Vectors

Our very first topic is unusual in that we will start with a brief written presentation. More typically we will begin each topic with a videotaped lecture by Professor Auroux and follow that with a brief written presentation.

As we pointed out in the introduction, vectors will be used throughout the course. The basic concepts are straightforward, but you will have to master some new terminology. Another important point we made earlier is that we can view vectors in two different ways: geometrically and algebraically. We will start with the geometric view and introduce terminology along the way.

Geometric view

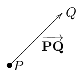
A vector is defined as having a magnitude and a direction. We represent it by an arrow in the plane or in space. The length of the arrow is the vector's magnitude and the direction of the arrow is the vector's direction.

In this way, two arrows with the same magnitude and direction represent the same vector.

(same vector)

We will refer to the start of the arrow as the *tail* and the end as the *tip* or *head*.

The vector between two points will be denoted $\overrightarrow{\mathbf{PQ}}$.

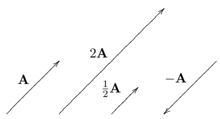


We call P the initial point and Q the terminal point of $\overrightarrow{\mathbf{PQ}}$.

The *magnitude* of the vector $\mathbf{A}$ will be denoted $|\mathbf{A}|$. Magnitude will also be called *length* or *norm*

Scaling, adding and subtracting vectors

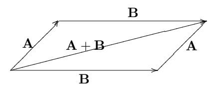
Scaling a vector means changing its length by a scale factor. For example,



(scaling a vector)

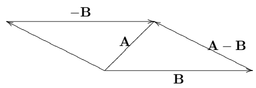
Because we use numbers to scale a vector we will often refer to real numbers as *scalars*

You add vectors by placing them head to tail. As the figure shows, this can be done in either order



It is often useful to think of vectors as *displacements*. In this way, $\mathbf{A}+\mathbf{B}$ can be thought of as the displacement $\mathbf{A}$ followed by the displacement $\mathbf{B}$.

You subtract vectors either by placing the tail to tail or by adding $\mathbf{A}+(-\mathbf{B})$.



Thought of as displacements $\mathbf{A}-\mathbf{B}$ is the displacement from the end of $\mathbf{B}$ to the end of $\mathbf{A}$.

Algebraic view

As is conventional, we label the origin O. In the plane $O=(0,0)$ and in space $O=(0,0,0)$.

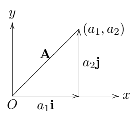
In the xy-plane if we place the tail of $\mathbf{A}$ at the origin, its head will be at the point with coordinates, say, (a_1,a_2). In this way, the coordinates of the head determine the vector $\mathbf{A}$. When we draw $\mathbf{A}$ from the origin we will refer to it as an *origin vector*.

Using the coordinates we write

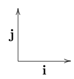
$$\mathbf{A}=\langle a_1,\,a_2\rangle.$$

Addition, subtraction and scaling using coordinates is discussed below.

Graphically:



The vectors $\mathbf{i}$ and $\mathbf{j}$ used in the figure above have coordinates $\mathbf{i}=\langle 1,\,0\rangle$, $\mathbf{j}=\langle 0,\,1\rangle$. We use them so often that they get their own symbols.



Notation and terminology

1. (a_1,a_2) indicates a point in the plane.
2. $\langle a_1,\,a_2\rangle=a_1\mathbf{i}+a_2\mathbf{j}$. This is equal to the vector drawn from the origin to the point (a_1,a_2).
3. For $\mathbf{A}=a_1\mathbf{i}+a_2\mathbf{j}$, a_1 and a_2 are called the $\mathbf{i}$ and $\mathbf{j}$ *components* of $\mathbf{A}$. (Note that they are scalars.)
5. $\overrightarrow{\mathbf{P}}=\overrightarrow{\mathbf{OP}}$ is the vector from the origin to P.

6. On the blackboard vectors will usually have an arrow above the letter. In print we will often drop the arrow and just use the bold face to indicate a vector, i.e. $\mathbf{P} \equiv \overrightarrow{\mathbf{P}}$.
7. A real number is a *scalar*, you can use it to scale a vector.

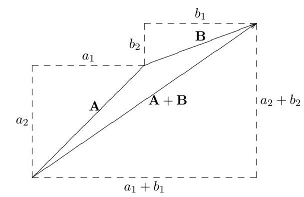
Vector algebra using coordinates

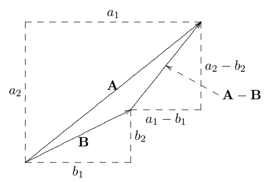
For the vectors $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j}$ we have the following algebraic rules. The figures below connect these rules to the geometric viewpoint.

Magnitude: $|\mathbf{A}| = \sqrt{a_1^2 + a_2^2}$ (this is just the Pythagorean theorem)

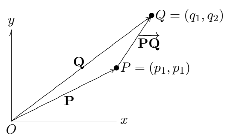
Addition: $\mathbf{A} + \mathbf{B} = (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j}$, that is, $\langle a_1, a_2\rangle + \langle b_1, b_2\rangle = \langle a_1 + b_1, a_2 + b_2\rangle$

Subtraction: $\mathbf{A} - \mathbf{B} = (a_1 - b_1)\mathbf{i} + (a_2 - b_2)\mathbf{j}$, that is, $\langle a_1, a_2\rangle - \langle b_1, b_2\rangle = \langle a_1 - b_1, a_2 - b_2\rangle$





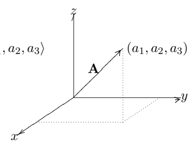
For two points P and Q the vector $\overrightarrow{\mathbf{PQ}} = \overrightarrow{\mathbf{Q}} - \overrightarrow{\mathbf{P}}$ i.e., $\overrightarrow{\mathbf{PQ}}$ is the *displacement* from P to Q.



Vectors in three dimensions

We represent a three dimensional vector as an arrow in space. Using coordinates we need three numbers to represent a vector.

$\mathbf{A} = \langle a_1, a_2, a_3\rangle$



Geometrically nothing changes for vectors in three dimensions. They are scaled and added exactly as above.

