Topics In Linear Algebra and Its Applications

James Emery

Edition 11/29/2015

Contents

1 Introduction

Linear algebra is the study of finite dimensional vector spaces and linear transformations. A vector space is a quadruple $(V, F, +, *)$. V is a set of vectors, F is a field of scalars, + is the operation of vector addition, and * is the operation of scalar multiplication. We usually do not write the multiplication operator, that is, we write $\alpha * v$ as αv. Let $\alpha, \beta \in F$ and $u, v, w \in V$. The following axioms are satisfied:

1. $u + v = v + u$
2. $u + (v + w) = (u + v) + w$
3. There is a zero element so that $0 \in V$ so that $u + 0 = u$.
4. For each $u \in V$, there is an inverse element -u so that $u + (-u) = 0$
5. $\alpha(u + v) = \alpha u + \alpha v$.
6. $(\alpha + \beta)u = \alpha u + \beta u$.
7. $(\alpha\beta)u = \alpha(\beta u)$.
8. $1u = u$

A finite set of vectors $v_1, v_2, ...v_n$ is linearly independent if

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_2 = 0$$

implies that each α_i is zero. Otherwise the set is called linearly dependent. A subset S of V is a subspace of V if the sum of any two elements in S is in S and the scalar product of any element in F with any element in S is in S. That is S is closed under addition and scalar multiplication. The subspace spanned by vectors

$$v_1, v_2, ...v_n$$

is

$$S = \{\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n : \alpha_i \in F\}$$

Theorem The nonzero vectors $v_1, v_2, ...v_n$ are linearly dependent if and only if some one of them is a linear combination of the preceding ones.
Proof. Suppose v_k can be written as a linear combination of $v_1, ...v_{k-1}$. Then we have a linear combination of $v_1, ...v_k$ set equal to zero with $\alpha_k = -1$, so that these vectors are linearly dependent. Conversely, suppose $v_1, ...v_n$ are dependent. Then we can find a set of α_i so that

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n = 0,$$

and at least one of them is not zero. Let k be the largest index for which α_i is not zero, then dividing by α_k we find that v_k is a linear combination of the preceding vectors.
Corollary. Any finite set of vectors contains a linear independent subset that spans the same space.
Theorem. Let vectors $v_1, v_2, ..., v_n$ span V. Suppose the vectors $u_1, u_2, ..., u_k$ are linearly independent. Then $n \geq k$.
Proof. The set $u_1, v_1, v_2, ..., v_n$ is linearly dependent and spans V, so some v_j is dependent on its predecessors. Then the set $u_1, v_1, v_2, ..., v_{j-1}, v_{j+1}, ..., v_n$ spans V, and is dependent. We may continue this, adding a u_i while removing a v_j and still having a set that spans V and is dependent. This can be continued until the u_i are exhausted, otherwise the v_j would be exhausted first and some subset of the $u_1, u_2, ..., u_k$ would then be dependent, which is not possible. Therefore there are more v_j than u_i, which forces

$$n \geq k.$$

Definition A *basis* of a vector space V is a set of linearly independent vectors that spans V. A vector space is finite dimensional if it has a finite basis.

Theorem Suppose a vector space has a finite basis

$$A = \{v_1, v_2, ..., v_n\}.$$

Then any other basis also has n elements.
Proof Let $B = \{u_1, u_2, u_3, ...\}$, which is possibly infinite, be a second basis of V. By the previous theorem $n \geq k$ for any subset

$$u_1, u_2, ..., u_k$$

of B. It follows that

$$B = \{u_1, u_2, u_3, ..., u_m\},$$

for some m, and that $n \geq m$. Reversing the role of A and B, we apply the previous theorem again to get $m \geq n$, which proves the corollary. We conclude that the dimension of a finite dimensional vector space can be well defined as the number of elements in any basis. Any vector $v \in V$ can be represented as a linear combination of the basis elements:

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + ... + \alpha v_n v_n.$$

This representation is unique, because if we subtract two different representations we would get a representation of the zero vector as a linear combination of the basis vectors with at least one nonzero coefficient, which contradicts that the vectors are linearly independent.

The scalar coefficients are called the coordinates of v, and form a element in the cartesian n-product of the scalar field F. These n-tuples of scalars themselves form an n dimensional vector space and are isomorphic to the original vector space. Let U and V be two vector spaces. A function

$$T : U \rightarrow V$$

is called a linear transformation if

1. For $u_1, u_2 \in U, T(u_1 + u_2) = T(u_1) + T(u_2)$.

2. For $\alpha \in F, u \in U, T(\alpha u) = \alpha T(u)$.

A linear transformation from U to itself, is called a linear operator. Associated with every linear transformation is a matrix. Let

$$\{u_1, u_2, u_3, ..., u_n\}$$

be a bases of U and let

$$\{v_1, v_2, ..., v_m\}$$

a bases of V. For each u_j, $T(u_j)$ is in V, so that it may be written as a linear combination of the basis vectors, we have

$$T(u_j) = \sum_{i=1}^{m} a_{ij} v_i.$$

Now let $u \in U$ and let its coordinates be $x_1, ..., x_n$. Vector u is represented by the coordinate vector

$$\begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$$

We have

$$\begin{aligned} T(u) = T(\sum_{j=1}^{n} x_j u_j) &= \sum_{j=1}^{n} x_j T(u_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} v_i \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} x_j) v_i \\ &= \sum_{i=1}^{m} y_i v_i, \end{aligned}$$

where the $y_1, y_2, ..., y_m$ are the components of the vector $T(u)$ in vector space V. We have shown that the coordinate vector x of u is mapped to the coordinate vector y of $T(u)$ by matrix multiplication,

$$\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & ...... & a_{1n} \\ a_{21} & ...... & a_{2n} \\ ... & ...... & ... \\ a_{m1} & ...... & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$$

Now suppose we have two linear transformations

$$T : U \to V,$$

and

$$S : V \to W.$$

The composite transformation is

$$ST : U \to W,$$

defined by $ST(u) = S(T(u))$.

Theorem. If A is the matrix of linear transformation S, and B is the matrix of linear transformation T, then the matrix multiplication product AB is the matrix of ST.

2 Linear Transformations and Linear Operators

Examples:

Finite dimensional linear transformation

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & a_{32} & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

Integral linear transformation

$$f \to g,$$

where

$$g(y) = \int k(x, y) f(x) dx$$

Differential operator

$$f \to g,$$

$$g(x) = (5\frac{d^2}{dx^2} + 2\frac{d}{dx} - 3) f(x)$$

3 Determinants

A determinant is a multilinear functional defined on a square matrix. A linear functional is a mapping from a vector space to a field of scalars. A multilinear functional is a function defined on a cartesian product of the vector space. The column vectors of a matrix may be considered to be vectors of the vector space V. The set of n column vectors constitute a point in the cartesian product. The functional is linear in the sense that

$$f(v_1, v_2, ..., v_k + v'_k, ..., v_n) = f(v_1, v_2, ..., v_k, ..., v_n) + f(v_1, v_2, ..., v'_k, ..., v_n),$$

and

$$f(v_1, v_2, ..., \alpha v_k, ..., v_n) = \alpha f(v_1, v_2, ..., v_k, ..., v_n).$$

A multilinear functional is alternating if interchanging a pair of variables changes the sign of the function.
Definition. A determinant $D(A)$ of an n dimensional square matrix A is the unique alternating multilinear functional defined on the n column vectors of A, which takes value 1 on the identity matrix.
Properties.

1. If two column vectors of a matrix are identical, then the determinant is zero. This follows because interchanging the columns changes the sign of the determinant, but the new matrix has not changed, so the value of the determinant is the same. The determinant must be zero.

2. Adding a multiple of one column to a second does not change the value of the determinant. This is clear from

$$D(v_1, ..., v_i, ..., v_j + \alpha v_i, ..., v_n)$$
$$= D(v_1, ..., v_i, ..., v_j, ..., v_n) + \alpha D(v_1, ..., v_i, ..., v_i, ..., v_n)$$
$$= D(v_1, ..., v_i, ..., v_j, ..., v_n) + 0.$$

Example To compute the determinant of

$$\begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}$$

subtract the first column from the second, then three times the second from the first, getting

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

So $D(A) = 2D(I) = 2$, where I is the identity matrix. Once we have a matrix in diagonal form, we see from its definition as a multilinear functional, that the determinant is equal to the product of each multiplier of each column times the determinant of the identity. That is, the value equals the product of the diagonal elements.

Example *Cramers's Rule* Suppose we have a system of n equations in n unknowns $x_1, x_2, ..., x_n$ written in the form

$$x_1 v_1 + x_2 v_2 + ..., x_n v_n = v.$$

We have

$$D(v, v_2, v_3, ..., v_n) = D(x_1 v_1 + x_2 v_2 + ... + x_n v_n, v_2, ..., v_n)$$

$$= x_1 D(v_1, v_2, v_3, ..., v_n).$$

So that unknown x_1 is given by

$$x_1 = \frac{D(v, v_2, v_3, ..., v_n)}{D(v_1, v_2, v_3, ..., v_n)}.$$

There is clearly a similar expression for each of $x_2, ...x_n$.

To compute a determinant we can perform permutations on the columns and add scalar multiples of columns to other columns, to put the matrix into diagonal form. Once in diagonal form, (or triangular form) the determinant equals the product of the diagonal elements.

There is an alternate definition of the determinant involving permutations. Let a be a n by n matrix. Consider the sum

$$\sum_{\sigma} s(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} ... a_{n\sigma(n)}$$

where σ is a permutation of the integers $1, 2, 3, 4, ..., n$ and $s(\sigma)$ is the sign of the permutation. The sign of the identity permutation is one, and interchanging a pair of elements, a transposition, changes the sigh of the permutation.

Notice that this is a multilinear functional, and is alternating, and further equals one on the identity matrix. Therefore it is the unique such functional, and so is equal to the determinant.

Properties.

1. $D(A^T) = D(T)$.
2. $D(AB) = D(A)D(B)$.
3. Expansion by minors about row i. Let A_{ij} be the matrix obtained from A by deleting row i and column j. Then

$$D(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} D(A_{ij}).$$

4 Solving Linear Equations

Suppose we want to solve a linear equation of the form

$$AX = B,$$

where A is a square matrix of size n

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix},$$

X is a column vector to be solved for

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix},$$

and a B is a right side vector

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_n \end{bmatrix}.$$

This can be written equivalently with the $x_1, x_2, ..., x_n$ coefficient values multiplying column vectors of A. That is,

$$x_1 A_1 + x_2 A_2 + ... + x_n A_n = B,$$

where A_j is the jth column vector of A,

$$A_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ ... \\ a_{nj} \end{bmatrix}.$$

This can be accomplished by Gaussian elimination, by repeatedly adding multiples of rows, so as to zero values in the matrix. This puts the set of equations in triangular form.

If the determinant of A is not zero, then the column vectors of A are linearly independent, and they form an n dimensional bases so that the each x_j is a coefficient, and such values of the coefficents can be found so that a linear combination of the column vectors gives the right side vector B. Thus if the determinant is nonzero, there is a unique solution vector

$$X = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$$

It is given by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} = A^{-1} B,$$

where A^{-1} is the inverse of A.

On the other hand if $\det(A)$ is zero, the equation may not have a solution. But consider the homogeneous equation where $B = 0$. Then the column vectors of A are not linearly independent, they are dependent because the determinant of A is zero. This means that there is a linear combination of vectors, with coefficients $x_1, x_2, ...x_n$, not all zero, so that

$$x_1 A_1 + x_2 A_2 + ... + x_n A_n = 0.$$

That is, there is always a nonzero solution to the equation in the homogeneous case, when the vector B is zero. There always exist a solution to the equation when B is zero, and the determinant of A is zero.

On the otherhand, if the determinant of A is not sero, then the inverse of A exists, so the unique solution to this homogeneous equation is given by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} = A^{-1}B = 0.$$

5 Permutations

See the paper **Group Theory** by James Emery (Document groups.tex).

6 Kernel and Range

Let U be a n dimensional vector space. Let T be a linear transformation from U to a vector space V. Let $K(T)$ be the kernel of T and $R(T)$ the range of T. Then

$$dim(K(T)) + dim(R(T)) = n$$

Let $u_1, u_2, ..., u_p$ be a basis of the kernel of T. It can be extended to be a full basis of U. Suppose V is also n dimensional. Let A be the matrix of T with respect to this basis. Then the first column of A must be all zeroes, so that $D(A)$ is zero. Let U have a second basis and let S be the linear transformation mapping the first bases to the second. Let B be the matrix of S. Then B has an inverse B^{-1} and

$$D(B)D(B^{-1}) = D(BB^{-1}) = D(I) = 1.$$

Then $D(B)$ is not zero. The matrix of T with respect to the second basis is $D(AB) = D(A)D(B)$, so that in general the determinant of the matrix of T with respect to bases of U and V is zero if and only if the kernel of T is not zero, and this is true if and only if T has an inverse.

7 Quotient Spaces

8 Direct Sums

9 The Inner Product

The inner product of two vectors x and y in a complex or real vector space is written (x, y). The dot product of vector analysis is an inner product. The inner product has the following properties:

1. $(x, y) = \overline{(y, x)}$,
2. $(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1(x_1, y) + \alpha_2(x, y)$,
3. $(x, x) \geq 0; (x, x) = 0$, if and only if $x = 0$.

From these properties we get

$$(x, \alpha y) = \overline{(\alpha y, x)} = \overline{\alpha(y, x)} = \overline{\alpha}\overline{(y, x)} = \overline{\alpha}(x, y)$$

An example of an inner product on a vector space of complex valued continuous functions of a real variable, say defined on an interval $[a, b]$ is

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

10 Inner Product Spaces

Let V be a vector space with an inner product (u, v) . A basis can be used to construct an orthonormal basis (Gram-Schmidt Orthogonalization). An inner product space is also a normed linear space with the norm

$$\|u\| = (u, u)^{1/2}.$$

The Cauchy-Schwarz inequality is

$$|(u, v)| \leq \|u\|\|v\|.$$

The triangle inequality is

$$\|u + v\|^2 \leq \|u\|^2 + \|v\|^2.$$

We look more at these below.

11 Normed Linear Spaces

A norm of a vector in Euclidean space is the square root of the sum of its components squared, which of course comes from the Pythagorean Theorem. A norm for a vector x in a general vector space, written as $\|x\|$, satisfies the following properties

(i) $\|x\| > 0$ if x is not zero.

(ii) $\| \ ax\| = |a|\|x\|$ if x is a scalar.

(iii) $\|x+y\| \leq \|x\| + \|y\|$.
Property (iii) is called the triangle inequality. A vector space with a norm is called a normed linear space. A normed linear space has a distance function or metric defined by

$$d(x,y) = \|x-y\|,$$

and thus is a metric space. An inner product space has a norm defined by

$$\|x\| = \sqrt{(x,x)}.$$

These things are proved in the subsection below called **The Triangle Inequality**.

