

Because someone asked.....

Context of Ordinary Differential Equations

Ordinary differential equations (ODEs) are fundamental in various fields, including mathematics, natural sciences, and social sciences. They are used to model situations where change occurs over time or space.

How ODEs Arise

Dynamically Changing Phenomena

- ODEs describe processes that evolve over time, such as:
 - Motion of objects (e.g., projectile motion)
 - Population growth in biology
 - Changes in economic indicators

Relationships Between Quantities

- They express relationships involving:
 - Derivatives, which represent rates of change
 - Functions that depend on one independent variable

Do you want to look a bit at ODEs ?

Partial Derivatives

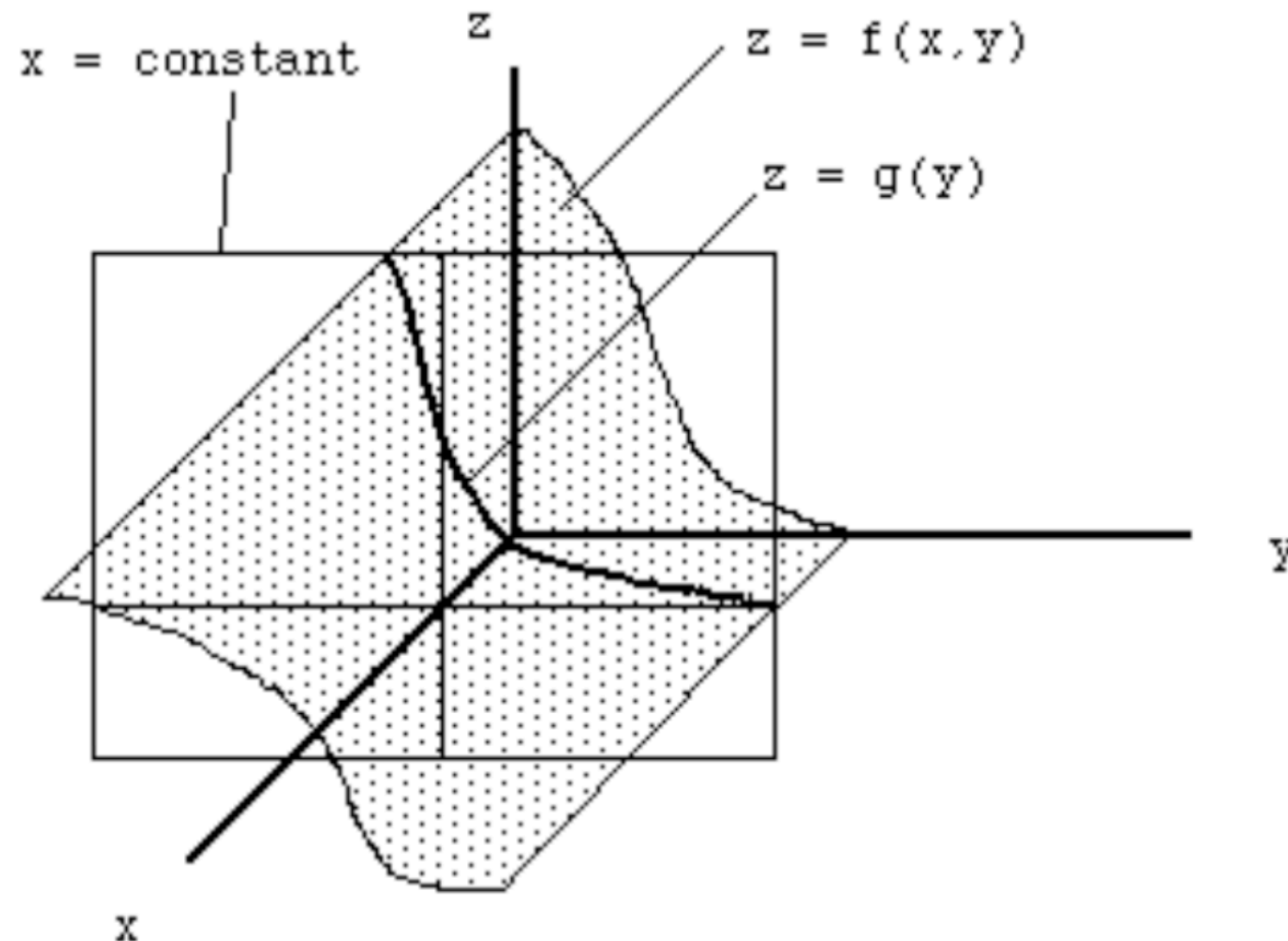
Partial Differentiation

Given the relationship $y = f(x)$, which represents a curve in 2-dimensions, then

$$\frac{dy}{dx} = \frac{df(x)}{dx} = \text{slope of the curve } y = f(x) \text{ or the rate of change of } y \text{ wrt } x$$

Now, if instead, we are given the relationship $z = f(x, y)$, which represents a surface in 3-dimensions, then differentiation is a bit more complicated.

Suppose $x = \text{constant}$, which is a plane intersecting the surface as shown:



The points satisfying $z = f(x, y)$ and $x = \text{constant}$ lie on a curve (as shown in the figure).

To learn about this curve, where $z = g(y)$, we might write dz/dy , but this does not indicate that z really is a function of 2 variables and we are temporarily holding one of the variables constant.

So, instead, we write

$$\frac{\partial z}{\partial y} = \text{partial derivative of } z \text{ wrt } y$$

where

$$y = f(x) \rightarrow \frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

$$z = f(x, y) \rightarrow \left(\frac{\partial z}{\partial x} \right)_y = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

If the subscript indicating which variable(s) are constant is left off, then it means that all other variables are constant.

It is important to understand that the notation means z has to be written as a function of x and y only and then differentiated wrt x holding y constant (last equation).

Notation

$$\frac{\partial}{\partial x} \frac{\partial z}{\partial x} \equiv \frac{\partial^2 z}{\partial x^2} , \quad \frac{\partial}{\partial x} \frac{\partial z}{\partial y} \equiv \frac{\partial^2 z}{\partial x \partial y} , \quad \frac{\partial}{\partial x} \frac{\partial^2 z}{\partial x \partial y} \equiv \frac{\partial^3 z}{\partial x^2 \partial y} , \quad \text{etc.....}$$

Some examples:

(1)

$$z = x^2 - y^2$$

Using

$$x = r \cos \theta , \quad y = r \sin \theta , \quad x^2 + y^2 = r^2$$

we have

$$z = x^2 - y^2 = r^2 \cos^2 \theta - r^2 \sin^2 \theta \rightarrow \left(\frac{\partial z}{\partial r} \right)_{\theta} = 2r(\cos^2 \theta - \sin^2 \theta)$$

$$z = 2x^2 - r^2 \rightarrow \left(\frac{\partial z}{\partial r} \right)_x = -2r$$

$$z = r^2 - 2y^2 \rightarrow \left(\frac{\partial z}{\partial r} \right)_y = 2r$$

(2)

$$z = f(x, y) = xy + x^2$$

$$\rightarrow \frac{\partial z}{\partial x} = y + 2x = \left(\frac{\partial z}{\partial x} \right)_y$$

$$\rightarrow \frac{\partial z}{\partial y} = x = \left(\frac{\partial z}{\partial y} \right)_x$$

Suppose that $g = xy$, then we have

$$z = f(x, y) = xy + x^2$$

$$z = h(x, g) = g + x^2$$

$$\left(\frac{\partial f}{\partial x} \right)_g = \left(\frac{\partial h}{\partial x} \right)_g = 2x$$

We now digress to look at **total differentials** and find some interesting facts.

$$\begin{aligned}\Delta f(x, y) &= f(x + \Delta x, y + \Delta y) - f(x, y) \\ &= f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) + f(x + \Delta x, y) - f(x, y)\end{aligned}$$

Taking the limit as

$$\Delta x \rightarrow 0 \quad \text{and} \quad \Delta y \rightarrow 0$$

we get

$$df(x, y) = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dy$$

This result easily generalizes to any number of cartesian coordinates.

Now for some function $g(x, y)$, pick dx, dy s.t. $g(x, y) = \text{constant}$ as $(x \rightarrow x + \Delta x, y \rightarrow y + \Delta y)$.

This implies that

$$g(x + \Delta x, y + \Delta y) = g(x, y)$$

which in turn implies that

$$\begin{aligned} \left. \frac{df}{dx} \right|_{dg=0} &= \left(\frac{\partial f}{\partial x} \right)_y \left. \frac{dx}{dx} \right|_{dg=0} + \left(\frac{\partial f}{\partial y} \right)_x \left. \frac{dy}{dx} \right|_{dg=0} \\ &\rightarrow \left(\frac{\partial f}{\partial x} \right)_g = \left(\frac{\partial f}{\partial x} \right)_y + \left(\frac{\partial f}{\partial y} \right)_x \left(\frac{\partial y}{\partial x} \right)_g \end{aligned}$$

Going back to our old example we have

$$f = xy + x^2 \quad , \quad g = xy$$

$$\begin{aligned}
 \left(\frac{\partial f}{\partial x} \right)_g &= \left(\frac{\partial f}{\partial x} \right)_y + \left(\frac{\partial f}{\partial y} \right)_x \left(\frac{\partial y}{\partial x} \right)_g \\
 &= (y + 2x) + (x) \left(\frac{\partial \left(\frac{g}{x} \right)}{\partial x} \right)_g \\
 &= (y + 2x) + (x) \left(-\frac{g}{x^2} \right) \\
 &= y + 2x - \frac{g}{x} \\
 &= y + 2x - y = 2x \quad \text{as before}
 \end{aligned}$$

Chain Rule

Suppose we have a function $u(x(t), y(t))$.

For example,

$$u = xy \quad , \quad x(t) = t \quad , \quad y(t) = e^{\alpha t}$$

This implies that

$$\frac{du}{dt} = \frac{d}{dt}(te^{\alpha t}) = e^{\alpha t} + \alpha te^{\alpha t}$$

Nothing new arises if we use the chain rule

$$du(x, y) = \left(\frac{\partial u}{\partial x} \right)_y dx + \left(\frac{\partial u}{\partial y} \right)_x dy$$

$$\frac{du(x, y)}{dt} = \left(\frac{\partial u}{\partial x} \right)_y \frac{dx}{dt} + \left(\frac{\partial u}{\partial y} \right)_x \frac{dy}{dt} = y \frac{dx}{dt} + x \frac{dy}{dt} = e^{\alpha t} + \alpha te^{\alpha t}$$

General Chain Rule

Given $u(x, y)$, $x(s, t)$, $y(s, t)$ we have

$$du = \left(\frac{\partial u}{\partial x} \right)_y dx + \left(\frac{\partial u}{\partial y} \right)_x dy$$

$$dx = \left(\frac{\partial x}{\partial s} \right)_t ds + \left(\frac{\partial x}{\partial t} \right)_s dt$$

$$dy = \left(\frac{\partial y}{\partial s} \right)_t ds + \left(\frac{\partial y}{\partial t} \right)_s dt$$

which gives

$$\begin{aligned}
du &= \left(\frac{\partial u}{\partial x} \right)_y \left(\left(\frac{\partial x}{\partial s} \right)_t ds + \left(\frac{\partial x}{\partial t} \right)_s dt \right) \\
&+ \left(\frac{\partial u}{\partial y} \right)_x \left(\left(\frac{\partial y}{\partial s} \right)_t ds + \left(\frac{\partial y}{\partial t} \right)_s dt \right) \\
&= \left(\left(\frac{\partial u}{\partial x} \right)_y \left(\frac{\partial x}{\partial s} \right)_t + \left(\frac{\partial u}{\partial y} \right)_x \left(\frac{\partial y}{\partial s} \right)_t \right) ds \\
&+ \left(\left(\frac{\partial u}{\partial x} \right)_y \left(\frac{\partial x}{\partial t} \right)_s + \left(\frac{\partial u}{\partial y} \right)_x \left(\frac{\partial y}{\partial t} \right)_s \right) dt
\end{aligned}$$

Since we can also write

$$du = \left(\frac{\partial u}{\partial s} \right)_t ds + \left(\frac{\partial u}{\partial t} \right)_s dt$$

we have

$$\left(\frac{\partial u}{\partial s}\right)_t = \left(\frac{\partial u}{\partial x}\right)_y \left(\frac{\partial x}{\partial s}\right)_t + \left(\frac{\partial u}{\partial y}\right)_x \left(\frac{\partial y}{\partial s}\right)_t$$

$$\left(\frac{\partial u}{\partial t}\right)_s = \left(\frac{\partial u}{\partial x}\right)_y \left(\frac{\partial x}{\partial t}\right)_s + \left(\frac{\partial u}{\partial y}\right)_x \left(\frac{\partial y}{\partial t}\right)_s$$

Example:

$$u = x^2 + y, \quad x = s + t, \quad x - y = s^2$$

These give

$$x = \frac{s + t + s^2}{2}, \quad y = \frac{s + t - s^2}{2}$$

and

$$\left(\frac{\partial x}{\partial s}\right)_t = \frac{1 + 2s}{2}, \quad \left(\frac{\partial y}{\partial s}\right)_t = \frac{1 - 2s}{2}$$

$$\begin{aligned} \left(\frac{\partial u}{\partial s}\right)_t &= (2x) \frac{1 + 2s}{2} + (1) \frac{1 - 2s}{2} = (s + t)(1 - 2s) + \frac{1}{2} - s \\ &= \frac{1}{2} + t - 2ts - 2s^2 \end{aligned}$$

Repeat

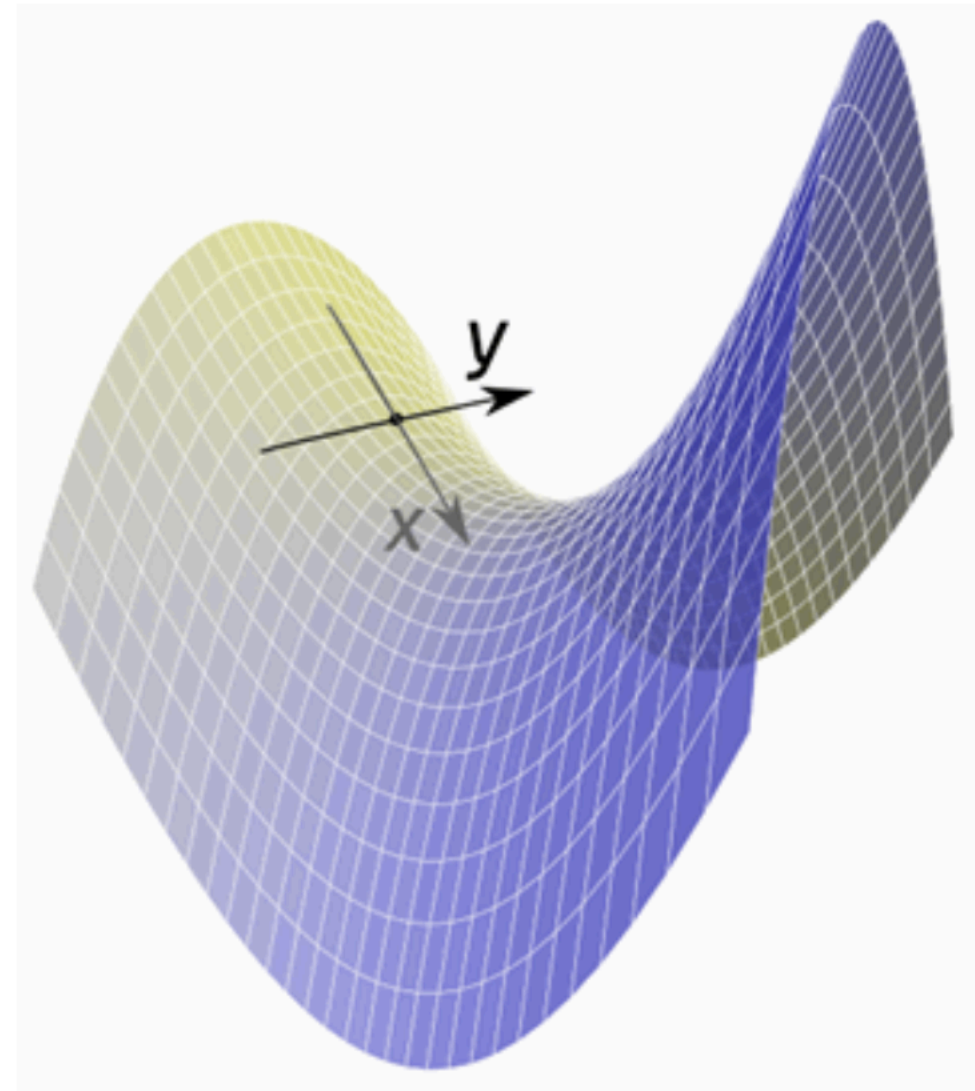
Partial Derivatives

A Partial Derivative is a derivative where we hold some variables constant. Like in this example:

Example: a function for a surface that depends on two variables **x** and **y**

When we find the slope in the **x** direction (while keeping **y** fixed) we have found a partial derivative.

Or we can find the slope in the **y** direction (while keeping **x** fixed).



Let's first think about a function of **one variable** (x):

$$f(x) = x^2$$

We can find its derivative using the Power Rule:

$$f'(x) = 2x$$

But what about a function of **two variables** (x and y):

$$f(x, y) = x^2 + y^3$$

We can find its **partial** derivative **with respect to x** when we treat **y as a constant** (imagine y is a number like 7 or something):

$$f'_x = 2x + 0 = 2x$$

Explanation:

- *the derivative of x^2 (with respect to x) is $2x$*
- *we **treat y as a constant**, so y^3 is also a constant (imagine $y=7$, then $7^3=343$ is also a constant), and the derivative of a constant is 0*

To find the partial derivative **with respect to y**, we treat **x as a constant**:

$$f'_y = 0 + 3y^2 = 3y^2$$

Explanation:

- we now **treat x as a constant**, so x^2 is also a constant, and the derivative of a constant is 0
- the derivative of y^3 (with respect to y) is $3y^2$

That is all there is to it. Just remember to treat **all other variables as if they are constants**.

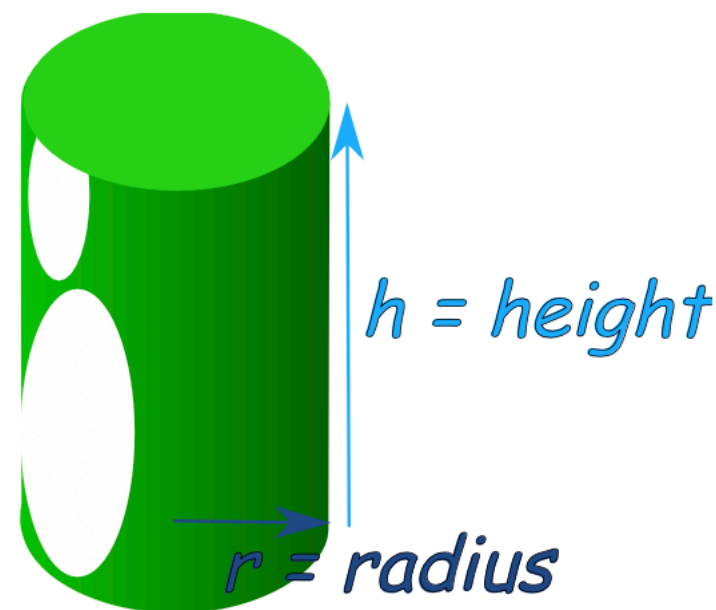
Holding A Variable Constant

So what does "holding a variable constant" look like?

