Mathematical and Theoretical Biology Institute

Linear Algebra Review

1. Taylor Expansions

1.1. Taylor Polynomials. The linear approximation to a function $f(x)$ at a point x_0, is the linear function which has the same value and the same derivative as $f(x)$ at x_0, that is, we look for a polynomial of degree 1 such that

$$P_1(x) = c_0 + c_1 x,$$

and

$$P_1(x_0) = f(x_0), \qquad P_1(x_0) = f'(x_0).$$

Two equations are required to determine the coefficients c_0 and c_1. The condition $P_1(x_0) = f(x_0)$ gives

$$c_0 + c_1 x_0 = f(x_0) \tag{1.1}$$

and since $P_1'(x) = c_1$ the condition $P_1'(x_0) = f'(x_0)$ gives

$$c_1 = f'(x_0). \tag{1.2}$$

Substitution of (1.2) into (1.1) gives

$$c_0 = f(x_0) - c_1 x_0 = f(x_0) - x_0 f'(x_0)$$

and then we obtain

$$\begin{aligned} P_1(x) &= c_0 + c_1 x = f(x_0) - x_0 f'(x_0) + f'(x_0)x \\ &= f(x_0) + (x - x_0) f'(x_0), \end{aligned} \tag{1.3}$$

which is the desired linear approximation. The algebra would have been simpler if we had written the linear function $P_1(x)$ in the form

$$P_1(x) = a_0 + a_1(x - x_0).$$

Then the condition $P_1(x_0) = f(x_0)$ would give

$$a_0 = f(x_0) \tag{1.4}$$

and since $P_1'(x) = a_1$, the condition $P_1'(x_0) = f'(x_0)$ would give

$$a_1 = f'(x_0). \tag{1.5}$$

In this section, we consider polynomial approximations to a function $f(x)$ obtained by specifying conditions at a base point x_0. For example, let us find the quadratic (of degree 2) polynomial $P_2(x)$ which has the same value, derivative, and second derivative as $f(x)$ at x_0. If we write

$$P_2(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2. \tag{1.6}$$

The condition $P_2(x_0) = f(x_0)$ gives

$$a_0 = f(x_0). \tag{1.7}$$

Since $P_2'(x) = a_1 + 2a_2(x - x_0)$, the condition $P_2'(x_0) = f'(x_0)$ gives

$$a_1 = f'(x_0) \tag{1.8}$$

Since $P_2''(x) = 2a_2$, the condition $P_2''(x_0) = f''(x_0)$ gives $2a_2 = f''(x_0)$ or

$$a_2 = \frac{1}{2} f''(x_0). \tag{1.9}$$

Substitution of (1.7), (1.8), and (1.9) into (1.6) gives

$$P_2(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0). \tag{1.10}$$

Because the conditions (1.7) and (1.8) are the same as (1.4) and (1.5) respectively, comparison of (1.10) with (1.3) shows that the constant and linear terms in $P_2(x)$ are the same as the constant and linear terms in $P_1(x)$. Thus the quadratic approximation $P_2(x)$ is the linear approximation $P_1(x)$ plus an additional second-degree term.

We now define the *Taylor polynomial of degree n for $f(x)$ at the base point x_0* as the polynomial $P_n(x)$ of degree n such that

$$P_n(x_0) = f(x_0), P_n'(x_0) = f'(x_0), \ldots, P_n^{(n)}(x_0) = f^{(n)}(x_0). \tag{1.11}$$

As a polynomial of degree n has $(n + 1)$ coefficients the $(n + 1)$ conditions (1.11) determine the polynomial $P_n(x)$ completely. If we follow the same process for determining coefficients as we have used in the cases $n = 1, 2, 3$ we obtain

$$\begin{aligned} P_n(x) =& f(x_0) + (x - x_0)\frac{f'(x_0)}{1!} + (x - x_0)^2 \frac{f''(x_0)}{2!} \\ &+ \cdots + (x - x_0)^n \frac{f^{(n)}(x_0)}{n!}, \end{aligned} \tag{1.12}$$

where $k!$ means the product of the integers from 1 to k, so that $1! = 1$, $2! = 1 \cdot 2 = 2$, $3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$, etc.; it is conventional to define $0! = 1$.

EXAMPLE 1. Calculate the Taylor polynomials of degrees 1, 2, 3 and 4 for the function $f(x) = e^x$ at the base point 0 and use the approximations obtained to estimate e.

Solution: Since the derivative of every order of e^x is e^x, with value 1 at $x = 0$, the formula (1.12) gives

$$
\begin{aligned}
P_1(x) &= 1 + \frac{x}{1!},\\
P_2(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!},\\
P_3(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!},\\
P_4(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!}.
\end{aligned}
$$

In fact, it is easy to see that

$$
P_n(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \cdots + \frac{x^n}{n!} = \sum_{k=0}^{n} \frac{x^k}{k!}.
$$

Since e is the value of the function e^x for $x = 1$, we calculate as approximations $P_1(1) = 1 + \frac{1}{1!} = 2$, $P_2(x) = 1 + \frac{1}{1!} + \frac{1}{2!} = 2\frac{1}{2}$, $P_3(1) = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} = 2\frac{2}{3}$, $P_4(x) = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} = 2\frac{17}{24} = 2.7083$.