A normed linear space that is a complete metric space, is called a Banach Space after Stefan Banach. Polish mathematician Sefan Banach (March 30, 1892 - August 31, 1945) was one of the greatest 20th century mathematicians.

12 Orthonormal Vectors

A set of vectors $S = \{x_i\}$ is an orthonormal set if each pair is orthogonal

$$(x_i, x_j) = 0,$$

and each vector has unit norm

$$(x_i x_i) = \|x_i\| = 1.$$

If x is an arbitrary vector and $S = \{x_i\}$ is an orthonormal set, then the coefficients

$$c_i = (x, x_i),$$

are called the Fourier coefficients of x with respect to the set S. If x can be expressed as a linear combination of the orthonormal vectors S, then there is a Fourier expansion of x

$$x = \sum_{i=1}^{n} c_i x_i.$$

Examples of orthonormal sets include the trigonometric functions and the various orthogonal polynomials used widely in applied mathematics.

13 Gram-Schmidt Orthogonalization

Given a vector v and a vector u, the projection of v onto u has length

$$\|v\| \cos(\theta),$$

where θ is the angle between v and u. This can be written as

$$\frac{(v, u)}{\|u\|}$$

The projection of v in the direction of u, or the component of v in the direction u is

$$\frac{(v, u)}{\|u\|} \frac{u}{\|u\|} = \frac{(v, u)}{\|u\|^2} u.$$

So suppose we have an orthogonal set $W = w_1, w_2, ...w_k$ and a vector v, not in W. Let V be the space spanned by W and v. We can find a replacement for v, w_{k+1}, so that $w_1, w_2, ...w_k, w_{k+1}$ is an orthogonal basis for V. We subtract from v its projections in the direction of each of the $w_1, w_2, ...w_k$. So

$$w_{k+1} = v - \frac{(v, w_1)}{\|w_1\|^2} w_1 - \frac{(v, w_2)}{\|w_2\|^2} w_2 - \frac{(v, w_k)}{\|w_k\|^2} w_k$$

Using this process, given a basis of a space, we can compute from this basis an orthogonal basis. See **Numerical Methods,** Dahlquist and Bjorck, 1974, Prentice-Hall, for the modified Gram-Schmidt process, which is an equivalent process with better numerical stability.

14 The Pythagorean Theorem for Inner Product Spaces

Proposition If two vectors u and v in an inner product space are orthogonal, then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof.

$$\|u+v\|^2 = (u+v, u+v) = (u,u) + (v,u) + (u,v) + (v,v) = \|u\|^2 + \|v\|^2.$$

15 Inequalities

15.1 The Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality. If x and y are two vectors in an inner product space, then

$$|(x,y)| \leq \|x\|\|y\|,$$

and we have equality if and only if x and y are dependent.
Proof. Given two vectors x and y, we have for any real or complex number λ

$$\begin{aligned} 0 &\leq \|x - \lambda y\| \\ &= (x - \lambda y, x - \lambda y) \\ &= (x,x) + (x, -\lambda y) + (-\lambda y, x) + (-\lambda y, -\lambda y) \\ &= \|x\|^2 - \bar{\lambda}(x,y) - \lambda(y,x) + |\lambda|^2\|y\|^2. \end{aligned}$$

Let

$$\lambda = \frac{(x,y)}{(y,y)}$$

Then

$$\bar{\lambda} = \frac{(y,x)}{(y,y)}.$$

The inequality becomes

$$0 \leq \|x\|^2 - \frac{|(x,y)|^2}{\|y\|^2}.$$

So

$$|(x,y)|^2 \le \|x\|^2\|y\|^2,$$

so

$$|(x,y)| \le \|x\|\|y\|.$$

If x is a multiple of y, say

$$x = \alpha y,$$

then

$$|(x,y)| = |\alpha|\|y\|^2 = |\alpha|\|y\|\|y\| = \|x\|\|y\|.$$

So we have equality.

Conversely, if we have equality, then x is a multiple of y. For if x and y are not dependent then for any λ

$$x - \lambda y$$

is not zero, and so in the prove above we would start with

$$0 < \|x - \lambda y\|,$$

and carrying through the steps of the proof above we find

$$|(x,y)| < \|x\|\|y\|.$$

So if we have equality, then x must be a multiple of y.

Real Vector Space Proof. If we are in a real vector space, we can give an alternate proof.

The equation above

$$0 \le \|x\|^2 - \bar{\lambda}(x,y) - \lambda(y,x) + |\lambda|^2\|y\|^2,$$

becomes

$$\begin{aligned} 0 &\le \|x\|^2 - 2\lambda(x,y) + \lambda^2\|y\|^2 \\ &= A\lambda^2 + B\lambda + C, \end{aligned}$$

where $A = \|y\|^2$, $B = 2(x,y)$, and $C = \|x\|^2$. So we obtain a quadratic function in λ of the form

$$f(\lambda) = A\lambda^2 + B\lambda + C \ge 0.$$

Because this function is nonnegative for any λ, it does not have a pair of real roots. So the discriminate

$$B^2 - 4AC$$

is not positive. Thus

$$B^2 \leq 4AC,$$

so

$$4|(x,y)|^2 \leq 4\|x\|^2\|y\|^2.$$

Thus

$$|(x,y)| \leq \|x\|\|y\|,$$

which is the Cauchy-Schwarz inequality.

A Third Proof. Given u and v, let us perform a Gram-Schmidt orthogonalization, to get a vector w that is orthogonal to v

$$w = u - \frac{(u,v)v}{\|v\|^2}$$

Then

$$u = \frac{(u,v)v}{\|v\|^2} + w$$

is an orthogonal decomposition of u. By the pythagorean theorem

$$\|u\|^2 = \|\frac{(u,v)v}{\|v\|^2}\|^2 + \|w\|^2 \geq \|\frac{(u,v)v}{\|v\|^2}\|^2 = \frac{|(u,v)|^2}{\|v\|^2},$$

with equality iff $w = 0$, that is iff u and v are dependent, that is iff u is a multiple of v.

A Fourth Proof. Given vectors x, y, there exists a λ and a vector z orthogonal to x so that

$$y = \lambda x + z,$$

so that

$$z = y - \lambda x.$$

Then the condition $(x,z) = 0$ gives

$$\bar{\lambda}(x,x) = (x,y),$$

and

$$\bar{\lambda} = \frac{(x, y)}{(x, x)}.$$

The pythagorean theorem gives

$$(y, y) = |\lambda|^2 (x, x) + (z, z)$$

and so

$$|\lambda|^2 (x, x) = (y, y) - (z, z).$$

Then squaring

$$|\bar{\lambda}(x, x)| = |(x, y)|,$$

we have

$$\begin{aligned} |(x, y)|^2 &= |\lambda|^2 (x, x)^2 = |\lambda|^2 (x, x)(x, x) \\ &= [(y, y) - ((z, z)](x, x) \leq (x, x)(y, y). \end{aligned}$$

The inequality reduces to equality iff $(z, z) = 0$ iff $y = \lambda x$.

15.2 The Sequence Space ℓ^2

The space ℓ^2 consists of sequences $s = \{s_n\}_1^\infty$ of real or complex numbers such that

$$\sum_{n=1}^{\infty} |s_n|^2 < \infty.$$

Theorem *(Cauchy-Schwartz Inequality)*. If s and t are in ℓ^2 then

$$|(s, t)| = \left| \sum_{i=1}^{\infty} s_i \overline{t_i} \right| \leq \left[\sum_{i=1}^{\infty} |s_i|^2 \right]^{1/2} \left[\sum_{i=1}^{\infty} |t_i|^2 \right]^{1/2} = \|s\| \|t\|.$$

Proof. The finite sequences of length n clearly form an inner product space, with inner product

$$(s, t) = \sum_{i=1}^{n} s_i \overline{t_i},$$

and norm

$$\|s\| = \left[\sum_{i=1}^{n} |s_i|^2\right]^{1/2},$$

and so satisfy the Cauchy-Schwarz inequality

$$\left|\sum_{i=1}^{n} s_i\overline{t_i}\right| = |(s,t)| \leq \|s\|\|t\| = \left[\sum_{i=1}^{n} |s_i|^2\right]^{1/2} \left[\sum_{i=1}^{n} |t_i|^2\right]^{1/2}.$$

Replacing s_i by $|s_i|$ and $\overline{t_i}$ by $|t_i|$, we get

$$\sum_{i=1}^{n} |s_i t_i| \leq \left[\sum_{i=1}^{n} |s_i|^2\right]^{1/2} \left[\sum_{i=1}^{n} |t_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{\infty} |s_i|^2\right]^{1/2} \left[\sum_{i=1}^{\infty} |t_i|^2\right]^{1/2} < \infty,$$

which shows that

$$\sum_{i=1}^{\infty} s_i\overline{t_i}$$

is absolutely convergent, and so convergent. Letting $n \to \infty$, for $s, t \in \ell^2$ we have the **Cauchy-Schwarz inequality for** ℓ^2

$$|(s,t)| = \left|\sum_{i=1}^{\infty} s_i\overline{t_i}\right| \leq \left[\sum_{i=1}^{\infty} |s_i|^2\right]^{1/2} \left[\sum_{i=1}^{\infty} |t_i|^2\right]^{1/2} = \|s\|\|t\|.$$

This will have shown that ℓ^2 is an inner product space, with this inner product, after we have shown that it is indeed a vector space. We shall do this next by proving the **Minkowski inequality**.

Theorem *(The Minkowski Inequality).* If s and t are in ℓ^2 then

$$\left[\sum_{i=1}^{\infty} |s_i + t_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{\infty} |s_i|^2\right]^{1/2} + \left[\sum_{i=1}^{\infty} |t_i|^2\right]^{1/2}.$$

Proof. We have

$$|s_i + t_i|^2 = (s_i + t_i)\overline{(s_i + t_i)} = |s_i|^2 + s_i\overline{t_i} + \overline{s_i}t_i + |t_i|^2.$$

So we have using the Cauchy-Swartz inequality

$$\sum_{i=1}^{\infty} |s_i + t_i|^2 = \sum_{i=1}^{\infty} (|s_i|^2 + s_i \overline{t_i} + \overline{s_i} t_i + |t_i|^2)$$

$$\leq \sum_{i=1}^{\infty} |s_i|^2 + 2 \left(\sum_{i=1}^{\infty} |s_i|^2 \right)^{1/2} \left(\sum_{i=1}^{\infty} |t_i|^2 \right)^{1/2} + \sum_{i=1}^{\infty} |t_i|^2.$$

So

$$\sum_{i=1}^{\infty} |s_i + t_i|^2 = \left[\left(\sum_{i=1}^{\infty} |s_i|^2 \right)^{1/2} + \left(\sum_{i=1}^{\infty} |t_i|^2 \right)^{1/2} \right]^2$$

Taking the square root of both sides we have proved the result.

The Minkowski inequality shows that If s and t are in ℓ^2 then $s + t$ is in ℓ^2. So ℓ^2 is a vector space and is an inner product space. Also we can show that it is complete, so is a Hilbert space. A sequence is a function defined on the domain of natural numbers

$$s(k) = s_k,$$

so ℓ^2 is a simple function space.

15.3 Bessel's Inequality

Proposition If $S = \{x_i, i = 1..n\}$ is an orthonormal set (they are unit vectors that are mutually orthogonal), then

$$\sum_{i=1}^{n} |(x, x_i)|^2 \leq \|x\|^2.$$

Proof. We have

$$0 \leq \|x - \sum_{i=1}^{n} (x, x_i) x_i\|^2$$

$$= (x - \sum_{i=1}^{n} (x, x_i) x_i, x - \sum_{j=1}^{n} (x, x_j) x_j)$$

$$= \|x\|^2 - \sum_{j=1}^{n} (x, x_j)(x, x_j) - \sum_{i=1}^{n} (x, x_i)(x_i, x) + \sum_{i=1}^{n} \sum_{j=1}^{n} ((x, x_i)(x, x_j)(x_i, x_j)$$

$$= \|x\|^2 - \sum_{j=1}^{n}(x, x_j)(x, x_j) - \sum_{i=1}^{n}(x, x_i)(x_i, x) + \sum_{j=1}^{n}((x, x_j)(x, x_j)$$

$$= \|x\|^2 - \sum_{i=1}^{n}(x, x_i)(x_i, x)$$

$$= \|x\|^2 - \sum_{i=1}^{n} |(x, x_i)|^2$$

That is

$$\sum_{i=1}^{n} |(x, x_i)|^2 \leq \|x\|^2.$$

The coefficients

$$(x, x_i)$$

are called the Fourier coefficients of x with respect to the orthonormal set $\{x_i, i = 1..n\}$. If x is a linear combination of the set S,

$$x = \sum_{i=1}^{n} \alpha_i x_i,$$

then

$$(x, x_i) = \alpha_i(x_i, x_i) = \alpha_i.$$

So Bessel's inequality is an equality if and only if x is a linear combination of elements of S.

The Cauchy-Schwarz inequality can be proved from the Bessel inequality. Indeed, given x and y, let

$$x_1 = \frac{y}{\|y\|}.$$

Then the set containing just this vector is an orthonormal set, so we have by the Bessel inequality

$$\|x\|^2 \geq |(x, x_1)|^2 = |(x, y/\|y\|^2 = |(x, y)|^2/\|y\|^2.$$

Thus

$$|(x, y)| \leq \|x\|\|y\|.$$

15.4 The Triangle Inequality: an Inner product Space is a Normed Linear Space, and a Metric Space

$$\|x+y\|^2 = (x+y, x+y)$$

$$= \|x\|^2 + (x,y) + (y,x) + \|y\|^2$$
$$= \|x\|^2 + 2RE(x,y) + \|y\|^2$$
$$\leq \|x\|^2 + 2|(x,y)| + \|y\|^2$$
$$\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$$
$$= (\|x\| + \|y\|)^2.$$

So

$$\|x+y\| \leq \|x\| + \|y\|$$

So $\sqrt{(x,x)}$ is a norm, and with this norm an inner product space is a normed linear space. Further, a normed linear space is a metric space. Define a metric by

$$d(x,y) = \|x-y\|$$

This is a metric, a distance function. Indeed we have

$$\|x-y\| = \|x-z+z-y\| \leq \|x-z\| + \|z-y\|.$$

16 The Parallelogram Law

Consider a parallelogram in the plane. Let s_1 and s_2 be vectors on the sides of the parallelogram. Let vectors d_1 and d_2 be the diagonals. Let x and y be the half diagonals. Then

$$s_1 = x+y,$$

and

$$s_2 = x-y$$

Draw a figure to see this. We find in general for any x and y in an inner product space that

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2,$$

which follows by expanding the inner products. In the case of our parallelogram, we have

$$\|s\|^2 + \|s_2\|^2 = \frac{\|d_1\|^2 + \|d_2\|^2}{2}.$$

That is *The sum of the squares of the sides of a parallelogram is equal to the average of the squares of the two diagonals.* Thus the general identity

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

is called the parallelogram law.