Algebraically the origin vector $\mathbf{A} = \langle a_1, a_2, a_3 \rangle$ starts at the origin and extends to the point (a_1, a_2, a_3). We have the special vectors $\mathbf{i} = \langle 1, 0, 0 \rangle$, $\mathbf{j} = \langle 0, 1, 0 \rangle$, $\mathbf{k} = \langle 0, 0, 1 \rangle$. Using them

$$\langle a_1, a_2, a_3 \rangle = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}.$$

Then, for $\mathbf{A} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3 \rangle$ we have

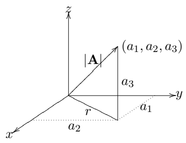
$$\langle a_1, a_2, a_3 \rangle + \langle b_1, b_2, b_3 \rangle = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle.$$

exactly as in the two dimensional case.

Magnitude in three dimensions also follows from the Pythagorean theorem.

$$|a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}| = |\langle a_1, a_2, a_3 \rangle| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

You can see this in the figure below, where $r = \sqrt{a_1^2 + a_2^2}$ and $|\mathbf{A}| = \sqrt{r^2 + a_3^2} = \sqrt{a_1^2 + a_2^2 + a_3^2}$.



Unit vectors

A unit vector is any vector with unit length. When we want to indicate that a vector is a unit vector we put a hat (circumflex) above it, e.g., $\hat{\mathbf{u}}$.

The special vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors.

Since vectors can be scaled, any vector can be rescaled to be a unit vector.

Example: Find a unit vector that is parallel to $\langle 3, 4 \rangle$.

Answer: Since $|\langle 3, 4 \rangle| = 5$ the vector $\frac{1}{5}\langle 3, 4 \rangle = \left\langle \frac{3}{5}, \frac{4}{5} \right\rangle$ has unit length and is parallel to $\langle 3, 4 \rangle$.

Dot Product

The dot product is one way of combining ("multiplying") two vectors. The output is a scalar (a number). It is called the dot product because the symbol used is a dot. Because the dot product results in a scalar it, is also called the scalar product.

As with most things in 18.02, we have a geometric and algebraic view of dot product.

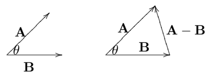
Algebraic definition (for 2D vectors):
If $\mathbf{A} = \langle a_1, a_2 \rangle$ and $\mathbf{B} = \langle b_1, b_2 \rangle$ then

$$\mathbf{A} \cdot \mathbf{B} = a_1 b_1 + a_2 b_2.$$

Example: $\langle 6, 5 \rangle \cdot \langle 1, 2 \rangle = 6 \cdot 1 + 5 \cdot 2 = 16.$

Geometric view:
The figure below shows $\mathbf{A}$, $\mathbf{B}$ with the angle θ between them. We get

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta$$



Showing the two views (algebraic and geometric) are the same requires the law of cosines

$$|\mathbf{A} - \mathbf{B}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2|\mathbf{A}||\mathbf{B}| \cos\theta$$

$\Rightarrow$ $(a_1^2 + a_2^2) + (b_1^2 + b_2^2) - ((a_1 - b_1)^2 + (a_2 - b_2)^2) = 2|\mathbf{A}||\mathbf{B}| \cos\theta$
$\Rightarrow$ $a_1 b_1 + a_2 b_2 = |\mathbf{A}||\mathbf{B}| \cos\theta.$
Since $\langle a_1, a_2 \rangle \cdot \langle b_1, b_2 \rangle = a_1 b_1 + a_2 b_2$, we have shown $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta.$

From the algebraic definition of dot product we easily get the the following algebraic law

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}.$$

Example: Find the dot product of $\mathbf{A}$ and $\mathbf{B}$.
i) $|\mathbf{A}| = 2$, $|\mathbf{B}| = 5$, $\theta = \pi/4$.
Answer: (draw the picture yourself) $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta = 10\sqrt{2}/2 = 5\sqrt{2}.$
ii) $\mathbf{A} = \mathbf{i} + 2\mathbf{j}$, $\mathbf{B} = 3\mathbf{i} + 4\mathbf{j}$.
Answer: $\mathbf{A} \cdot \mathbf{B} = 1 \cdot 3 + 2 \cdot 4 = 11.$

Three dimensional vectors
The dot product works the same in 3D as in 2D. If $\mathbf{A} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3 \rangle$ then

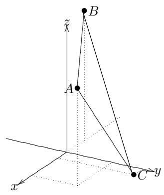
$$\mathbf{A} \cdot \mathbf{B} = a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3.$$

The geometric view is identical and the same proof shows

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta$$

Example:
Show $A = (4, 3, 6)$, $B = (-2, 0, 8)$, $C = (1, 5, 0)$
are the vertices of a right triangle.

__**Answer:**__ Two legs of the triangle are $\overrightarrow{\mathbf{AC}} = \langle -3,\, 2,\, -6\rangle$ and $\overrightarrow{\mathbf{AB}} = \langle -6,\, -3,\, 2\rangle \;\Rightarrow$ $\overrightarrow{\mathbf{AC}} \cdot \overrightarrow{\mathbf{AB}} = 18 - 6 - 12 = 0$. The geometric view of dot product implies the angle between the legs is $\pi/2$ (i.e $\cos\theta = 0$).



Definition of the term orthogonal and the test for orthogonality

When two vectors are perpendicular to each other we say they are *orthogonal*.

As seen in the example, since $\cos(\pi/2) = 0$, the dot product gives a test for orthogonality between vectors:

$$\mathbf{A} \perp \mathbf{B} \;\Leftrightarrow\; \mathbf{A} \cdot \mathbf{B} = 0.$$

Dot product and length

Both the algebraic and geometric formulas for dot product show it is intimately connected to length. In fact, they show for a vector $\mathbf{A}$

$$\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2.$$

Let's show this using both views.
Algebraically: suppose $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ then

$$\mathbf{A} \cdot \mathbf{A} = \langle a_1, a_2, a_3\rangle \cdot \langle a_1, a_2, a_3\rangle = a_1^2 + a_2^2 + a_3^2 = |\mathbf{A}|^2.$$

Geometrically: the angle θ between $\mathbf{A}$ and itself is 0. Therefore,

$$\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}||\mathbf{A}|\cos\theta = |\mathbf{A}||\mathbf{A}| = |\mathbf{A}|^2.$$

As promised both views give the formula.