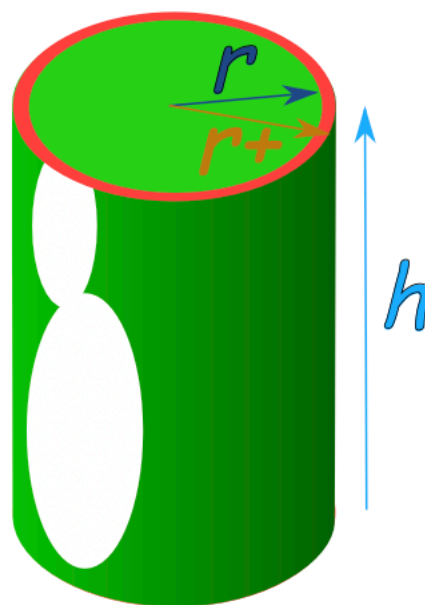
Example: the volume of a cylinder is $V = \pi r^2 h$

We can write that in "multi variable" form as

$$f(r, h) = \pi r^2 h$$



For the partial derivative with respect to r we hold **h constant**, and r changes:



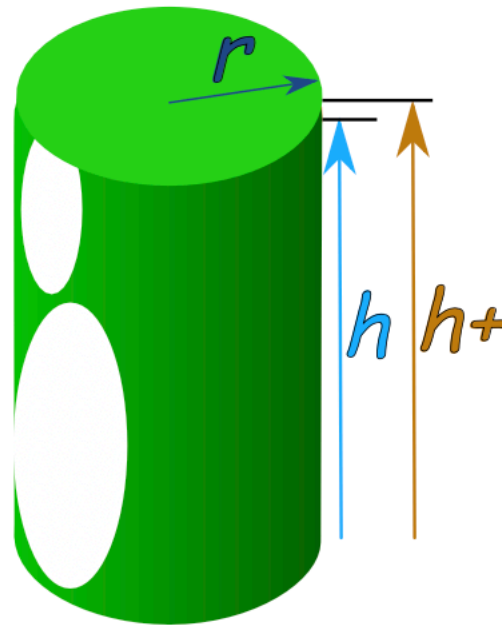
$$f'_r = \pi (2r) h = 2\pi r h$$

(The derivative of r^2 with respect to r is $2r$, and π and h are constants)

It says "as only the radius changes (by the tiniest amount), the volume changes by $2\pi r h$ "

It is like we add a skin with a circle's circumference ($2\pi r$) and a height of h .

For the partial derivative with respect to h we hold **r constant**:



$$f'_h = \pi r^2 (1) = \pi r^2$$

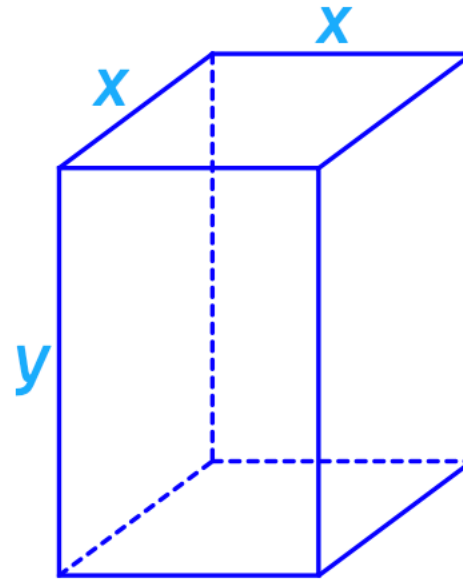
(π and r^2 are constants, and the derivative of h with respect to h is 1)

It says "as only the height changes (by the tiniest amount), the volume changes by πr^2 "

It is like we add the thinnest disk on top with a circle's area of πr^2 .

Let's see another example.

Example: The surface area of a square prism.



The surface includes the top and bottom with areas of x^2 each, and 4 sides of area xy each:

$$f(x, y) = 2x^2 + 4xy$$

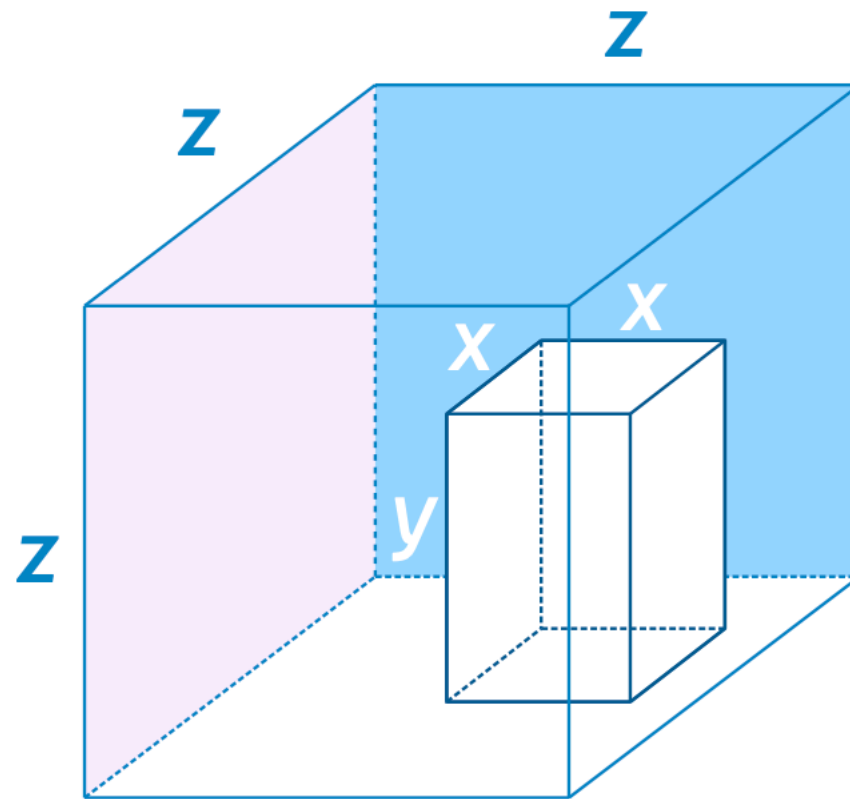
$$f'_x = 4x + 4y$$

$$f'_y = 0 + 4x = 4x$$

Three or More Variables

We can have 3 or more variables. Just find the partial derivative of each variable in turn while treating **all other variables as constants**.

Example: The volume of a cube with a square prism cut out from it.



$$f(x, y, z) = z^3 - x^2y$$

$$f'_x = 0 - 2xy = -2xy$$

$$f'_y = 0 - x^2 = -x^2$$

$$f'_z = 3z^2 - 0 = 3z^2$$

When there are many x's and y's it can get confusing, so a mental trick is to change the "constant" variables into letters like "c" or "k" that *look* like constants.

$$\text{Example: } f(x, y) = y^3 \sin(x) + x^2 \tan(y)$$

It has x's and y's all over the place! So let us try the letter change trick.

With respect to x we can change "y" to "k":

$$f(x, y) = k^3 \sin(x) + x^2 \tan(k)$$

$$f'_x = k^3 \cos(x) + 2x \tan(k)$$

But remember to turn it back again!

$$f'_x = y^3 \cos(x) + 2x \tan(y)$$

Likewise with respect to y we turn the "x" into a "k":

$$f(x, y) = y^3 \sin(k) + k^2 \tan(y)$$

$$f'_y = 3y^2 \sin(k) + k^2 \sec^2(y)$$

$$f'_y = 3y^2 \sin(x) + x^2 \sec^2(y)$$

But only do this if you have trouble remembering, as it is a little extra work.

Notation: we have used $\mathbf{f}'_{\mathbf{x}}$ to mean "the partial derivative with respect to x", but another very common notation is to use a funny backwards d (∂) like this:

$$\frac{\partial f}{\partial x} = 2x$$

Which is the same as:

$$\mathbf{f}'_{\mathbf{x}} = 2x$$

∂ is called "del" or "dee" or "curly dee"

So $\frac{\partial f}{\partial x}$ can be said "del f del x"

Example: find the partial derivatives of $\mathbf{f(x, y, z) = x^4 - 3xyz}$ using "curly dee" notation

$$f(x, y, z) = x^4 - 3xyz$$

$$\frac{\partial f}{\partial x} = 4x^3 - 3yz$$

$$\frac{\partial f}{\partial y} = -3xz$$

$$\frac{\partial f}{\partial z} = -3xy$$

Do you want to look a bit at PDEs ?

Linear Algebra

Basic Linear Algebra

2.1 Matrices and Vectors

Matrices

DEFINITION ■ A **matrix** is any rectangular array of numbers. ■

For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad [2 \quad 1]$$

are all matrices.

If a matrix A has m rows and n columns, we call A an $m \times n$ matrix. We refer to $m \times n$ as the **order** of the matrix. A typical $m \times n$ matrix A may be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

DEFINITION ■ The number in the i th row and j th column of A is called the **ij th element** of A and is written a_{ij} . ■

For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

then $a_{11} = 1$, $a_{23} = 6$, and $a_{31} = 7$.

Sometimes we will use the notation $A = [a_{ij}]$ to indicate that A is the matrix whose ij th element is a_{ij} .

DEFINITION ■ Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are **equal** if and only if A and B are of the same order and for all i and j , $a_{ij} = b_{ij}$. ■

For example, if

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} x & y \\ w & z \end{bmatrix}$$

then $A = B$ if and only if $x = 1$, $y = 2$, $w = 3$, and $z = 4$.

Vectors

Any matrix with only one column (that is, any $m \times 1$ matrix) may be thought of as a **column vector**. The number of rows in a column vector is the **dimension** of the column vector. Thus,

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

may be thought of as a 2×1 matrix or a two-dimensional column vector. R^m will denote the set of all m -dimensional column vectors.

In analogous fashion, we can think of any vector with only one row (a $1 \times n$ matrix) as a **row vector**. The dimension of a row vector is the number of columns in the vector. Thus, $[9 \ 2 \ 3]$ may be viewed as a 1×3 matrix or a three-dimensional row vector. In this book, vectors appear in boldface type: for instance, vector **v**. An m -dimensional vector (either row or column) in which all elements equal zero is called a **zero vector** (written **0**). Thus,

$$[0 \ 0] \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

are two-dimensional zero vectors.

Any m -dimensional vector corresponds to a directed line segment in the m -dimensional plane. For example, in the two-dimensional plane, the vector

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

corresponds to the line segment joining the point

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

to the point

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

The directed line segments corresponding to

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

are drawn in Figure 1.

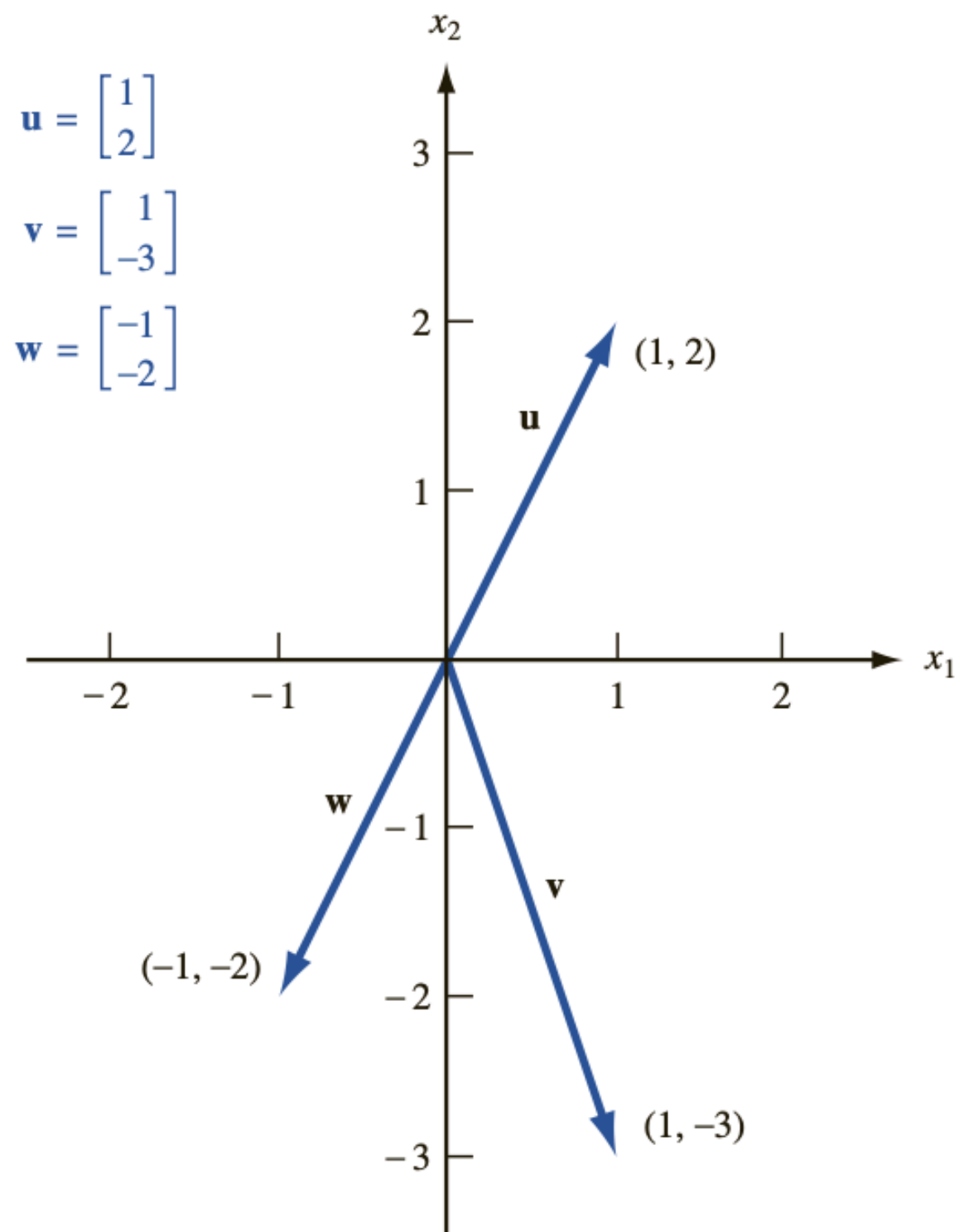


FIGURE 1
Vectors Are Directed
Line Segments

The Scalar Product of Two Vectors

An important result of multiplying two vectors is the *scalar product*. To define the scalar product of two vectors, suppose we have a row vector $\mathbf{u} = [u_1 \ u_2 \ \cdots \ u_n]$ and a column vector

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

of the same dimension. The **scalar product** of $\mathbf{u}$ and $\mathbf{v}$ (written $\mathbf{u} \cdot \mathbf{v}$) is the number $u_1v_1 + u_2v_2 + \cdots + u_nv_n$.

For the scalar product of two vectors to be defined, the first vector must be a row vector and the second vector must be a column vector. For example, if

$$\mathbf{u} = [1 \ 2 \ 3] \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

then $\mathbf{u} \cdot \mathbf{v} = 1(2) + 2(1) + 3(2) = 10$. By these rules for computing a scalar product, if

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = [2 \ 3]$$

then $\mathbf{u} \cdot \mathbf{v}$ is not defined. Also, if

$$\mathbf{u} = [1 \ 2 \ 3] \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

then $\mathbf{u} \cdot \mathbf{v}$ is not defined because the vectors are of two different dimensions.

Note that two vectors are perpendicular if and only if their scalar product equals 0. Thus, the vectors $[1 \ -1]$ and $[1 \ 1]$ are perpendicular.

We note that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$, where $\|\mathbf{u}\|$ is the length of the vector $\mathbf{u}$ and θ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$.

Matrix Operations

We now describe the arithmetic operations on matrices that are used later in this book.

The Scalar Multiple of a Matrix

Given any matrix A and any number c (a *number* is sometimes referred to as a *scalar*), the matrix cA is obtained from the matrix A by multiplying each element of A by c . For example,

$$\text{if } A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}, \quad \text{then } 3A = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$$

For $c = -1$, scalar multiplication of the matrix A is sometimes written as $-A$.

Addition of Two Matrices

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices with the same order (say, $m \times n$). Then the matrix $C = A + B$ is defined to be the $m \times n$ matrix whose ij th element is $a_{ij} + b_{ij}$. Thus, to obtain the sum of two matrices A and B , we add the corresponding elements of A and B . For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 1 & -1 \end{bmatrix}$$

then

$$A + B = \begin{bmatrix} 1 - 1 & 2 - 2 & 3 - 3 \\ 0 + 2 & -1 + 1 & 1 - 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

This rule for matrix addition may be used to add vectors of the same dimension. For example, if $\mathbf{u} = [1 \ 2]$ and $\mathbf{v} = [2 \ 1]$, then $\mathbf{u} + \mathbf{v} = [1 + 2 \ 2 + 1] = [3 \ 3]$. Vectors may be added geometrically by the parallelogram law (see Figure 2).

We can use scalar multiplication and the addition of matrices to define the concept of a line segment. A glance at Figure 1 should convince you that any point u in the m -dimensional plane corresponds to the m -dimensional vector $\mathbf{u}$ formed by joining the origin to the point u . For any two points u and v in the m -dimensional plane, the **line segment** joining u and v (called the line segment uv) is the set of all points in the m -dimensional plane that correspond to the vectors $c\mathbf{u} + (1 - c)\mathbf{v}$, where $0 \leq c \leq 1$ (Figure 3). For example, if $u = (1, 2)$ and $v = (2, 1)$, then the line segment uv consists of the points corresponding to the vectors $c[1 \ 2] + (1 - c)[2 \ 1] = [2 - c \ 1 + c]$, where $0 \leq c \leq 1$. For $c = 0$ and $c = 1$, we obtain the endpoints of the line segment uv ; for $c = \frac{1}{2}$, we obtain the midpoint $(0.5\mathbf{u} + 0.5\mathbf{v})$ of the line segment uv .