17 Quadratic Forms

A symmetric matrix defines a quadratic form, a homogeneous second degree algebraic function. So, if matrix M is symmetric, and x is a column vector, then

$$q(x) = x^T M x,$$

is a quadratic form. For example,

$$(x_1 + 5x_2 + 2x_3)^2 = x_1^2 + 10x_1x_2 + 4x_1x_3 + 25x_2^2 + 20x_2x_3 + 4x_3^2,$$

is a quadratic form. Each term in the expansion is of the second degree. As a matrix equation this is

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & 5 & 2 \\ 5 & 25 & 10 \\ 2 & 10 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Every quadratic form can be diagonalized. That is there is a change of basis so that the matrix M becomes a diagonal matrix and the quadratic form becomes

$$q(x) = \sum_{i=1}^{n} a_i x_i^2.$$

Quadratic forms arise in various practical applications such as in vibration problems and in expressions for energy. Diagonalizing corresponds to the introduction of the so called normal coordinates. The elements on the diagonal of the diagonal matrix are the eigenvalues.

18 Canonical Forms

When matrices can not be diagonalized, there exists various canonical forms, which characterize the matrices and its properties. Among these are: Jordon Normal form, Rational Normal form, Triangular form, and LU decomposition.

19 Upper Triangular Form

For every matrix M, there is a a matrix T so that

$$T^{-1}MT$$

is upper triangular. One may make an eigenvector of M the first column of T, to get a partitioned matrix where the first column has zeroes below the first element of the first column. Then one may apply induction to the remaining lower dimensional submatrix. See Richard Bellman **The Stability Theory of Differential Equations** Dover, for a simple proof.

20 Isometries, Rotations, Orthogonal Matrices

An isometry is a transformation that preserves length.

$$||T(v)|| = ||v||.$$

Let T be a linear transformation that is an isometry. Let T have a matrix representation M in some orthogonal basis. An orthogonal matrix is a matrix whose column vectors are orthogonal and each has norm equal to one. So an orthogonal matrix has its transpose as its inverse. This also implies that its row vectors are orthonormal. We shall show that M is an orthogonal matrix.

We have

$$Mv \cdot Mv = v \cdot v,$$

because T is an isometry. So if v is a column vector, then

$$(Mv)^T(Mv) = v^T M^T Mv = v^T v.$$

Let v_i be the column vector whose ith element is one, and all other elements are zero. To show that M is an orthogonal matrix we will use a Lemma.
Lemma. If P is any n by n matrix, then

$$v_i^T P v_j = P_{ij},$$

where v_i be the column vector with n elements, where the ith element is 1 and all are others 0.

Proof. We will make this obvious with a simple example. Suppose we have a 4 by 4 matrix P, and let $i = 2$ and $j = 3$, then

$$v_2^T P v_3 =$$

$$[0100] \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} p_{21} & p_{22} & p_{23} & p_{24} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = p_{23}.$$

Proposition. Linear transformation T is an isometry, if and only if its matrix M is an orthogonal matrix.
Proof. From above we have

$$(Mv)^T(Mv) = v^T M^T M v = v^T v.$$

Let $P = M^T M$ (we write the elements of P using a lowercase p, so the element in row i and column j is p_ij). we have

$$v_i^T P v_i = v_i^T v_i = 1.$$

From the lemma we have $p_{ii} = 1$. For i not equal to j, we have

$$p_{ij} = v_i^T P v_j = v_i^T v_j = 0.$$

We can verify this again with a little different argument

$$(v_i + v_j)^T P(v_i + v_j) = (v_i + v_j)^T (v_i + v_j) = v_i^T v_i + v_j^T v_j = 2.$$

And

$$(v_i + v_j)^T P(v_i + v_j)$$

$$= p_{ii} + p_{ji} + p_{ij} + p_{jj} = p_{ji} + p_{ij} + 2.$$

Thus

$$p_{ji} + p_{ij} = 0,$$

or

$$p_{ji} = -p_{ij}.$$

But

$$P = M^T M$$

is symmetric, so

$$p_{ij} = 0$$

and so P is the identity.

We have shown that The transpose of M is the inverse of M, so M is an orthogonal matrix.

Conversely, if M is an orthogonal matrix then for any v

$$\|Mv\|^2 = (Mv)^T(Mv) = v^T M^T M v = v^T I v = v^T v = \|v\|^2,$$

so matrix M represents an isometry. **This completes the proof.**

For any orthogonal matrix M, we have

$$I = M^T M,$$

so

$$1 = \det M^T M = \det M^T \det M = (\det M)^2.$$

The determinant must be 1, or -1.

In two space or three space, a rotation matrix, representing a rotation about a fixed axis, is clearly an isometry. Hence it is orthogonal. Any real eigenvalue of an isometry must be 1 or -1, for suppose λ is an eigenvalue and v an eigenvector of isometry transformation Y. Then both

$$T(v) = \lambda v$$

and

$$||T(v)|| = ||v||.$$

So

$$||v|| = ||T(v)|| = ||\lambda v|| = |\lambda|||v||.$$

An orthogonal matrix whose determinant is 1, is called proper. A proper orthogonal transformation represents a rotation.

A proper orthogonal matrix defines three Euler angles, which explicitly shows that the matrix represents a rotation. See Ben Noble, **Applied Linear Algebra.** Also for the determination of the rotation axis of the matrix, which is an eigenvector, see the discussion in the paper on rotation matrices contained in an issue of the IEEE Journal on Computational Geometry and Graphics. Also see my **Rotations**, **rotation.tex, rotation.pdf**, computer codes in the Emery Fortran and C libraries, **emerylib.for, emerylib.c**.

21 Rotation Matrices

See the paper **Rotations** by James Emery (Document rotation.tex).

22 Exponentials of Matrices and Operators in a Banach Algebra

See See the paper **Rotations** by James Emery (Document rotation.tex).

23 Eigenvalues

If a is a square matrix, and

$$av = \lambda v$$

then λ is called an eigenvalue and v the corresponding eigenvector. For example suppose

$$a = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$$

The characteristic equation is:

$$\begin{vmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{vmatrix} = 0$$

$$= (1-\lambda)(3-\lambda).$$

Eigenvalues are the roots of the characteristic equation.

Eigenvalues $\lambda_1 = 1, \lambda_2 = 3$. Corresponding eigenvectors:

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Matlab:

```
a =

     1     2
     0     3
```

```
lambda=eig(a)

lambda =

     1
     3

[v lambda]=eig(a)

v =

    1.0000    0.7071
         0    0.7071


lambda =

     1     0
     0     3
```

A symmetric matrix can always be diagonalized using the eigenvalues and the eigenvectors. The eigenvalues are the elements of the diagonal. This is the simplest example of what is called the spectral theorem. In vibration theory, this is equivalent to using normal coordinates. In quantum mechanics eigenanalysis plays a large role.

24 The Characteristic Polynomial and the Cayley-Hamilton Theorem

The characteristic polynomial of a n by n matrix M, $f(\lambda)$ is the polynomial

$$det(M - \lambda I),$$

where I is the identity matrix. The roots of the characteristic polynomial are the eigenvalues of M.

The Cayley-Hamilton theorem says that a matrix is a root of its characteristic polynomial. That is suppose

$$a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$$

is the characteristic polynomial of a matrix M. Then

$$a_3M^3 + a_2M^2 + a_1M + a_0I = 0$$

Example

25 Unitary Transformations

In a Hilbert space, a unitary transformation U has the property that

$$(Ux, Uy) = (x, y).$$

A complex n by n matrix M is unitary if its inverse is the conjugate transpose of M.

A real orthogonal matrix M is a matrix so that the its inverse is the transpose.

A unitary transformation is

26 Transpose, Trace, Self-Adjoint Operators

The trace of a matrix is the sum of the diagonal elements. The trace has the following property. If A and B are square matrices then

$$trace(AB) = trace(BA).$$

This follows because

$$trace(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}b_{ji}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} b_{ji} a_{ij}$$

$$= trace(BA).$$

If

$$M' = PMP^{-1},$$

then

$$trace(M') = trace(PMP^{-1}) = trace(P^{-1}PM) = trace(M).$$

27 The Spectrum

The root word of spectrum is a word meaning appearance or image. One might speculate that spectrum was first used for the image of lines obtained from a spectroscope. I suppose the use of spectrum in mathematical analysis must have come after the invention of quantum mechanics.

A bounded linear operator T is an operator such that there exists some number M so that

$$\|Tx\| \leq M\|x\|$$

for all x.

The spectrum of a bounded linear operator T is the set λ in which the operator

$$T - \lambda I$$

has an inverse. In general the spectrum may consist of several parts, the continuous spectrum, the residual spectrum, and the point spectrum (the eigenvalues). A finite linear transformation has only a point spectrum.

28 The Spectral Theorem

Various versions of the spectral theorem in linear algebra show how a linear transformation may be characterized by its eigenvalues and eigenvectors. In the most simple form of a symmetric matrix it is shown that the matrix may be diagonalized by introducing a basis of eigenvectors. The eigenvectors are orthogonal to each other and the eigenvalues are real. A very good survey

of the spectral theorem, in both finite dimensional and infinite dimensional spaces is
Lorch Edward R, **The Spectral Theorem**, In *Studies in Modern Analysis* Volume I, R. C. Buck editor, MAA Studies in Mathematics, The Mathematical Association of America, 1962.

29 Tensors

A tensor is a multilinear functional. That is, it is a function defined on a cartesian product of vector spaces. It is a linear transformation in each individual vector space v_i. Thus

$$f(v_1, v_2, ..., v_i + u_i, ...v_n) = f(v_1, v_2, ..., v_i, ...v_n) + f(v_1, v_2, ..., u_i, ...v_n)$$

and so on.

30 Application of Linear Algebra to Vibration Theory, Normal Coordinates

See the document Vibration, (vibra.tex).

30.1 A Simple Spring Example

Spring force:

$$F = kx.$$

Spring Potential Energy:

$$V = k\frac{x^2}{2}.$$

Kinetic Energy:

$$T = m\frac{\dot{x}^2}{2}$$

Lagrangian:

$$L = T - V.$$

Lagrange's Equation:

$$\frac{d}{dt}\frac{\partial L}{d\dot{x}} - \frac{\partial L}{\partial x} = 0.$$

We have

$$\frac{\partial L}{d\dot{x}} = m\dot{x}.$$

$$\frac{d}{dt}\frac{\partial L}{d\dot{x}} = m\ddot{x}.$$

$$\frac{\partial L}{\partial x} = -kx$$

So Lagrange's equation is

$$m\ddot{x} + kx = 0.$$

Trying $e^{\alpha t}$ as a solution, we find

$$(m\alpha^2 + k)e^{\alpha t} = 0$$

or

$$\alpha = i\sqrt{k/m} = i\omega,$$

where

$$\omega^2 = \frac{k}{m}.$$

So

$$x = e^{i\omega t}$$

The differential equation can be written as

$$\frac{\partial^2}{\partial t^2}x = -\omega^2 x.$$

So $x = e^{i\omega t}$ and $x = e^{-i\omega t}$ are eigenfunctions, and $-\omega^2$ is an eigenvalue of the linear operator

$$\frac{\partial^2}{\partial t^2}.$$

30.2 Decoupling Equations

Consider the equation

$$M\ddot{X} + KX = F,$$

where M is a square n-dimensional symmetric mass matrix, K is a square n-dimensional symmetric stiffness matrix, and F is a n-dimensional force. Both M and K must be positive definite because they represent the kinetic and strain quadratic forms. If either of them were not positive definite, there would be a nodal displacement or a nodal displacement velocity giving a negative energy. We assume a solution of the form

$$X = X_0 \exp(i\omega t).$$

Consider the homogeneous problem with $F = 0$. We have

$$-M\omega^2 X_0 + KX_0 = 0.$$

Let us write X for X_0. Then we have

$$(K - \omega^2 M)X = 0$$

This is a generalized eigenvalue problem. This homogeneous problem has a nonzero solution if and only if the determinant is zero:

$$|K - \omega^2 M| = 0.$$

This determinant is called the characteristic function, which is an nth degree polynomial in the variable ω^2. The roots are called eigenvalues. In the case of symmetric matrices the eigenvalues are real. To show this, let the inner product be

$$(u, v) = \sum_{i=1}^{n} u_i \bar{v}_i,$$

where the subscripted values are the vector components, and where the bar indicates the complex conjugate. Then if λ is an eigenvalue, with eigenvector v, because M and K are self-adjoint (real symmetric), we have

$$\lambda(Mv, v) = (\lambda Mv, v) = (Kv, v) = (v, \bar{K}^T v)$$

$$= (v, Kv) = (v, \lambda Mv) = \bar{\lambda}(v, Mv) = \bar{\lambda}(\bar{M}^T v, v) = \bar{\lambda}(Mv, v).$$

Therefore

$$\lambda = \bar{\lambda},$$

and so λ is real.

In this vibration problem the eigenvalues are not only real, but positive, because the matrices are positive definite. Under certain conditions, (1) If they are distinct, or (2) If one of the matrices is positive definite, then the vibration problem can be decoupled. Suppose that the eigenvalues are distinct. That is, the eigenfrequencies $\omega_1, ..., \omega_n$ are distinct. Since the determinant of the matrix

$$K - \omega_i^2 M$$

is zero for each i, its column vectors are linearly dependent, so that there exists a nonzero vector X_i such that

$$(K - \omega_i^2 M)X_i = 0.$$

The vectors $X_1, ..., X_n$ are called eigenvectors or modal vectors. Let P be a matrix which has as columns the eigenvectors

$$P = [X_1|X_2|...|X_n]$$

By the definition of the eigenvectors, we have

$$K[X_1|X_2|...|X_n] = [\omega_1^2 M X_1|\omega_2^2 M X_2|...|\omega_n^2 M X_n]$$

$$= M[X_1|X_2|...|X_n]\Lambda = MP\Lambda,$$

where Λ is the diagonal matrix of eigenvalues. Then

$$P^T K P = P^T M P \Lambda.$$

We will show that the eigenvectors have certain orthogonality properties. We will show that if ω_i^2 is not equal to ω_j^2, then vector X_i is perpendicular to X_j. Let A be a matrix, then we write A^T for the transpose. Because K and M are symmetric

$$K^T = K$$

and

$$M^T = M$$

We have

$$X_i^T K X_j = X_i^T \omega_j^2 M X_j = \omega_j^2 X_i^T M X_j$$

On the other hand

$$X_i^T K X_j = (K^T X_i)^T X_j = (K X_i)^T X_j = (\omega_i^2 M X_i)^T X_j$$
$$= \omega_i^2 X_i^T M^T X_j = \omega_i^2 X_i^T M X_j.$$

So

$$\omega_j^2 X_i^T M X_j = \omega_i^2 X_i^T M X_j.$$

But ω_j and ω_i are not equal, so that we must have

$$X_i^T M X_j = 0,$$

and then

$$X_i^T K X_j = 0.$$

Now let

$$X_i^T M X_i = \mu_i^2.$$

Any multiple of an eigenvector is still an eigenvector, so we scale the eigenvectors so that each

$$\mu_i^2 = 1$$

Then

$$P^T M P = I.$$

Then we have

$$P^T K P = P^T M P \Lambda = I \Lambda = \Lambda.$$

Both matrices have been diagonalized.