As we know a value for e, namely $2.71828\ldots$, we might conjecture that the approximation $P_n(x)$ to $f(x)$ becomes more accurate in general as n is increased, and that we may approximate the value of a function as accurately as we wish by making the degree of the approximating Taylor polynomial large enough. To validate (or disprove) such a conjecture we would need an estimate of the error in the approximation (the difference between the function and the Taylor polynomial) and how it depends on the degree n of the Taylor polynomial.

Example 2. Find the Taylor polynomial $P_n(x)$ of the function $f(x) = \frac{1}{1-x}$ at the base point 0.

Solution. We have

$$
\begin{aligned}
f(x) &= \frac{1}{1-x} & f(0) &= 1\\
f'(x) &= \frac{1}{(1-x)^2} & f'(0) &= 1\\
f''(x) &= \frac{2}{(1-x)^3} & f''(0) &= 2\\
f'''(x) &= \frac{3\cdot 2}{(1-x)^4} & f'''(0) &= 3
\end{aligned}
$$

and we may see that in general,

$$
f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}} \qquad f^{(k)}(0) = k!
$$

Thus the Taylor polynomial of degree n is

$$
\begin{aligned}
P_n(x) &= 1 + x + \frac{2!}{2!}x^2 + \frac{3!}{3!}x^3 + \cdots + \frac{n!}{n!}x^n \\
&= 1 + x + x^2 + \cdots + x^n.
\end{aligned}
\tag{1.13}
$$

It is possible to obtain a different but equivalent expression for $P_n(x)$ in Example 2 above by a device used for summing a geometric series. We multiply the equation (1.13) by x, obtaining

$$xP_n(x) = x + x^2 + x^3 + \cdots + x^n + x^{n+1} \tag{1.14}$$

and subtract (1.14) from (1.13) obtaining

$$(1-x)P_n(x) = 1 - x^{n+1}.$$

Thus

$$P_n(x) = \frac{1}{1-x} - \frac{x^{n+1}}{1-x}. \tag{1.15}$$

The relation (1.15) says that the difference between the function $f(x) = \frac{1}{1-x}$ and the Taylor polynomial $P_n(x)$ is $\frac{x^{n+1}}{1-x}$. From this explicit formula for the error we see that if $|x| < 1$, so that $x^{n+1} \to 0$ as $n \to \infty$, the error decreases to zero as $n \to \infty$. Thus on the interval $-1 < x < 1$, the Taylor polynomial $P_n(x)$ approximates the function $f(x) = \frac{1}{1-x}$ with an error which approaches zero as $n \to \infty$.

The main uses of Taylor polynomial approximation are not to estimate the value of a function at a given point. Currently available technology, such as inexpensive electronic pocket calculators, provides easier methods. The importance of Taylor polynomials is in approximating functions over an interval, and this will require error estimates which are valid over such an interval.

EXAMPLE 3. The Taylor polynomial $P_n(x)$ of the function $f(x) = \ln(1+x)$ at the base point 0 is

$$\begin{aligned} P_n(x) &= \frac{x}{1!} - \frac{x^2}{2!} + 2!\frac{x^3}{3!} - \cdots + (-1)^{n-1}\frac{(n-1)!}{n!}x^2 \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n}. \end{aligned}$$

EXAMPLE 4. The Taylor polynomial $P_n(x)$ of the function $f(x) = \sin x$ at the base point 0 is

$$P_n(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}, \text{etc.}$$

Although we speak of $P_n(x)$ as a polynomial of degree n, in this case $P_n(x)$ is actually a polynomial degree $(n-1)$ when n is even.

EXAMPLE 5. The Taylor polynomial $P_n(x)$ of the function $f(x) = (1+x)^p$ at the base point 0

$$P_3(x) = 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3, \text{etc.}$$

Note that if p is a positive integer, $f^{(k)}(0) = 0$ for $k \geq p+1$ because

$$f^{(k)}(0) = p(p-1)(p-2)\cdots(p-k+1).$$

Thus $P_n(x)$ is a polynomial of fixed degree p for every $n \geq p$. However, if p is not a positive integer, $P_n(x)$ will always have degree n no matter how large n is.

EXAMPLE 6. Find the Taylor polynomials of degree 0,1,2, and 3 of the function $f(x) = \int_0^x e^{-t^2} dt$ at the base point 0, and use them to estimate $f(0.2)$.

Solution. We have

$$\begin{aligned}
f(x) &= \int_0^x e^{-t^2} dt, \qquad f(0) = 0 \\
f'(x) &= e^{-x^2}, \qquad f'(0) = 1 \\
f''(x) &= -2xe^{-x^2}, \qquad f''(0) = 0 \\
f'''(x) &= -2xe^{-x^2} - (2x)(-2x)e^{-x^2}, \qquad f'''(0) = -2
\end{aligned}$$

Thus

$$\begin{aligned}
P_0(x) &= 0, & P_0(0.2) &= 0 \\
P_1(x) &= x, & P_1(0.2) &= 0.2 \\
P_2(x) &= x, & P_2(0.2) &= 0.2 \\
P_3(x) &= x - \frac{x}{3!}x^3 = x - \frac{x^3}{3!}, & P_3(0.2) &= 0.2 - 0.00267 = 0.19733.
\end{aligned}$$

The values $P_0(0.2)$, $P_1(0.2)$, $P_2(0.2)$, $P_3(0.2)$ are the respective approximations to $f(0.2) = \int_0^{0.2} e^{-t^2} dt$.