Components and Projection

If $\mathbf{A}$ is any vector and $\widehat{\mathbf{u}}$ is a unit vector then the *component* of $\mathbf{A}$ in the direction of $\widehat{\mathbf{u}}$ is

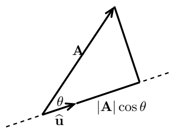
$$\mathbf{A} \cdot \widehat{\mathbf{u}}.$$

(Note: the component is a scalar.)

If θ is the angle between $\mathbf{A}$ and $\widehat{\mathbf{u}}$ then since $|\widehat{\mathbf{u}}| = 1$

$$A \cdot \widehat{\mathbf{u}} = |\mathbf{A}||\widehat{\mathbf{u}}|\cos\theta = |\mathbf{A}|\cos\theta.$$

The figure shows that geometrically this is the length of the leg of the right triangle with hypotenuse $\mathbf{A}$ and one leg parallel to $\widehat{\mathbf{u}}$.



We also call the leg parallel to $\widehat{\mathbf{u}}$ the *orthogonal projection* of $\mathbf{A}$ on $\widehat{\mathbf{u}}$.

For a non-unit vector: the component of $\mathbf{A}$ in the direction of $\mathbf{B}$ is simply the component of $\mathbf{A}$ in the direction of $\widehat{\mathbf{u}} = \frac{\mathbf{B}}{|\mathbf{B}|}$. ($\widehat{\mathbf{u}}$ is the unit vector in the same direction as $\mathbf{B}$.)

Example: Find the component of $\mathbf{A}$ in the direction of $\mathbf{B}$.

i) $|\mathbf{A}| = 2$, $|\mathbf{B}| = 5$, $\theta = \pi/4$.

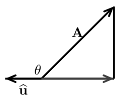
Answer: Referring to the figure above: the component is $|\mathbf{A}|\cos\theta = 2\cos(\pi/4) = \sqrt{2}$. Note, the length of $\mathbf{B}$ given is irrelevant, since we only care about the unit vector parallel to $\mathbf{B}$.

ii) $\mathbf{A} = \mathbf{i} + 2\mathbf{j}$, $\mathbf{B} = 3\mathbf{i} + 4\mathbf{j}$.

Answer: Unit vector in direction of $\mathbf{B}$ is $\frac{\mathbf{B}}{|\mathbf{B}|} = \frac{3}{5}\mathbf{i} + \frac{4}{5}\mathbf{j} \;\Rightarrow\;$ component is $\mathbf{A} \cdot \mathbf{B}/|\mathbf{B}| = 3/5 + 8/5 = 11/5$.

iii) Find the component of $\mathbf{A} = \langle 2, 2\rangle$ in the direction of $\widehat{\mathbf{u}} = \langle -1, 0\rangle$

Answer: The vector $\widehat{\mathbf{u}}$ is a unit vector, so the component is $\mathbf{A} \cdot \widehat{\mathbf{u}} = \langle 2, 2\rangle \cdot \langle -1, 0\rangle = -2$. The negative component is okay, it says the projection of $\mathbf{A}$ and $\widehat{\mathbf{u}}$ point in opposite directions.



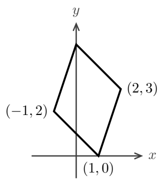
We emphasize one more time that the component of a vector is a *scalar*.

Areas and Determinants

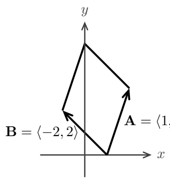
1. Compute $\begin{vmatrix} 6 & 5 \\ 1 & 2 \end{vmatrix}$.

<u>Answer:</u> $\begin{vmatrix} 6 & 5 \\ 1 & 2 \end{vmatrix} = 6 \cdot 2 - 5 \cdot 1 = 7.$

2. Compute the area of the parallelogram shown.



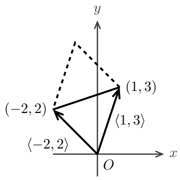
<u>Answer:</u> The area is given by the determinant of the vectors determining the parallelogram.



Area $= |\det(\mathbf{A}, \mathbf{B})| = \left|\det\begin{pmatrix} 1 & 3 \\ -2 & 2 \end{pmatrix}\right| = 2 + 6 = 8.$

3. Find the area of the triangle with vertices $(0,0)$, $(-2,2)$ and $(1,3)$.

<u>Answer:</u> The triangle is half a parallelogram. So the area is $\frac{1}{2}\left|\det\begin{pmatrix} 1 & 3 \\ -2 & 2 \end{pmatrix}\right| = 2.$



Determinants 1.

Given a square array A of numbers, we associate with it a number called the **determinant** of A, and written either $\det(A)$, or $|A|$. For 2×2

$$(1) \qquad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Do not memorize this as a formula — learn instead the pattern which gives the terms. The 2×2 case is easy: the product of the elements on one diagonal (the "main diagonal"), minus the product of the elements on the other (the "antidiagonal").

Below we will see how to compute 3×3 determinants $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$. First, try the following 2×2 example on your own, then check your work against the solution.

Example 1.1 Evaluate $\begin{vmatrix} 1 & -2 \\ -1 & 3 \end{vmatrix}$ using (1).

Solution. Using the same order as in (1), we get $12 + (-8) + 1 - 6 - 8 - (-2) = -7$.

Important facts about $|A|$:

D-1. $|A|$ is multiplied by -1 if we interchange two rows or two columns.

D-2. $|A| = 0$ if one row or column is all zero, or if two rows or two columns are the same.

D-3. $|A|$ is multiplied by c, if every element of some row or column is multiplied by c.

D-4. The value of $|A|$ is unchanged if we add to one row (or column) a constant multiple of another row (resp. column).