30.3 Example: Coupled Oscillators

Consider two masses m_1 and m_2. They can each move in a horizontal direction. Let the masses be connected with springs, k_1, k_2, and k_3. Spring k_1 is connected on the left with a fixed point, and on the right with mass m_1. Spring k_2 is connected on the left with m_1, and on the right with mass m_2. Spring k_3 is connected on the left with m_2, and on the right with a fixed point. Let u_1 and u_2 be the displacement from equilibrium of masses m_1 and m_2 respectively. The kinetic energy of the system is

$$T = \frac{1}{2}m_1\dot{u}_1^2 + \frac{1}{2}m_2\dot{u}_2^2$$

and the potential energy is

$$V = \frac{1}{2}k_1u_1^2 + \frac{1}{2}k_2(u_2 - u_1)^2 + \frac{1}{2}k_3u_2^2.$$

The Lagrangian is

$$L = T - V.$$

The equations of motion are given by Lagrange's equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{u}_1} - \frac{\partial L}{\partial u_1} = 0$$

and

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{u}_2} - \frac{\partial L}{\partial u_2} = 0.$$

We have

$$\frac{\partial L}{\partial \dot{u}_1} = \frac{\partial T}{\partial \dot{u}_1} = m_1\dot{u}_1$$

and

$$\frac{\partial L}{\partial \dot{u}_2} = \frac{\partial T}{\partial \dot{u}_2} = m_2\dot{u}_2.$$

Also

$$-\frac{\partial L}{\partial u_1} = \frac{\partial V}{\partial u_1} = k_1u_1 - k_2(u_2 - u_1)$$

and

$$-\frac{\partial L}{\partial u_2} = \frac{\partial V}{\partial u_2} = k_2(u_2 - u_1) + k_3u_2.$$

Thus Lagrange's equations are

$$m_1\ddot{u}_1 + k_1u_1 - k_2(u_2 - u_1) = 0$$

and

$$m_2\ddot{u}_2 + k_2(u_2 - u_1) + k_3u_2 = 0.$$

In matrix form the equations become

$$M \begin{bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{bmatrix} + K \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

where

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix},$$

and

$$K = \begin{bmatrix} (k_1 + k_2) & -k_2 \\ -k_2 & (k_3 + k_2) \end{bmatrix}.$$

Let

$$u_1 = U_1 e^{i\omega t}$$

and

$$u_2 = U_2 e^{i\omega t}.$$

Then

$$(K - \omega^2 M) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

This equation has a nonzero solution if and only if

$$\det(K - \omega^2 M) = 0.$$

Let $\lambda = \omega^2$. The determinant is

$$(k_1 + k_2 - \lambda m_1)(k_3 + k_2 - \lambda m_2) - k_2^2,$$

which we equate to zero to find the eigenvalues λ. Now let us specialize the problem and set the spring constants k_1 and k_3, to a common value k. And also set the two masses m_1 and m_3, to a common value m. Let the coupling spring be a weak spring.

That is, let spring constant k_2 be equal to a fraction of k. The equation for the eigenvalues is the quadratic equation

$$(k + k_2 - \lambda m)(k + k_2 - \lambda m) - k_2^2 = 0.$$

That is, it is

$$A\lambda^2 + B\lambda + C = 0,$$

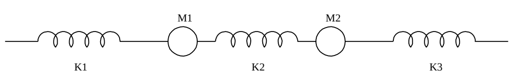

Figure 1: Coupled springs, With masses m_1 and m_1. The spring constants are k_1, k_2 and k_3. k_2 is the constant for the coupling spring and is much smaller than the other two constants.

with

$$A = m^2$$

$$B = -2(k + k_2)m$$

and

$$C = (k + k_2)^2 - k_2^2 = k^2 + 2kk_2.$$

Let $a = (k + k_2 - \lambda m)$ and $b = -k_2$. Then since the determinant is 0, we must have

$$a^2 = b^2$$

So either

$$a = b,$$

or

$$a = -b.$$

The equations for the eigenvectors are

$$aU_1 + bU_2 = 0,$$

and

$$bU_1 + aU_2 = 0.$$

If $a = b$, then the eigenvector is

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

On the other hand, if $a = -b$, then the eigenvector is

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The eigenvalues are λ_1 and λ_2. Let a_i be the value of a for eigenvalue λ_i, and let b_i be the value of b for eigenvalue λ_i. Now let us compute with values assigned to the constants. Let $m = 1$, $k = 1$, and $k_2 = 1/10$. The computation is done with a Matlab script (Octave script). Call it **coupled.m**:

```
m=1.;
k=1.;
k_2=1/10;
A= m*m;
B= -2*(k+k_2)*m ;
C=  k^2 + 2 * k * k_2;
lambda_1=(-B + sqrt(B*B - 4*A*C))/(2*A)
omega_1=sqrt(lambda_1)
a_1 = (k + k_2 - lambda_1 * m)
b_1 = -k_2
lambda_2=(-B - sqrt(B*B - 4*A*C))/(2*A)
omega_2=sqrt(lambda_2)
a_2 = (k + k_2 - lambda_2 * m)
b_2 = -k_2
omega=(omega_1+omega_2)/2
t1=2*pi/omega
phi=(omega_1 -omega_2)/2
t2=2*pi/phi
```

The output is:

```
octave-3.0.0.exe:5> coupled
lambda_1 =  1.2000
omega_1 =  1.0954
a_1 = -0.10000
b_1 = -0.10000
lambda_2 =  1.00000
omega_2 =  1.00000
a_2 =  0.10000
b_2 = -0.10000
omega =  1.0477
t1 =  5.9970
phi =  0.047723
t2 =  131.66
octave-3.0.0.exe:6> exit
```

The first eigenvalue is

$$\lambda_1 = 1.2$$

a_1 and b_1 are equal, so the eigenvector is

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The corresponding frequencies are

$$\omega_1 = \pm\sqrt{\lambda_1} = \pm 1.0954$$

Hence two linearly independent solutions are

$$e^{i\omega_1 t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$e^{-i\omega_1 t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

So any linear combination of these solutions is also a solution. That is, the space spanned by these two solutions is a solution. But both $cos(\omega_1 t)$ and $sin(\omega_1 t)$ can be written as a linear combination of $e^{i\omega_1 t}$ and $e^{-i\omega_1 t}$. Hence this solution space is the same as the space spanned by

$$\cos(\omega_1 t) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$\sin(\omega_1 t) \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The second eigenvalue is

$$\lambda_2 = 1.0$$

a_2 and b_2 are not equal, so the eigenvector is

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The corresponding frequencies are

$$\omega_1 = \pm\sqrt{\lambda_2} = \pm 1.0$$

Hence two linearly independent solutions are

$$e^{i\omega_2 t}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$e^{-i\omega_2 t}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Again the solution space is spanned by

$$\cos(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\sin(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Suppose the initial conditions are

$$u_1(0) = A$$

$$\frac{du_1(0)}{dt} = 0$$

$$u_2(0) = 0$$

$$\frac{du_2(0)}{dt} = 0$$

Let the solution be

$$u = c_{11}\cos(\omega_1 t)\begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_{12}\sin(\omega_1 t)\begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_{21}\cos(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_{22}\sin(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where the constants are to be determined by the initial conditions.

From the first initial condition we have

$$c_{11} + c_{21} = A$$

From the third initial condition we have

$$-c_{11} + c_{21} = 0$$

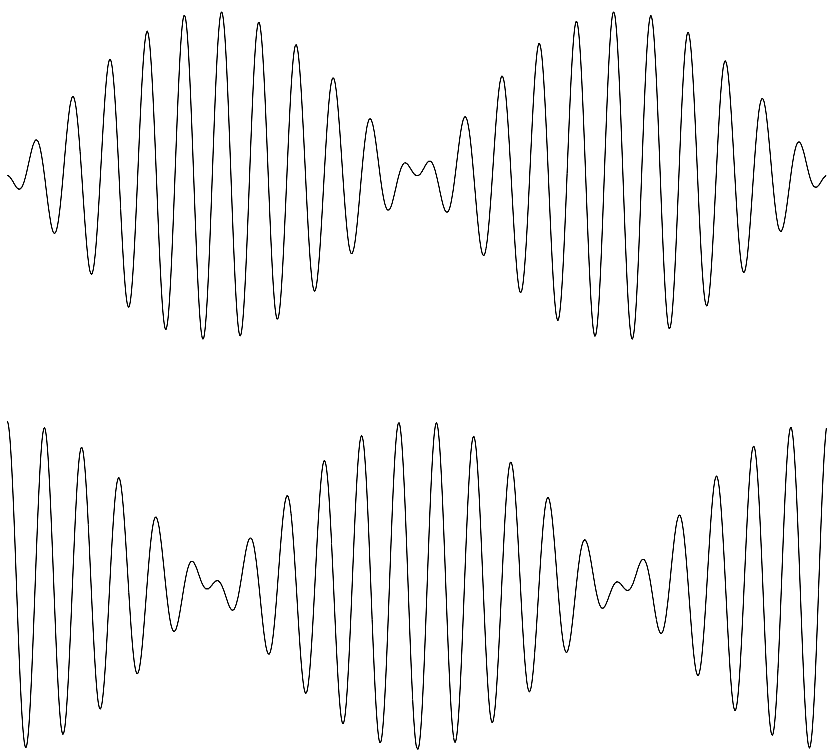

Figure 2: Coupled oscillations. The lower curve is the oscillation of mass m_1. The upper curve is the oscillation of mass m_2.

Thus

$$c_{21} = A/2$$

and

$$c_{11} = A/2.$$

Differentiating our solution, we have

$$\frac{du}{dt} = -c_{11}\omega_1 \sin(\omega_1 t)\begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_{12}\omega_1 \cos(\omega_1 t)\begin{bmatrix} 1 \\ -1 \end{bmatrix} - c_{21}\omega_2 \sin(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_{22}\omega_2 \cos(\omega_2 t)\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

From the second initial condition we have

$$c_{12}\omega_1 + c_{22}\omega_2 = 0.$$

From the forth initial condition we have

$$-c_{12}\omega_1 + c_{22}\omega_2 = 0.$$

Therefore $c_{22} = 0$ and $c_{12} = 0$. Therefore our solution is

$$u_1 = \frac{A}{2}\cos(\omega_1 t) + \frac{A}{2}\cos(\omega_2 t).$$

$$u_2 = -\frac{A}{2}\cos(\omega_1 t) + \frac{A}{2}\cos(\omega_2 t).$$

Now if $\omega = (\omega_1 + \omega_2)/2$ and $\phi = (\omega_1 - \omega_2)/2$, then

$$\cos(\omega_1 t) + \cos(\omega_2 t) = 2\cos(\omega t)\cos(\phi t)$$

So

$$u_1 = A\cos(\omega t)\cos(\phi t).$$

Similarly

$$u_2 = -A\sin(\omega t)\sin(\phi t).$$

For the numbers we have used here the oscillating frequency is

$$\omega = 1.0477,$$

and the corresponding period is

$$T_1 = 5.9970.$$

The modulating frequency is

$$\phi = 0.047723,$$

and the modulating period is

$$t_2 = 131.66.$$

The oscillators operate as follows. The first oscillator starts at maximum amplitude A, while the second oscillator is stopped at zero amplitude. The amplitude of the first oscillator begins to decrease, according to the modulating factor $\cos(\phi t)$. At $t = t_2/4$, the amplitude of the first oscillator has decreased to zero, whereas the amplitude of the second oscillator has increased to its maximum amplitude A, according to the modulating factor $\sin(\phi t)$. The energy of the first oscillator has been transferred to the second. This behavior repeats, as energy flows back and forth between the two oscillators.

30.4 The General Problem of Linear Vibration

The reference for this section is:
Bradbury T C, **Theoretical Mechanics**, John Wiley, 1968, p352.

The kinetic and potential energies are given as quadratic forms

$$T = \frac{1}{2} g_{ij} \dot{q}_i \dot{q}_j,$$

and

$$V = \frac{1}{2} k_{ij} q_i q_j.$$

The Lagrange equations of motion become

$$g_{ij} \ddot{q}_j + k_{ij} q_j = 0.$$

Write this as a matrix equation

$$G\ddot{Q} + KQ = 0,$$

where G and K are symmetric n by n matrices. This equation has a solution

$$Q = X \exp(\omega t).$$

Then

$$(K - \lambda G)X = 0,$$

where

$$\lambda = \omega^2.$$

This is a generalized eigenvalue problem. The quadratic forms are positive definite. The eigenvalues are real, and the eigenvectors (normal modes) are orthogonal. Let the rows of matrix S be constructed out of eigenvectors, which are scaled to be of unit length with metric G (S, or S^T is called the modal matrix in Thomson "Theory of Vibration."). That is,

$$X^T G X = \delta_{ij}.$$

Then

$$S G S^T = I.$$

Let Λ be a diagonal matrix of the eigenvalues. Then

$$K S^T = \Lambda G S.$$

Define normal coordinates Q' by

$$Q = Q' S^T.$$

Then the equations of motion become

$$G S^T \ddot{Q}' + K S^T Q' = 0.$$

Multiplying by S and simplifying, we get

$$\ddot{Q}' + \Lambda Q' = 0.$$

The problem could also be solved by diagonalizing the original energy quadratic forms. Any two positive definite quadratic forms can be simultaneously diagonalized. See for example, D. E. Littlewood **A University Algebra**, 2nd ed. 1958 Dover, p53.

31 Polynomial Roots, The Frobenius Companion Matrix

The roots of a polynomial $p(x)$ can be computed as the eigenvalues of the Frobenious companion matric. This is because $p(x)$ is the characteristic polynomial of the companion matrix, which can be proved using mathematical induction. See MatLab help for the polynomial roots function, called **roots**.

For example if $p(x)$ is a third degee polynomial, then

$$p(x) = c_0 + c_1 x + c_2 x^2 + x^3,$$

Then the companion matrix is the following 3×3 matrix:

$$\begin{bmatrix} 0 & 0 & -c_0 \\ 1 & 0 & -c_1 \\ 0 & 1 & -c_2 \end{bmatrix}$$

The characteristic polynomial is

$$\begin{vmatrix} -\lambda & 0 & -c_0 \\ 1 & -\lambda & -c_1 \\ 0 & 1 & -c_2 - \lambda \end{vmatrix}$$

$$= -\lambda \begin{vmatrix} -\lambda & -c_1 \\ 1 & -c_2 - \lambda \end{vmatrix} - c_0 \begin{vmatrix} 1 & -\lambda \\ 0 & 1 \end{vmatrix}$$

$$= -c_0 - \lambda(c_1 + c_2\lambda + \lambda^2)$$
$$= -c_0 - c_1\lambda - c_2\lambda^2 - \lambda^3,$$

by induction. And so it goes.