1.2. Taylor's Theorem. We have calculated Taylor polynomials for a given function, thinking of them as approximations to the function. The questions which we explore in this section is how good an approximation to $f(x)$ is the Taylor polynomial $P_n(x)$. The Taylor polynomial $P_n(x)$ at base point x_0 is designed to approximate the function $f(x)$ as well as possible at the base point x_0, being defined by the conditions that $P_n(x)$ and its derivatives up to order $(n-1)$ at the point x_0. We shall obtain an estimate for the difference between $f(x)$ and $P_n(x)$ on an interval containing x_0, and this will measure the accuracy of the approximation on that interval.

Let us define

$$R_n(x) = f(x) - P_n(x).$$

Then $R_n(x)$ is the error made in approximating $f(x)$ by $P_n(x)$, and is the quantity which we wish to estimate. We may interpret the mean value theorem as saying

$$R_0(x) = (x - x_0)f'(c)$$

for some point c between x_0 and x, and

$$R_1(x) = (x - x_0)(c - x_0)f''(d)$$

for points c and d between x_0 and x and d between x_0 and c. Taylor's theorem expresses $R_n(x)$ in terms of the values of $f^{(n+1)}(t)$ as t ranges over the interval from x_0 to x.

For a given positive integer n, suppose that $f(t), f'(t), \ldots, f^{(n+1)}(t)$ are continuous on $x_0 \le t \le x$. Let

$$P_n(x) = f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2!} + \cdots + (x - x_0)^n \frac{f^{(n)}(x_0)}{n!} \tag{1.16}$$

and

$$R_n(x) = f(x) - P_n(x), \tag{1.17}$$

then

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1} \tag{1.18}$$

for some point c between x_0 and x.
If

$$|f^{(n+1)}(t)| \le M \tag{1.19}$$

for all t between x_0 and x, then the remainder $R_n(x)$ satisfies the estimate

$$|R_n(x)| \le \frac{M}{(n+1)!}|x - x_0|^{n+1}. \tag{1.20}$$

EXAMPLE. Obtain a bound for the error in the estimate 0.19733 for $\int_0^{0.2} e^{-t^2} dt$.

Solution. The estimate 0.19733 was obtained from the Taylor polynomial of degree 3. Thus we must estimate $R_3(0.2)$. In Example 6 with $f(x) = \int_0^x e^{-t^2} dt$, we found

$$f'''(x) = -2e^{-x^2} + 4x^2 e^{-x^2}.$$

Therefore,

$$\begin{aligned} f^{(4)}(x) &= 4xe^{-x^2} + 8xe^{-x^2} + (4x^2)(-2x)e^{-x^2} \\ &= 12xe^{-x^2} - 8x^3 e^{-x^2} = e^{-x^2}(12x - 8x^3). \end{aligned}$$

For $0 \le x \le 0.2$ we have

$$|f^{(4)}(x)| = e^{-x^2}|12x - 8x^3| \le |12x - 8x^3|.$$

It is not difficult to verify that the function $12x - 8x^3$ is monotone decreasing for $0 \le x \le 0.2$, from 0 at $x = 0$ to -2.335 at $x = 0.2$. Thus $|f^{(4)}(x)| \le 2.336$ for $0 \le x \le 0.2$. Now the estimate (1.20) with $n = 3$ gives

$$|R_3(0.2)| \le \frac{2.336}{4!}(0.2)^4 = 0.0001557.$$

This shows that

$$|\int_0^{0.2} e^{-t^2} dt - 0.19733| \leq 0.0001557,$$

or that $\int_0^{0.2} e^{-t^2}$ is between $0.19733 - 0.00016 = 0.19717$, and $0.19733 + 0.00016 = 0.19749$.

1.3. T. aylor's Theorem for Functions of Two Variables

To find maxima and minima of a differentiable function f of one variable, we look for *critical points* - points x_0 for which $f'(x_0) = 0$. In order to decide whether a critical point x_0 is a maximum, a minimum, or neither, we look at the quadratic approximation to $f(x)$ near x_0 given by Taylor's theorem,

$$f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + R_2. \tag{1.21}$$

The linear term $f'(x_0)(x-x_0)$ in the Taylor polynomial is not present in (1.21) because of the assumption $f'(x_0) = 0$. if x is close enough to x_0 for the remainder term R_2 to be smaller than the quadratic term $\frac{1}{2}f''(x_0)(x-x_0)^2$, then $f(x)-f(x_0)$ has the same sign as $f''(x_0)$. Thus if $f''(x_0) > 0, f(x) > f(x_0)$ for x near x_0, and x_0 is a relative minimum. This is the justification for the familiar second derivative test for a function of one variable. However if $f''(x_0) = 0$, we can not determine the nature of the funcion near x_0 without looking at higher order approximations.

In order to obtain an analogous criterion for critical points of a function of two variables, we must first show how to formulate Taylor's theorem for a function of two variables.