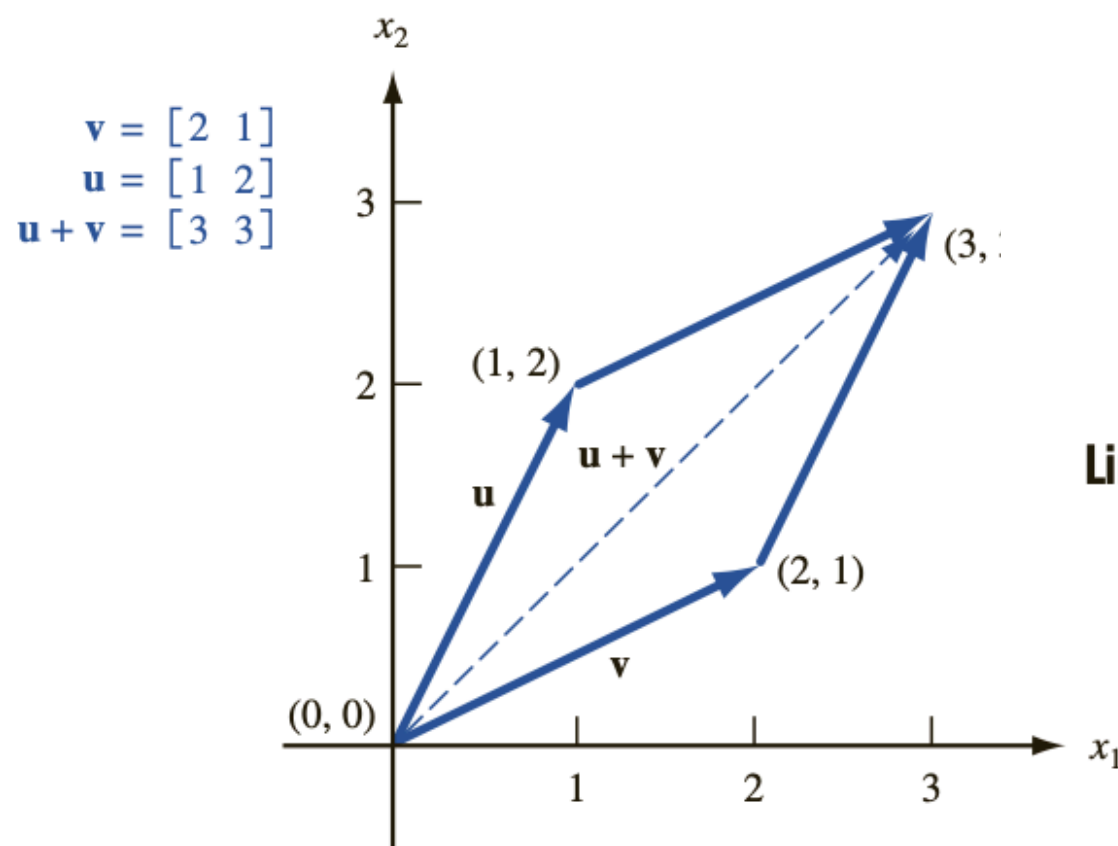
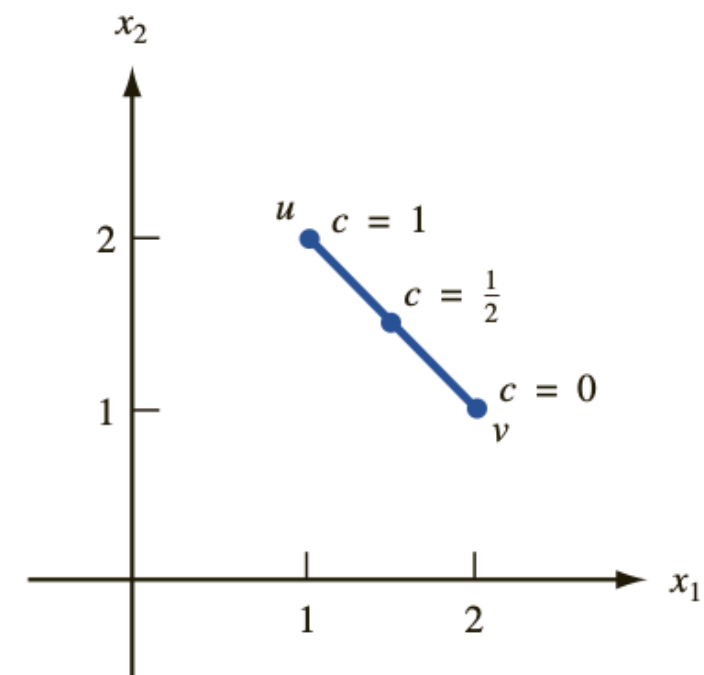


FIGURE 2
Addition of Vectors

FIGURE 3
Line Segment Joining
 $u = (1, 2)$ and
 $v = (2, 1)$



Using the parallelogram law, the line segment uv may also be viewed as the points corresponding to the vectors $\mathbf{u} + c(\mathbf{v} - \mathbf{u})$, where $0 \leq c \leq 1$ (Figure 4). Observe that for $c = 0$, we obtain the vector $\mathbf{u}$ (corresponding to point u), and for $c = 1$, we obtain the vector $\mathbf{v}$ (corresponding to point v).

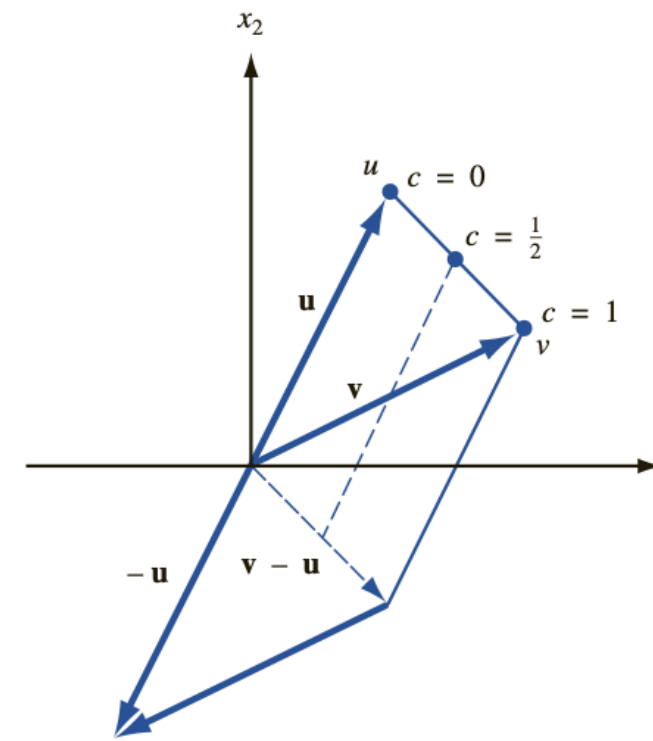


FIGURE 4
Representation of Line
Segment uv

The Transpose of a Matrix

Given any $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

the **transpose** of A (written A^T) is the $n \times m$ matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Thus, A^T is obtained from A by letting row 1 of A be column 1 of A^T , letting row 2 of A be column 2 of A^T , and so on. For example,

$$\text{if } A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{then } A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Observe that $(A^T)^T = A$. Let $B = [1 \ 2]$; then

$$B^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad (B^T)^T = [1 \ 2] = B$$

As indicated by these two examples, for any matrix A , $(A^T)^T = A$.

Matrix Multiplication

Given two matrices A and B , the matrix product of A and B (written AB) is defined if and only if

$$\text{Number of columns in } A = \text{number of rows in } B \quad (1)$$

For the moment, assume that for some positive integer r , A has r columns and B has r rows. Then for some m and n , A is an $m \times r$ matrix and B is an $r \times n$ matrix.

DEFINITION ■

The **matrix product** $C = AB$ of A and B is the $m \times n$ matrix C whose ij th element is determined as follows:

$$ij\text{th element of } C = \text{scalar product of row } i \text{ of } A \times \text{column } j \text{ of } B \quad \blacksquare \quad (2)$$

If Equation (1) is satisfied, then each row of A and each column of B will have the same number of elements. Also, if (1) is satisfied, then the scalar product in Equation (2) will be defined. The product matrix $C = AB$ will have the same number of rows as A and the same number of columns as B .

EXAMPLE 1**Matrix Multiplication**

Compute $C = AB$ for

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}$$

Solution Because A is a 2×3 matrix and B is a 3×2 matrix, AB is defined, and C will be a 2×2 matrix. From Equation (2),

$$c_{11} = [1 \quad 1 \quad 2] \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 1(1) + 1(2) + 2(1) = 5$$

$$c_{12} = [1 \quad 1 \quad 2] \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 1(1) + 1(3) + 2(2) = 8$$

$$c_{21} = [2 \quad 1 \quad 3] \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 2(1) + 1(2) + 3(1) = 7$$

$$c_{22} = [2 \quad 1 \quad 3] \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 2(1) + 1(3) + 3(2) = 11$$

$$C = AB = \begin{bmatrix} 5 & 8 \\ 7 & 11 \end{bmatrix}$$

EXAMPLE 2 Column Vector Times Row Vector

Find AB for

$$A = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{and} \quad B = [1 \quad 2]$$

Solution Because A has one column and B has one row, $C = AB$ will exist. From Equation (2), we know that C is a 2×2 matrix with

$$\begin{aligned} c_{11} &= 3(1) = 3 & c_{21} &= 4(1) = 4 \\ c_{12} &= 3(2) = 6 & c_{22} &= 4(2) = 8 \end{aligned}$$

Thus,

$$C = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix}$$

EXAMPLE 3**Row Vector Times Column Vector**

Compute $D = BA$ for the A and B of Example 2.

Solution In this case, D will be a 1×1 matrix (or a scalar). From Equation (2),

$$d_{11} = [1 \quad 2] \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 1(3) + 2(4) = 11$$

Thus, $D = [11]$. In this example, matrix multiplication is equivalent to scalar multiplication of a row and column vector.

Recall that if you multiply two real numbers a and b , then $ab = ba$. This is called the *commutative property of multiplication*. Examples 2 and 3 show that for matrix multiplication, it may be that $AB \neq BA$. Matrix multiplication is not necessarily commutative. (In some cases, however, $AB = BA$ will hold.)

EXAMPLE 4**Undefined Matrix Product**

Show that AB is undefined if

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}$$

Solution This follows because A has two columns and B has three rows. Thus, Equation (1) is not satisfied.

TABLE 1

Gallons of Crude Oil Required to Produce 1 Gallon
of Gasoline

Crude Oil	Premium Unleaded	Regular Unleaded	Regular Leaded
1	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$

Many computations that commonly occur in operations research (and other branches of mathematics) can be concisely expressed by using matrix multiplication. To illustrate this, suppose an oil company manufactures three types of gasoline: premium unleaded, regular unleaded, and regular leaded. These gasolines are produced by mixing two types of crude oil: crude oil 1 and crude oil 2. The number of gallons of crude oil required to manufacture 1 gallon of gasoline is given in Table 1.

From this information, we can find the amount of each type of crude oil needed to manufacture a given amount of gasoline. For example, if the company wants to produce 10 gallons of premium unleaded, 6 gallons of regular unleaded, and 5 gallons of regular leaded, then the company's crude oil requirements would be

$$\text{Crude 1 required} = \left(\frac{3}{4}\right)(10) + \left(\frac{2}{3}\right)(6) + \left(\frac{1}{4}\right)5 = 12.75 \text{ gallons}$$

$$\text{Crude 2 required} = \left(\frac{1}{4}\right)(10) + \left(\frac{1}{3}\right)(6) + \left(\frac{3}{4}\right)5 = 8.25 \text{ gallons}$$

More generally, we define

p_U = gallons of premium unleaded produced

r_U = gallons of regular unleaded produced

r_L = gallons of regular leaded produced

c_1 = gallons of crude 1 required

c_2 = gallons of crude 2 required

Then the relationship between these variables may be expressed by

$$\begin{aligned}c_1 &= \left(\frac{3}{4}\right) p_U + \left(\frac{2}{3}\right) r_U + \left(\frac{1}{4}\right) r_L \\c_2 &= \left(\frac{1}{4}\right) p_U + \left(\frac{1}{3}\right) r_U + \left(\frac{3}{4}\right) r_L\end{aligned}$$

Using matrix multiplication, these relationships may be expressed by

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{2}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{3} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} p_U \\ r_U \\ r_L \end{bmatrix}$$

Properties of Matrix Multiplication

To close this section, we discuss some important properties of matrix multiplication. In what follows, we assume that all matrix products are defined.

1 Row i of $AB = (\text{row } i \text{ of } A)B$. To illustrate this property, let

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}$$

Then row 2 of the 2×2 matrix AB is equal to

$$[2 \quad 1 \quad 3] \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix} = [7 \quad 11]$$

This answer agrees with Example 1.

2 Column j of $AB = A(\text{column } j \text{ of } B)$. Thus, for A and B as given, the first column of AB is

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

Properties 1 and 2 are helpful when you need to compute only *part* of the matrix AB .

3 Matrix multiplication is associative. That is, $A(BC) = (AB)C$. To illustrate, let

$$A = [1 \quad 2], \quad B = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Then $AB = [10 \quad 13]$ and $(AB)C = 10(2) + 13(1) = [33]$.

On the other hand,

$$BC = \begin{bmatrix} 7 \\ 13 \end{bmatrix}$$

so $A(BC) = 1(7) + 2(13) = [33]$. In this case, $A(BC) = (AB)C$ does hold.

4 Matrix multiplication is distributive. That is, $A(B + C) = AB + AC$ and $(B + C)D = BD + CD$.

2.2 Matrices and Systems of Linear Equations

Consider a system of linear equations given by

[illegible]

In Equation (3), $x_1, x_2, \dots, x_n$ are referred to as **variables**, or unknowns, and the a_{ij} 's and b_i 's are **constants**. A set of equations such as (3) is called a linear system of m equations in n variables.

DEFINITION ■ A **solution** to a linear system of m equations in n unknowns is a set of values for the unknowns that satisfies each of the system's m equations. ■

To understand linear programming, we need to know a great deal about the properties of solutions to linear equation systems. With this in mind, we will devote much effort to studying such systems.

We denote a possible solution to Equation (3) by an n -dimensional column vector $\mathbf{x}$, in which the i th element of $\mathbf{x}$ is the value of x_i . The following example illustrates the concept of a solution to a linear system.

Show that

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is a solution to the linear system

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ 2x_1 - x_2 &= 0 \end{aligned} \tag{4}$$

and that

$$\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

is not a solution to linear system (4).

Solution To show that

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is a solution to Equation (4), we substitute $x_1 = 1$ and $x_2 = 2$ in both equations and check that they are satisfied: $1 + 2(2) = 5$ and $2(1) - 2 = 0$.

The vector

$$\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

is not a solution to (4), because $x_1 = 3$ and $x_2 = 1$ fail to satisfy $2x_1 - x_2 = 0$.

Using matrices can greatly simplify the statement and solution of a system of linear equations. To show how matrices can be used to compactly represent Equation (3), let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Then (3) may be written as

$$A\mathbf{x} = \mathbf{b} \tag{5}$$

Observe that both sides of Equation (5) will be $m \times 1$ matrices (or $m \times 1$ column vectors). For the matrix $A\mathbf{x}$ to equal the matrix $\mathbf{b}$ (or for the vector $A\mathbf{x}$ to equal the vector $\mathbf{b}$), their corresponding elements must be equal. The first element of $A\mathbf{x}$ is the scalar product of row 1 of A with $\mathbf{x}$. This may be written as

$$[a_{11} \quad a_{12} \quad \cdots \quad a_{1n}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$$

This must equal the first element of $\mathbf{b}$ (which is b_1). Thus, (5) implies that $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$. This is the first equation of (3). Similarly, (5) implies that the scalar

product of row i of A with $\mathbf{x}$ must equal b_i , and this is just the i th equation of (3). Our discussion shows that (3) and (5) are two different ways of writing the same linear system. We call (5) the **matrix representation** of (3). For example, the matrix representation of (4) is

$$\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Sometimes we abbreviate (5) by writing

$$A|\mathbf{b} \tag{6}$$

If A is an $m \times n$ matrix, it is assumed that the variables in (6) are $x_1, x_2, \dots, x_n$. Then (6) is still another representation of (3). For instance, the matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{array} \right]$$

represents the system of equations

$$x_1 + 2x_2 + 3x_3 = 2$$

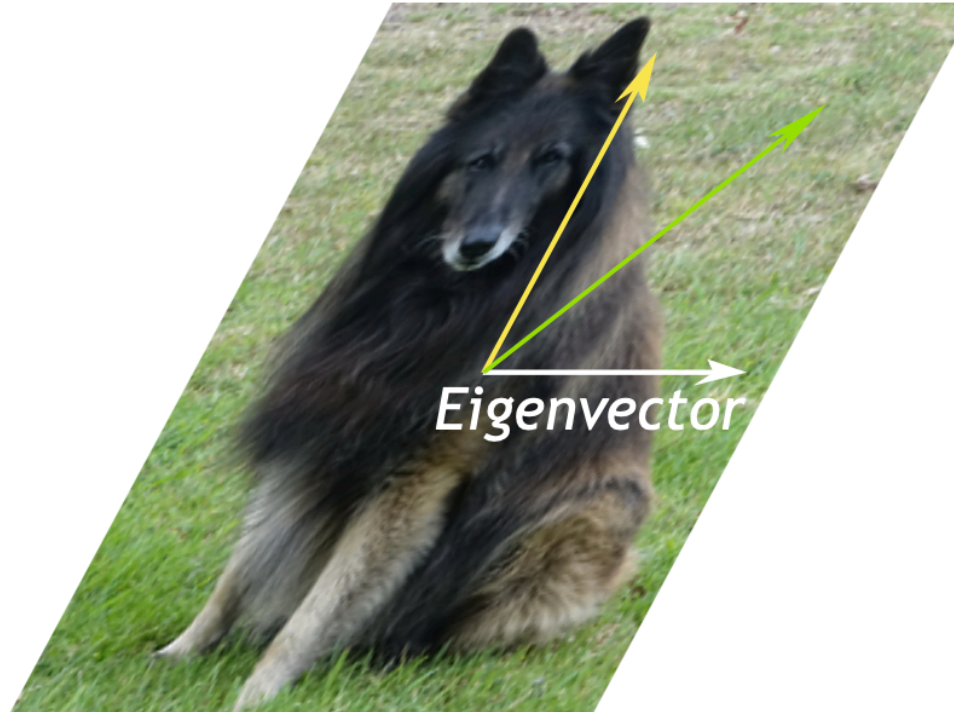
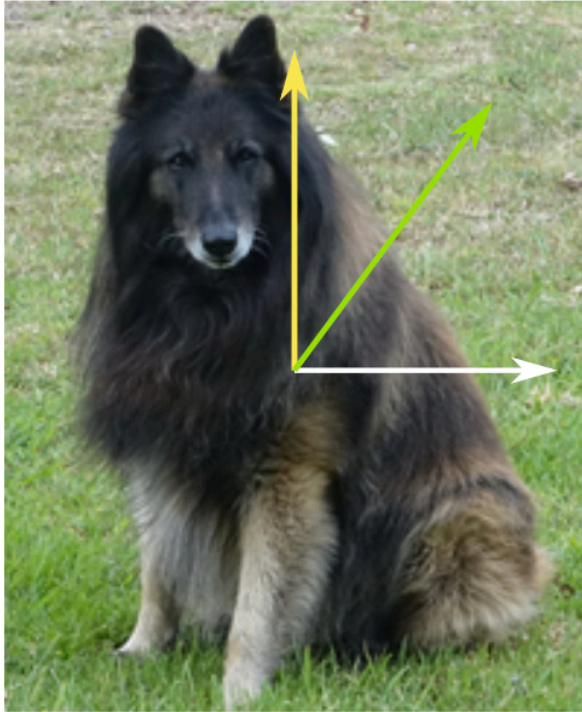
$$x_2 + 2x_3 = 3$$

$$x_1 + x_2 + x_3 = 1$$

More Linear Algebra

Eigenvector and Eigenvalue

A simple example is that an eigenvector **does not change direction** in a transformation:



How do we find that vector?

The Mathematics Of It

For a square matrix **A**, an Eigenvector and Eigenvalue make this equation true:

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

Diagram illustrating the equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ with labels and arrows:

- Matrix** (brown text) points to $\mathbf{A}$ (brown text).
- Eigenvector** (orange text) points to $\mathbf{v}$ (orange text) in both $\mathbf{A}\mathbf{v}$ and $\lambda\mathbf{v}$.
- Eigenvalue** (blue text) points to λ (blue text).

Let us see it in action:

Example: For this matrix

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}$$

an eigenvector is

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

with a matching eigenvalue of 6

Let's do some matrix multiplies to see if that is true.

Av gives us:

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} -6 \times 1 + 3 \times 4 \\ 4 \times 1 + 5 \times 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

λv gives us :

$$6 \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

Yes they are equal!

So we get $Av = \lambda v$ as promised.

Notice how we multiply a **matrix** by a **vector** and get the same result as when we multiply a **scalar** (just a number) by that **vector**.

How do we find these eigen things?

We start by finding the **eigenvalue**. We know this equation must be true:

$$Av = \lambda v$$

Next we put in an identity matrix so we are dealing with matrix-vs-matrix:

$$Av = \lambda Iv$$

Bring all to left hand side:

$$Av - \lambda Iv = 0$$

If **v** is non-zero then we can (hopefully) solve for λ using just the determinant:

$$| A - \lambda I | = 0$$

Let's try that equation on our previous example:

Example: Solve for λ

Start with $|A - \lambda I| = 0$

$$\left| \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

Which is:

$$\begin{vmatrix} -6-\lambda & 3 \\ 4 & 5-\lambda \end{vmatrix} = 0$$

Calculating that determinant gets:

$$(-6-\lambda)(5-\lambda) - 3 \times 4 = 0$$

Which simplifies to this Quadratic Equation:

$$\lambda^2 + \lambda - 42 = 0$$

And solving it gets:

$$\lambda = -7 \text{ or } 6$$

And yes, there are **two** possible eigenvalues.