Notice that if n is odd then the characteristic equation computed from the companion matrix gives

$$-(c_0 + c_1\lambda + c_2\lambda^2 + ... + \lambda^n)$$

whereas if n is even then the characteristic equation computed from the companion matrix gives

$$c_0 + c_1\lambda + c_2\lambda^2 + ... + \lambda^n,$$

which takes care of the alternating signs in Cramers rule in the induction step.

32 Projection Operators

A projection operator is a linear operator P that has the property that

$$P(P(v)) = P(v).$$

So for example in the simple geometric case of projecting a 3-d vector to a 2-d subspace, that is to a plane, the projection then lies in the plane, hence projecting it again to the plane does not alter it because it is already in the plane.

33 Functional Analysis

When our vector spaces are infinite dimensional, for example consider the vector space of all continuous functions on an interval $[a, b]$, we enter the realm of functional analysis. This area of mathematics arose out of the problems of solving differential and integral equations, and in such areas as Fourier Analysis. So one studies Hilbert Spaces, which are complete inner product spaces, and Banach Spaces, which are complete normed linear spaces. In a further increase in abstraction one studies Topological Vector spaces in which there may be no inner product or norm.

34 Hamel Basis

Hamel constructed a basis for the vector space of real numbers over the rational field. He did this using Zorn's Lemma. So for example the vectors $\sqrt{2}$ and $\sqrt{3}$ are linearly independent. Otherwise

$$\sqrt{\frac{3}{2}}$$

would be a rational number.

For if we assumed that

$$\sqrt{\frac{3}{2}}$$

were rational we could write

$$\sqrt{\frac{3}{2}} = \frac{n}{m},$$

where n and m have no common factor. Then

$$\frac{3}{2} = \frac{n^2}{m^2},$$

$$3m^2 = 2n^2,$$

so n^2 and thus n must have a factor of 3. Hence m^2 and hence m must have a factor of 3. This contradicts our assumption.

See Angus Taylor **Introduction to Functional Analysis** for a treatment of Hamel basis.

35 Numerical Linear Algebra

See Golub and Van Loan, and books on Matlab.

36 Quantum Mechanics

Quantum Mechanics was treated by John Von Neumann in terms of Hilbert Spaces and Linear Operators in Hilbert Space, and involving eigenvalues and eigenvectors of such spaces of functions. One runs into things like delta functions, which are approximations to functions, but not actually functions at all. Thus we use mathematical objects called distributions. These distributions are functionals and not to be confused with probability distributions. Von Neumann made great contributions to the study of Banach Algebras.

37 The Schrödinger Wave Equation

From the photoelectric effect, the energy of a photon is

$$E = h\nu,$$

where h is Planck's constant, and ν is the frequency of the photon. From relativity theory, the energy of a particle is given by the equation

$$E^2 = p^2v^2 + m^2c^4,$$

where v is the particle velocity and m is its mass. For a photon the velocity is the velocity of light c, and the mass is zero. Thus for a photon

$$E^2 = p^2 c^2.$$

This implies that the wavelength is

$$\lambda = c/\nu = \frac{E/p}{E/h} = h/p.$$

Let $\mathbf{k}$ be the wave number vector. By definition

$$k\lambda = 2\pi.$$

We have

$$\mathbf{p} = \frac{\mathbf{k}h}{k\lambda} = \hbar\mathbf{k},$$

where

$$\hbar = \frac{h}{2\pi}.$$

A plane wave takes the form

$$e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} = e^{(i/\hbar)(\mathbf{p}\cdot\mathbf{r}-Et)}$$

A wave packet is represented as a sum of such plane waves, written as the Fourier transform

$$\psi(\mathbf{r}, t) = \frac{1}{\sqrt{2\pi h}} \int a(\mathbf{p}) e^{(i/\hbar)(\mathbf{p}\cdot\mathbf{r}-Et)} d\mathbf{p}.$$

This wave function satisfies the equation

$$(\hbar^2/2m)\nabla^2\psi - (\hbar/i)\partial\psi/\partial t = 0,$$

because $p^2/(2m) - E = 0$. It is called the Schrödinger wave equation. The group velocity of the wave packet is

$$v_g = \frac{d\omega}{dk} = \frac{dE}{dp}.$$

Using

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}},$$

and

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}.$$

We find that

$$\frac{dE}{dp} = \frac{pc^2}{E} = v.$$

Thus the group velocity of the wave packet corresponds to the particle velocity. $|\psi^2|$ is proportional to the probability density function for the location of the particle. This is the Born postulate.

38 The Postulates of Quantum Mechanics

The dynamical variables of a mechanical system are interpreted as operators, operating on wave functions which are elements of a Hilbert space. The identification with classical mechanics is done somewhat in the manner of the mathematical theory of distributions. The expected values of a variable (i.e. the measured values) form a discrete set which is the set of eigenvalues of the operator.

39 The Bra and Ket Notation of Dirac

A ket $v >$ vector is an element of a Hilbert space (a complete complex inner product space). A bra vector $< u$ is an element of the dual space, that is, it is a linear operator. Thus the value of u on v is

$$u(v) =< u|v >,$$

which also may be interpreted as an inner product.

In another way of looking at it, a bra vector $< v$ is a row matrix of components with respect to some basis, of a vector v, that is if

$$v = \sum_{i=1}^{n} v_i b_i$$

then

$$v >= \left[\begin{array}{cccc} v_1 & v_2 & ... & v_n \end{array} \right]$$

and a ket vector $v >$ is the conjugate transpose, and thus a column vector of components.

From a mathematical point of view, this bra and ket business seems rather complicated, confused, and is not clearly necessary.

Chapter VII in Messiah outlines some linear algebra in this bra ket notation.

Section 2.3 of Shankar gives another explanation of this notation.

40 Example: The Hydrogen Atom

Schrödinger's equation is a partial differential equation and can be solved by the method of separation of variable. That is, a solution can be written as a product, where each product function uses independent variables. Then usually the partial differential equation becomes a set of ordinary differential equations and eigenvalue problems. These eigenvalues than characterize quantum energy levels, such as in the case of the hydrogen atom. See Emery, **The Hydrogen Atom**.

41 The Relation Between Linear Algebra, Functional Analysis, and Abstract Solutions to Problems

Ultimately the solution of applied problems must involve a finite set of numbers rather than an abstract infinite theoretical solution. But the abstract solution can define properties of the solution and indeed existence of a solution and uniqueness. This is important, for finite approximations of a solution that does not exist will fail. So finally we typically will use the results of linear algebra to find arbitrary close approximations to applied problems. Specifically we require results from numerical linear algebra. That is we need algorithms that are robust, accurate, and render roundoff and truncation error small.

42 Appendix A: Rotation Matrices

43 Rotation Matrix Defined by Axis and Angle

Let a unit vector $\mathbf{n}$ specify a rotation axis, and let α be a rotation angle in the right hand rule sense. We shall show that the rotation of a vector $\mathbf{x}$ to a vector $\mathbf{y}$, around an axis in the direction of a vector $\mathbf{n}$, by an angle α, can be accomplished by multiplying $\mathbf{x}$ by a rotation matrix $\mathbf{M}$. To show this we shall write vectors as column vectors. So for example we shall write vector $\mathbf{n}$ as

$$\mathbf{n} = n_1\mathbf{i} + n_2\mathbf{j} + n_3\mathbf{k},$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the standard basis unit vectors directed along the cartesian axes. We shall also write each such vector as a 3 by 1 matrix known as a column vector,

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}.$$

We shall find a matrix $\mathbf{M}$, so that

$$\mathbf{y} = \mathbf{M}\mathbf{x}.$$

(This derivation is a solution to exercise 12.51, p 421, **Applied Linear Algebra**, Ben Noble, 1969, Prentice Hall.)

Let the origin be O. Refer to the figures called **R3 View** and **Plane View**. The figure **Plane View** shows the plane perpendicular to the rotation axis, defined by vector $\mathbf{n}$, with points P, Q, R and S in this plane.
Vector $\mathbf{x}$ is the vector $\vec{OP}$ from the origin O in 3-space to point P.

$$\mathbf{x} = \vec{OP}.$$

Vector $\mathbf{y}$ is the vector $\vec{OR}$ from the origin to point R.

$$\mathbf{y} = \vec{OR}.$$

Vector $\vec{OQ}$ is the rotation axis, the direction of $\mathbf{n}$.

Let us project point P to the axis defined by $\mathbf{n}$, getting point Q. Then

$$\vec{PQ} = (\mathbf{x} \cdot \mathbf{n})\mathbf{n} - \mathbf{x}.$$

We have a plane triangle PQR, where the measure of angle PQR is α. Although the derivation is valid for any angle α, we have drawn the figure so that $\alpha < \pi/2$. We construct a point S on line PQ, so that SR is perpendicular to PQ. Then vector y is

$$\mathbf{y} = \vec{OP} + \vec{PS} + \vec{SR}.$$

We shall show that each of these vectors is a product of a matrix and $\mathbf{x}$. Then the matrix we are looking for is the sum of these matrices.

Let β be the angle between x and n. Let r be the length of PQ and of QR. We have

$$r = \|\mathbf{x}\| \sin(\beta).$$

Considering the triangle SQR, we see that the length of $\vec{SQ}$ is $r\cos(\alpha)$. Thus

$$\vec{PS} = \frac{r(1-\cos(\alpha))}{r}\vec{PQ} = (1-\cos(\alpha))((\mathbf{x} \cdot \mathbf{n})\mathbf{n} - \mathbf{x}).$$

Vector $\vec{SR}$ is perpendicular to the plane containing $\mathbf{x}$ and $\mathbf{n}$. Such a unit vector is

$$\frac{\mathbf{n} \times \mathbf{x}}{\|\mathbf{x}\| \sin(\beta)} = \frac{\mathbf{n} \times \mathbf{x}}{r}.$$

From the figure the length of SR is $r\sin(\alpha)$. Hence

$$\vec{SR} = \mathbf{n} \times \mathbf{x} \sin(\alpha).$$

So

$$\mathbf{y} = \vec{OP} + \vec{PS} + \vec{SR} = \mathbf{x} + (1-\cos(\alpha))((\mathbf{x} \cdot \mathbf{n})\mathbf{n} - \mathbf{x}) + \sin(\alpha)\mathbf{n} \times \mathbf{x}$$

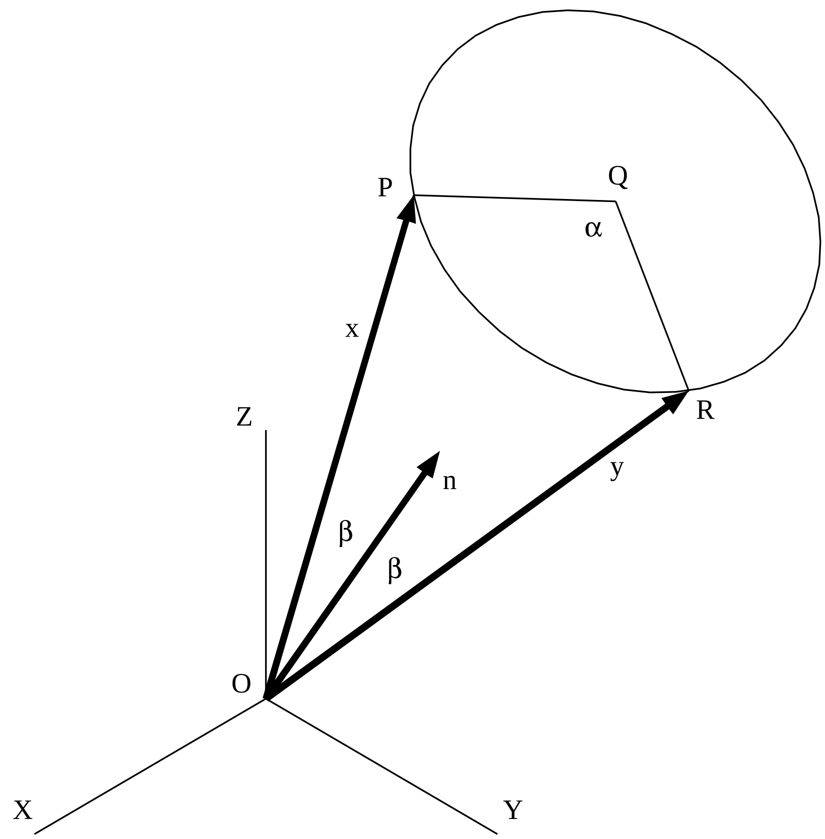

Figure 3: **3-D View of Rotation.** In the three dimensional view we see that vector **n** is in the direction of the rotation axis directed toward the center of the rotation circle, which is point Q. Vector **x**, which is $\vec{OP}$, is rotated by angle α to vector **y**, which is $\vec{OR}$. Angle β is the measure of both angle POQ and ROQ.

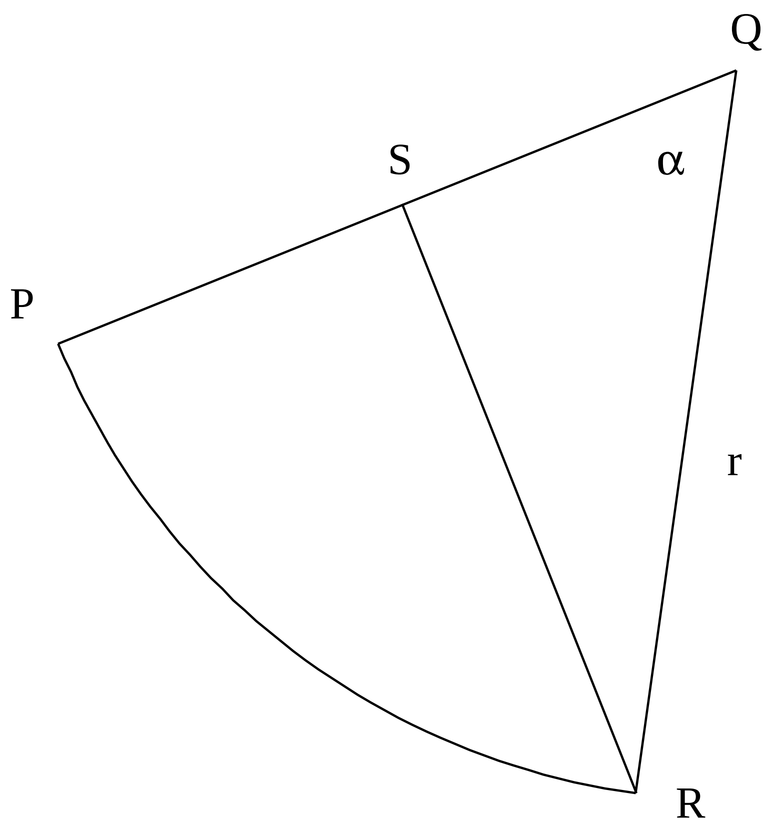

Figure 4: **Plane View of Rotation.** The plane is perpendicular to the rotation axis through Q. P is rotated to R. The rotation angle is α. The length of $\vec{SQ}$ is $r\cos(\alpha)$, so the length of $\vec{PS}$ is $r(1-\cos(\alpha))$.

$$= \cos(\alpha)\mathbf{x} + (1 - \cos(\alpha))(\mathbf{x} \cdot \mathbf{n})\mathbf{n}) + \sin(\alpha)\mathbf{n} \times \mathbf{x}.$$

Each of these terms is a linear transformation of vector $\mathbf{x}$. It can be written as a matrix equation.