For a function $f(x, y)$ of two variables, we let $x = x_0 + h$, $y = y_0 + k$ and define a function F of one variable t by

$$F(t) = f(x_0 + ht, y_0 + kt).$$

Then $F(1) = f(x, y)$ and $F(0) = f(x_0, y_0)$ Taylor's theorem for one variable applied to the function $F(t)$ gives

$$F(1) = F(0) + \frac{F'(0)}{1!} + \frac{F''(0)}{2!} + R_2. \tag{1.22}$$

By the chain rule

$$F'(t) = hf_x(x_0 + ht, y_0 + kt) + kf_y(x_0 + ht, y_0 + kt), \tag{1.23}$$

and a second differentiation gives

$$\begin{aligned} F''(t) &= h[hf_{xx}(x_0 + ht, y_0 + kt) + kf_{xy}(x_0 + ht, y_0 + kt)] \\ &\quad + k(hf_{yx}(x_0 + ht, y_0 + kt) + kf_{yy}(x_0 + ht, y_0 + kt)] \\ &= h^2 f_{xx}(x_0 + ht, y_0 + kt) + kf_{yy}(x_0 + ht, y_0 + kt) \\ &\quad k^2 f_{yy}(x_0 + ht, y_0 + kt). \end{aligned} \tag{1.24}$$

Substituting $t = 0$ in (1.23) and (1.24) we obtain

$$F'(0) = hf_x(x_0, y_0) + kf_y(x_0, y_0)$$
$$F''(0) = h^2 f_{xx}(x_0, y_0) + 2hkf_{xy}(x_0, y_0) + k^2 f_{yy}(x_0, y_0) \tag{1.25}$$

Now, substitution of (1.25) into (1.22) gives Taylor's theorem for functions of two variables

$$\begin{aligned} f(x, y) =& f(x_0, y_0) + (x - x0)f_x(x_0, y_0) + (y - y_0)f(x_0, y_0) \\ &+ \frac{1}{2}[(x - x_0)^2 f_{xx}(x_0, y_0) + 2(x - x_0)(y - y_0)f_{xy}(x_0, y_0) \\ &+ (y - y_0)^2 f_{yy}(x_0, y_0)] + R_2 \end{aligned} \tag{1.26}$$

Similar expansions can be computed for a function of n-variables.

2. Linear alebra review

2.1. Vectors and matrices. A real n-vector is an ordered n-tuple of real numbers of the form

$$v = (a_1, a_2, \ldots, a_n).$$

We also write v in the form

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{2.1}$$

In these notes, we rather us the notation (2.4), since this notation will simplify calculations later. The rela numbers will be called *scalars.* If v is a real n-vector, then we say $v \in \mathbb{R}^n$.

We have the following two vector operations:

(1) *Addition*, which is given by the formula

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix};$$

(2) *Scalar multiplication*, defined by

$$\lambda \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

We define the *scalar product* (also called *inner* or *dot* product) of two vectors by

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + \cdots + a_n b_n.$$

EXAMPLE 1. Consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}.$$

Then

$$v_1 + v_2 = \begin{pmatrix} -1 \\ 4 \\ 0 \end{pmatrix}, \qquad 3v_1 = \begin{pmatrix} 3 \\ 9 \\ -6 \end{pmatrix},$$

$$v_1 \cdot v_2 = (1)(-2) + (3)(1) + (-2)(2) = -3.$$

□

It is not hard to see that the operations defined above have the following properties:

$$\begin{aligned} u+v &= v+u \\ (u+v)+w &= u+(v+w) \\ \lambda(u+v) &= \lambda u + \lambda v \\ u\cdot v &= v\cdot u \\ u\cdot(v+w) &= u\cdot v + u\cdot w \\ u\cdot(\lambda v) &= \lambda u\cdot v \end{aligned}$$

An $m \times n$ matrix is an array of mn real numbers with m rows and n columns:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a2n \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{2.2}$$

The matrix A is also written as $A = (a_{ij})$. The *size* of A is $m \times n$. Matrices of the same size can be added, and matrices can be multiplied by scalars, according to the following rules:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots\ddots\vdots & & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix},$$

$$\lambda \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

EXAMPLE 2. Consider the matrices

$$A = \begin{pmatrix} 2 & -1 & 3 \\ -2 & 5 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & -1 & 3 \\ 4 & 4 & -2 \end{pmatrix}.$$

Then we have

$$\begin{aligned} A+B &= \begin{pmatrix} 2 & -2 & 6 \\ 2 & 9 & -2 \end{pmatrix}, \\ 5A &= \begin{pmatrix} 10 & -5 & 15 \\ -10 & 25 & 0 \end{pmatrix}, \\ -B &= \begin{pmatrix} 0 & 1 & -3 \\ -4 & -4 & 2 \end{pmatrix}. \end{aligned}$$

□

The matrix operations satisfy the following properties:

$$\begin{aligned} A+B &= B+A \\ A+(B+C) &= (A+B)+C \\ \lambda(A+B) &= \lambda A+\lambda B. \end{aligned}$$

If $A=(a_{ij})$ is an $m\times n$ matrix, and $B=(b_{jk})$ is an $n\times p$ matrix, then the product AB is defined as the $m\times p$ matrix given by $C=(c_{ik})$ where

$$c_ik=\sum_{j=1}^{n} a_{ij}b_{jk}$$

That is, the (i,k)-th element of C is the scalar product of the ith row of A and the kth column of B. Note that the product of two matrices A and B is defined only when the number of columns of A is equal to the number of rows of B.
EXAMPLE 3. Let

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

then

$$AB=\begin{pmatrix} -2 & 12 & 9 \\ -4 & 7 & -13 \end{pmatrix}.$$

□

We can check that the matrix product satisfies the following properties

$$\begin{aligned} A(BC) &= (AB)C \\ A(B+C) &= AB+AC \\ A(\lambda B) &= \lambda AB \end{aligned}$$