Now we know **eigenvalues**, let us find their matching **eigenvectors**.

Example (continued): Find the Eigenvector for the Eigenvalue $\lambda = 6$:

Start with:

$$Av = \lambda v$$

Put in the values we know:

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 6 \begin{bmatrix} x \\ y \end{bmatrix}$$

After multiplying we get these two equations:

$$-6x + 3y = 6x$$

$$4x + 5y = 6y$$

Bringing all to left hand side:

$$-12x + 3y = 0$$

$$4x - 1y = 0$$

Either equation reveals that $y = 4x$, so the **eigenvector** is any non-zero multiple of this:

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

And we get the solution shown at the top of the page:

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} -6 \times 1 + 3 \times 4 \\ 4 \times 1 + 5 \times 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

... and also ...

$$6 \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

So $Av = \lambda v$, and we have success!

Now it is **your turn** to find the eigenvector for the other eigenvalue of -7

Why?

What is the purpose of these?

One of the cool things is we can use matrices to do transformations in space, which is used a lot in computer graphics.

In that case the eigenvector is "the direction that doesn't change direction" !

And the eigenvalue is the scale of the stretch:

- **1** means no change,
- **2** means doubling in length,
- **-1** means pointing backwards along the eigenvalue's direction
- etc

There are also many applications in physics, etc.

Why "Eigen"

Eigen is a German word meaning "own" or "typical"

*"das ist ihnen **eigen**" is German for "that is **typical** of them"*

Sometimes in English we use the word "characteristic", so an eigenvector can be called a "characteristic vector".

Not Just Two Dimensions

Eigenvectors work perfectly well in 3 and higher dimensions.

Example: find the eigenvalues for this 3x3 matrix:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 5 \\ 0 & 4 & 3 \end{bmatrix}$$

First calculate $A - \lambda I$:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 5 \\ 0 & 4 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2-\lambda & 0 & 0 \\ 0 & 4-\lambda & 5 \\ 0 & 4 & 3-\lambda \end{bmatrix}$$

Now the determinant should equal zero:

$$\begin{vmatrix} 2-\lambda & 0 & 0 \\ 0 & 4-\lambda & 5 \\ 0 & 4 & 3-\lambda \end{vmatrix} = 0$$

Which is:

$$(2-\lambda) [(4-\lambda)(3-\lambda) - 5 \times 4] = 0$$

This ends up being a cubic equation, but just looking at it here we see one of the roots is **2** (because of $2-\lambda$), and the part inside the square brackets is Quadratic, with roots of **-1** and **8**.

So the Eigenvalues are **-1**, **2** and **8**

Example (continued): find the Eigenvector that matches the Eigenvalue **-1**

Put in the values we know:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 5 \\ 0 & 4 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = -1 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

After multiplying we get these equations:

$$2x = -x$$

$$4y + 5z = -y$$

$$4y + 3z = -z$$

Bringing all to left hand side:

$$3x = 0$$

$$5y + 5z = 0$$

$$4y + 4z = 0$$

So **$x = 0$** , and **$y = -z$** and so the **eigenvector** is any non-zero multiple this:

$$\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

TEST Av :

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 5 \\ 0 & 4 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 4-5 \\ 4-3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

And λv :

$$-1 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

So $Av = \lambda v$, yay!

(You can try your hand at the eigenvalues of **2** and **8**)

Rotation

Back in the 2D world again, this matrix will do a rotation by θ :

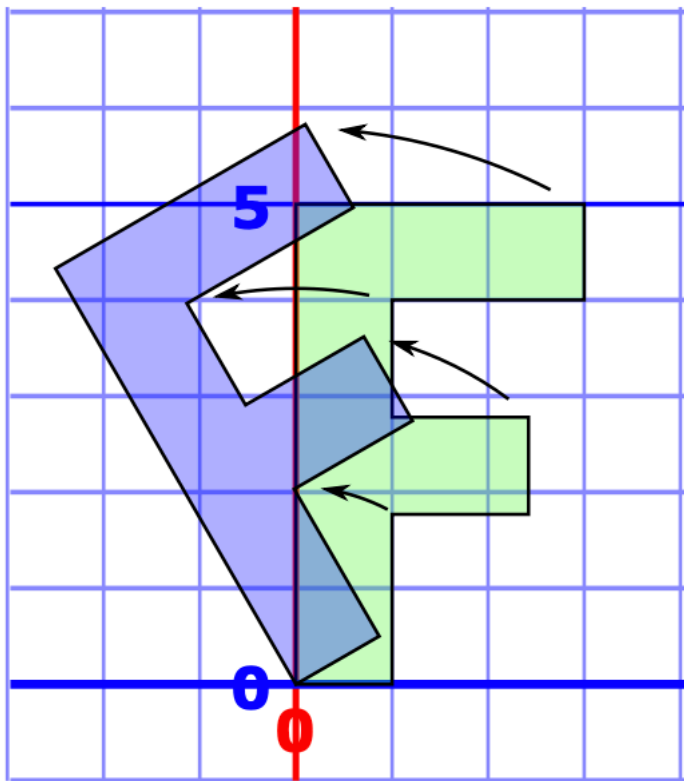
$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Example: Rotate by 30°

$\cos(30^\circ) = \frac{\sqrt{3}}{2}$ and $\sin(30^\circ) = \frac{1}{2}$, so:

$$\begin{bmatrix} \cos(30^\circ) & -\sin(30^\circ) \\ \sin(30^\circ) & \cos(30^\circ) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

But if we **rotate all points**, what is the "direction that doesn't change direction"?



Let us work through the mathematics to find out:

First calculate $A - \lambda I$:

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} - \lambda & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} - \lambda \end{bmatrix}$$

Now the determinant should equal zero:

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

Which is:

$$\left(\frac{\sqrt{3}}{2} - \lambda\right)\left(\frac{\sqrt{3}}{2} - \lambda\right) - \left(\frac{-1}{2}\right)\left(\frac{1}{2}\right) = 0$$

Which becomes this Quadratic Equation:

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

Whose roots are:

$$\lambda = \frac{\sqrt{3}}{2} \pm \frac{i}{2}$$

The eigenvalues are complex!

I don't know how to show you that on a graph, but we still get a solution.

Eigenvector

So, what is an eigenvector that matches, say, the $\frac{\sqrt{3}}{2} + \frac{i}{2}$ root?

Start with:

$$Av = \lambda v$$

Put in the values we know:

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & -1 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \left(\frac{\sqrt{3}}{2} + \frac{i}{2} \right) \begin{bmatrix} x \\ y \end{bmatrix}$$

After multiplying we get these two equations:

$$\frac{\sqrt{3}}{2}x - \frac{1}{2}y = \frac{\sqrt{3}}{2}x + \frac{i}{2}x$$

$$\frac{1}{2}x + \frac{\sqrt{3}}{2}y = \frac{\sqrt{3}}{2}y + \frac{i}{2}y$$

Which simplify to:

$$-y = ix$$

$$x = iy$$

And the solution is any non-zero multiple of:

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

or

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

Wow, such a simple answer!

Is this just because we chose 30°? Or does it work for any rotation matrix? I will let you work that out! Try another angle, or better still use "cos(θ)" and "sin(θ)".

Oh, and let us **check** at least one of those solutions:

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} i\frac{\sqrt{3}}{2} - \frac{1}{2} \\ \frac{i}{2} + \frac{\sqrt{3}}{2} \end{bmatrix}$$

Does it match this?

$$\left(\frac{\sqrt{3}}{2} + \frac{i}{2}\right) \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} i\frac{\sqrt{3}}{2} - \frac{1}{2} \\ \frac{\sqrt{3}}{2} + \frac{i}{2} \end{bmatrix}$$

Oh yes it does!

Solving Systems of Linear Equations Using Matrices

Example: Solve

- $x + y + z = 6$
- $2y + 5z = -4$
- $2x + 5y - z = 27$

But first we need to write the question in Matrix form.

In Matrix Form?

OK. A Matrix is an array of numbers, right?

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}$$

A Matrix

Well, think about the equations:

$$\begin{array}{cccccc} x & + & y & + & z & = & 6 \\ & & 2y & + & 5z & = & -4 \\ 2x & + & 5y & - & z & = & 27 \end{array}$$

They could be turned into a table of numbers like this:

$$\begin{array}{rrcr} 1 & 1 & 1 & = 6 \\ 0 & 2 & 5 & = -4 \\ 2 & 5 & -1 & = 27 \end{array}$$

We could even separate the numbers before and after the "=" into:

$$\begin{array}{rrr} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 2 & 5 & -1 \end{array} \quad \text{and} \quad \begin{array}{r} 6 \\ -4 \\ 27 \end{array}$$

Now it looks like we have 2 Matrices.

In fact we have a third one, which is $[x \ y \ z]$:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ -4 \\ 27 \end{bmatrix}$$

Why does $[x \ y \ z]$ go there? Because when we Multiply Matrices the left side becomes:

"Dot Product"

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + y + z \\ 2y + 5z \\ 2x + 5y - z \end{bmatrix}$$

Which is the original left side of our equations above (you might like to check that).

The Matrix Solution

We can write this:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ -4 \\ 27 \end{bmatrix}$$

like this:

$$AX = B$$

where

- **A** is the 3x3 matrix of x, y and z **coefficients**
- **X** is **x, y and z**, and
- **B** is **6, -4 and 27**

Then (as shown on the [Inverse of a Matrix](#) page) the solution is this:

$$X = A^{-1}B$$

What does that mean?

It means that we can find the values of x, y and z (the X matrix) by multiplying the **inverse of the A matrix** by the **B matrix**.

So let's go ahead and do that.

First, we need to find the **inverse of the A matrix** (assuming it exists!)

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 2 & 5 & -1 \end{bmatrix}^{-1} = \frac{1}{-21} \begin{bmatrix} -27 & 6 & 3 \\ 10 & -3 & -5 \\ -4 & -3 & 2 \end{bmatrix}$$

Then multiply A^{-1} by B (we can use the Matrix Calculator again):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{-21} \begin{bmatrix} -27 & 6 & 3 \\ 10 & -3 & -5 \\ -4 & -3 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -4 \\ 27 \end{bmatrix} = \frac{1}{-21} \begin{bmatrix} -105 \\ -63 \\ 42 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$$

And we are done! The solution is:

$$\begin{aligned} x &= 5, \\ y &= 3, \\ z &= -2 \end{aligned}$$

Group Theory

Group Theory and Elementary Particles

Introduction to Group Theory

The theory of finite groups and continuous groups is a very useful tool for studying symmetry and invariance.

In order to use group theory we need to introduce some definitions and concepts.

A **group** G is defined as a set of objects or operations (called **elements**) that may be combined or **multiplied** to form a well-defined product and that satisfy the following four conditions.

If we label the elements $a, b, c, \dots$, then the conditions are:

1. If a and b are any two elements, then the product ab is an element.
2. The defined multiplication is associative, $(ab)c = a(bc)$
3. There is a unit element I , with $Ia = aI = a$ for all elements a .
4. Each element has an inverse $b = a^{-1}$, with $a^{-1} = a^{-1}a = I$ for all elements a .

In physics, these abstract conditions will take on direct physical meaning in terms of transformations of vectors and tensors.

As a very simple, but not trivial, example of a group, consider the set $\{I, a, b, c\}$ that combine according to the group multiplication table below.

	I	a	b	c
I	I	a	b	c
a	a	b	c	I
b	b	c	I	a
c	c	I	a	b

Clearly, the four conditions of the definition of group are satisfied.

The elements $\{I, a, b, c\}$ are abstract mathematical entities, completely unrestricted except for the above multiplication table.

A **representation** of the group is a set of particular objects $\{I, a, b, c\}$ that satisfy the multiplication table.

Some examples are:

$$\{I = 1, a = i, b = -1, c = -i\}$$

where the combination rules=ordinary multiplication.

This group representation is labelled C_4 .

Since multiplication of the group elements is “commutative”, i.e., $ab = ba$ for any pair of elements, the group is labelled “commutative or abelian”.

This group is also a “cyclic” group, since all of the elements can be written as successive powers of one element. In this case,

$$\{I = i^4, a = i, b = i^2, c = i^3\}$$

Another representation is given by the successive 90° rotations in a plane.

Remember the matrix that represents a rotation through angle ϕ in 2-dimensions is given by

$$R = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

With $\phi = 0, \pi/2, \pi, 3\pi/2$, we have

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad a = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Thus, we have seen two explicit representations.

One with numbers(complex) using ordinary multiplication and the other with matrices using matrix multiplication.

There maybe a correspondence between the elements of two groups.

The correspondence can be one-to-one , two-to-one or, many-to-one.

If the correspondence satisfies the same group multiplication table, then it is said to be **homomorphic**.

If it is also one-to-one, then it is said to be **isomorphic**.

The two representations that we discussed above are one-to-one and preserve the multiplication table and hence, are **isomorphic**.

The representation of group elements by matrices is a very powerful technique and has been almost universally adopted among physicists.

It can be proven that all the elements of finite groups and continuous groups of the type important in physics can be represented by unitary matrices.

For unitary matrices we have

$$A^\dagger = A^{-1}$$

or, in words, the **complex conjugate/transposed** is equal to the **inverse**.

If there exists a transformation that will transform our original representation matrices into diagonal or block-diagonal form, for example,

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \Rightarrow \begin{pmatrix} P_{11} & P_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{11} & Q_{12} \\ 0 & 0 & Q_{21} & Q_{22} \end{pmatrix}$$

such that the smaller portions or submatrices are no longer coupled together, then the original representation is said to be **reducible**.

Equivalently, we have

$$R = SAS^{-1} = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}$$

which is a “similarity” transformation.

We write

$$R = P \oplus Q$$

and say that R has been **decomposed** into the representations P and Q .

The **irreducible** representations play a role in group theory that is roughly analogous to the unit vectors of vector analysis.

They are the simplest representations → all others may be built up from them.

If a group element x is transformed into another group element y by means of a similarity transformation involving a third group element g_i

$$g_i x g_i^{-1} = y$$

then y is said to be **conjugate** to x .

A **class** is a set of mutually conjugate group elements.

A class is generally not a group.

The **trace** of each group element (each matrix of our representation) is invariant under unitary transformations. We define

$$\text{Tr}(A) = \sum_{i=1}^n A_{ii} = \text{Trace}(A)$$

Then we have

$$\begin{aligned}
\text{Tr}(SAS^{-1}) &= \sum_{i=1}^n (SAS^{-1})_{ii} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n s_{ik} a_{kj} s_{ji}^{-1} \\
&= \sum_{j=1}^n \sum_{k=1}^n a_{kj} \sum_{i=1}^n s_{ji}^{-1} s_{ik} = \sum_{j=1}^n \sum_{k=1}^n a_{kj} (S^{-1}S)_{jk} \\
&= \sum_{j=1}^n \sum_{k=1}^n a_{kj} (I)_{jk} = \sum_{j=1}^n \sum_{k=1}^n a_{kj} \delta_{jk} \\
&= \sum_{j=1}^n a_{jj} = \text{Tr}(A)
\end{aligned}$$

We now relabel the trace as the **character**.

Then we have that all members of a given class (in a given representation) have the same character.

This follows directly from the invariance of the trace under similarity or unitary transformations.

If a subset of the group elements (including the identity element I) satisfies the four group requirements (means that the subset is a group also), then we call the subset a **subgroup**.

Every group has two trivial subgroups: the identity element alone and the group itself.

The group C_4 we discussed earlier has a nontrivial subgroup comprised of the elements $\{I, b\}$.

The **order** of a group is equal to the number of group elements.

Consider a subgroup H with elements $\{h_i\}$ and a group element x not in H .

Then xh_i and h_ix are not in the subgroup H .

The sets generated by

$$xh_i, \quad i = 1, 2, \dots \quad \text{and} \quad h_ix, \quad i = 1, 2, \dots$$

are called the left and right **cosets**, respectively, of the subgroup H with respect to x .

The coset of a subgroup has the same number of distinct elements as the subgroup.

This means that the original group G can be written as the sum of H and its cosets:

$$G = H + x_1H + x_2H + \dots$$

Since H and all of its cosets have the same order (same number of elements), the order (number of elements) of any subgroup is a divisor of the order (number of elements) of the group.

The similarity transformation of a subgroup H by a fixed group element x **not** in H

$$xHx^{-1}$$

yields a subgroup.

If this subgroup is identical to H for all x , i.e., if

$$xHx^{-1} = H$$

then H is called an **invariant, normal or self-conjugate** subgroup.

In physics, groups usually appear as a set of operations that leave a system unchanged or invariant.

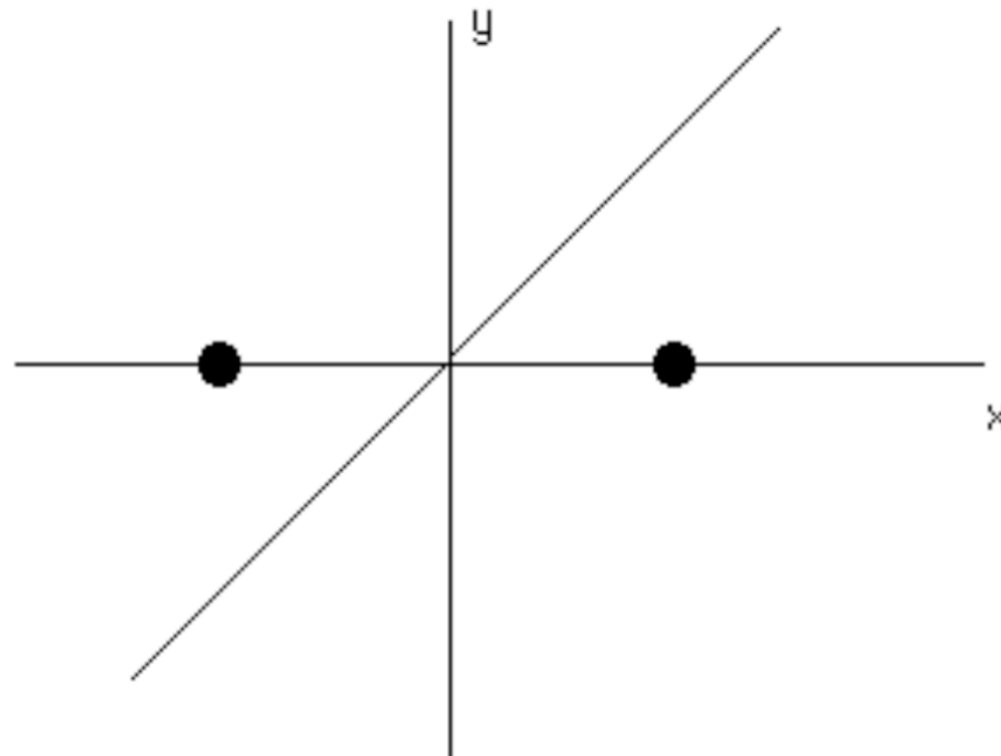
This is an expression of symmetry.

A symmetry is usually defined as the invariance of the Hamiltonian of a system under a group of transformations.

We now investigate the symmetry properties of sets of objects such as the atoms in a molecule or a crystal.

Two Objects - Twofold Symmetry Axis

Consider first the two-dimensional system of two identical atoms in the xy -plane located at $(1, 0)$ and $(-1, 0)$ as shown below.