Indeed we have

$$\mathbf{n} \times \mathbf{x} = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Let

$$\mathbf{N} = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix}$$

Because $\mathbf{x}$ and $\mathbf{n}$ are column vectors, we have

$$(\mathbf{x} \cdot \mathbf{n})\mathbf{n} = \mathbf{n}(\mathbf{n} \cdot \mathbf{x}) = \mathbf{n}\mathbf{n}^T\mathbf{x}.$$

Then we have

$$\mathbf{n}\mathbf{n}^T = \begin{bmatrix} n_1n_1 & n_1n_2 & n_1n_3 \\ n_2n_1 & n_2n_2 & n_2n_3 \\ n_3n_1 & n_3n_2 & n_3n_3 \end{bmatrix}.$$

We may write the equation as

$$\mathbf{y} = [\cos(\alpha)\mathbf{I} + (1 - \cos(\alpha))\mathbf{n}\mathbf{n}^T + \sin(\alpha)\mathbf{N}]\mathbf{x}.$$

We may simplify this a little further, because

$$\mathbf{N}^2 = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -n_3^2 - n_2^2 & n_1n_2 & n_1n_3 \\ n_2n_1 & -n_2^2 - n_1^2 & n_2n_3 \\ n_3n_1 & n_3n_2 & -n_2^2 - n_1^2 \end{bmatrix}$$

$$= \begin{bmatrix} n_1n_1 - 1 & n_1n_2 & n_1n_3 \\ n_2n_1 & n_2n_2 - 1 & n_2n_3 \\ n_3n_1 & n_3n_2 & n_3n_3 - 1 \end{bmatrix}$$

$$= \mathbf{n}\mathbf{n}^T - \mathbf{I}.$$

We have shown that

$$\mathbf{N}^2 + I = \mathbf{n}\mathbf{n}^T.$$

So

$$\mathbf{y} = [\cos(\alpha)I + (1 - \cos(\alpha))\mathbf{n}\mathbf{n}^T + \sin(\alpha)\mathbf{N}]\mathbf{x}$$

$$= [\cos(\alpha)\mathbf{I} + (1 - \cos(\alpha))(\mathbf{N}^2 + \mathbf{I}) + \sin(\alpha)\mathbf{N}]\mathbf{x}$$

$$= [\mathbf{I} + (1 - \cos(\alpha))\mathbf{N}^2 + \sin(\alpha)\mathbf{N}]\mathbf{x}.$$

So our rotation matrix

$$\mathbf{M} = \mathbf{I} + (1 - \cos(\alpha))\mathbf{N}^2 + \sin(\alpha)\mathbf{N}.$$

Notice that $\mathbf{Nn}$ is really the cross product of two parallel vectors, so is zero. Explicitly

$$\mathbf{Nn} = \begin{bmatrix} -n_3n_2 + n_2n_3 \\ n_3n_1 - n_1n_3 \\ -n_2n_1 + n_1n_2 \end{bmatrix} = 0$$

Hence multiplying

$$\mathbf{N}^2 + I = \mathbf{n}\mathbf{n}^T.$$

by $\mathbf{N}$, gives

$$\mathbf{N}^3 + \mathbf{N} = \mathbf{N}\mathbf{n}\mathbf{n}^T = 0,$$

so

$$\mathbf{N}^3 = -\mathbf{N}.$$

Continuing we find

$$\mathbf{N}^4 = -\mathbf{N}^2,$$

$$\mathbf{N}^5 = \mathbf{N},$$

$$\mathbf{N}^6 = \mathbf{N}^2,$$

$$\mathbf{N}^7 = -\mathbf{N},$$

$$\mathbf{N}^8 = -\mathbf{N}^2$$

....

and so on. We will use this later to show that the rotation matrix $\mathbf{M}$ is an exponential. Let us show explicitly that $\mathbf{M}$ is an orthogonal matrix. Using the facts that

$$\mathbf{N}^T = -\mathbf{N},$$

and

$$\mathbf{N}^4 = -\mathbf{N}^2,$$

we have

$$\mathbf{M}\mathbf{M}^T = [\mathbf{I} + (1-\cos(\alpha))\mathbf{N}^2 + \sin(\alpha)\mathbf{N}][\mathbf{I} + (1-\cos(\alpha))\mathbf{N}^2 - \sin(\alpha)\mathbf{N}].$$

Expanding this expression we find that

$$\mathbf{M}\mathbf{M}^T = \mathbf{I}.$$

See the references: Jay P Filmore, **A Note On Rotation Matrices**, IEEE Computer Graphics and Applications, February 1984, the orthogonal matrix subroutine **orthgm** in libraries **emerylib.ftn** and **emerylib.c**, as well as **axisang**, and **v2rot**, and the program **trnsf.c**. This latter program allows one to construct a general transformation matrix in steps, and then to apply it to a file containing points. See the batch files, **viewffun.bat**, **viewcfun.bat**, **getffun.bat** , and **getcfun.bat**, which allow one to see the subroutines and functions available in the libraries, and to extract such functions and subroutines so that they may be incorporated into programs. The library **emerylib.c** contains both C functions and C++ functions.

44 Axis and Angle of a Proper Rotation Matrix

Suppose we have a set of orthogonal basis vectors u_1, u_2, u_3. Clearly the matrix of a rotation transformation with respect to this basis, about axis u_3 by angle θ, is

$$M = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The trace of this matrix is the sum of the diagonal elements. The trace of M is

$$trace(M) = 2\cos(\theta) + 1.$$

The matrix of this rotation with respect to any other orthogonal basis is

$$M' = PMP^{-1},$$

where P is the change of basis matrix.

The trace has the property that for n by n matrices A and B,

$$trace(AB) = trace(BA).$$

This follows because

$$\begin{aligned} trace(AB) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}b_{ji} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} \\ &= trace(BA). \end{aligned}$$

Then

$$trace(M') = trace(PMP^{-1}) = trace(P^{-1}PM) = trace(M) = 2\cos(\theta) + 1.$$

Hence the trace of a rotation matrix determines the rotation angle. A rotation matrix is an orthogonal matrix in which the column vectors (also the row vectors) are unit orthogonal vectors. Hence

$$MM^T = I,$$

That is M^T is the inverse of M.

$$1 = det(I) = det(MM^T) = det(M)det(M^T) = det(M)^2.$$

It follows that

$$|det(M)| = 1.$$

For a proper orthogonal matrix $det(M) = 1$.

A vector v in the direction of the rotation axis is transformed to itself. Thus it is an eigenvector with eigenvalue $\lambda = 1$, that is

$$Mv = \lambda v.$$

Clearly this is the only eigenvector if M is not the identity. For any other vector that is not in the direction of the axis, is rotated. Hence the axis of the rotation may be obtained by calculating an eigenvector of M. Given a proper orthogonal transformation M, we can use the explicit formula for M calculated above to find an eigenvector of M.

We have

$$M = I + (1 - \cos(\alpha))N^2 + \sin(\alpha)N$$

and

$$M^T = I + (1 - \cos(\alpha))N^2 - \sin(\alpha)N.$$

So

$$\frac{1}{2}(M + M^T) = I + (1 - \cos(\alpha))N^2.$$

We have

$$N^2 = nn^T - I,$$

so

$$\frac{1}{2}(M + M^T) = I + (1 - \cos(\alpha))(nn^T - I)$$
$$= \cos(\alpha)I + (1 - \cos(\alpha))nn^T$$

Now

$$\cos(\alpha) = \frac{trace(M) - 1}{2},$$

and

$$(1 - \cos(\alpha)) = 1 - \frac{trace(M) - 1}{2} = \frac{3 - trace(M)}{2}.$$

Thus

$$(M + M^T) - (trace(M) - 1)I = (3 - trace(M))nn^T.$$

Any row or column of matrix nn^T is a multiple of vector n. So any row or column of matrix

$$(M + M^T) - (trace(M) - 1)I$$

is an eigenvector for eigenvalue $\lambda = 1$, and gives the rotation axis. Here is a Fortran subroutine called **axisang** for finding the axis and angle of a rotation matrix:

```
c+ axisang axis and angle of a rotation matrix, January 2004
      subroutine axisang(a,x,t)
      implicit real*8(a-h,o-z)
c Input:
c  a  3 by 3 orthogonal rotation matrix
c Output:
c  x  unit vector in the direction of the rotation axis
c  t  rotation angle, 0 <= t <= pi (right hand rule)
c References:
c  (1) Rotations, James Emery, January 2004, (rotations.tex)
c  (2) A Note on Rotation Matrices, Jay P Fillmore,
c      IEEE Computer Graphics and Applications, February, 1984.
c  (3) Applied Linear Algebra, B. Noble, 1969.
      real*8 a(3,3),b(3,3),x(3),y(3),z(3),w(3)
      real*8 c(3,3)
      zero=0.
c     compute the trace of a.
      trc=a(1,1)+a(2,2)+a(3,3)
c     Compute the positive angle of rotation
      cs=(trc-1.0d0)/2.0d0
      sn=sqrt(1.0d0 - cs*cs)
      t=atan2(sn,cs)
c     Find the transpose of a
      call mattrn(a,3,3,3,b,3)
c     Add a to its transpose.
      call mata(a,3,3,3,b,3,c,3)
      s=trc-1
      do i=1,3
       c(i,i)=c(i,i)-s
      enddo
c     We have computed a matrix c whose row and column vectors are
c     multiples of the required eigenvector.
      amax=0.
      do i=1,3
       anorm=c(i,1)**2+c(i,2)**2+c(i,3)**2
       if(anorm .gt. amax)then
        k=i
        amax=anorm
       endif
      enddo
      anorm=sqrt(amax)
      xmax=0.
      do j=1,3
       y(j)=1.
       x(j)=c(k,j)/anorm
```

```
        if(abs(x(j)).gt.xmax)then
         m=j
         xmax=abs(x(j))
        endif
      enddo
      y(m)=0.
      if(m .eq. 1)then
       y(2)=x(3)
       y(3)=-x(2)
      endif
      if(m .eq. 2)then
       y(1)=x(3)
       y(3)=-x(1)
      endif
      if(m .eq. 3)then
       y(1)=x(2)
       y(2)=-x(1)
      endif
c     We have found a unit eigenvector x and
c     we have found a vector y perpendicular to x.
c     Rotate y to z,  z=a*y.
      do i=1,3
       z(i)=0.
       do j=1,3
        z(i)=z(i) + a(i,j)*y(j)
       enddo
      enddo
c     Compute the cross product of y and z, w=y cross z
      call crsspr(y,z,w)
c     If the right hand rule is satisfied, w should be in the
c     direction of the axis x.
      s=dotpr(x,w)
c     If the right hand rule is not satisfied, reverse the direction
c     of the axis.
      if(s .lt. zero)then
       do j=1,3
        x(j)=-x(j)
       enddo
      endif
      return
      end
```

45 Obtaining the Rotation As The Exponential of an Element of a Banach Algebra

A Banach Algebra is a normed linear vector space in which a multiplication is defined, and which is complete. A complete space is one in which every Cauchy sequence converges to an element of the space. Let B be an element of the space of n by n matrices. One can take as norm the sum of the absolute values of the elements of the matrix. A norm satisfies the triangle inequality.

$$\|B_1 + B_2\| \leq \|B_1\| + \|B_2\|.$$

Also for a Banach Algebra the norm satisfies

$$\|B_1 B_2\| \leq \|B_1\|\|B_2\|.$$

We can take the exponential of the norm of B

$$e^{\|B\|} = \sum_{n=0}^{\infty} \frac{\|B\|^n}{n!}.$$

This converges for every B because the real exponential function is an entire function. Since it converges, the partial sums are a Cauchy sequence. This means that given some $\epsilon > 0$ there exists some integer N so that for every $m, n > N$, say $n > m$, the partial sums S_m, S_n differ by less than ϵ, that is

$$\frac{\|B\|^{m+1}}{(m+1)!} + \frac{\|B\|^{m+2}}{(m+2)!} + ... + \frac{\|B\|^n}{n!} < \epsilon$$

But using the norm inequalities this shows that

$$\left\| \frac{B^{m+1}}{(m+1)!} + \frac{B^{m+2}}{(m+2)!} + ... + \frac{B^n}{n!} \right\| < \epsilon.$$

It follows that the partial sums of the series

$$e^B = \sum_{n=0}^{\infty} \frac{B^n}{n!}$$

form a Cauchy sequence in the Banach space. Since the Banach space is complete, the series converges to some element, here a n by n matrix. So the exponential of the matrix (operator) is defined. Let us take the 3 by 3 matrix N above defined by our unit axis vector n. That is,

$$N = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix}$$

We use the properties of N found above, namely

$$N^3 = -N.$$

$$N^4 = -N^2,$$
$$N^5 = N,$$
$$N^6 = N^2,$$
$$N^7 = -N,$$
$$N^8 = -N^2$$
$$.......$$

Then we write

$$e^{tN} = \sum_{n=0}^{\infty} \frac{t^n N^n}{n!}$$
$$= I + (tN + \frac{t^3}{3!}N^3 + \frac{t^5}{5!}N^5 + ...)$$
$$+(\frac{t^2}{2!}N^2 + \frac{t^4}{4!}N^4 + \frac{t^6}{6!}N^6 + ...)$$
$$= I + (tN - \frac{t^3}{3!}N + \frac{t^5}{5!}N + ...)$$
$$+(\frac{t^2}{2!}N^2 - \frac{t^4}{4!}N^2 + \frac{t^6}{6!}N^2 + ...)$$
$$= I + (t - \frac{t^3}{3!} + \frac{t^5}{5!} + ...)N$$
$$+(\frac{t^2}{2!} - \frac{t^4}{4!} + \frac{t^6}{6!} + ...)N^2$$
$$= I + \sin(t)N + (1 - \cos(t))N^2.$$

This is the formula for the rotation matrix M derived above. Hence the rotation matrix for rotation about axis n by angle α is the exponential

$$M = e^{\alpha N}.$$

46 Properties of The Exponential of a Matrix

If a matrix A is upper triangular

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1n} \\ 0 & a_{22} & a_{23} & ... & a_{2n} \\ 0 & 0 & a_{33} & .... & a_{3n} \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & a_{nn} \end{bmatrix}$$

Then A^n is an upper triangular matrix of the form

$$\begin{bmatrix} a_{11}^n & ... & ... & .... & ... \\ 0 & a_{22}^n & ... & ... & ... \\ 0 & 0 & a_{33}^n & .... & ... \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & a_{nn}^n, \end{bmatrix}$$