It is important to note that the matrix product **is not commutative**, that is, in general $AB\neq BA$. Moreover, if the product AB is defined, in general BA is not defined.
A matrix is called a *square matrix* if the number of its rows is equal to the number of its columns. The *main diagonal* of an $n\times n$ square matrix $A=(a_{ij})$ is the n-tuple $(a_{11},a_{22},\dots,a_{nn})$. A square matrix D is called a *diagonal matrix* if all its elemnts are zero with the exception of its diagonal:

$$D=\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

we also write D as $D=\operatorname{diag}(\lambda_1,\lambda_2,\dots,\lambda_n)$.
EXAMPLE 4. Let

$$A=\begin{pmatrix} 2 & -2 & 4 \\ 0 & 1 & -1 \\ -2 & 5 & 0 \end{pmatrix}, \qquad D=\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{pmatrix}=\operatorname{diag}(2,-1,5).$$

Then

$$AD = \begin{pmatrix} 4 & 1 & 20 \\ 0 & -1 & -5 \\ -4 & -5 & 0 \end{pmatrix},$$

$$DA = \begin{pmatrix} 4 & -2 & 8 \\ 0 & -1 & 1 \\ -10 & 25 & 0 \end{pmatrix}.$$

□

The example above suggests the follow fact, which is indeed true: If $A = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$ is a square matrix whose columns are the n-vectors $v_1, v_2, \ldots, v_n$, and $B = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ is a square matrix whose rows are the n-vectors $u_1, u_2, \ldots, u_n$, then

$$A\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{pmatrix} \tag{2.3}$$

$$\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)B = \begin{pmatrix} \lambda_1 u_1 \\ \lambda_2 u_2 \\ \vdots \\ \lambda_n u_n \end{pmatrix}. \tag{2.4}$$

The $n \times n$ *identity matrix* is the matrix $I = \text{diag}(1, 1, \ldots, 1)$, and satisfies

$$AI = IA = A \tag{2.5}$$

for all $n \times n$ matrices A.

2.2. Systems of linear equations. A system of m linear equations in n variables is a system of equations of the form

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m. \end{array}$$

We can write this system in the form $Ax = b$, where

$$A = (a_{ij}), \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

The above system can be solved by reducing the matrix A to a triangular matrix using the standard row operations, a method called the Gauss-Jordan method. We show this method with an example.

EXAMPLE 1. Solve the system

$$\begin{aligned} x - 3y - 5z &= -8 \\ -x + 2y + 4z &= 5 \\ 2x - 5y - 11z &= -9. \end{aligned}$$

We will solve this system by the Gauss-Jordan method. Using the standard row operations:

$$\begin{pmatrix} 1 & -3 & -5 & | & -8 \\ -1 & -2 & 4 & | & 5 \\ 2 & -5 & -11 & | & -9 \end{pmatrix} \sim \begin{pmatrix} 1 & -3 & -5 & | & -8 \\ 0 & 1 & 1 & | & 3 \\ 0 & 1 & -1 & | & 7 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & -2 & | & 1 \\ 0 & 1 & 1 & | & 3 \\ 0 & 0 & -2 & | & 4 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & 0 & | & -3 \\ 0 & 1 & 0 & | & 5 \\ 0 & 0 & 1 & | & -2 \end{pmatrix}.$$

Therefore the solution is given by $x = -3, y = 5, z = -2$.

□

2.3. Inverse matrix. Let A be a square $n \times n$ matrix. The inverse of A, if it exists, is a matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I.$$

Not every matrix A has an inverse. If A has a matrix, then it is called *invertible*, and its inverse is unique. **Example 1.** Find the inverse of the matrix

$$A = \begin{pmatrix} 1 & -3 & -5 \\ -1 & 2 & 4 \\ 2 & -5 & -11 \end{pmatrix}.$$

The inverse of A must be of the form (2.4), so we have to solve the system of 9 linear equations

$$A \cdot (x_{ij}) = I.$$

As in the previous example, we use the Gauss-Jordan method to solve this system of equations. The first and last steps of the algorithm are

$$\begin{pmatrix} 1 & -3 & -5 & | & 1 & 0 & 0 \\ -1 & -2 & 4 & | & 0 & 1 & 0 \\ 2 & -5 & -11 & | & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & | & -1 & -4 & -1 \\ 0 & 1 & 0 & | & -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & | & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Therefore the inverse of A is the matrix

$$A^{-1} = \begin{pmatrix} -1 & -4 & -1 \\ -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

□

2.4. Determinant. The *determinant* of a square matrix is defined inductively by

$$\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{aligned}\det\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix} =& a_{11}\det\begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn}\end{pmatrix} \\ &- a_{12}\det\begin{pmatrix} a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn}\end{pmatrix} + \cdots \\ &+ (-1)^{n+1}a_{1n}\det\begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)}\end{pmatrix}.\end{aligned}$$

EXAMPLE 1. Calculate the determinant of the matrix

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 4 & 1 \\ 0 & -2 & -1\end{pmatrix}.$$

We have

$$\det A = 2\det\begin{pmatrix} 4 & 1 \\ -2 & -1\end{pmatrix} - 0 + \det\begin{pmatrix} 1 & 4 \\ 0 & -2\end{pmatrix} = 2(-4+2) + (-2) = -6.$$

