a point $X_n$ such that

$$\|X_n - X_0\|_V < \frac{1}{n} \quad \text{but} \quad \|f(X_n) - f(X_0)\|_W \geqslant \varepsilon$$

The sequence $(X_n)$ constructed in this way is convergent to $X_0$

$(\|X_n - X_0\|_V < \frac{1}{n} \to 0$ as $n \to \infty)$ but $f(X_n)$ does not converge to $f(X_0)$

(for all n $\|f(X_n) - f(X_0)\|_W \geq \varepsilon > 0$) and the result is proved.

EXERCISE.

(a) For any pair of inner-product spaces $\mathbb{V}$, $(\ ,\ )_V$ and $\mathbb{W}$, $(\ ,\ )_W$ show that a mapping $f : \mathbb{V} \to \mathbb{W}$ which satisfies a Lipschitz condition

$$\|f(X) - f(Y)\|_W \leq k\|X - Y\|_V^{\alpha}$$

(where $k \geq 0$ and $\alpha > 0$ are given constants) is continuous.

(b) For $\mathbb{R}$ with absolute value as norm show that any real valued function of a real variable with a bounded derivative satisfies a Lipschitz condition (with exponent $\alpha = 1$). [Hint: use the mean value theorem.]

(c) Show that the only real valued functions of a real variable satisfying Lipschitz conditions of the form

$$|f(x) - f(y)| \leq k|x - y|^2$$

are constant functions.

(d) Show that a linear function from $\mathbb{V}$ to $\mathbb{W}$ satisfies the Lipschitz condition of (a) if and only if

$$\|T(X)\|_W \leq k \|X\|_V^\alpha \quad \text{for all } X \text{ in } \mathbb{V}$$

and

(**) show that this is only possible for T not identically 0 if $\alpha = 1$.