This becomes apparent by just calculating an example such as say A^2 for a 3 by 3 upper triangular matrix. Now if a sequence of matrices converges to a matrix M, then a sequence consisting of say the ijth element of each matrix in the sequence, converges to the ijth element of M. It follows that if A is upper triangular, then

$$e^A = \begin{bmatrix} e^{a_{11}} & ... & ... & .... & ... \\ 0 & e^{a_{22}} & ... & ... & ... \\ 0 & 0 & e^{a_{33}} & .... & ... \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & e^{a_{nn}} \end{bmatrix}$$

So

$$det(A) = e^{a_{11}} e^{a_{22}} ... e^{a_{nn}}.$$

$$= e^{a_{11} + a_{22} + ... + a_{nn}}$$

$$= e^{trace(A)}.$$

If A is a general matrix, then there exists a matrix T such that

$$T^{-1} A T = B,$$

and B is upper triangular. For a simple proof, see Bellman Page 21. Note that T might be a complex matrix. We have

$$T^{-1}(e^A)T = I + \frac{1}{1!}T^{-1}AT + \frac{1}{2!}(T^{-1}AT)(T^{-1}AT) + \frac{1}{3!}(T^{-1}AT)(T^{-1}AT)(T^{-1}AT) + ...$$

$$= e^{T^{-1}AT}.$$

Hence because $T^{-1}AT$ is upper triangular, we have

$$det(e^A) = det(T^{-1}e^A T) = det(e^{T^{-1}AT}) = e^{trace(T^{-1}AT)} = e^{trace(A)}.$$

We have proven:

Proposition For any square matrix

$$det(e^A) = e^{trace(A)}.$$

For our matrix N above,

$$N = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix}$$

the trace is 0. It follows that

$$det(e^N) = e^{trace(N)} = e^0 = 1.$$

Because the transpose of N is - N, we see that

$$(e^{\alpha N})^T = e^{-\alpha N}.$$

It may be shown that

$$e^{A+B} = e^A e^B$$

if and only if

$$AB = BA.$$

So we find that

$$e^{\alpha N}(e^{\alpha N})^T = e^{\alpha N}(e^{-\alpha N}) = I.$$

So we see directly that

$$e^{\alpha N}$$

is a proper orthogonal matrix, and so represents a rotation. Also directly $n = (n_1, n_2, n_3)$ is an eigenvector because $Nn = 0$ and the only nonzero term in the exponential series for $e^{\alpha N} n$ is $In = n$. So n is the rotation axis. Because

$$N^2 = \begin{bmatrix} n_1 n_1 - 1 & n_1 n_2 & n_1 n_3 \\ n_2 n_1 & n_2 n_2 - 1 & n_2 n_3 \\ n_3 n_1 & n_3 n_2 & n_3 n_3 - 1 \end{bmatrix}$$

$$trace(N^2) = \|n\|^2 - 3 = -2.$$

Hence

$$trace(e^{\alpha N}) = trace(I + \sin(\alpha) N + (1 - \cos(\alpha)) N^2)$$
$$= 3 + 0 + (1 - \cos(\alpha))(-2) = 1 + 2\cos(\alpha),$$

So we see directly that α is the rotation angle.

47 A Test Program *rotations.ftn* With Subroutines *orthgm* and *axisang*

Here is a test program using subroutines orthgm and axisang.

```
c rotations.ftn\ Compute orthogonal rotation matrix
c Formerly called orthgm.ftn
c Revised slightly 10/16/15, changed name, added a couple of prints.
c 9/30/15
c 1/7/04
c Test of revised subroutines orthgm and axisang.
c Computes rotation matrix from from a supplied axis and angle.
c Computes axis and angle from the just computed orthogonal
c rotation matrix as a check on the computation.
      implicit real*8(a-h,o-z)
      dimension a(3,3)
      dimension x(3)
      dimension ain(3)
      dimension tmp(3)
      pi=4.0d0*atan(1.0d0)
      zero=0.
      a(1,1)=0.
      a(1,2)=0.
```

```
      a(1,3)=-1.
      a(2,1)=0.
      a(2,2)=-1.
      a(2,3)=0.
      a(3,1)=-1.
      a(3,2)=0.
      a(3,3)=0.
      write(*,*)' Enter an angle t (degrees), -180 <= t <= 180 '
      call readr(nf,ain, nr)
      t=ain(1)*pi/180.0d0
      write(*,'(a,g15.8,a)')' Angle  = ', ain(1),' degrees'
      write(*,*)' Enter a direction vector '
      call readr(nf,ain, nr)
      do i=1,3
       x(i)=ain(i)
      enddo
      write(*,'(a,3(1x,g15.8))')' Direction= ',(x(k),k=1,3)
      call orthgm(x,t,a)
      write(*,*)' rotation matrix a = '
      do i=1,3
       write(*,'(3(1x,g15.8))')(a(i,j),j=1,3)
      enddo
      ia=3
      call det3(a,ia,det)
      write(*,'(1x,a,g22.14)')' determinant=',det
      if(det .le. zero)then
       write(*,*)' This matrix is an improper rotation matrix'
      endif
      call axisang(a,x,t)
      write(*,*)' Computed rotation axis and angle of matrix a'
      write(*,'(1x,a,g22.14,g22.14,g22.14)')' axis=',x(1),x(2),x(3)
      write(*,'(1x,a,g22.14)')' angle= ',t*180./pi
      write(*,*)
      write(*,*)' Now we interchange columns 1 and 2 of matrix a'
      write(*,*)' to create an orthogonal matrix,'
      write(*,*)' which is an improper rotation matrix. '
      do i=1,3
       tmp(i)=a(i,1)
       a(i,1)=a(i,2)
       a(i,2)=tmp(i)
      enddo
c
      write(*,*)' a = '
      do i=1,3
       write(*,'(3(1x,g15.8))')(a(i,j),j=1,3)
      enddo
      call axisang(a,x,t)
      call det3(a,ia,det)
      write(*,'(1x,a,g22.14)')' determinant=',det
      if(det .le. zero)then
       write(*,*)' This matrix is an improper rotation matrix'
      endif
      write(*,*)' Attempt to compute an axis and angle'
      write(*,*)' Computed axis and angle of this improper rotation'
      write(*,'(1x,a,g22.14,g22.14,g22.14)')' axis=',x(1),x(2),x(3)
```

```
      write(*,'(1x,a,g22.14)')' angle= ',t*180./pi
      end
c+ axisang axis and angle of a rotation matrix, January 2004
      subroutine axisang(a,x,t)
      implicit real*8(a-h,o-z)
c Input:
c  a  3 by 3 orthogonal rotation matrix
c Output:
c  x  unit vector in the direction of the rotation axis
c  t  rotation angle, 0 <= t <= pi (right hand rule)
c References:
c  (1) Rotations, James Emery, January 2004, (rotations.tex)
c  (2) A Note on Rotation Matrices, Jay P Fillmore,
c      IEEE Computer Graphics and Applications, February, 1984.
c  (3) Applied Linear Algebra, B. Noble, 1969.
      real*8 a(3,3),b(3,3),x(3),y(3),z(3),w(3)
      real*8 c(3,3)
      zero=0.
c     compute the trace of a.
      trc=a(1,1)+a(2,2)+a(3,3)
c     Compute the positive angle of rotation
      cs=(trc-1.0d0)/2.0d0
      sn=sqrt(1.0d0 - cs*cs)
      t=atan2(sn,cs)
c     Find the transpose of a
      call mattrn(a,3,3,3,b,3)
c     Add a to its transpose.
      call mata(a,3,3,3,b,3,c,3)
      s=trc-1
      do i=1,3
       c(i,i)=c(i,i)-s
      enddo
c     We have computed a matrix c whose row and column vectors are
c     multiples of the required eigenvector.
      amax=0.
      do i=1,3
       anorm=c(i,1)**2+c(i,2)**2+c(i,3)**2
       if(anorm .gt. amax)then
        k=i
        amax=anorm
       endif
      enddo
      anorm=sqrt(amax)
      xmax=0.
      do j=1,3
       y(j)=1.
       x(j)=c(k,j)/anorm
       if(abs(x(j)).gt.xmax)then
        m=j
        xmax=abs(x(j))
       endif
      enddo
      y(m)=0.
      if(m .eq. 1)then
       y(2)=x(3)
```

```
       y(3)=-x(2)
      endif
      if(m .eq. 2)then
       y(1)=x(3)
       y(3)=-x(1)
      endif
      if(m .eq. 3)then
       y(1)=x(2)
       y(2)=-x(1)
      endif
c     We have found a unit eigenvector x and
c     we have found a vector y perpendicular to x.
c     Rotate y to z,  z=a*y.
      do i=1,3
       z(i)=0.
       do j=1,3
        z(i)=z(i) + a(i,j)*y(j)
       enddo
      enddo
c     Compute the cross product of y and z, w=y cross z
      call crsspr(y,z,w)
c     If the right hand rule is satisfied, w should be in the
c     direction of the axis x.
      s=dotpr(x,w)
c     If the right hand rule is not satisfied, reverse the direction
c     of the axis.
      if(s .lt. zero)then
       do j=1,3
        x(j)=-x(j)
       enddo
      endif
      return
      end
c+ mattrn   matrix transpose
      subroutine mattrn(a,ia,m,n,b,ib)
        implicit real*8(a-h,o-z)
c arguments
c   a-matrix
c  ia-row dimension of a in calling program
c   m-number of rows in a
c   n-number of columns in a
c   b-transpose of a
c  ib-row dimension of b in calling program
c
      dimension a(ia,*),b(ib,*)
      do 10 i=1,m
      do 10 j=1,n
10    b(j,i)=a(i,j)
      return
      end
c+ mata matrix addition
      subroutine mata(a,ia,m,n,b,ib,c,ic)
        implicit real*8(a-h,o-z)
c arguments
c  a-matrix
```

```
c ia-row dimension of a in calling program
c  m-number of rows
c  n-number of columns
c  b-matrix
c ib-row dimension of b in calling program
c  c-sum matrix: c=a*b
c ic-row dimension of c in calling program
      dimension a(ia,*),b(ib,*),c(ic,*)
c     c=a+b
      do 10 i=1,m
      do 10 j=1,n
      c(i,j)=a(i,j)+b(i,j)
10    continue
      return
      end
c+ crsspr  vector cross product.
      subroutine crsspr(a,b,c)
      implicit real*8(a-h,o-z)
c     c=product of a and b
      dimension a(3),b(3),c(3)
      c(1)=a(2)*b(3)-a(3)*b(2)
      c(2)=a(3)*b(1)-a(1)*b(3)
      c(3)=a(1)*b(2)-a(2)*b(1)
      return
      end
c+ dotpr scalar product of 3-space vectors
      function dotpr(a,b)
      implicit real*8(a-h,o-z)
c     2/5/97
      dimension a(*),b(*)
      s=0.
      do i=1,3
       s=s+a(i)*b(i)
      enddo
      dotpr=s
      return
      end
c+ readr read a row of numbers and return in double precision array
      subroutine readr(nf, a, nr)
      implicit real*8(a-h,o-z)
c Input:
c nf    unit number of file to read
c       nf=0 is the standard input file (keyboard)
c Output:
c a     array containing double precision numbers found
c nr    number of values in returned array,
c     or 0 for empty or blank line,
c       or -1 for end of file on unit nf.
c   Numbers are separated by spaces.
c   Examples of valid numbers are:
c   12.13 34 45e4 4.78e-6 4e2,5.6D-23,10000.d015
c requires subroutine valsub and function lenstr
c a semicolon and all characters following are ignored.
c This can be used for comments.
c modified 6/16/97 added semicolon feature
```

```
      dimension a(*)
      character*200  b
      character*200  c
      character*1 d
        c=' '
      if(nf.eq.0)then
         read(*,'(a)',end=99)b
      else
         read(nf,'(a)',end=99)b
      endif
      nr=0
      lsemi=index(b,';')
      if(lsemi .gt. 0)then
       if(lsemi .gt. 1)then
        b=b(1:(lsemi-1))
       else
        return
       endif
      endif
      l=lenstr(b)
      if(l.ge.200)then
        write(*,*)' error in readr subroutine '
        write(*,*)' record is too long '
      endif
      do 1 i=1,l
       d=b(i:i)
       if (d.ne.' ') then
        k=lenstr(c)
        if (k.gt.0)then
           c=c(1:k)//d
        else
           c=d
        endif
       endif
       if( (d.eq.' ').or.(i.eq.l)) then
        if (c.ne.' ') then
          nr=nr+1
           call valsub(c,a(nr),ier)
          c=' '
        endif
       endif
1     continue
      return
99    nr=-1
      return
      end
c+ valsub converts string to floating point number (r*8)
      subroutine valsub(s,v,ier)
      implicit real*8(a-h,o-z)
c     examples of valid strings are: 12.13 34 45e4 4.78e-6 4E2
c     the string is checked for valid characters,
c     but the string can still be invalid.
c      s-string
c      v-returned value
c      ier- 0 normal
```

```
c           1 if invalid character found, v returned 0
c
      logical p
      character s*(*),c*50,t*50,ch*15
      character z*1
      data ch/'1234567890+-.eE'/
      v=0.
      ier=1
      l=lenstr(s)
      if(l.eq.0)return
      p=.true.
      do 10 i=1,l
      z=s(i:i)
      if((z.eq.'D').or.(z.eq.'d'))then
          s(i:i)='e'
      endif
      p=p.and.(index(ch,s(i:i)).ne.0)
 10   continue
      if(.not.p)return
      n=index(s,'.')
      if(n.eq.0)then
          n=index(s,'e')
          if(n.eq.0)n=index(s,'E')
          if(n.eq.0)n=index(s,'d')
          if(n.eq.0)n=index(s,'D')
          if(n.eq.0)then
             s=s(1:l)//'.'
          else
            t=s(n:l)
            s=s(1:(n-1))//'.'//t
          endif
          l=l+1
      endif
      write(c,'(a30)')s(1:l)
      read(c,'(g30.23)')v
      ier=0
      return
      end
c+ lenstr   nonblank length of string
      function lenstr(s)
c  length of the substring of s obtained by deleting all
c  trailing blanks from s.  thus the length of a string
c  containing only blanks will be 0.
      character   s*(*)
      lenstr=0
      n=len(s)
      do 10 i=n,1,-1
      if(s(i:i) .ne. ' ')then
         lenstr=i
         return
      endif
10    continue
      return
      end
c+ orthgm  generate a rotation matrix (orthogonal) from axis and angle
```

```
      subroutine orthgm(x,t,a)
      implicit real*8(a-h,o-z)
c   a is the rotation matrix for column vectors
c   x-axis vector
c   t-rotation angle
c   a-output 3 by 3 matrix
c References:
c  (1) Rotations, James Emery, January 2004, (rotations.tex)
c  (2) A Note on Rotation Matrices, Jay P Fillmore,
c      IEEE Computer Graphics and Applications, February, 1984.
c  (3) Applied Linear Algebra, B. Noble, 1969.
      real*8 id(3,3),l(3,3),l2(3,3),a(3,3),x(3)
      real*8 lambda
      data (id(1,j),j=1,3)/1.d0,0.d0,0.d0/
      data (id(2,j),j=1,3)/0.d0,1.d0,0.d0/
      data (id(3,j),j=1,3)/0.d0,0.d0,1.d0/
      lambda=sqrt(x(1)**2+x(2)**2+x(3)**2)
      do i=1,3
       l(i,i)=0.
      enddo
      l(1,2)=-x(3)
      l(1,3)=x(2)
      l(2,3)=-x(1)
      l(2,1)=-l(1,2)
      l(3,1)=-l(1,3)
      l(3,2)=-l(2,3)
      call matm(l,3,3,3,l,3,3,l2,3)
      c1=sin(t)/lambda
      c2=(1.-cos(t))/(lambda**2)
      do i=1,3
       do j=1,3
        a(i,j)=id(i,j)+c1*l(i,j)+c2*l2(i,j)
       enddo
      enddo
      return
      end
c+ matm  matrix multiplication
      subroutine matm(a,ia,m,n,b,ib,l,c,ic)
        implicit real*8(a-h,o-z)
c arguments
c  a-matrix
c ia-row dimension of a in calling program
c  m-number of rows of a
c  n-number of columns of a
c  b-matrix
c ib-row dimension of b in calling program
c  l-number of columns of b
c  c-product matrix: c=a*b
c ic-row dimension of c in calling program
c
      dimension a(ia,*),b(ib,*),c(ic,*)
c     c=a*b
      do 10 i=1,m
      do 10 j=1,l
      c(i,j)=0.
```

```
      do 5 k=1,n
5     c(i,j)=c(i,j)+a(i,k)*b(k,j)
10    continue
      return
      end
c
c+ det3   determinant of 3 by 3 matrix
       subroutine det3(b,ib,det)
c b-3 by 3 matrix
c ib-row dimension of b
c det-computed determinant
       implicit real*8(a-h,o-z)
       dimension b(ib,*)
       det=b(1,1)*(b(2,2)*b(3,3)-b(2,3)*b(3,2))
       det=det-b(1,2)*(b(2,1)*b(3,3)-b(2,3)*b(3,1))
       det=det+b(1,3)*(b(2,1)*b(3,2)-b(2,2)*b(3,1))
       return
       end
```