□

If $A = \begin{pmatrix} v_1 & \dots & v_n\end{pmatrix}$ is a square $n \times n$ matrix whose columns are the vectors v_i, then det A is an alternating multilinear form on the columns of A, *i.e.*

$$\begin{aligned}\det(v_1 \dots \alpha u + \beta v \dots v_n) &= \alpha\det(v_1 \dots u \dots v_n) \\ &\quad +\beta\det(v_1 \dots v \dots v_n) \\ \det(v_1 \dots v_i \dots v_j \dots v_n) &= -\det(v_1 \dots v_j \dots v_i \dots v_n)\end{aligned} \tag{2.6}$$

From (2.6), we can deduce the following properties of the determinant.

$$\det(v_1 \dots 0 \dots v_n) = 0 \tag{2.7}$$

$$\det(v_1 \dots u \dots u \dots v_n) = 0 \tag{2.8}$$

$$\det(v_1 \dots u \dots v + \alpha u \dots v_n) = \det(v_1 \dots u \dots v \dots v_n). \tag{2.9}$$

The transpose of a matrix $A = (a_{ij})$ is the matrix $A^T = (a_{ji})$, *i.e.* the matrix whose columns are the rows of A , and whose rows are the columns of A. If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.
One can show inductively that

$$\det A^T = \det A. \tag{2.10}$$

From (2.10) we conclude that the alternating multilinear properties (2.6 - 2.9) also hold for the rows of a matrix.
EXAMPLE 2. Calculate the determinant of the matrix

$$A = \begin{pmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ -2 & 1 & -1 & 3 & 2 \\ 10 & 4 & -2 & 0 & 1 \\ 4 & 6 & -1 & 0 & 3 \end{pmatrix}$$

We use properties (2.6 - 2.9) applied to the rows of A to calculate this determinant.

$$\det A = \det \begin{pmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{pmatrix} = -\det \begin{pmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{pmatrix}$$

$$= -\det \begin{pmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 0 & -26 & -38 & 0 \\ 0 & 0 & -70 & -66 & 11 \\ 0 & 0 & -59 & -66 & 7 \end{pmatrix} = -\det \begin{pmatrix} -26 & -38 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{pmatrix}$$

$$= 2\det \begin{pmatrix} 13 & 19 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{pmatrix} = 2\left(13\det \begin{pmatrix} -66 & 11 \\ -66 & 7 \end{pmatrix} - 19\det \begin{pmatrix} -70 & 11 \\ -59 & 7 \end{pmatrix}\right)$$

(2.11)

$$= 2(13(-462 + 726) - 19(-490 + 649)) = 822.$$

□

If A and B are square matrices of the same size, then

$$\det(AB) = (\det A)(\det B). \tag{2.12}$$

If the matrix A has an inverse A^{-1}, then the equation (2.12) implies

$$(\det A)(\det A^{-1}) = 1. \tag{2.13}$$

Thus, from equation (2.13), we can conclude that if A has an inverse, then $\det A \neq= 0$. The converse is true and is contained in the following theorem. Let A be an $n \times n$ matrix. Then the following are equivalent

(1) The system $Ax = b$ has a unique solution for each n-vector b.
(2) The matrix A is invertible.
(3) $\det A \neq 0$.

Note that if $\det A = 0$, then the system $Ax = 0$ has nonzero solutions x. We use this fact in the following section.

2.5. Eigenvalues and Eigenvectors. Let A be a square matrix. We say that $v \neq 0$ is an *eigenvector* of A if

$$Av = \lambda v \tag{2.14}$$

for some scalar[1] $\lambda \in \mathbb{C}$. The scalar λ is called the *eigenvalue* of A with respect to v[2].

EXAMPLE 1. Consider

$$A = \begin{pmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \qquad v = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

[1] We now also consider complex numbers as scalars.

[2] We also say that v is an eigenvector of A with respect to λ. Note that for each eigenvector of a matrix A corresponds a unique eigenvalue; however, several eigenvectors may correspond to a single eigenvalue.

Then v is an eigenvector of A with respect to the eigenvalue $\lambda = 2$:

$$Av = \begin{pmatrix} 5 & 8 & -3 \\ 0 & -3 & 9 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

□

We now describe an algorithm to find the eigenvalues of a matrix. First, note that if v is an eigenvector of A with respect to the eigenvalue λ, then λ and v are solutions to the equation (2.14) with $v \neq 0$. Equation 2.14 can be written in the form

$$(A - \lambda I)v = 0, \tag{2.15}$$

where I is the identity matrix of the same size as A. By theorem 2.4, equation 2.15 has a nonzero solution in v if and only if

$$\det(A - \lambda I) = 0. \tag{2.16}$$

Equation 2.16 is called the *characteristic equation* of the matrix A. Observe that if A is an $n \times n$ matrix, then the characteristic equation 2.16 is a polynomial equation in λ of degree n.

EXAMPLE 2. Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

The characteristic equation of A is given by

$$0 = \det(A - \lambda I) = \det \begin{pmatrix} 5-\lambda & 8 & -3 \\ 0 & -3-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{pmatrix} = (5-\lambda)(-3-\lambda)(2-\lambda).$$

Thus the eigenvalues of A are $\lambda_1 = 5$, $\lambda_2 = -3$, and $\lambda_3 = 2$. To calculate the eigenvectors, we solve the equation 2.15 for each λ_i. Using the Gauss-Jordan method we obtain that

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

are the eigenvectors of A with respect to $\lambda_1 = 5$, $\lambda_2 = -3$, and $\lambda_3 = 2$.