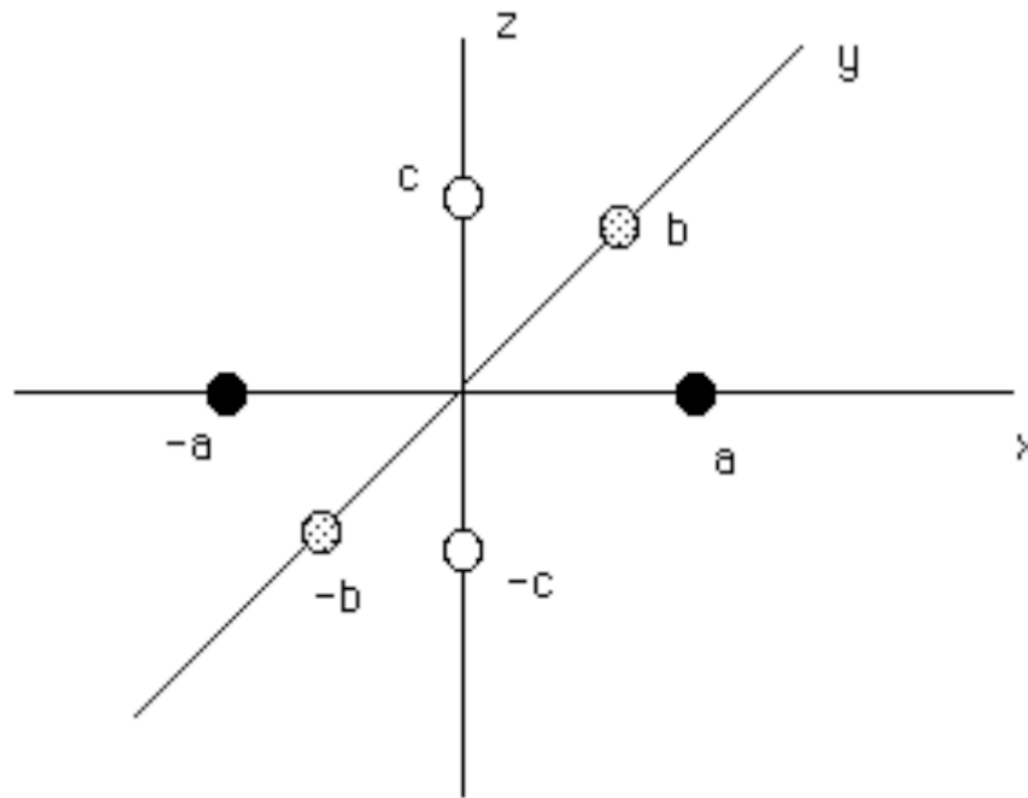
What rotations can be carried out (keeping both atoms in the xy -plane) that will leave the system invariant?

The only ones are I (always works) and a rotation by π about the z -axis.

So we have a rather uninteresting group of two elements.

The z -axis is labelled a **twofold** symmetry axis, corresponding to the two rotation angles 0 and π that leave the system invariant.

Our system becomes more interesting in 3-dimensions. Now imagine a molecule (or part of a crystal) with atoms of element X at $\pm a$ on the x -axis, atoms of element Y at $\pm b$ on the y -axis, and atoms of element Z at $\pm c$ on the z -axis as shown below:



Clearly, each axis is now a twofold symmetry axis.

The matrix representation of the corresponding rotations are:

$$R_x(\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, R_y(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$R_z(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

These four elements $\{I, R_x(\pi), R_y(\pi), R_z(\pi)\}$ form an abelian group with a group multiplication table:

	I	R _x	R _y	R _z
I	I	R _x	R _y	R _z
R _x	R _x	I	R _z	R _y
R _y	R _y	R _z	I	R _x
R _z	R _z	R _y	R _x	I

The products in this table can be obtained in either of two distinct ways:

- (1) We may analyze the operations themselves - a rotation of π about the x -axis followed by a rotation of π about the y -axis is equivalent to a rotation of π about the z -axis

$$R_y(\pi)R_x(\pi) = R_z(\pi)$$

and so on.

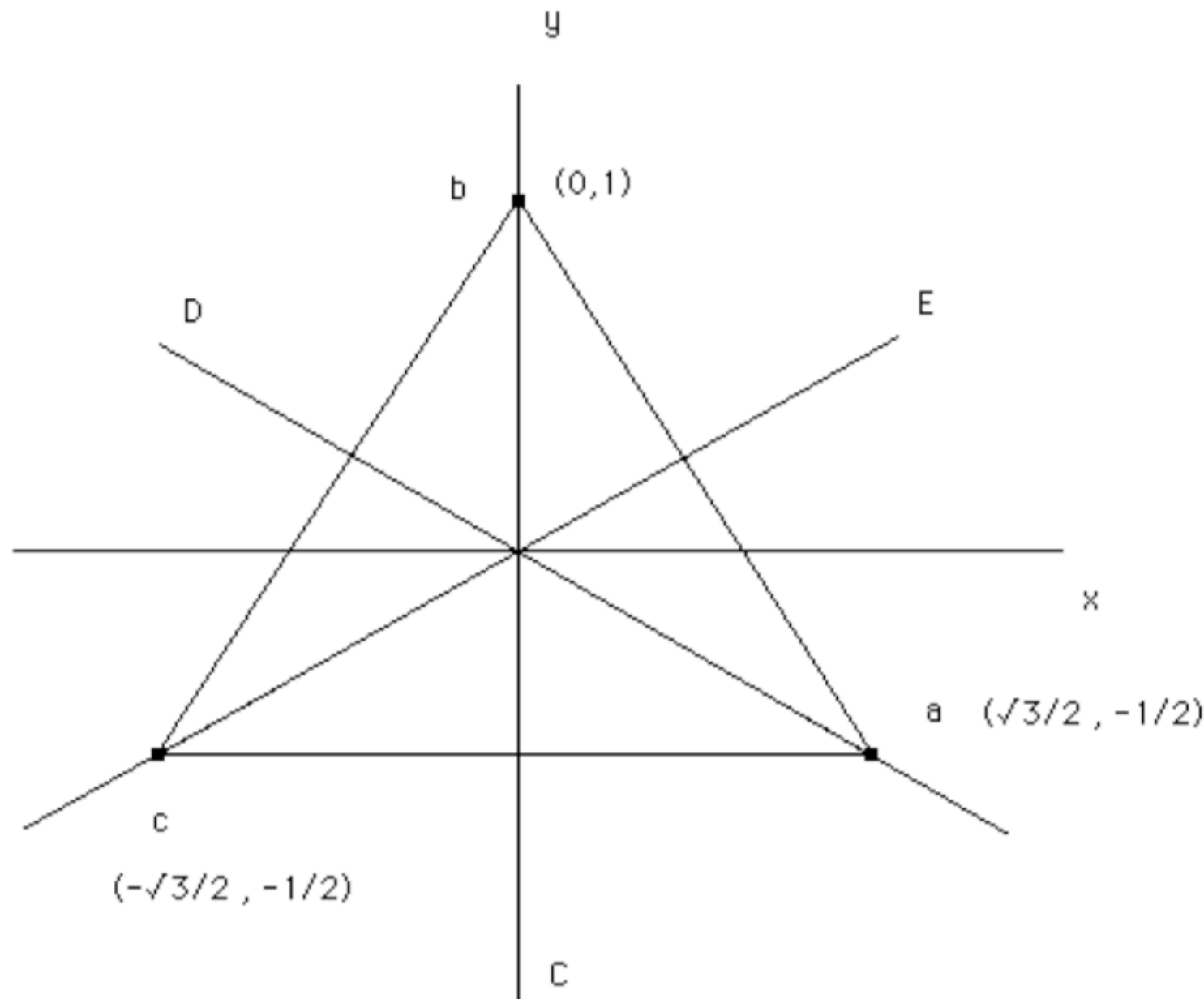
- (2) Alternatively, once the matrix representation is established, we can obtain the products by direct matrix multiplication.

This latter method is especially important when the system is too complex for direct physical interpretation.

This symmetry group is often labelled D_2 , where the D signifies a **dihedral** group and the subscript 2 signifies a twofold symmetry axis (and that no higher symmetry axis exists).

Three Objects – Threefold Symmetry Axis

Consider now three identical atoms at the vertices of an equilateral triangle as shown below:



Rotations (in this case we assume counterclockwise) of the triangle of 0 , $2\pi/3$, and $4\pi/3$ leave the triangle invariant.

In matrix form, we have

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A = R_z(2\pi/3) = \begin{pmatrix} \cos(2\pi/3) & -\sin(2\pi/3) \\ \sin(2\pi/3) & \cos(2\pi/3) \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

$$B = R_z(4\pi/3) = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}$$

The z-axis is a threefold symmetry axis.

$\{I, A, B\}$ form a cyclic group, which is a subgroup of the complete 6-element group that we now discuss.

In the $x - y$ plane there are three additional axes of symmetry - each atom (vertex) and the origin define the twofold symmetry axes C , D , and E .

We note that twofold rotation axes can also be described via reflections through a plane containing the twofold axis.

In this case, we have for rotation of π about the C -axis.

$$C = R_C(\pi) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

The rotation of π about the D -axis can be replaced by a rotation of $4\pi/3$ about the z -axis followed by a reflection in the $y - z$ plane ($x \rightarrow -x$) or

$$D = R_D(\pi) = CB = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}$$

Similarly the rotation of π about the E -axis can be replaced by a rotation of $2\pi/3$ about the z -axis followed by a reflection in the $y - z$ plane ($x \rightarrow -x$) or

$$E = R_E(\pi) = CA = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

The complete group multiplication table is:

	I	A	B	C	D	E
I	I	A	B	C	D	E
A	A	B	I	D	E	C
B	B	I	A	E	C	D
C	C	E	D	I	B	A
D	D	C	E	A	I	B
E	E	D	C	B	A	I

Notice that each element of the group appears only once in each row and in each column. It is clear from the multiplication table that the group is not abelian.

We have explicitly constructed a 6-element group and a 2×2 irreducible matrix representation of it.

The only other distinct 6-element group is the cyclic group $\{I, R, R^2, R^3, R^4, R^5\}$ with

$$R = \begin{pmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

The group $\{I, A, B, C, D, E\}$ is labeled D_3 in crystallography, the dihedral group with a threefold axis of symmetry.

The three axes (C, D, E) in the $x - y$ plane are twofold symmetry axes and, as a result, (I, C) , (I, D) , and (I, E) all form two-element subgroups.

None of these two-element subgroups are invariant.

There are two other irreducible representations of the symmetry group of the equilateral triangle.

They are

1. the trivial $\{1, 1, 1, 1, 1, 1\}$
2. the almost as trivial $\{1, 1, 1, -1, -1, -1\}$

Both of these representations are homomorphic to D_3 .

A general and most important result for finite groups of h elements is that

$$\sum_i n_i^2 = h$$

where n_i is the dimension of the matrices of the i^{th} irreducible representation. equality is called the **dimensionality theorem** and is very useful in establishing the irreducible representations of a group. Here for D_3 we have

$$1^2 + 1^1 + 2^2 = 6$$

for our three representations.

This means that no other representations of the symmetry group of three objects exists.

We note from the examples above that for the transformations involving rotations and reflections, the transformations involving only pure rotations have determinant $= 1$ and the those involving a rotation and a reflection have determinant $= -1$.

Continuous Groups

The groups that we have been discussing all have a finite number of elements.

If we have a group element that contains one or more parameters that vary continuously over some range, then variation of the parameter will produce a continuum of group elements (an infinite number).

These groups are called continuous groups.

If continuous group has a rule for determining combination of elements or the transformation of an element or elements into a different element which is an analytic function of the parameters, where analytic means having derivatives of all orders, then it is called a **Lie** group.

For example, suppose we have the transformation rule

$$x'_i = f_i(x_1, x_2, x_3, \theta)$$

then for this transformation group to be a Lie group the functions f_i must be analytic functions of the parameter θ .

This analytic property will allow us to define infinitesimal transformations thereby reducing the study of the whole group to a study of the group elements when they are only slightly different than the identity element I .

Orthogonal Groups

We begin our study by looking at the orthogonal group O_3 .

In particular, we consider the set of $n \times n$ real, orthogonal matrices with determinant $= +1$ (no reflections).

The defining property of a real orthogonal matrix is

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

where T represents the **transpose** operation.

In terms of matrix elements this gives

$$\sum_i (A^T)_{ji} A_{ik} - \sum_i (A^{-1})_{ji} A_{ik} = I_{jk} - \delta_{jk} = \mathbf{0}$$

The ordinary rotation matrices are an example of real 3×3 orthogonal matrices:

$$R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{pmatrix}, \quad R_y(\theta) = \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix}$$

$$R_z(\gamma) = \begin{pmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

These matrices follow the convention that the rotation is a counterclockwise rotation of the coordinate system to a new orientation.

Each $n \times n$ real orthogonal matrix with determinant $= +1$ has $n(n - 1)/2$ independent parameters.

For example, 2-dimensional rotations are described by 2×2 real orthogonal matrices and they need only one independent parameter ($2(2 - 1)/2 = 1$), namely the rotation angle.

On the other hand, 3-dimensional rotations require three independent angles ($3(3 - 1)/2 = 3$).

Let us see this explicitly for the 2×2 case: Suppose we have the general 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where all the elements are real.

Imposing the condition determinant $= +1$ gives

$$ad - bc = 1$$

Imposing the condition that the transpose is the inverse gives

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{pmatrix} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

These equations imply the solution

$$\begin{pmatrix} a & \sqrt{1-a^2} \\ -\sqrt{1-a^2} & a \end{pmatrix}$$

which shows that there is only one independent parameter.

Special Unitary Groups, $SU(2)$

The set of complex $n \times n$ unitary matrices also forms a group.

This group is labeled $U(n)$.

If we impose an additional restriction that the determinant of the matrices be $+1$, then we get the special unitary group, labeled $SU(n)$.

The defining relation for complex unitary matrices is

$$A^\dagger = A^{-1}$$

where $\dagger$ represents the complex-conjugate-transpose or adjoint operation.

In this case we have $n^2 - 1$ independent parameters.

This can be seen explicitly below: Suppose we have the general 2×2 matrix representing the transformation U

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where all the elements are complex.

Imposing the condition determinant = +1 gives

$$ad - bc = 1$$

Imposing the condition that the adjoint is the inverse gives

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} |a|^2 + |b|^2 & ac^* + bd^* \\ a^*c + b^*d & |c|^2 + |d|^2 \end{pmatrix} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

These equations imply the solution

$$\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$$

where

$$|a|^2 + |b|^2 = aa^* + bb^* = 1$$

Since each complex number has two independent components, this shows that there are three $(2^2 - 1)$ independent parameters.

For $n = 3$, there will be eight $(3^2 - 1)$ independent parameters.

This will become the famous **eightfold way** of elementary particle physics(later).

Generators

In all cases, rotations about a **common axis** combine as

$$R_z(\phi_2)R_z(\phi_1) = R_z(\phi_1 + \phi_2)$$

that is, multiplication of these matrices is equivalent to addition of the arguments.

This suggests that we look for an **exponential** representation of the rotations, i.e.,

$$\exp(\phi_2)\exp(\phi_1) = \exp(\phi_1 + \phi_2)$$

Suppose we define the exponential function of a matrix by the following relation:

$$U = e^{iaH} = I + iaH + \frac{(iaH)^2}{2!} + \dots$$

where a is a real parameter independent of matrix(operator) H .

It is easy to prove that if H is Hermitian, then U is unitary and if U is unitary, the H is Hermitian.

In group theory, H is labeled a **generator**, the generator of U .

The following matrix describes a finite rotation of the coordinates through an angle ϕ counterclockwise about the z-axis:

$$R_z(\phi) = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now let R_z be an infinitesimal rotation through an angle $\delta\phi$.

We can then write

$$R_z(\delta\phi) = I + i\delta\phi M_z$$

where

$$M_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Similarly, we find

$$M_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad M_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}$$

A finite rotation ϕ can be compounded out of successive infinitesimal rotations $\delta\phi$

$$R_z(\delta\phi_2 + \delta\phi_1) = (I + i\delta\phi_2 M_z)(I + i\delta\phi_1 M_z)$$

Letting $\delta\phi = \phi/N$ for N rotations, with $N \rightarrow \infty$, we get

$$R_z(\phi) = \lim_{N \rightarrow \infty} (I + (i\phi/N)M_z)^N = \exp(i\phi M_z)$$

This implies that M_z is the generator of the group $R_z(\phi)$, a subgroup of O_3 .

Before proving this result, we note:

(a) M_z is Hermitian and thus, $R_z(\phi)$ is unitary

(b) $\text{Tr}(M_z) = 0$ and $\det(R_z(\phi)) = +1$

In direct analogy with M_z , M_x may be identified as the generator of R_x , the subgroup of rotations about the x-axis, and M_y as the generator of R_y .

$$\begin{aligned} R_z(\phi) &= \exp(i\phi M_z) = I + i\phi M_z + \frac{(i\phi M_z)^2}{2!} + \frac{(i\phi M_z)^3}{3!} + \dots \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &\quad + i\phi M_z + \frac{(i\phi)^2}{2!} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{(i\phi)^3}{3!} M_z + \dots \end{aligned}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \left(1 - \frac{\phi^2}{2!} + \dots\right) + i\left(\phi - \frac{\phi^3}{3!} + \dots\right)M_z + \dots$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos \phi + i \sin \phi M_z + \dots$$

$$= \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where we have used the relations

$$M_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad M_z^3 = M_z , \quad , \quad \text{etc}$$

and we recognized the two series as $\cos \phi$ and $\sin \phi$.

Thus, we get the same matrix $R_z(\phi)$ as earlier.

Other relations we can write are:

1. all infinitesimal rotations commute

$$[R_i(\delta\phi_i), R_j(\delta\phi_j)] = 0$$

2. an infinitesimal rotation about an axis defined by unit vector $\hat{n}$ is

$$R(\delta\phi) = I + i\delta\phi \hat{n}M$$

3. the generators satisfy the commutation rules

$$[M_i, M_j] = \sum_k i\epsilon_{ijk} M_k$$

This will turn out to be the commutation relations for angular momentum operators in quantum mechanics.

In a similar way the elements (U_1, U_2, U_3) of the two-dimensional unitary group, $SU(2)$, may be generated by

$$\exp(a\sigma_1/2) \ , \ \exp(a\sigma_2/2) \ , \ \exp(a\sigma_3/2)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli spin matrices and a, b , and c are real parameters.

We note that the σ 's are Hermitian and have zero trace and thus, the elements of $SU(2)$ are unitary and have determinant $= +1$.

We also note that the generators in diagonal form such as σ_3 will lead to conserved quantum numbers.

Finally, if we define $s_i = \frac{1}{2}\sigma_i$ then we have

$$[s_i, s_j] = \sum_k i\epsilon_{ijk} s_k$$

which are the angular momentum commutation rules implying that the s_i are angular momentum operators. (Quantum mechanics course)

Can go deeper.....

ODEs

PDEs

Ordinary Differential Equations(ODE)

ODEs are equations involving derivatives in various ways. The highest derivative specifies the order of the equation. Single variable equations are ODEs and multi-variable equations are Partial Differential Equations(PDE)s.

An ODE

$$L(y(x)) = R(x)$$

is linear if the differential operator L is a linear operator satisfying the linearity property

$$L(a_1 y_1(x) + a_2 y_2(x)) = a_1 L(y_1(x)) + a_2 L(y_2(x))$$

where a_1 and a_2 are constants.

Examples:

The 1st order ODE

$$L(y(x)) = \frac{dy(x)}{dx}$$

is linear since

$$L(ay(x)) = \frac{d(ay(x))}{dx} = a \frac{dy(x)}{dx}$$

but the 1st order ODE

$$L(y(x)) = \sqrt{\frac{dy(x)}{dx}}$$

is nonlinear since

$$L(ay(x)) = \sqrt{\frac{d(ay(x))}{dx}} = \sqrt{a} \sqrt{\frac{dy(x)}{dx}} \neq a \sqrt{\frac{dy(x)}{dx}}$$

The 2nd order ODE

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{L} \sin \theta(t) = 0$$

which describes the large amplitude motion of a pendulum is non-linear because

$$\sin(\lambda\theta_1 + \mu\theta_2) \neq \lambda \sin(\theta_1) + \mu \sin(\theta_2)$$

We note that if θ is always small, then $\sin \theta \approx \theta$ and the new ODE

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{L} \theta(t) = 0$$

is a linear ODE.