All of these facts are easy to check for 2×2 determinants from the formula (1); from this, their truth also for 3×3 determinants will follow from the Laplace expansion.

Though the letters $a, b, c, \ldots$ can be used for very small determinants, they can't for larger ones; it's important early on to get used to the standard notation for the entries of determinants. This is what the common software packages and the literature use. The determinants of order two and three would be written respectively

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad\qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

In general, the **ij-entry**, written a_{ij}, is the number in the i-th row and j-th column.

Its **ij-minor**, written $|A_{ij}|$, is the determinant that's left after deleting from $|A|$ the row and column containing a_{ij}.

Its **ij-cofactor**, written here A_{ij}, is given as a formula by $A_{ij} = (-1)^{i+j}|A_{ij}|$. For a 3×3 determinant, it is easier to think of it this way: we put $+$ or $-$ in front of the ij-minor, according to whether $+$ or $-$ occurs in the ij-position in the checkerboard pattern

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix} . \tag{2}$$

Example 1.2 $|A| = \begin{vmatrix} 1 & 0 & 3 \\ 1 & 2 & -1 \\ 2 & 1 & -1 \end{vmatrix}$. Find $|A_{12}|,\ A_{12},\ |A_{22}|,\ A_{22}$.

Solution. $|A_{12}| = \begin{vmatrix} 1 & -1 \\ 2 & -1 \end{vmatrix} = 1, \quad A_{12} = -1.$ $\qquad |A_{22}| = \begin{vmatrix} 1 & 3 \\ 2 & -1 \end{vmatrix} = -7, \quad A_{22} = -7.$

Laplace expansion by cofactors

This is another way to evaluate a determinant; we give the rule for a 3×3. It generalizes easily to an $n \times n$ determinant.

Select any row (or column) of the determinant. Multiply each entry a_{ij} in that row (or column) by its cofactor A_{ij}, and add the three resulting numbers; you get the value of the determinant.

As practice with notation, here is the formula for the Laplace expansion of a third order (i.e., a 3×3) determinant using the cofactors of the first row:

$$a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \ = \ |A| \tag{3}$$

and the formula using the cofactors of the j-th column:

$$a_{1j}A_{1j} + a_{2j}A_{2j} + a_{3j}A_{3j} \ = \ |A| \tag{4}$$

Example 1.3 Evaluate the determinant in Example 1.2 using the Laplace expansions by the first row and by the second column, and check by also using (1).

Solution. The Laplace expansion by the first row is

$$\begin{vmatrix} 1 & 0 & 3 \\ 1 & 2 & -1 \\ 2 & 1 & -1 \end{vmatrix} \ = \ 1\cdot\begin{vmatrix} 2 & -1 \\ 1 & -1 \end{vmatrix} - 0\cdot\begin{vmatrix} 1 & -1 \\ 2 & -1 \end{vmatrix} + 3\cdot\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = 1\cdot(-1) - 0\cdot 1 + 3\cdot(-3) = -10.$$

The Laplace expansion by the second column would be

$$\begin{vmatrix} 1 & 0 & 3 \\ 1 & 2 & -1 \\ 2 & 1 & -1 \end{vmatrix} \ = \ -0\cdot\begin{vmatrix} 1 & -1 \\ 2 & -1 \end{vmatrix} + 2\cdot\begin{vmatrix} 1 & 3 \\ 2 & -1 \end{vmatrix} - 1\cdot\begin{vmatrix} 1 & 3 \\ 1 & -1 \end{vmatrix} = 0 + 2\cdot(-7) - 1\cdot(-4) = -10.$$

Checking by (1), we have $|A| \ = \ -2 + 0 + 3 - 12 - 0 - (-1) \ = \ -10.$

Example 1.4 Show the Laplace expansion by the first row gives the following formula (which you may have seen before).

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + chd - gec - hfa - ibd$$

Solution. We have

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \cdot \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \cdot \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \cdot \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg), \end{aligned}$$

whose six terms agree with the six terms on the right of the formula above.

(A similar argument can be made for the Laplace expansion by any row or column.)

For $n \times n$ determinants, the **minor** $|A_{ij}|$ of the entry a_{ij} is defined to be the determinant obtained by deleting the i-th row and j-th column; the **cofactor** A_{ij} is the minor, prefixed by a + or − sign according to the natural generalization of the checkerboard pattern (2). Then the Laplace expansion by the i-th row would be

$$|A| \;=\; a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}\ .$$

This is an inductive calculation — it expresses the determinant of order n in terms of determinants of order $n-1$. Thus, since we can calculate determinants of order 3, it allows us to calculate determinants of order 4; then determinants of order 5, and so on. If we take for definiteness $i = 1$, then the above Laplace expansion formula can be used as the basis of an inductive definition of the $n \times n$ determinant.

Example 1.5 Evaluate $\begin{vmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 1 & 4 \\ -1 & 4 & 1 & 0 \\ 0 & 4 & 2 & -1 \end{vmatrix}$ by its Laplace expansion by the first row.

Solution. $1 \cdot \begin{vmatrix} -1 & 1 & 4 \\ 4 & 1 & 0 \\ 4 & 2 & -1 \end{vmatrix} \;-\; 0 \cdot A_{12} \;+\; 2 \cdot \begin{vmatrix} 2 & -1 & 4 \\ -1 & 4 & 0 \\ 0 & 4 & -1 \end{vmatrix} \;-\; 3 \cdot \begin{vmatrix} 2 & -1 & 1 \\ -1 & 4 & 1 \\ 0 & 4 & 2 \end{vmatrix}$

$$= \; 1 \cdot 21 \;+\; 2 \cdot (-23) \;-\; 3 \cdot 2 \;=\; -31.$$

Cross Product

The cross product is another way of multiplying two vectors. (The name comes from the symbol used to indicate the product.) Because the result of this multiplication is *another vector* it is also called the *vector product*.

As usual, there is an algebraic and a geometric way to describe the cross product. We'll define it algebraically and then move to the geometric description.