□

EXAMPLE 3. Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

The characteristic equation of A is given by

$$0 = \det\begin{pmatrix} 1-\lambda & 1 & 0 \\ 2 & -\lambda & 0 \\ 0 & 0 & 3-\lambda \end{pmatrix} = (1-\lambda)\det\begin{pmatrix} -\lambda & 0 \\ 0 & 3-\lambda \end{pmatrix} - \det\begin{pmatrix} 2 & 0 \\ 0 & 3-\lambda \end{pmatrix}$$
$$= (1-\lambda)(-\lambda)(3-\lambda) - 2(3-\lambda) = -(\lambda-3)(\lambda+1)(\lambda-2).$$

Thus, the eigenvalues of A are then $\lambda_1 = 3$, $\lambda_2 = -1$, and $\lambda_3 = 2$. The eigenvectors of A with respect to these eigenvalues are, respectively,

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

EXAMPLE 4. Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & 3 \\ -4 & 2 \end{pmatrix}.$$

The characteristic equation of A is the equation

$$0 = \det\begin{pmatrix} 1-\lambda & 3 \\ -4 & 2-\lambda \end{pmatrix} = \lambda^2 - 3\lambda + 14.$$

Then the eigenvalues of A are

$$\lambda_1 = \frac{3}{2} + \frac{\sqrt{47}}{2}i \text{ and } \lambda_2 = \frac{3}{2} - \frac{\sqrt{47}}{2}i.$$

The eigenvectors of A corresponding to λ_1 and λ_2 are, respectively,

$$v_1 = \begin{pmatrix} 3 \\ \frac{1}{2} + \frac{\sqrt{47}}{2}i \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 3 \\ \frac{1}{2} - \frac{\sqrt{47}}{2}i \end{pmatrix}.$$

□

Let A be an $n \times n$ matrix, and suppose that the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all different. If $v_1, v_2, \ldots, v_n$ are the eigenvectors of A with respect to the λ_i, let

$$D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$
$$S = (v_1, v_2, \ldots, v_n)$$

The matrix D is the diagonal matrix whose diagonal elements are the eigenvalues of A, and S is the matrix whose columns are the eigenvectors of A. Hence

$$AS = A\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} Av_1 & Av_2 & \cdots & Av_n \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{pmatrix}$$
$$= \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = SD,$$

where we have used (2.3). One can show that the matrix S has an inverse. Then, we factor A in the form

$$A = SDS^{-1}. \tag{2.17}$$

The expression (2.17) is called the *diagonalization* of A. Not every matrix has a diagonalization. If the matrix A has a diagonalization, then we say that A is

diagonalizable. We have seen, in the case where all the eigenvalues of A are distinct, that A is diagonalizable.
EXAMPLE 5. Diagonalize the matrix

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

The eigenvalues of A are $\lambda_1 = 3$, $\lambda_2 = 2$, and $\lambda_3 = -1$, and the eigenvectors with respect to these eigenvalues are

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}.$$

Let

$$S = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{pmatrix},$$
$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Then we have $A = SDS^{-1}$. Thus, the diagonalization of A is

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{3} & 0 \end{pmatrix}.$$

□

From (2.3) and (2.4) one can check that

$$(\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n))^k = \mathrm{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k).$$

Thus, if $A = SDS^{-1}$,

$$\begin{aligned} A^k &= (SDS^{-1})(SDS^{-1})\cdots(SDS^{-1}) \qquad (k\text{factors}) \\ &= (SD^kS^{-1}). \end{aligned}$$

If $p(x) = a_kx^k + \cdots + a_1x + a_0$ is a polynomial and A is an $n \times n$ matrix, we define

$$p(A) = a_kA^k + \cdots + a_1A + a_0I.$$

By the above observations, we see that if A is diagonalizable, then

$$p(A) = p(SDS^{-1}) = S{\cdot}p(D){\cdot}S^{-1} = S{\cdot}\mathrm{diag}\,(p(\lambda_1), p(\lambda_2)){\cdot}S^{-1} = S{\cdot}\mathrm{diag}(0, 0){\cdot}S^{-1} = 0.$$

□

The above example suggests the following fact, which is indeed true: Every diagonalizable matrix satisfies its charateristic equation[3].
If f is an analytic function, and A is diagonalizable, we define

$$f(A) = f(SDS^{-1}) = S\cdot f(D)\cdot S^{-1} = S\cdot \text{diag}\left(f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)\right)\cdot S^{-1}. \tag{2.18}$$

In the next section we study applications of these results to systems of differential equations.

[3] In fact, one can show that every matrix satisfies its characteristic equation.

2.6. Systems of differential equations. We begin this section with a study of the matrix e^{tA}. By the definition (2.18), we have that if A is a diagonalizable matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$e^{tA} = S \cdot \text{diag}\left(e^{t\lambda_1}, \ldots, e^{t\lambda_n}\right) \cdot S^{-1}.$$