It describes the small amplitude motion of a pendulum.

All physical systems not far from **EQUILIBRIUM** (small displacements) are linear systems.

Linear ODEs are simpler to solve than non-linear ODEs.

This is a consequence of the following two superposition principles which apply only to linear ODEs:

- (1) If $y_1(x)$ and $y_2(x)$ are **any** two solutions of a **homogeneous linear ODE**

$$L(y(x)) = 0$$

then

$$y_h(x) = c_1 y_1(x) + c_2 y_2(x) \quad c_1, c_2 = \text{constants}$$

is also a solution, i.e.,

$$Ly_h(x) = L(c_1 y_1(x) + c_2 y_2(x)) = c_1 L(y_1(x)) + c_2 L(y_2(x)) = 0$$

(2) If $y_h(x)$ is a (i.e., any) solution of the homogeneous linear ODE, $L(y_h(x)) = 0$ and $y_p(x)$ is a solution of the inhomogeneous linear ODE

$$L(y_p(x)) = R(x)$$

then the linear combination

$$y(x) = y_p(x) + Ay_h(x) \quad A = \text{constant}$$

is a solution of the inhomogeneous linear ODE

$$L(y(x)) = R(x)$$

i.e.,

$$L(y(x)) = L(y_p(x) + Ay_h(x)) = L(y_p(x)) + AL(y_h(x)) = R(x)$$

Solutions of ODEs

1st order ODEs can be solved by **direct integration** if the equation is **SEPARABLE**:

$$\frac{dy}{dx} = \frac{p(x)}{q(y)} \rightarrow \int q(y)dy = \int p(x)dx$$

Example:

$$\frac{dy}{dx} - x\sqrt{1-y^2} = 0 \rightarrow \frac{dy}{\sqrt{1-y^2}} = xdx$$

$$\int \frac{dy}{\sqrt{1-y^2}} = \sin^{-1} y = \int xdx = \frac{1}{2}x^2 + C \rightarrow y = \sin \left(\frac{1}{2}x^2 + C \right)$$

This works for both linear and nonlinear ODEs.

2nd-Order Homogeneous ODEs with Constant Coefficients

Consider the ODE

$$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 4y = 0$$

If we let

$$D = \frac{d}{dx}$$

then we have the equation

$$(D^2 + 5D + 4)y = 0$$

These equations can be solved by **substitution(guessing)** and converting the ODE into an **algebraic (quadratic)** equation.

We choose(guess) the solution

$$y = ce^{ax}$$

Substitution gives the allowed values of a (corresponding to possible solutions)

$$(a^2 + 5a + 4)ce^{ax} = 0 \rightarrow a^2 + 5a + 4 = 0 \rightarrow a = -1 \text{ and } a = -4$$

The most general solution is a superposition of all possible solutions with arbitrary multiplicative constants (number of possible solutions = order of ODE) so

$$y = c_1 e^{-x} + c_2 e^{-4x}$$

The arbitrary constants c_1 and c_2 are determined by boundary conditions where we specify

$$y(x_0) \text{ and } \frac{dy(x_0)}{dx}, \quad x_0 = \text{arbitrary point}$$

For example suppose,

$$x_0 = 0 \text{ and } y(0) = 2 \text{ and } \frac{dy(0)}{dx} = -3$$

then we have

$$y(0) = 2 = c_1 + c_2$$

$$\frac{dy(0)}{dx} = -3 = -c_1 - 4c_2$$

$$\rightarrow c_1 = \frac{5}{3} \text{ and } c_2 = \frac{1}{3}$$

and the general solution with these boundary conditions is

$$y(x) = \frac{5}{3}e^{-x} + \frac{1}{3}e^{-4x}$$

The solutions for a can be imaginary numbers (trigonometric solutions) or complex numbers (mixed trigonometric and exponential solutions).

Example: Simple Harmonic Oscillator

In this case:

$$\frac{d^2y}{dt^2} + \omega^2 y = 0 \quad , \quad D = \frac{d}{dt} \rightarrow (D^2 + \omega^2)y(t) = 0$$

Choosing

$$y = c_1 e^{at} \rightarrow a^2 + \omega^2 = 0 \rightarrow a = \pm i\omega$$

which gives the general solution

$$y(t) = c_1 e^{i\omega t} + c_2 e^{-i\omega t}$$

With boundary conditions

$$y(0) = 0 \quad \text{and} \quad \frac{dy(0)}{dt} = 4\omega \quad (\text{what is the oscillator doing?})$$

we get

$$c_1 + c_2 = 0$$

$$i\omega(c_1 - c_2) = 4\omega \rightarrow c_1 - c_2 = -4i$$

$$\rightarrow c_1 = -2i \quad \text{and} \quad c_2 = 2i$$

so that

$$y(t) = -2i(e^{i\omega t} - e^{-i\omega t}) = 4 \sin(\omega t)$$

Example: Equal Roots

Suppose we have

$$(D^2 - 6D + 9)y = 0 \rightarrow a = 3, 3$$

→ These solutions are linearly dependent.

The rule is then to choose a solution of the form

$$y(x) = (Ax + B)e^{3x}$$

since we **must have** 2 constants because it is a 2nd-order ODE.

Series Solution of Homogeneous 2nd Order Linear ODEs: Method of Frobenius

We have seen

- (1) given one solution of a homogeneous 2nd order ODE we can generate a second linearly independent solution integration
- (2) if we are given both solutions, then a particular solution of the inhomogeneous ODE also can be calculated by integration

Thus, it remains for us to obtain at least 1 solution for the homogeneous equation.

A method, with which it is usually possible to do so, is to obtain the solution $y(x)$ of the homogeneous ODE

$$Ly(x) = 0$$

using a **power series** in x .

We choose

$$y(x) = x^s \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda} \quad a_0 \neq 0$$

We then have

$$\frac{dy}{dx} = \sum_{\lambda=0}^{\infty} a_{\lambda}(\lambda + s)x^{\lambda+s-1}$$

$$\frac{d^2y}{dx^2} = \sum_{\lambda=0}^{\infty} a_{\lambda}(\lambda + s)(\lambda + s - 1)x^{\lambda+s-2}$$

Substitution back into the homogeneous ODE produces a powers series which sums to zero.

In this case, because each term in the series is linearly independent, each coefficient must separately $= 0$.

This allows us to determine the values of s and the a -coefficients and hence the solution to the problem.

Let us illustrate the method with some examples.

Consider the equation

$$\frac{d^2 u}{dx^2} + x^2 u = 0$$

Let us first use the simpler method which is valid for certain equations where $s = \text{integer}$.

In this case, we can choose

$$y(x) = x^s \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda} \quad a_0 \neq 0$$

$$u(x) = \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda} \quad a_0 \neq 0$$

i.e., we can use the simpler series when $s = \text{integer}$.

We then have

$$\frac{d^2 u}{dx^2} = \sum_{\lambda=0}^{\infty} a_{\lambda} (\lambda)(\lambda - 1) x^{\lambda-2}$$

Substitution gives

$$\sum_{\lambda=0}^{\infty} a_{\lambda} (\lambda)(\lambda - 1) x^{\lambda-2} + \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda+2} = 0$$

or

$$a_2(1)(2)x^0 + a_3(2)(3)x^3 + [a_4(3)(4) + a_0]x^2 \\ + \dots + [a_{k-2} + a_{k+2}(k+2)(k+1)]x^k = \dots = 0$$

We then have

$$a_0 \neq 0$$

$$a_1 \neq 0$$

$$a_2 = 0$$

$$a_3 = 0$$

$$a_k + a_{k+4}(k+3)(k+4) = 0$$

$$a_{k+4} = -\frac{a_k}{(k+3)(k+4)}$$

$$a_2 = 0 \rightarrow a_6 = a_{10} = a_{14} = \dots = 0$$

$$a_3 = 0 \rightarrow a_7 = a_{11} = a_{15} = \dots = 0$$

Starting with a_0 we get

$$a_4 = -\frac{a_0}{(3)(4)}$$

$$a_8 = -\frac{a_4}{(7)(8)} = \frac{a_0}{(3)(4)(7)(8)}$$

Starting with a_1 we get

$$a_5 = -\frac{a_1}{(4)(5)}$$

$$a_9 = -\frac{a_5}{(8)(9)} = \frac{a_1}{(4)(5)(8)(9)}$$

and the general solution is

$$u(x) = c_1 \left(1 - \frac{1}{12}x^4 + \frac{1}{672}x^8 - \dots \right) + c_2 \left(x - \frac{1}{20}x^5 + \frac{1}{1440}x^9 - \dots \right)$$

Let us now do this same equation with s left in (more general).

We choose

$$u(x) = \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda+s} \quad a_0 \neq 0$$

We then have

$$\frac{d^2 u}{dx^2} = \sum_{\lambda=0}^{\infty} a_{\lambda} (s + \lambda)(s + \lambda - 1) x^{\lambda-2+s}$$

Substitution gives

$$\sum_{\lambda=0}^{\infty} a_{\lambda} (s + \lambda)(s + \lambda - 1) x^{\lambda-2+s} + \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda+s+2} = 0$$

or

$$\begin{aligned}
& a_0 s(s-1)x^{s-2} + a_1 s(s+1)x^{s-1} + a_2(s+1)(s+2)x^s \\
& + a_3(s+2)(s+3)x^{s+1} \\
& + [a_4(s+3)(s+4) + a_0]x^{s+2} + \dots \\
& + [a_{k-2} + a_{k+2}(s+k+2)(s+k+1)]x^{s+k} + \dots = 0
\end{aligned}$$

We then have

$$a_0 s(s-1) = 0 \rightarrow a_0 \neq 0, s = 0, 1$$

$$a_1 s(s+1) = 0 \rightarrow a_1 \neq 0, s = 0, -1 \quad \text{and} \quad a_1 = 0, s = 0, -1$$

$$a_2(s+1)(s+2) = 0 \rightarrow a_2 = 0$$

$$a_3(s+2)(s+3) = 0 \rightarrow a_3 = 0$$

$$[a_{k-2} + a_{k+2}(s+k+2)(s+k+1)] = 0$$

$$a_{k+4} = -\frac{a_k}{(s+k+3)(s+k+4)}$$

$$a_2 = 0 \rightarrow a_6 = a_{10} = a_{14} = \dots = 0$$

$$a_3 = 0 \rightarrow a_7 = a_{11} = a_{15} = \dots = 0$$

Starting with $s = 0$ and choosing $a_0 = a_1 = 1$ we get

$$a_4 = -\frac{a_0}{(3)(4)}, \quad a_8 = -\frac{a_4}{(7)(8)} = \frac{a_0}{(3)(4)(7)(8)}, \quad \dots\dots$$

$$a_5 = -\frac{a_1}{(4)(5)}, \quad a_9 = -\frac{a_5}{(8)(9)} = \frac{a_1}{(4)(5)(8)(9)}, \quad \dots\dots$$

The case $s = 1$ does not generate any new series (always happens if $s = \text{integer}$).

The general solution is the same as earlier

$$u(x) = c_1 \left(1 - \frac{1}{12}x^4 + \frac{1}{672}x^8 - \dots \right) + c_2 \left(x - \frac{1}{20}x^5 + \frac{1}{1440}x^9 - \dots \right)$$

We now consider the equation

$$8x^2 \frac{d^2 u}{dx^2} + 6x \frac{du}{dx} + (x - 1)u = 0$$

Assume

$$u(x) = x^s \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda} \quad a_0 \neq 0$$

Substitution gives

$$\begin{aligned} 8 \sum_{\lambda=0}^{\infty} a_{\lambda} (\lambda + s)(\lambda + s - 1) x^{\lambda+s} + 6 \sum_{\lambda=0}^{\infty} a_{\lambda} (\lambda + s) x^{\lambda+s} + \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda+s+1} \\ - \sum_{\lambda=0}^{\infty} a_{\lambda} x^{\lambda+s} = 0 \end{aligned}$$

$$\begin{aligned}
& [8a_0s(s-1) + 6a_0s - a_0]x^s + [8a_1s(s+1) + 6a_1(s+1) - a_1 + a_0]x^{s+1} \\
& + [8a_2(s+1)(s+2) + 6a_2(s+2) - a_2 + a_1]x^{s+2} + \dots \\
& + [8a_n(s+n-1)(s+n) + 6a_n(s+n) - a_n + a_{n-1}]x^{s+n} + \dots = 0
\end{aligned}$$

The coefficient of x^s is called the **indicial equation**.

It determines the allowed values of s .

We choose $a_0 = 1$ for simplicity.

We get

$$8s(s-1) + 6s - 1 = 0 = 8s^2 - 2s - 1 \rightarrow s = \frac{2 \pm \sqrt{4 + 32}}{16} = \frac{1}{2}, -\frac{1}{4}$$

These two allowed s values will generate the two linearly independent homogeneous solutions.

We have from the other coefficients

$$8a_n(s+n-1)(s+n) + 6a_n(s+n) = a_n + a_{n-1} = 0$$

$$a_n = -\frac{a_{n-1}}{8(s+n-1)(s+n) + 6(s+n) - 1}$$

This last relation is a recursion relation and determines all higher order coefficients(for each allowed s value) in terms of a_0 .

We get

$$s = \frac{1}{2}$$

$$\begin{aligned} u_1(x) &= x^{1/2} - \frac{1}{8(1/2)(3/2) + 6(3/2) - 1} x^{3/2} \\ &\quad + \frac{1}{(8(3/2)(5/2) + 6(5/2) - 1)(8(1/2)(3/2) + 6(3/2) - 1)} x^{5/2} - \dots \\ &= x^{1/2} - \frac{1}{14} x^{3/2} + \frac{1}{616} x^{5/2} - \dots \end{aligned}$$

$$s = -\frac{1}{4}$$

$$u_2(x) = x^{-1/4} - \frac{1}{2}x^{3/4} + \frac{1}{40}x^{7/4} - \dots$$

and the most general solution is

$$u(x) = au_1(x) + bu_2(x)$$

A second example is slightly trickier (even though equation is simpler).

Consider the equation

$$\left(\frac{d^2}{dx^2} + k^2 \right) y(x) = 0$$

Remember we already know the solutions to this equation and that will help us to figure out what is going on with the method.

Substitution gives

$$\sum_{\lambda=0}^{\infty} a_{\lambda}(\lambda + s)(\lambda + s - 1)x^{\lambda+s-2} + k^2 \sum_{\lambda=0}^{\infty} a_{\lambda}x^{\lambda+s} = 0$$

or

$$a_0s(s-1)x^{s-2} + a_1s(s+1)x^{s-1} + [a_0k^2 + a_2(s+2)(s+1)]x^s + \dots \\ + [a_{\lambda}k^2 + a_{\lambda+2}(s+\lambda+2)(s+\lambda+1)]x^{s+\lambda} + \dots = 0$$

Since the coefficient of each power must separately be $= 0$, we have

$$a_0s(s-1) = 0 \rightarrow s = 0 \text{ or } 1 \text{ since } a_0 \neq 0$$

This is always STEP #1: determine the possible values of s .

It is at this point that we use the assumption $a_0 \neq 0$.

Then look at next coefficient

$$a_1 s(s+1) = 0 \rightarrow \begin{cases} \text{if } s = 0, \text{ then } a_1 \text{ can be nonzero} \\ \text{if } s = 1, \text{ then } a_1 = 0 \end{cases}$$

In general, the coefficient of the other higher powers are given by

$$a_{\lambda+2} = -\frac{k^2}{(s+\lambda+2)(s+\lambda+1)} a_{\lambda}$$

This is a **recursion relation**, which allows an ordered step-by-step determination of all coefficients in terms of a_0 or a_1 i.e.,

$$a_2 = -\frac{k^2}{(s+2)(s+1)} a_0$$

$$a_3 = -\frac{k^2}{(s+3)(s+2)} a_1$$

and so on.

The particular recurrence relation for this equation contains the special feature that it steps λ by 2.

As a result, the coefficients a_λ are separated into two disjoint groups, one with even λ and another with odd λ .

For the even chain we find

(a) For $s = 0$

$$a_2 = -\frac{k^2}{2!}a_0, \quad a_4 = -\frac{k^2}{12}a_2 = \frac{k^2}{4!}a_0, \quad \dots\dots\dots$$

(b) For $s = 1$

$$a_2 = -\frac{k^2}{3!}a_0, \quad a_4 = -\frac{k^2}{20}a_2 = \frac{k^2}{5!}a_0, \quad \dots\dots\dots$$

or we get solutions

(a)

$$y_1(x) = y_{\text{even}}(x) = 1 - \frac{k^2}{2!}x^2 + \frac{k^2}{4!}x^4 - \dots = \cos kx$$

(b)

$$\begin{aligned} y_2(x) = y_{\text{odd}}(x) &= x \left(1 - \frac{k^2}{3!}x^2 + \frac{k^2}{5!}x^4 - \dots \right) \\ &= \frac{1}{k} \left(x - \frac{k^2}{3!}x^3 + \frac{k^2}{5!}x^5 - \dots \right) = \frac{1}{k} \sin kx \end{aligned}$$

where we have used $a_0 = 1$.

What about the odd chain of coefficients? There is no solution in that case for $s = 1$, since $a_1 = 0$ and therefore all higher order odd coefficients vanish.

For $s = 0$ however, $a_1 \neq 0$ and this gives

$$a_3 = -\frac{k^2}{3!}a_1, \quad a_5 = -\frac{k^2}{20}a_3 = \frac{k^2}{5!}a_1, \quad \dots\dots\dots$$

which is the same solution as $y_{odd}(x)$ from before, so we get nothing new.

In fact, we should not have expected anything new since we already have found the 2 allowed linearly independent solutions.

The solutions are, of course, the same as we found earlier.

Lots more in full lectures

Partial Differential Equations (PDEs) In Physics, there are differential equations of motion that describe the response of systems to external disturbances.

These are called ordinary differential equations (ODEs).

There are also differential equations of states, or field equations, whose solutions give the space-time dependence of physical properties.

These are called partial differential equations (PDEs) in the 4 variables x, y, z, t .

In general, the PDE's we will discuss describe three-dimensional situations. The independent variables are the position vector $\vec{r}$ and the time t .

The actual variables used to specify $\vec{r}$ are dictated by the coordinate system in use, i.e., (x, y, z) , (ρ, ϕ, z) , (r, θ, ϕ) , etc.

The most important PDEs are:

(1) The wave equation

$$\nabla^2 u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

This equation describes as a function of position and time the displacement from equilibrium, $u(\vec{r}, t)$, of a vibrating string or membrane, or a vibrating solid, gas or liquid.

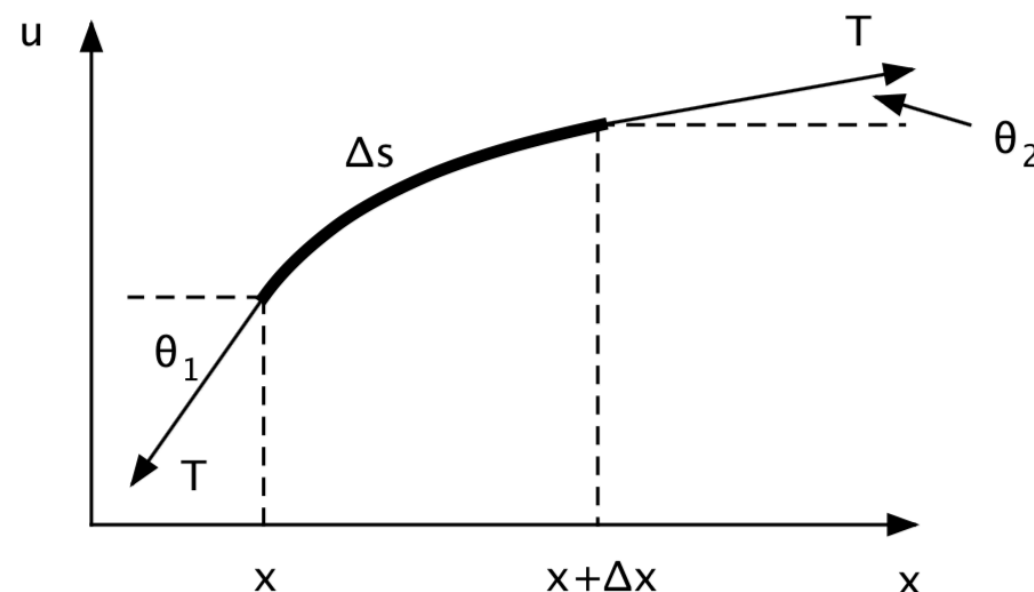
The equation also occurs in electromagnetism, where $u(\vec{r}, t)$ may be a component of the electric or magnetic field in an electromagnetic wave, or the current or voltage along a transmission line.