Determinant definition for cross product

For the vectors $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$ we define the cross product by the following formula

$$\begin{aligned}
\mathbf{A}\times\mathbf{B} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\
&= \mathbf{i}\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \\
&= (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \\
&= \langle a_2b_3 - a_3b_2,\, a_3b_1 - a_1b_3,\, a_1b_2 - a_2b_1\rangle.
\end{aligned}$$

The bottom three equations above are easily seen to be equivalent and should be taken as the definition of the cross product. The top line is technically flawed because we are not really allowed to use vectors as entries in a determinant. Nonetheless is is an excellent way to remember how to compute the cross product.

Example: $\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 3 & 0 \\ 3 & -2 & 0 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 3 & -2 \end{vmatrix}\mathbf{k} = -13\,\mathbf{k}$

Example: Compute $\mathbf{i}\times\mathbf{j}$.

<u>Answer:</u> $\mathbf{i} = \langle 1, 0, 0\rangle$ and $\mathbf{j} = \langle 0, 1, 0\rangle$ therefore

$$\mathbf{i}\times\mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{i}(0) - \mathbf{j}(0) + \mathbf{k}(1) = \mathbf{k}.$$

Algebraic facts: (these follow easily from properties of determinant).

1. $\mathbf{A}\times\mathbf{A} = \mathbf{0}$
2. Anti-commutivity: $\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A}$
3. Distributive law: $\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C}$
4. Non-associativity: $(\mathbf{A}\times\mathbf{B})\times\mathbf{C} \neq \mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ (example in a moment).

For the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ we have

$$\mathbf{i}\times\mathbf{j} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}.$$

Example: (non-associativity) $(\mathbf{i}\times\mathbf{j})\times\mathbf{j} = -\mathbf{i}$ but $\mathbf{i}\times(\mathbf{j}\times\mathbf{j}) = 0$.

Example: It is possible to compute a cross product using the algebraic facts and the known products of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. For example,

$$(2\mathbf{i}+3\mathbf{j})\times(3\mathbf{i}-2\mathbf{j}) = (6\mathbf{i}\times\mathbf{i}) - (4\mathbf{i}\times\mathbf{j}) + (9\mathbf{j}\times\mathbf{i}) - (6\mathbf{j}\times\mathbf{j}) = -13\mathbf{k}.$$

The first equation follows from the distributive law. In the second, we used $\mathbf{i}\times\mathbf{i} = \mathbf{j}\times\mathbf{j} = 0$ (algebraic fact 1), $\mathbf{i}\times\mathbf{j} = \mathbf{k}$ (computed above) and $\mathbf{j}\times\mathbf{i} = -\mathbf{k}$ (anti-commutivity).

Geometric description

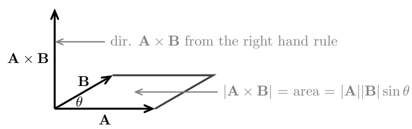
To describe the cross product geometrically we need to describe its magnitude and direction. This is done in the following theorem.

Theorem: The magnitude of $\mathbf{A}\times\mathbf{B}$ is

$$\begin{aligned} |\mathbf{A}\times\mathbf{B}| &= |\mathbf{A}||\mathbf{B}|\sin\theta, \text{ where } \theta \text{ is the angle between them} \\ &= \text{area of the parallelogram spanned by } \mathbf{A} \text{ and } \mathbf{B}. \end{aligned}$$

The direction of $\mathbf{A}\times\mathbf{B}$ is determined as follows.

$\mathbf{A}\times\mathbf{B}$ is perpendicular to the plane of $\mathbf{A}$ and $\mathbf{B}$. In the figure below there are two directions perpendicular to the plane –up and down. The choice is made by the *right hand rule.* This rule says to take your right hand and point your fingers in the direction of $\mathbf{A}$ so that they curl towards $\mathbf{B}$; then your thumb points in the direction of $\mathbf{A}\times\mathbf{B}$.

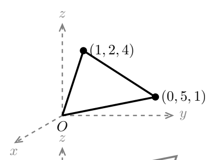


We will not go through the proof of this theorem. It makes use of the Lagrange identity

$$|\mathbf{A}\times\mathbf{B}|^2 = |\mathbf{A}|^2|\mathbf{B}|^2 - (\mathbf{A}\cdot\mathbf{B})^2.$$

This identity is easily show by expanding both sides using components.

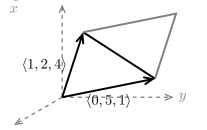
Example: Find the area of the triangle shown.



<u>Answer:</u>

The area of the triangle is half the area of the parallelogram (see figure).

So, area triangle $= \frac{1}{2}|\langle 1,2,4\rangle \times \langle 0,5,1\rangle|$.



$$\langle 1,2,4\rangle \times \langle 0,5,1\rangle = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 4 \\ 0 & 5 & 1 \end{vmatrix} = \mathbf{i}(-18) - \mathbf{j} + 5\mathbf{k}.$$

Area triangle $= \frac{1}{2}\sqrt{18^2+1^2+5^2} = \frac{1}{2}\sqrt{350}.$

DON'T FORGET THE GEOMETRY -it will be used to solve problems.

Equation of a Plane

1. Later we will return to the topic of planes in more detail. Here we will content ourself with one example.

Find the equation of the plane containing the three points $P_1 = (1,3,1)$, $\quad P_2 = (1,2,2)$, $P_3 = (2,3,3)$.

Answer:

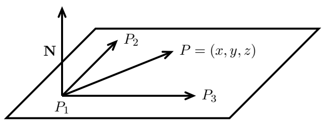
The vectors $\overrightarrow{\mathbf{P_1P_2}}$ and $\overrightarrow{\mathbf{P_1P_3}}$ are in the plane, so

$$\mathbf{N} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & -1 & 1 \\ 1 & 0 & 2 \end{vmatrix} = \mathbf{i}(-2) - \mathbf{j}(-1) + \mathbf{k}(1) = \langle -2, 1, 1 \rangle.$$

is orthogonal to the plane.