48 Running Some Examples

Example 1:

We will compute the rotation matrix for a 30 degree rotation about the z axis. This is the matrix

$$\mathbf{A} = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) & 0 \\ \sin(\pi/6) & \cos(\pi/6) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We also shall show that if we interchange a pair of column vectors, the matrix becomes an improper rotation with determinant -1. This matrix is improper because it will convert a basis for a right handed coordinate system into a left handed coordinate basis.

We run program **rotations.ftn** and compute:

```
 Enter an angle t (degrees), -180 <= t <= 180
Angle  =   30.000000     degrees
 Enter a direction vector
Direction=    .00000000       .00000000       1.0000000
 rotation matrix a =
  .86602540      -.50000000       .00000000
```

```
 .50000000       .86602540       .00000000
 .00000000       .00000000      1.0000000
determinant=   1.0000000000000
Computed rotation axis and angle of matrix a
axis=   .00000000000000       .00000000000000       1.0000000000000
angle=    30.000000000000

Now we interchange columns 1 and 2 of matrix a
to create an orthogonal matrix\index{orthogonal matrix},
which is an improper rotation matrix.
a =
-.50000000       .86602540       .00000000
 .86602540       .50000000       .00000000
 .00000000       .00000000      1.0000000
determinant=  -1.0000000000000
Attempt to compute an axis and angle
Computed axis and angle of this improper rotation
axis=  -.50000000000000       .86602540378444       .00000000000000
angle=    90.000000000000
```

Although the computation on the improper matrix to compute a rotation angle and axis is carried out, the result has no meaning, since it only has meaning for a proper rotation matrix whose determinant equals one.

For the proper orthogonal matrix

$$\mathbf{A} = \begin{bmatrix} .86602540 & -.50000000 & .00000000 \\ .50000000 & .86602540 & .00000000 \\ .00000000 & .00000000 & 1.0000000 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{3}/2 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we must have a real eigenvalue $\lambda = 1$, and a corresponding eigenvector in the axis direction, because the matrix represents an isometry and the axis is mapped to itself.

Such an eigenvalue is a solution of the equation

$$det(\mathbf{A} - \lambda \mathbf{I}) = 0,$$

where $\mathbf{I}$ is the identity matrix. We have

$$det(\mathbf{A} - \lambda \mathbf{I}) = det \begin{bmatrix} \sqrt{3}/2 - \lambda & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{bmatrix}$$

$$= ((\sqrt{3}/2 - \lambda)^2 + 1/4)(1 - \lambda)$$
$$= (\lambda^2 - \sqrt{3}\lambda + 1)(1 - \lambda)$$

The characteristic equation for the eigenvalues therefore is

$$(\lambda^2 - \sqrt{3}\lambda + 1)(1 - \lambda) = 0.$$

Using the quadratic formula on the first factor, we obtain the roots

$$\frac{\sqrt{3}+i}{2}, \frac{\sqrt{3}-i}{2}$$

The three eigenvalues are

$$\lambda = \{1, \frac{\sqrt{3}+i}{2}, \frac{\sqrt{3}-i}{2}\}$$

So the real eigenvalue is 1. Using this value we have for the corresponding eigenvector the equations

$$\begin{bmatrix} \sqrt{3}/2 - 1 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 - 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

So z is arbitrary and because the upper left 2 by 2 submatrix has nonzero determinant equal to

$$(\sqrt{3}/2 - 1)^2 + 1/4 = 2 - \sqrt{3},$$

x and y are each zero. Therefore the eigenvector is parallel to the z axis, which is the axis of rotation.

Now let us use Octave (a clone of Matlab) to check this calculation.

```
octave-3.0.0.exe:9> [ .86602540 -.5 0;
> .5 .86602540 0;
> 0 0 1]
ans =
```

```
   0.86603  -0.50000   0.00000
   0.50000   0.86603   0.00000
   0.00000   0.00000   1.00000

octave-3.0.0.exe:10> a=ans
a =

   0.86603  -0.50000   0.00000
   0.50000   0.86603   0.00000
   0.00000   0.00000   1.00000

octave-3.0.0.exe:11> [v d] = eig(a)
v =

   0.00000 - 0.70711i   0.00000 + 0.70711i   0.00000 + 0.00000i
  -0.70711 + 0.00000i  -0.70711 - 0.00000i   0.00000 + 0.00000i
   0.00000 - 0.00000i   0.00000 + 0.00000i   1.00000 + 0.00000i

d =

   0.86603 + 0.50000i   0.00000 + 0.00000i   0.00000 + 0.00000i
   0.00000 + 0.00000i   0.86603 - 0.50000i   0.00000 + 0.00000i
   0.00000 + 0.00000i   0.00000 + 0.00000i   1.00000 + 0.00000i

octave-3.0.0.exe:12> exit
```

The Octave eigenvalues are on the diagonal of matrix **d**, and are

$$\{0.86603 + .5i, 0.86603 - .5i, 1\} = \{\sqrt{3}/2 + (1/2)i, \sqrt{3}/2 - (1/2)i, 1\},$$

Note. Notice first that the product of an n by n matrix A with a second n by n diagonal matrix D becomes a product matrix that has its kth column equal to the kth column of A multiplied by the k element of the diagonal of D. Therefore here

$$\mathbf{av} = \mathbf{vd}.$$

And notice second that if we multiply both sides of this equation by $\mathbf{v}^{-1}$ we have diagonalized matrix **a**, using matrix **v** as a change of basis matrix.

This is the reason that Matlab and Octave return the results of the eigenvalue computation in this form.

The Octave eigenvectors are the columns of matrix $\mathbf{v}$ which are normalized to each have norm equal to 1 and are

$$\mathbf{v}_1 = \begin{bmatrix} -(\sqrt{2}/2)i \\ -\sqrt{2}/2 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_2 = \begin{bmatrix} (\sqrt{2}/2)i \\ -\sqrt{2}/2 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Here is the script called eigenrot.m that does the eigenanalysis

```
% eigenrot.m, find the eigenvalues and eigenvectors of the rotation
%             matrix, which rotates by 30 degrees around the z axis.
% see program rotations.ftn, which defines rotation matrices
a=[sqrt(3)/2    -1/2   0;
      1/2    sqrt(3)/2 0;
       0          0    1]
[v d]=eig(a)
v1=v(1:3,1) % select the first column vector of v, which is the first eigenvector
v2=v(1:3,2)
v3=v(1:3,3)
e1=d(1,1)    % set e1 to the first eigenvalue
e2=d(2,2)
e3=d(3,3)
av1=a*v1     % show that a*v1 = e1*v1
e1v1=e1*v1
```

Example 2:

Now we shall compute an example where the result is not so obvious We shall compute the rotation matrix for a 65 degree rotation about the axis in vector direction $\mathbf{i}+\mathbf{j}+\mathbf{k}$. We run the program and compute:

```
 Enter an angle t (degrees), -180 <= t <= 180
Angle  =   65.000000     degrees
 Enter a direction vector
Direction=    1.0000000       1.0000000       1.0000000
 rotation matrix a =
  .61507884      -.33079647       .71571762
  .71571762       .61507884      -.33079647
 -.33079647       .71571762       .61507884
 determinant=   1.0000000000000
 Computed rotation axis and angle of matrix a
 axis=   .57735026918963       .57735026918963       .57735026918963
 angle=    65.000000000000

 Now we interchange columns 1 and 2 of matrix a
 to create an orthogonal matrix,
 which is an improper rotation matrix.
 a =
 -.33079647       .61507884       .71571762
  .61507884       .71571762      -.33079647
  .71571762      -.33079647       .61507884
 determinant=  -1.0000000000000
 Attempt to compute an axis and angle
 Computed axis and angle of this improper rotation
 axis=  -.33079646539450       .61507884116047       .71571762423403
 angle=    90.000000000000
```

Exercise: Write a program that computes a matrix, which rotates some specified vector to the x-axis. This can be used in mapping a coordinate bases to a standard basis

49 Bibliography for Appendex A

Some of these entries are included here because, either they are referred to in the text, or they have information on orthogonal and unitary matrices, or on exponentials of operators, or on exponentials of matrices.

[1] Bachman George, **Elements of Abstract Harmonic Analysis**, Academic Press, 1964.

[2] Bellman Richard, **Stability Theory of Differential Equations**, Dover, 1953

[3] Brauer Fred, Nohel John A, **The Qualitative Theory of Ordinary Differential Equations**, Dover, 1969.

[4] Douglas Ronald G, **Banach Algebra Techniques In Operator Theory**, Academic Press, 1972.

[5] Emery James D, **Topics In Linear Algebra and Its Applications**, 2015, www.stemsoc.org/je/lineara.tex, www.stemsoc.org/je/lineara.pdf.

[5] Filmore Jay, **A Note On Rotation Matrices**, IEEE Computer Graphics and Applications, February 1984, (Zerox Copy: Emery Papers, number 365).

[6] Golub Gene H, Van Loan Charles F, **Matrix Computations**, 2nd Edition, John Hopkins University Press 1989.

[7] Kolmogorov A. N., Fomin S. V., **Functional Analysis**, Dover?

[8] Moler Clive, Van Loan Charles, **Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twentyfive Years Later**, SIAM Review Vol. 54, No. 1, pp3-49, March 2003.

[9] Naimark M, **Normed Rings**, Noordhoff Groningen, 1959.

[11] Noble Ben, **Applied Linear Algebra**, Prentice-Hall, 1969.

[12] Taylor A E, **An Introduction to Functional Analysis**.

50 Bibliography

The number of books written on linear algebra and functional analysis is huge. So I list here some favorites of mine.

[1] Akhiezer N I, Glazman I M, **Theory of Linear Operators in Hilbert Space**, V I 1961, V II 1963, (two volumes Dover 1993).

[2] Bachman George, **Elements of Abstract Harmonic Analysis**, Academic Press, 1964.

[3] Bamberg Paul, Sternberg Shlomo, **A Course in Mathematics for Students of Physics,**, volumes 1 and 2 , Cambridge University Press, 1988..

[4] Banach Stefan, **Theorie Des Operations Lineares**, Chelsea, 1932.

[5] Banach Stefan, **Theory of Linear Operations**, English translation by F. Jellett, 1987, Elsevier..

[6] Courant Richard, Hilbert David, **Methods of Mathematical Physics**, Volumes I and 2, Wiley, Springer , 1937.

[7] Emery James D, **Group Theory**, (supply details) permutations.

[8] Emery James D, **The Hydrogen Atom** (supply details)

[9] Emery James D, **Vibration Theory**, (supply details)

[10] Fausett Laurene V, **Applied Numerical Analysis Using Matlab**, Prentice-Hall, 1999..

[11] Golub Gene H, Van Loan Charles F, **Matrix Computations**, John Hopkins University Press, 1989.

[12] Halmos Paul R, **Finite-Dimensional Vector Spaces**, Springer-Verlag, 1974.

[13] Halmos Paul R, **Introduction to Hilbert Space and the Theory of Spectral Multiplicity**, 2nd ed., Chelsea, 1951.

[14] Herstein I N, **Topics in Algebra**, Blaisdell, 1964., (Frobenius Theorem on Division Algebras, Quaternions).

[15] Lange Serge, **Linear Algebra** Addison Wesley, 1968

[16] Lawson and Hanson, **Solving Linear Least Squares Problems**, (JEL), 1974, (least squares, singular value decomposition).

[17] Lorch Edward R, **The Spectral Theorem**, In *Studies in Modern Analysis* Volume I, R. C. Buck editor, MAA Studies in Mathematics, The Mathematical Association of America, 1962.

[18] Messiah, **Quantum Mechanics**, Volumes I and II, (supply details).

[19] Nering Evar D, **Linear Algebra and Matrix Theory**, John Wiley, 1963.

[20] Noble Ben, **Applied Linear Algebra**, Prentice-Hall 1969.

[21] Powell John L, Crasemann Bernd **Quantum Mechanics** Addison-Wesley, 1965.

[22] Reed Michael, Simon Barry, 1972,, **Methods of Modern Mathematical Physics, Part 1. Functional Analysis**, 1972, Academic Press.

[23] Shankar, **Quantum Mechanics**, (supply details).

[24] Stoker J J, **Nonlinear Vibrations**, (JEL).

[25] Strang Gilbert, **Linear Algebra and Its Applications**, Academic Press, 1980, 2nd Edition..

[26] Taylor Angus E, **Introduction to Functional Analysis**, John Wiley, 1967.

[27] Widom Harold, **Lectures on Integral Equations**,1969, Van Nostrand Reinhold.

[28] Yosida Kosaku, **Functional Analysis**, 2nd Edition, 1968, Springer-Verlag.

51 Index