Now let A be a diagonalizable matrix, and define the function F by $F(t) = e^{tA}$. The derivative of F is given by

$$\begin{aligned} F'(t) = \frac{d}{dt}e^{tA} &= S \cdot \text{diag}\left(\lambda_1 e^{t\lambda_1}, \ldots, \lambda_n e^{t\lambda_n}\right) \cdot S^{-1} \\ &= S \cdot \text{diag}\left(\lambda_1, \ldots, \lambda_n\right) \cdot \text{diag}\left(e^{t\lambda_1}, \ldots, e^{t\lambda_n}\right) \cdot S^{-1} \\ &= S \cdot \text{diag}\left(\lambda_1, \ldots, \lambda_n\right) \cdot S^{-1} \cdot S \cdot \text{diag}\left(e^{t\lambda_1}, \ldots, e^{t\lambda_n}\right) \cdot S^{-1} \\ &= A \cdot e^{tA}, \end{aligned}$$

where we have used (2.3) - (2.4). Therefore we conclude

$$F'(t) = AF(t). \tag{2.19}$$

From equation (2.19) we conclude that the solution to the differential equation

$$X'(t) = AX(t), \qquad X(0) = X_0,$$

where $X(t)$ is an n-vector valued function and A is an $n \times n$ matrix, is given by

$$X(t) = e^{tA}X_0.$$

EXAMPLE 1. Solve the system of equations

$$\begin{aligned} \frac{dx}{dt} &= x - 9y \\ \frac{dy}{dt} &= -6x + 4y \\ x(0) = 1, &\qquad y(0) = 3. \end{aligned}$$

This system can be written in the form $X'(t) = AX(t)$, $X(0) = X_0$, where

$$\begin{aligned} X(t) &= \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \\ A &= \begin{pmatrix} 1 & -9 \\ -6 & 4 \end{pmatrix}, \\ X_0 &= \begin{pmatrix} 1 \\ 3 \end{pmatrix}. \end{aligned} \tag{2.20}$$

Thus, the solution is given by

$$X(t) = e^{tA} X_0.$$

The diagonalization of A is

$$\begin{pmatrix} 1 & -9 \\ -6 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} -5 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{pmatrix},$$

so

$$\begin{aligned} X(t) &= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} e^{-5t} & 0 \\ 0 & e^{10t} \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} e^{-5t} & 0 \\ 0 & e^{10t} \end{pmatrix} \begin{pmatrix} \frac{4}{5} \\ -\frac{7}{5} \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \frac{4}{5} e^{-5t} \\ -\frac{7}{5} e^{10t} \end{pmatrix} \\ &= \begin{pmatrix} \frac{12}{5} e^{-5t} - \frac{7}{5} e^{10t} \\ \frac{8}{5} e^{-5t} + \frac{7}{5} e^{10t} \end{pmatrix}. \end{aligned} \tag{2.21}$$

Therefore the solution to the system of equations is

$$\begin{aligned} x(t) &= \frac{12}{5} e^{-5t} - \frac{7}{5} e^{10t} \\ y(t) &= \frac{8}{5} e^{-5t} + \frac{7}{5} e^{10t}. \end{aligned}$$

□

From this example we can conclude that the general solution to the equation $X'(t) = AX(t)$ is of the form $X(t) = c_1 e^{\lambda_1 t} v_1 + c_2 e^{\lambda_2 t} v_2 + \cdots + c_n e^{\lambda_n t} v_n$, for some constants c_i, where the λ_i's are the eigenvalues and the v_i's are the eigenvectors of the matrix$A$.

Exercises

(1) Diagonalize the matrices
   (a)
$$A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix};$$
   (b)
$$A = \begin{pmatrix} 1 & 3 \\ -4 & 2 \end{pmatrix}.$$

(2) For the matrices 1a and 1b of Exercise 1 above, calculate $p(A)$ where
$$p(x) = x^2 - 3x + 14.$$

(3) Suppose that M is a diagonalizable $n \times n$ matrix. Show that
$$\sin^2(M) + cos^2(M) = I,$$
where I is the identity $n \times n$ matrix.

(4) Solve the system of differential equations

$$\frac{dx}{dt} = x - 9y$$
$$\frac{dy}{dt} = -6x + 4y,$$

with the initial conditions $x(0) = 1$, $y(0) = 3$.

(5) Consider the differential equation of the harmonic oscillator

$$\frac{d^2x}{dt^2} = -kx, \tag{2.22}$$

with $k > 0$. By the substitution $y(t) = \frac{dx}{dt}$, the equation (2.22) can be written in the form

$$\frac{dx}{dt} = y \tag{2.23}$$
$$\frac{dy}{dt} = -kx.$$

Give the explicit solution of the system (2.23) with the initial conditions $x(0) = x_0$, $y(0) = y_0$. Deduce from this solution the explicit solution to the equation (2.22).

(6) Consider the SIRS model:

$$\frac{dS}{dt} = -\beta\frac{SI}{N} + \theta R$$
$$\frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I - \theta R,$$

where $N = S + I + R$ is a constant. The linearization at the disease-free equilibrium $(N, 0, 0)$ is given by

$$\frac{dx}{dt} = -\beta y + \theta z \tag{2.24}$$
$$\frac{dy}{dt} = (\beta - \gamma)y$$
$$\frac{dz}{dt} = \gamma y - \theta z. \tag{2.25}$$

Show that for the initial conditions

$$\begin{pmatrix} x(0) \\ y(0) \\ z(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

if $\frac{\beta}{\gamma} < 1$, then

$$\lim_{t\to\infty}\begin{pmatrix}x(t)\\ y(t)\\ z(t)\end{pmatrix}=\begin{pmatrix}x_0+y_0+z_0\\ 0\\ 0\end{pmatrix}.$$

(Hint: Write the system (2.24) in matrix form $X'(t) = AX$. Diagonalize the matrix A and use the methods seen in this review to obtain the vector form of the solution).