Now for any point $P = (x,y,z)$ in the plane, the vector $\overrightarrow{\mathbf{P_1P}}$ is also in the plane and is therefore orthogonal to $\mathbf{N}$. Expressing this with the dot product we get

$$\begin{aligned} & \mathbf{N} \cdot \overrightarrow{\mathbf{P_1P}} = 0 \\ \Leftrightarrow \quad & \langle -2,1,1 \rangle \cdot \langle x-1, y-3, z-1 \rangle = 0 \\ \Leftrightarrow \quad & -2(x-1) + (y-3) + (z-1) = 0 \\ \Leftrightarrow \quad & -2x + y + z = 2. \end{aligned}$$



The equation of the plane is $-2x + y + z = 2$. You should check that the three points P_1, P_2, P_3 do, in fact, satisfy this equation.

The standard terminology for the vector $\mathbf{N}$ is to call it a *normal* to the plane.

Cross Product

The cross product is another way of multiplying two vectors. (The name comes from the symbol used to indicate the product.) Because the result of this multiplication is *another vector* it is also called the *vector product*.

As usual, there is an algebraic and a geometric way to describe the cross product. We'll define it algebraically and then move to the geometric description.

Determinant definition for cross product

For the vectors $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$ we define the cross product by the following formula

$$\begin{aligned}
\mathbf{A}\times\mathbf{B} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\
&= \mathbf{i}\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \\
&= (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \\
&= \langle a_2b_3 - a_3b_2,\ a_3b_1 - a_1b_3,\ a_1b_2 - a_2b_1\rangle.
\end{aligned}$$

The bottom three equations above are easily seen to be equivalent and should be taken as the definition of the cross product. The top line is technically flawed because we are not really allowed to use vectors as entries in a determinant. Nonetheless is is an excellent way to remember how to compute the cross product.

Example: $\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 3 & 0 \\ 3 & -2 & 0 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 3 & -2 \end{vmatrix}\mathbf{k} = -13\,\mathbf{k}$

Example: Compute $\mathbf{i}\times\mathbf{j}$.

<u>Answer:</u> $\mathbf{i} = \langle 1, 0, 0\rangle$ and $\mathbf{j} = \langle 0, 1, 0\rangle$ therefore

$$\mathbf{i}\times\mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{i}(0) - \mathbf{j}(0) + \mathbf{k}(1) = \mathbf{k}.$$

Algebraic facts: (these follow easily from properties of determinant).

1. $\mathbf{A}\times\mathbf{A} = \mathbf{0}$
2. Anti-commutivity: $\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A}$
3. Distributive law: $\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C}$
4. Non-associativity: $(\mathbf{A}\times\mathbf{B})\times\mathbf{C} \neq \mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ (example in a moment).

For the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ we have

$$\mathbf{i}\times\mathbf{j} = \mathbf{k},\quad \mathbf{j}\times\mathbf{k} = \mathbf{i},\quad \mathbf{k}\times\mathbf{i} = \mathbf{j}.$$

Example: (non-associativity) $(\mathbf{i}\times\mathbf{j})\times\mathbf{j} = -\mathbf{i}$ but $\mathbf{i}\times(\mathbf{j}\times\mathbf{j}) = 0$.

Example: It is possible to compute a cross product using the algebraic facts and the known products of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. For example,

$$(2\mathbf{i}+3\mathbf{j})\times(3\mathbf{i}-2\mathbf{j}) = (6\mathbf{i}\times\mathbf{i}) - (4\mathbf{i}\times\mathbf{j}) + (9\mathbf{j}\times\mathbf{i}) - (6\mathbf{j}\times\mathbf{j}) = -13\mathbf{k}.$$

The first equation follows from the distributive law. In the second, we used $\mathbf{i}\times\mathbf{i} = \mathbf{j}\times\mathbf{j} = 0$ (algebraic fact 1), $\mathbf{i}\times\mathbf{j} = \mathbf{k}$ (computed above) and $\mathbf{j}\times\mathbf{i} = -\mathbf{k}$ (anti-commutivity).

Geometric description

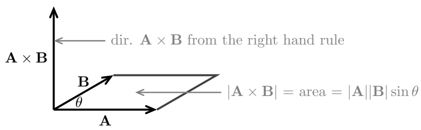
To describe the cross product geometrically we need to describe its magnitude and direction. This is done in the following theorem.

Theorem: The magnitude of $\mathbf{A}\times\mathbf{B}$ is

$$\begin{aligned} |\mathbf{A}\times\mathbf{B}| &= |\mathbf{A}||\mathbf{B}|\sin\theta, \text{ where } \theta \text{ is the angle between them} \\ &= \text{area of the parallelogram spanned by } \mathbf{A} \text{ and } \mathbf{B}. \end{aligned}$$

The direction of $\mathbf{A}\times\mathbf{B}$ is determined as follows.

$\mathbf{A}\times\mathbf{B}$ is perpendicular to the plane of $\mathbf{A}$ and $\mathbf{B}$. In the figure below there are two directions perpendicular to the plane –up and down. The choice is made by the *right hand rule.* This rule says to take your right hand and point your fingers in the direction of $\mathbf{A}$ so that they curl towards $\mathbf{B}$; then your thumb points in the direction of $\mathbf{A}\times\mathbf{B}$.

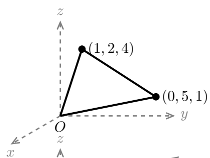


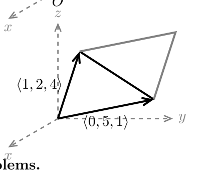
We will not go through the proof of this theorem. It makes use of the Lagrange identity

$$|\mathbf{A}\times\mathbf{B}|^2 = |\mathbf{A}|^2|\mathbf{B}|^2 - (\mathbf{A}\cdot\mathbf{B})^2.$$

This identity is easily show by expanding both sides using components.

Example: Find the area of the triangle shown.





Answer:

The area of the triangle is half the area of the parallelogram (see figure).