The quantity c is the speed of propagation of the waves.

We now derive the wave equation in a special case.

We consider the small transverse displacements $u(x, t)$ of a uniform string of mass per unit length ρ held under uniform tension T , assuming that the string is initially located along the x -axis in a Cartesian coordinate system.

The figure below shows the forces acting on an elemental length δs of the string.



If the tension T in the string is uniform along its length, the net upward vertical force on the element is

$$\Delta F = T \sin \theta_2 - T \sin \theta_1$$

Assuming that the angles θ_1 and θ_2 are both small, we may make the approximation $\sin \theta \approx \tan \theta$.

Since, at any point on the string the slope is

$$\tan \theta = \frac{\partial u}{\partial x}$$

the force can be written as

$$\Delta F = T \left(\frac{\partial u(x + \Delta x, t)}{\partial x} - \frac{\partial u(x, t)}{\partial x} \right) \approx T \frac{\partial^2 u(x, t)}{\partial x^2} \Delta x$$

where we have used the standard definition of the partial derivative.

The upward force may be equated, by Newton's second law, to the product of the mass of the element and its upward acceleration.

The element has mass $\rho \Delta s$, which is approximately equal to $\rho \Delta x$ if the vibrations of the string are small, and so we have

$$\rho \Delta x \frac{\partial^2 u(x, t)}{\partial t^2} = T \frac{\partial^2 u(x, t)}{\partial x^2} \Delta x$$

$$\rightarrow \frac{\partial^2 u(x, t)}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 u(x, t)}{\partial x^2} \quad , \quad c^2 = \frac{T}{\rho}$$

which is the one-dimensional form of the wave equation.

(2) The diffusion equation

$$\nabla^2 u = \frac{1}{D} \frac{\partial u}{\partial t}$$

This equation describes the temperature $u(\vec{r}, t)$ in a region containing no heat sources or sinks.

It also applies to the diffusion of a chemical that has concentration $u(\vec{r}, t)$.

The constant D is called the diffusivity.

We now derive now derive the diffusion equation satisfied by the temperature $u(\vec{r}, t)$ at time t for a material of uniform thermal conductivity k , specific heat capacity s , and density ρ .

Let us consider an arbitrary volume V lying within a solid, and bounded by a surface S .

At any point in the solid the rate of heat flow per unit area in any given direction $\vec{r}$ is proportional to minus the component of the temperature gradient in that direction and is given by

$$(-k\nabla u) \cdot \vec{r}$$

The total flux of heat out of the volume V per unit time is given by

$$-\frac{dQ}{dt} = \iint_S (-k\nabla u) \cdot \hat{n} dS = \iiint_V \nabla \cdot (-k\nabla u) dV$$

where Q is the total heat energy in V at time t , and $\hat{n}$ is the outward-pointing unit normal to S ; note that we have used the divergence theorem to convert the surface integral into a volume integral.

We can also express Q as the volume integral over V ,

$$Q = \iiint_V s\rho u dV$$

and so its rate of change is given by

$$\frac{dQ}{dt} = \iiint_V s\rho \frac{\partial u}{\partial t} dV$$

Comparing the two expressions for dQ/dt and remembering the volume V is arbitrary, we obtain the three-dimensional diffusion equation

$$\nabla^2 u = \frac{1}{D} \frac{\partial u}{\partial t} \quad , \quad D = \frac{k}{s\rho}$$

(3) Laplace equation

$$\nabla^2 u = 0$$

This equation is obtained by setting

$$\frac{\partial u}{\partial t}$$

in the diffusion equation and describes (for example) the **steady-state** temperature of a solid in which there are no heat sources - i.e., the temperature after a long time has elapsed.

Laplace's equation also describes the gravitational potential in a region containing no matter, or the electrostatic potential in a charge-free region.

It also applies to the flow of an incompressible fluid with no sources, sinks or vortices - in this case $u(\vec{r}, t)$ is the velocity potential, from which the velocity is given by $\vec{v} = \nabla u$.

(4) Poisson equation

$$\nabla^2 u = \rho(\vec{r})$$

This equation describes the same physical situations as Laplace's equation, but in regions containing matter, charges, or sources of heat or fluid. The function $\rho(\vec{r})$ is called the source density, and in physical applications usually contains some multiplicative physical constants.

For example, if u is the electrostatic potential in some region of space, in which $\rho(\vec{r})$ is the density of electric charge, then

$$\nabla^2 u = -\frac{1}{\epsilon_0} \rho(\vec{r})$$

where ϵ_0 is the permittivity of free space.

Alternatively, u might represent the gravitational potential in some region where the matter density is given by $\rho(\vec{r})$; then

$$\nabla^2 u = 4\pi G \rho(\vec{r})$$

where G is the gravitational constant.

(5) Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2 u + V(\vec{r})u = i\hbar\frac{\partial u}{\partial t}$$

This equation describes the quantum mechanical wave function $u(\vec{r}, t)$ of a non-relativistic particle of mass m ; $\hbar$ is Planck's constant divided by 2π . All of these equations are **linear**, They are all 2nd-order in the space variables and of 1st- or 2nd-order in time.

The use of these differential operators guarantees several things:

- (1) differential operators imply invariance with respect to space and time translations and hence conservation of energy and momentum
- (2) the differential operator ∇^2 is the simplest operator that will be invariant under the parity transformation (inversion)

(3) equations that are 2nd-order in time are invariant under time reversal ($t \rightarrow -t$) and hence, a movie of the system in time should represent a real physical system whether it is run forwards or backwards.

For example, the wave equation might have a solution representing a wave propagating to the right and if we run the movie backwards, we get a wave propagating to the left, which is also a valid solution.

In the diffusion or heat conduction, the field equation (for the density or temperature fields) is only 1st-order in time.

The equation does not, and should not, satisfy time-reversal invariance, since heat is known to flow from high temperature to low temperature and NEVER the other way around.

A movie of a pool of water solidifying into a block of ice on a hot day has obviously been run backwards.

The Schrödinger equation is a sort of diffusion equation with an imaginary diffusion constant; the wave function is a complex function.

There are many general methods for studying and solving for general aspects/properties of PDEs and their solutions.

We will concentrate in this course on one solution method, namely, **separation of variables**.

Separation of Variables and Eigenfunction Expansions

Under certain circumstances the solution of a PDE may be written as a sum of terms, each of which is the product of functions of only one of the variable.

This is called solution by **separation of variables (SOV)**.

Let us illustrate the procedure by an example:

Consider a 1-dimensional wave equation describing the transverse vibrations of a string

$$\frac{\partial^2 u(x, t)}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 u(x, t)}{\partial x^2}$$

In the SOV method we simply look for a solution of the form

$$u(x, t) = X(x) T(t)$$

Direct substitution then gives

$$\frac{\partial^2}{\partial t^2} u(x, t) = T(t) \frac{d^2 X(x)}{dx^2} = \frac{1}{c^2} X(x) \frac{d^2 T(t)}{dt^2}$$

or

$$\frac{1}{X(x)} \frac{d^2 X(x)}{dx^2} = \frac{1}{c^2} \frac{1}{T(t)} \frac{d^2 T(t)}{dt^2}$$

It is clear that we have **separated the variables**.

Since the LHS is a function of x only and the RHS is a function of t only, both sides of this equation must be equal for all x and t .

The only way this is possible is for both of them to be equal to the same constant, say λ .

$$\frac{1}{X(x)} \frac{d^2 X(x)}{dx^2} = \lambda = \frac{1}{c^2} \frac{1}{T(t)} \frac{d^2 T(t)}{dt^2}$$

which gives

$$\frac{d^2 X(x)}{dx^2} - \lambda X(x) = 0 \quad , \quad \frac{d^2 T(t)}{dt^2} - \lambda c^2 T(t) = 0$$

These are two separated ODEs (single variable).

They are not completely independent of each other since the same separation constant λ must appear in both.

They are both eigenvalue/eigenfunction equations.

The general solutions of these equations are:

$$X(x) = A \cos(\sqrt{-\lambda}x) + B \sin(\sqrt{-\lambda}x)$$

$$T(t) = D \cos(\sqrt{-\lambda}ct) + E \sin(\sqrt{-\lambda}ct)$$

As a rule, **all possible values** of the separation constant λ are allowed unless **explicitly forbidden** by the physics of the system. i.e., certain values of λ can be forbidden when the corresponding solution $X(x)$, which depends on λ , does not have the correct properties.

The properties in question, are the **boundary conditions** imposed by the physics of the system.

It may happen that one or more of these boundary conditions can be satisfied only when the separation constant takes on a set of special values. This set then contains the only permissible values, or **eigenvalues**, for the problem.

The corresponding solutions are called **eigenfunctions**.

Since the 1-dimensional wave equation is linear, the general solution periodic in x with period 2π is then the linear superposition

$$u(x, t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} X_n(x) T_n(t)$$

of all possible solutions.

Note that this is a **double Fourier series**.

Boundary and Initial Conditions

The complete determination of a solution of the PDE requires the specification of a suitable set of boundary and initial conditions.

The boundaries may not be just points, but, depending on the dimension of the system they can be lines or surfaces.

Let us return to the wave equation.

Real example of a vibrating string (we go back through everything again for practice):

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2}$$

gives

$$y(x, t) = X(x) T(t)$$

$$\frac{X''}{X} = -\alpha^2 = \frac{1}{c^2} \frac{\ddot{T}}{T}$$

$$X'' + \alpha^2 X = 0 \quad , \quad \ddot{T} + \alpha^2 c^2 T = 0$$

Define

$$\alpha c = \omega = 2\pi\nu = \frac{2\pi c}{\lambda} \rightarrow \alpha = \frac{2\pi}{\lambda} = k = \text{wave number}$$

The solution is (as before)

$$y = (a \sin(kx) + b \cos(kx))(d \sin(\omega t) + e \cos(\omega t))$$

The string is fastened at $x = 0$ and $x = L$, so that $y(0) = y(L) = 0$, which gives

$$y(0, t) = b(d \sin(\omega t) + e \cos(\omega t)) = 0 \rightarrow b = 0$$

$$y(L, t) = a \sin(kL)(d \sin(\omega t) + e \cos(\omega t)) = 0$$

$$\rightarrow \sin(kL) = 0 \rightarrow kL = n\pi$$

and

$$y_n(x, t) = a_n \sin\left(\frac{n\pi x}{L}\right) \left(d_n \sin\left(\frac{n\pi \nu t}{L}\right) + e_n \cos\left(\frac{n\pi \nu t}{L}\right) \right)$$

Now choose initial conditions at $t = 0$

$$y(x, 0) = f(x) = \begin{cases} x/2 & 0 \leq x \leq L/2 \\ L/2 - x/2 & L/2 \leq x \leq L \end{cases}$$

and

$$\left. \frac{\partial y(x, t)}{\partial t} \right|_{t=0} = 0$$

The condition

$$\left. \frac{\partial y(x, t)}{\partial t} \right|_{t=0} = 0$$

$$\rightarrow a_n \sin\left(\frac{n\pi x}{L}\right) \frac{n\pi \nu}{L} \left(d_n \sin\left(\frac{n\pi \nu t}{L}\right) + e_n \cos\left(\frac{n\pi \nu t}{L}\right) \right) \Big|_{t=0}$$

$$0 = a_n \sin\left(\frac{n\pi x}{L}\right) \frac{n\pi \nu}{L} (d_n) \rightarrow d_n = 0$$

This means we physically pull the string into a triangular shape and let it go from rest.

The most general solution is then is

$$y(x, t) = \left(\sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi}{L} x\right) \right) \cos\left(\frac{n\pi \nu}{L} t\right)$$

Now using the other initial condition, we have

$$y(x, 0) = f(x) = \sum_{n=1}^{\infty} A_n \sin \left(\frac{n\pi}{L} x \right)$$

or the A_n are the Fourier coefficients of the sine series for the triangular pulse.

We have

$$\begin{aligned} A_m &= \frac{2}{L} \int_0^L \sin \left(\frac{m\pi}{L} x \right) f(x) dx = \frac{1}{L} \int_0^L \sin \left(\frac{m\pi}{L} x \right) x dx \\ &\quad + \frac{1}{L} \int_0^L \sin \left(\frac{m\pi}{L} x \right) (L - x) dx \\ &= \frac{1}{L} \frac{L^2}{m^2 \pi^2} \int_0^{m\pi/2} y \sin y dy - \frac{1}{L} \frac{L^2}{m^2 \pi^2} \int_{m\pi/2}^{m\pi} y \sin y dy \\ &\quad + \frac{L}{m\pi} \int_{m\pi/2}^{m\pi} \sin y dy \end{aligned}$$

$$\begin{aligned}
&= \frac{L}{m^2 \pi^2} (\sin y - y \cos y) \Big|_0^{m\pi/2} - \frac{L}{m^2 \pi^2} (\sin y - y \cos y) \Big|_{m\pi/2}^{m\pi} \\
&\quad - \frac{L}{m\pi} (\cos y) \Big|_{m\pi/2}^{m\pi} \\
&= \frac{L}{m^2 \pi^2} \left(\sin \left(\frac{m\pi}{2} \right) - \frac{m\pi}{2} \cos \left(\frac{m\pi}{2} \right) - \sin(0) + (0) \cos(0) \right) \\
&\quad - \frac{L}{m^2 \pi^2} \left(\sin(m\pi) - m\pi \cos(m\pi) - \sin \left(\frac{m\pi}{2} \right) + \frac{m\pi}{2} \cos \left(\frac{m\pi}{2} \right) \right) \\
&\quad - \frac{L}{m\pi} \left(\cos(m\pi) - \cos \left(\frac{m\pi}{2} \right) \right)
\end{aligned}$$

or

$$A_m = \begin{cases} \left(\frac{2L}{m^2 \pi^2} \right) \sin \left(\frac{m\pi}{2} \right) & m \text{ odd} \\ 0 & m \text{ even} \end{cases}$$

and so on.

The other possible initial condition is

$$y(x, 0) = 0$$

and

$$\left. \frac{\partial y(x, t)}{\partial t} \right|_{t=0} = f(x) = \begin{cases} x/2 & 0 \leq x \leq L/2 \\ L/2 - x/2 & L/2 \leq x \leq L \end{cases}$$

which corresponds to hitting the string when it is flat.

This gives

$$y(x, t) = \left(\sum_{n=1}^{\infty} B_n \sin \left(\frac{n\pi x}{L} \right) \right) \sin \left(\frac{n\pi \nu}{L} t \right)$$

$$\left. \frac{\partial y(x, t)}{\partial t} \right|_{t=0} = f(x) = \sum_{n=1}^{\infty} B_n \sin \left(\frac{n\pi x}{L} \right) = \text{initial velocity profile}$$

For each n (in either case) there is a different frequency

$$f_n = \frac{\omega_n}{2\pi} = \frac{n\pi\nu}{2\pi L} = \frac{n\nu}{2L}$$

which are the **normal mode frequencies**.

These frequencies can be excited separately and would be **stable**.

The string would vibrate in a single mode or one term of the general sum

$$\sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{n\pi\nu}{L}t\right)$$

If we took a photograph at any given value of t , we get a picture of the string

$$y = \sin\left(\frac{n\pi x}{L}\right)$$

which are the shapes shown earlier.

At a fixed point x ,

$$y = \sin\left(\frac{n\pi\nu}{L}t\right)$$

or a particular x point oscillates up and down with the normal mode frequency.

Fast Image Sequence (like a movie)

```
function z=acoeff(m,L)
if (2*floor(m/2) == m)
    z=0;
else
    z=(2*L/(m^2*pi^2))*sin(m*pi/2);
end
endfunction

function z=aterm(m,L,v,x,t)
z=acoeff(m,L)*sin(m*pi*x/L).*cos(m*pi*v*t/L);
endfunction

% m-file waveqxt.m
L=1;
v=1;
x=0:0.01:1;
for j = 1:201
    t=(j-1)*0.01; sum=0;
    for k=1:100
        sum=sum+aterm(k,L,v,x,t);
    end
    plot(x,sum,'-k');
    axis([-1 2 -.5 .5]);
    pause(0.01)
end
```

Run code....

Vibrations of a Rectangular Drum

We now choose to look at a rectangular drum because we can handle this boundary easily with a simple extension of our 1-dimensional solutions in cartesian coordinates.

We will look at a circular drum later.