So, area triangle $= \dfrac{1}{2}|\langle 1,2,4\rangle \times \langle 0,5,1\rangle|$.

$$\langle 1,2,4\rangle \times \langle 0,5,1\rangle = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 4 \\ 0 & 5 & 1 \end{vmatrix} = \mathbf{i}(-18) - \mathbf{j} + 5\mathbf{k}.$$

Area triangle $= \dfrac{1}{2}\sqrt{18^2+1^2+5^2} = \dfrac{1}{2}\sqrt{350}$.

DON'T FORGET THE GEOMETRY -it will be used to solve problems.

Equation of a Plane

1. Later we will return to the topic of planes in more detail. Here we will content ourself with one example.

Find the equation of the plane containing the three points $P_1 = (1,3,1)$, $P_2 = (1,2,2)$, $P_3 = (2,3,3)$.

Answer:

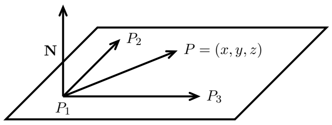
The vectors $\overrightarrow{\mathbf{P_1P_2}}$ and $\overrightarrow{\mathbf{P_1P_3}}$ are in the plane, so

$$\mathbf{N} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & -1 & 1 \\ 1 & 0 & 2 \end{vmatrix} = \mathbf{i}(-2) - \mathbf{j}(-1) + \mathbf{k}(1) = \langle -2, 1, 1 \rangle.$$

is orthogonal to the plane.

Now for any point $P = (x, y, z)$ in the plane, the vector $\overrightarrow{\mathbf{P_1P}}$ is also in the plane and is therefore orthogonal to $\mathbf{N}$. Expressing this with the dot product we get

$$\begin{aligned} & \mathbf{N} \cdot \overrightarrow{\mathbf{P_1P}} = 0 \\ \Leftrightarrow\quad & \langle -2, 1, 1 \rangle \cdot \langle x-1, y-3, z-1 \rangle = 0 \\ \Leftrightarrow\quad & -2(x-1) + (y-3) + (z-1) = 0 \\ \Leftrightarrow\quad & -2x + y + z = 2. \end{aligned}$$



The equation of the plane is $-2x + y + z = 2$. You should check that the three points P_1, P_2, P_3 do, in fact, satisfy this equation.

The standard terminology for the vector $\mathbf{N}$ is to call it a *normal* to the plane.

Matrices 1. Matrix Algebra

Matrix algebra.

Previously we calculated the determinants of square arrays of numbers. Such arrays are important in mathematics and its applications; they are called *matrices*. In general, they need not be square, only rectangular.

A rectangular array of numbers having m rows and n columns is called an $m \times n$ **matrix**. The number in the i-th row and j-th column (where $1 \leq i \leq m,\ 1 \leq j \leq n$) is called the **ij-entry**, and denoted a_{ij}; the matrix itself is denoted by A, or sometimes by (a_{ij}).

Two matrices of the same size are *equal* if corresponding entries are equal.

Two special kinds of matrices are the **row-vectors**: the $1 \times n$ matrices $(a_1, a_2, \ldots, a_n)$; and the **column vectors**: the $m \times 1$ matrices consisting of a column of m numbers.

> From now on, row-vectors or column-vectors will be indicated by boldface small letters; when writing them by hand, put an arrow over the symbol.

Matrix operations

There are four basic operations which produce new matrices from old.

1. Scalar multiplication: Multiply each entry by c : $cA = (ca_{ij})$

2. Matrix addition: Add the corresponding entries: $A + B = (a_{ij} + b_{ij})$; the two matrices must have the same number of rows and the same number of columns.

3. Transposition: The *transpose* of the $m \times n$ matrix A is the $n \times m$ matrix obtained by making the rows of A the columns of the new matrix. Common notations for the transpose are A^T and A'; using the first we can write its definition as $A^T = (a_{ji})$.

If the matrix A is square, you can think of A^T as the matrix obtained by flipping A over around its main diagonal.

Example 1.1 Let $A = \begin{pmatrix} 2 & -3 \\ 0 & 1 \\ -1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 5 \\ -2 & 3 \\ -1 & 0 \end{pmatrix}$. Find $A+B$, A^T, $2A-3B$.

Solution. $A + B = \begin{pmatrix} 3 & 2 \\ -2 & 4 \\ -2 & 2 \end{pmatrix}$; $A^T = \begin{pmatrix} 2 & 0 & -1 \\ -3 & 1 & 2 \end{pmatrix}$;

$$2A + (-3B) = \begin{pmatrix} 4 & -6 \\ 0 & 2 \\ -2 & 4 \end{pmatrix} + \begin{pmatrix} -3 & -15 \\ 6 & -9 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -21 \\ 6 & -7 \\ 1 & 4 \end{pmatrix}.$$

4. Matrix multiplication This is the most important operation. Schematically, we have

$$\begin{array}{ccccc} A & \cdot & B & = & C \\ m \times n & & n \times p & & m \times p \\ & & c_{ij} & = & \displaystyle\sum_{k=1}^{n} a_{ik} b_{kj} \end{array}$$

The essential points are:

1. For the multiplication to be defined, A must have as many *columns* as B has *rows*;

2. The ij-th entry of the product matrix C is the dot product of the i-th row of A with the j-th column of B.

Example 1.2 $\quad \begin{pmatrix} 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} -1 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2+4-2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix};$

$$\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \begin{pmatrix} 4 & 5 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 8 & 10 \\ -4 & -5 \end{pmatrix}; \qquad \begin{pmatrix} 2 & 0 & 1 \\ 1 & -1 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & -2 & -6 \\ 0 & 2 & 2 \end{pmatrix}$$

The two most important types of multiplication, for multivariable calculus and differential equations, are:

1. AB, where A and B are two *square* matrices of the same size — these can always be multiplied;
2. $A\mathbf{b}$, where A is a square $n \times n$ matrix, and $\mathbf{b}$ is a column n-vector.

Laws and properties of matrix multiplication

M-1. $A(B+C) = AB + AC, \qquad (A+B)C = AC + BC$ *distributive laws*

M-2. $(AB)C = A(BC); \qquad (cA)B = c(AB).$ *associative laws*

In both cases, the matrices must have compatible dimensions.