The vibration of a 2-dimensional membrane fixed at the boundaries

$$x = 0, x = a, y = 0, y = b$$

can be described as follows:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

We choose

$$u(x, y, t) = X(x)Y(y)T(t)$$

SOV substitution gives

$$\frac{1}{X} \frac{\partial^2 X}{\partial x^2} = \frac{1}{Y} \frac{\partial^2 Y}{\partial y^2} = \frac{1}{c^2} \frac{1}{T} \frac{\partial^2 T}{\partial t^2}$$

The separation constant assignment goes like

$$\frac{1}{X} \frac{\partial^2 X}{\partial x^2} = \lambda_x \quad , \quad \frac{1}{Y} \frac{\partial^2 Y}{\partial y^2} = \lambda_y \quad , \quad \frac{1}{c^2} \frac{1}{T} \frac{\partial^2 T}{\partial t^2} = \lambda$$

with

$$\lambda_x + \lambda_y = \lambda \quad , \quad \text{all constants}$$

The solutions are

$$X(x) = A \cos k_x x + B \sin k_x x \quad , \quad k_x^2 = -\lambda_x$$

$$Y(y) = C \cos k_y y + D \sin k_y y \quad , \quad k_y^2 = -\lambda_y$$

$$T(t) = E \cos \alpha t + D \sin \alpha t \quad , \quad \alpha^2 = -\lambda$$

with

$$k_x^2 + k_y^2 = \frac{\alpha^2}{c^2}$$

The boundary conditions give

$$X(0) = 0 \rightarrow A = 0 \quad , \quad Y(0) = 0 \rightarrow C = 0$$

$$X(a) = 0 \rightarrow k_x = \frac{m\pi}{a} = k_m \quad , \quad Y(b) = 0 \rightarrow k_y = \frac{n\pi}{a} = k_n$$

or

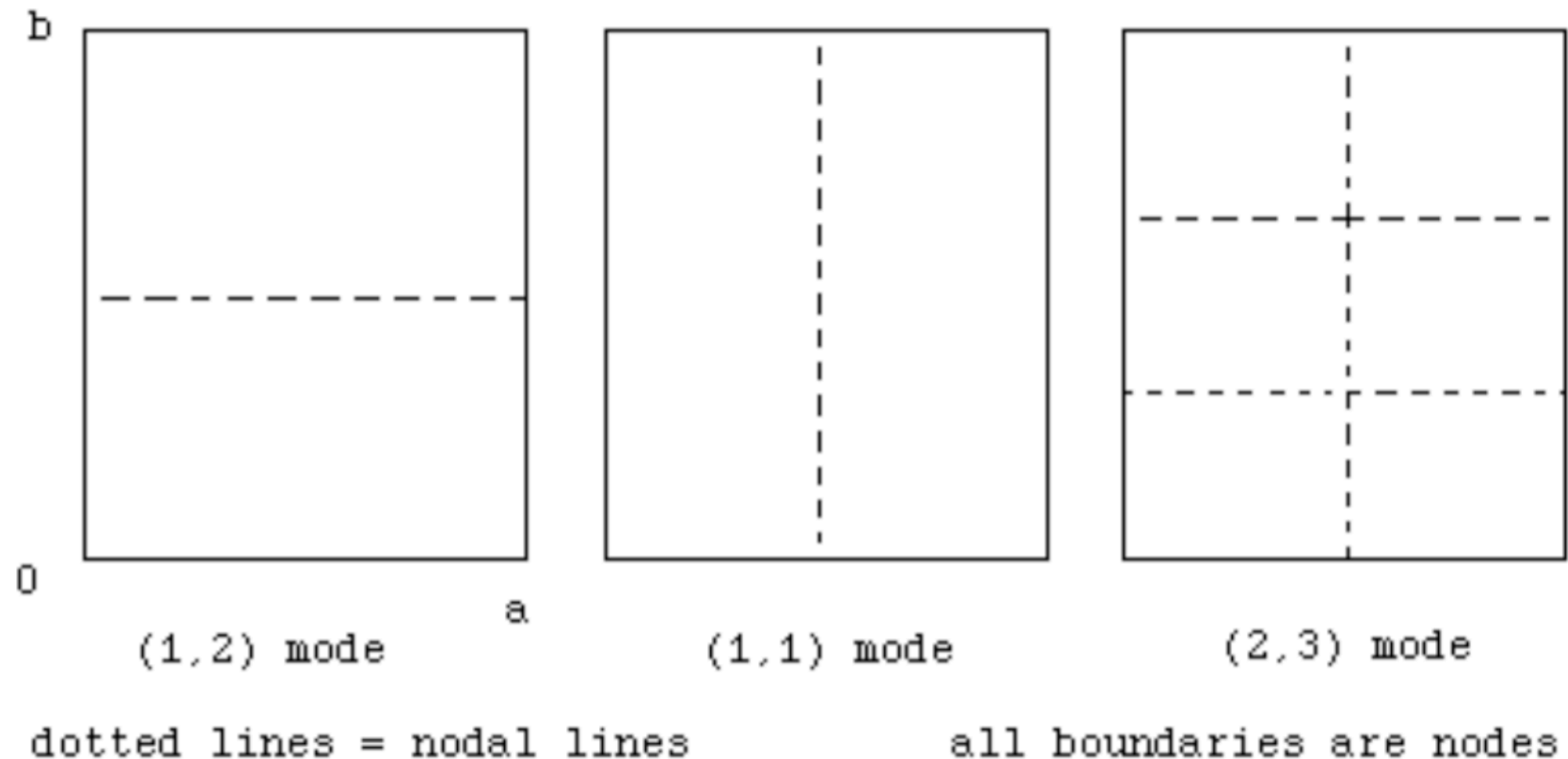
$$\frac{m^2\pi^2}{a^2} + \frac{n^2\pi^2}{a^2} = \frac{\alpha^2}{c^2} = \frac{\omega_{mn}^2}{c^2} \rightarrow \omega_{mn} = c\pi\sqrt{\frac{m^2}{a^2} + \frac{n^2}{a^2}}$$

The general solution is then a sum of all possible solutions (all m, n)

$$u(x, y, t) = \sum_{m,n=1}^{\infty} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} (c_{nmn} \cos \omega_{mn} t + c_{mn} \sin \omega_{mn} t)$$

where ω_{mn} = frequency of the (m, n) normal mode.

Some examples of modes are shown below:

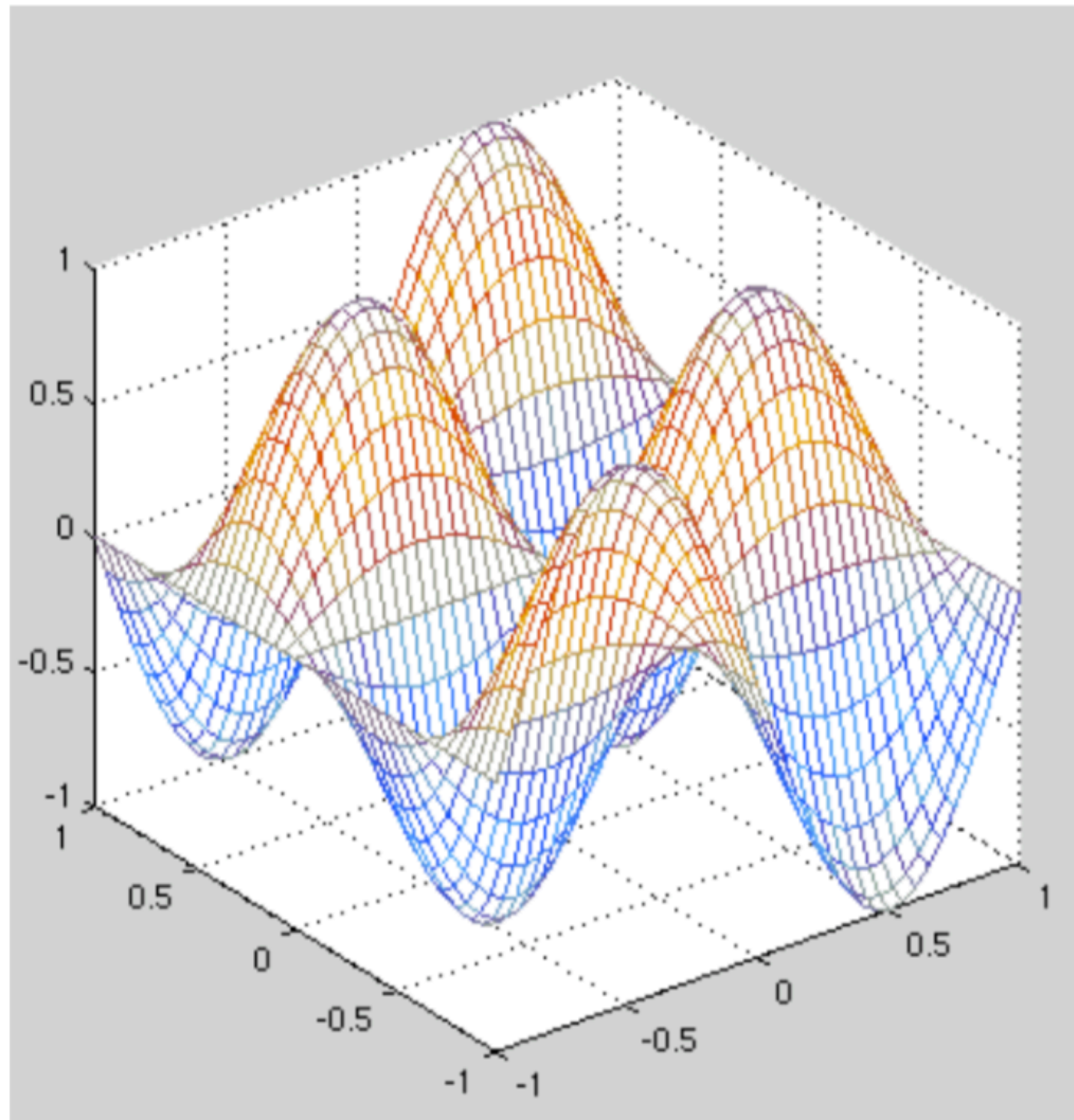


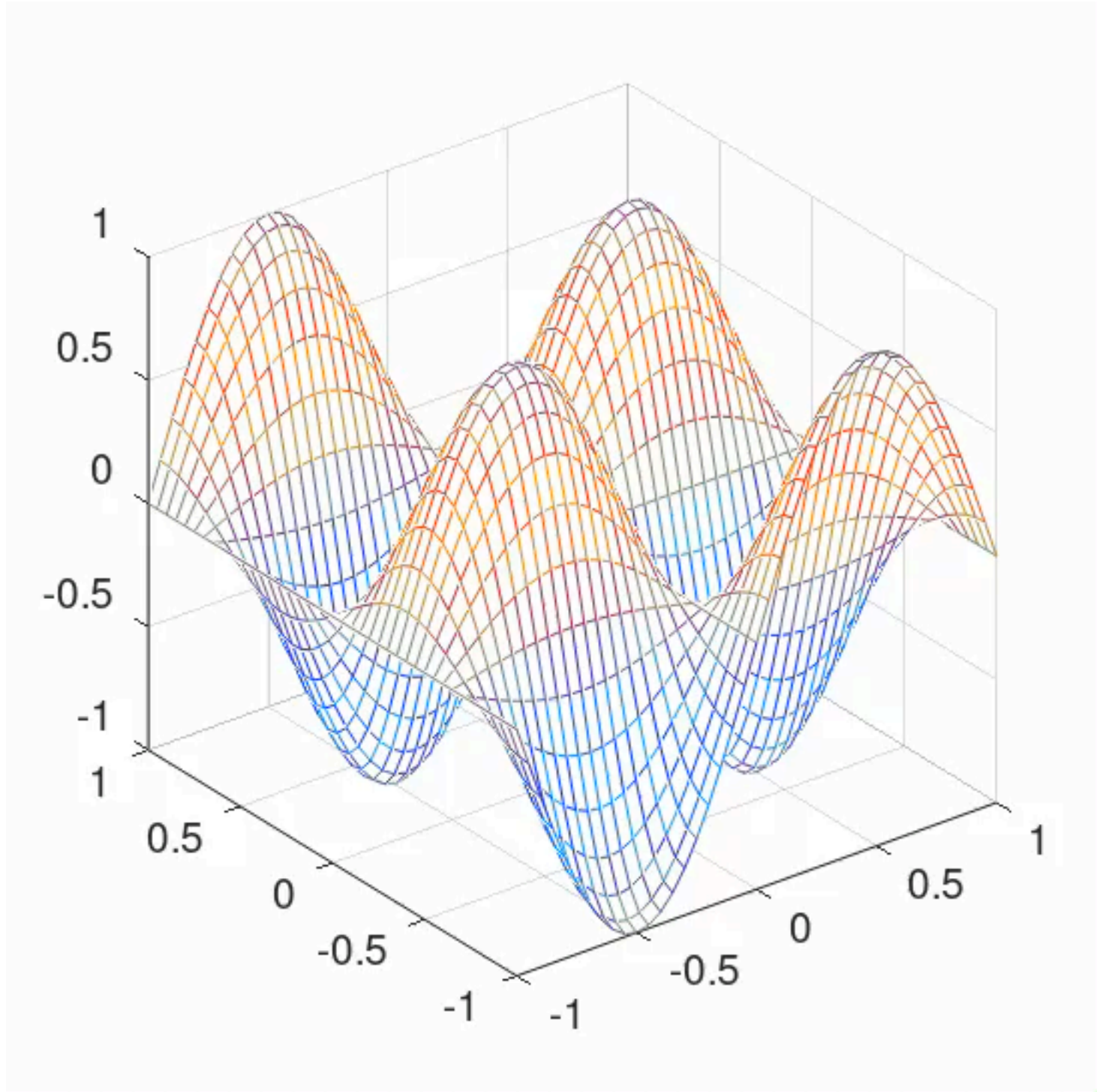
As always, the strength with which various normal modes are excited depends on the exact initial conditions.

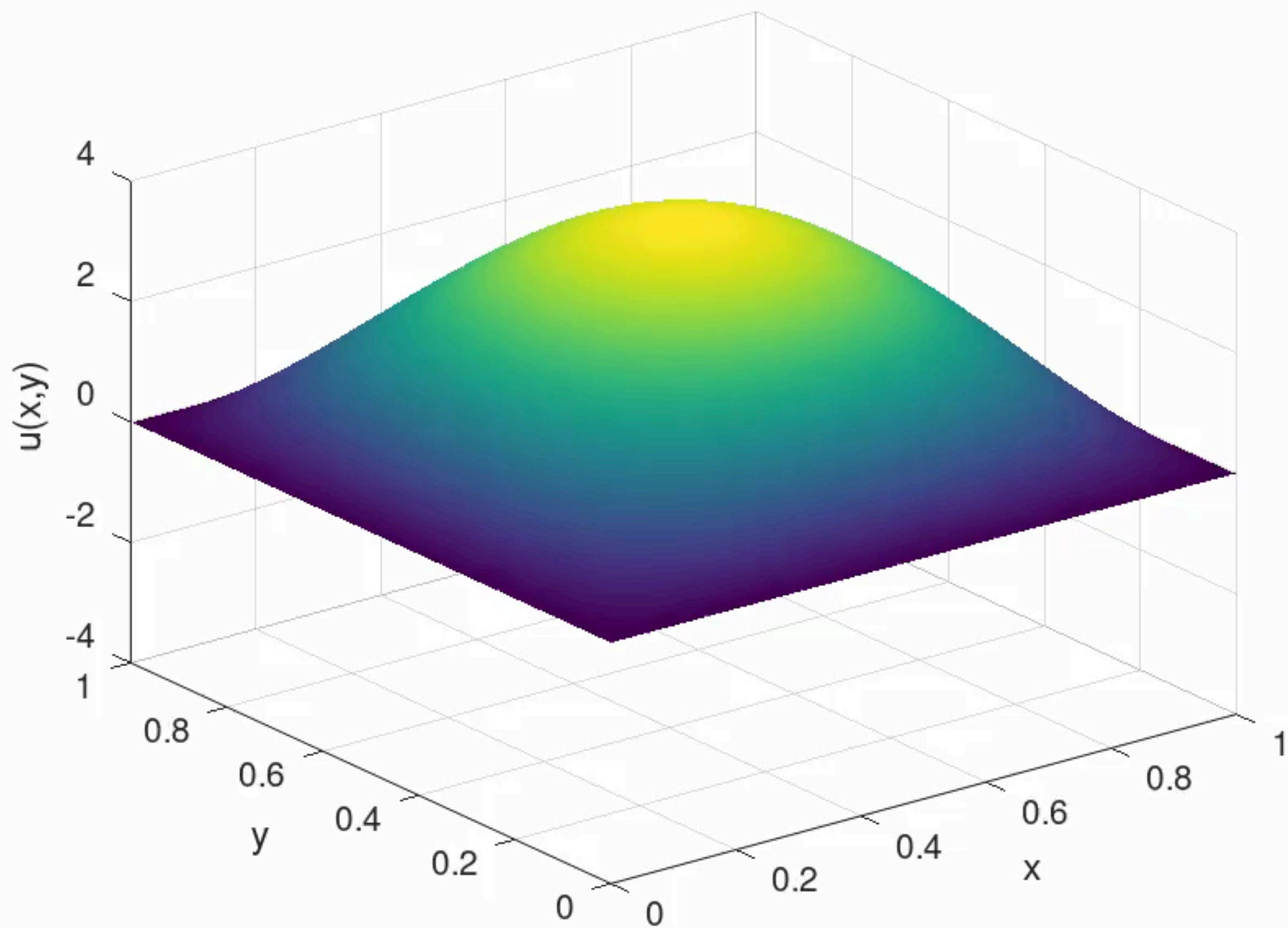
The OCTAVE program below `sq-drum.m` shows the (n, m) modes of the rectangular membrane (image sequence type movie).

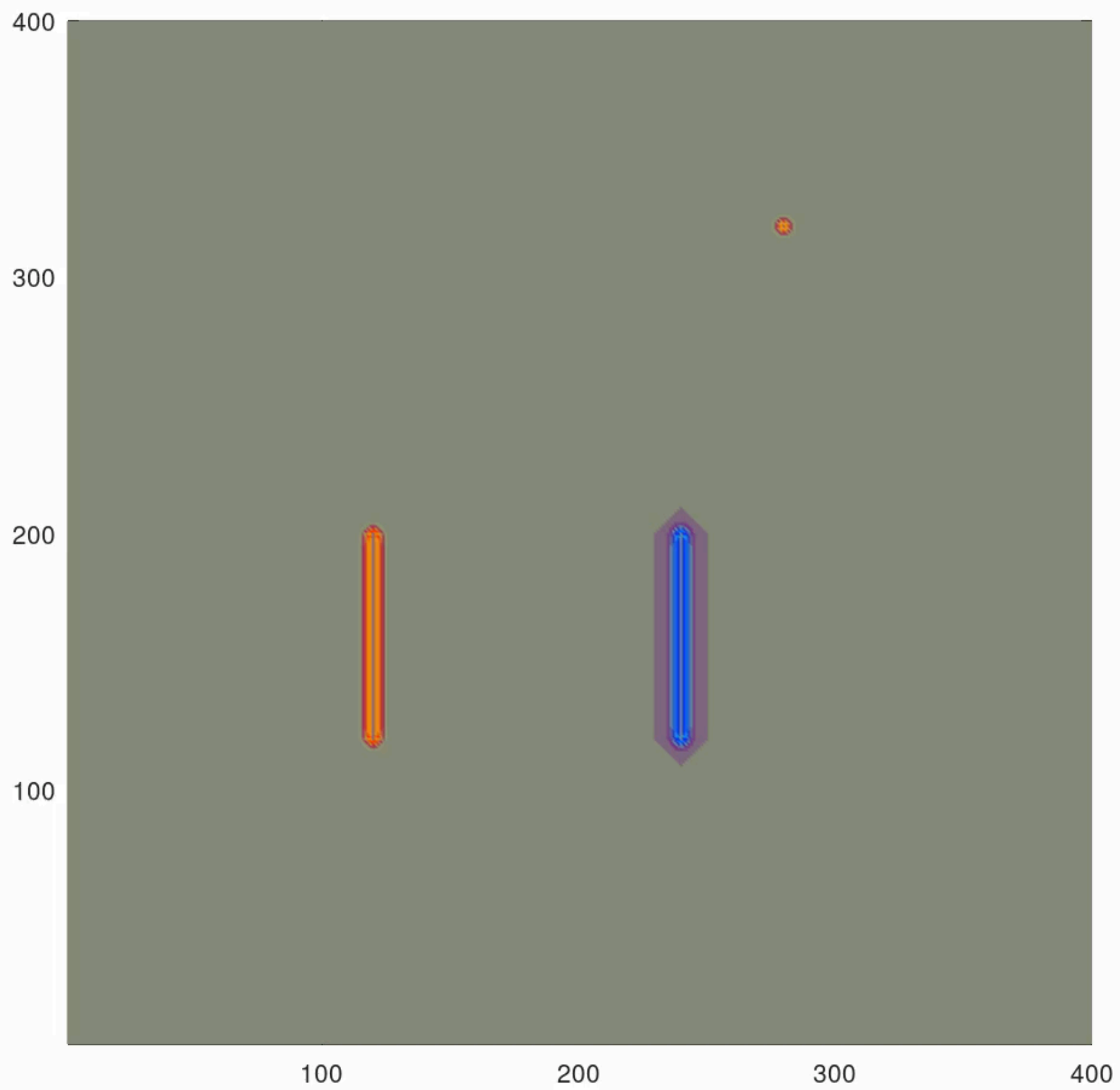
```
x=-1:0.05:1;
y=-1:0.05:1;
[X,Y]=meshgrid(x,y);
c=1;a=2;b=2;
% (1,2) mode
;m=1;n=1;
% (1,1) mode
;m=2;n=3;
% (2,3) mode
m=2;n=3;
w=c*pi*sqrt(m^2/a^2+n^2/b^2);
Z=sin(m*pi*X/a).*sin(n*pi*Y/b);
lim=[-1 1 -1 1 -1 1];
figure('Position', [200 200 400 400])
for j=1:500
    t=(j-1)*.01;
    mesh(X,Y,Z*cos(w*t))
    axis(lim);
    colormap(waves);
    pause(0.01);
end
```

An image from the sequence for the (2,3) mode looks like









Lots more in full lectures