M-3. Let $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; then $AI = A$ and $IA = A$ for any 3×3 matrix.

I is called the **identity** matrix of order 3. There is an analogously defined square identity matrix I_n of any order n, obeying the same multiplication laws.

M-4. In general, for two square $n \times n$ matrices A and B, $AB \neq BA$: *matrix multiplication is not commutative.* (There are a few important exceptions, but they are very special — for example, the equality $AI = IA$ where I is the identity matrix.)

M-5. For two square $n \times n$ matrices A and B, we have the *determinant law*:

$$|AB| = |A||B|, \qquad \text{also written} \qquad \det(AB) = \det(A)\det(B)$$

For 2×2 matrices, this can be verified by direct calculation, but this naive method is unsuitable for larger matrices; it's better to use some theory. We will simply assume it in these notes; we will also assume the other results above (of which only the associative law **M-2** offers any difficulty in the proof).

M-6. A useful fact is this: matrix multiplication can be used to pick out a row or column of a given matrix: you multiply by a simple row or column vector to do this. Two examples should give the idea:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} \qquad \text{the second column}$$

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \qquad \text{the first row}$$

Meaning of matrix multiplication

In these examples we will explore the effect of matrix multiplication on the xy-plane.

Example 1: The matrix $A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}$ transforms the unit square into a parallelogram as follows.

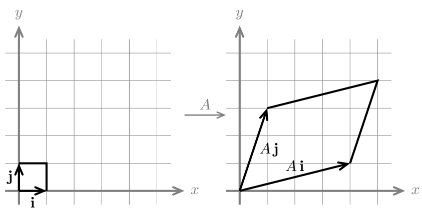
The unit square has sides $\mathbf{i}$ and $\mathbf{j}$. In order multiply a matrix times a vector we write them as column vectors. For example, $\mathbf{i} = \langle 1, 0 \rangle$, $\mathbf{j} = \langle 0, 1 \rangle$ and $\mathbf{v} = \langle a_1, a_2 \rangle$ are written

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

The matrix multiplication then becomes

$$A\,\mathbf{i} = A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}; \quad A\,\mathbf{j} = A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

We think of the all the points in the square as the endpoints of origin vectors. If we multiply A by all of these vectors we get the following picture.



The square is mapped to the parallelogram. We know that the area of the parallelogram is $|A| = 11$. (Think about the 2×2 determinant you would use to compute the area of the parallelogram.)

Matrices 2. Solving Square Systems of Linear Equations; Inverse Matrices

Solving square systems of linear equations; inverse matrices.

Linear algebra is essentially about solving systems of linear equations, an important application of mathematics to real-world problems in engineering, business, and science, especially the social sciences. Here we will just stick to the most important case, where the system is *square*, i.e., there are as many variables as there are equations. In low dimensions such systems look as follows (we give a 2×2 system and a 3×3 system):

$$
\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}
\qquad
\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}
\tag{7}
$$

In these systems, the a_{ij} and b_i are given, and we want to solve for the x_i.

As a simple mathematical example, consider the linear change of coordinates given by the equations

$$
\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 \\ x_2 &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 \\ x_3 &= a_{31}y_1 + a_{32}y_2 + a_{33}y_3 \end{aligned}
$$

If we know the y-coordinates of a point, then these equations tell us its x-coordinates immediately. But if instead we are given the x-coordinates, to find the y-coordinates we must solve a system of equations like (7) above, with the y_i as the unknowns.

Using matrix multiplication, we can abbreviate the system on the right in (7) by

$$
A\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},
\tag{8}
$$

where A is the square matrix of coefficients (a_{ij}). (The 2×2 system and the $n \times n$ system would be written analogously; all of them are abbreviated by the same equation $A\mathbf{x} = \mathbf{b}$, notice.)

You have had experience with solving small systems like (7) by *elimination*: multiplying the equations by constants and subtracting them from each other, the purpose being to eliminate all the variables but one. When elimination is done systematically, it is an efficient method. Here however we want to talk about another method more compatible with hand-held calculators and MatLab, and which leads more rapidly to certain key ideas and results in linear algebra.

Inverse matrices.

Referring to the system (8), suppose we can find a square matrix M, the same size as A, such that

$$
MA = I \qquad \text{(the identity matrix).}
\tag{9}
$$

We can then solve (8) by matrix multiplication, using the successive steps,

$$\begin{aligned} A\,\mathbf{x} &= \mathbf{b} \\ M(A\,\mathbf{x}) &= M\,\mathbf{b} \\ \mathbf{x} &= M\,\mathbf{b}; \end{aligned} \tag{10}$$

where the step $M(A\mathbf{x}) = \mathbf{x}$ is justified by

$$\begin{aligned} M(A\,\mathbf{x}) &= (MA)\mathbf{x}, && \text{by associative law;} \\ &= I\,\mathbf{x}, && \text{by (9);} \\ &= \mathbf{x}, && \text{because } I \text{ is the identity matrix .} \end{aligned}$$

Moreover, the solution is unique, since (10) gives an explicit formula for it.

The same procedure solves the problem of determining the inverse to the linear change of coordinates $\mathbf{x} = A\mathbf{y}$, as the next example illustrates.

Example 2.1 Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ and $M = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$. Verify that M satisfies (9) above, and use it to solve the first system below for x_i and the second for the y_i in terms of the x_i:

$$\begin{aligned} x_1 + 2x_2 &= -1 \qquad & x_1 &= \ y_1 + 2y_2 \\ 2x_1 + 3x_2 &= 4 & x_2 &= 2y_1 + 3y_2 \end{aligned}$$

Solution. We have $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, by matrix multiplication. To solve the first system, we have by (10), $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} -1 \\ 4 \end{pmatrix} = \begin{pmatrix} 11 \\ -6 \end{pmatrix}$, so the solution is $x_1 = 11, x_2 = -6$. By reasoning similar to that used above in going from $A\mathbf{x} = \mathbf{b}$ to $\mathbf{x} = M\mathbf{b}$, the solution to $\mathbf{x} = A\mathbf{y}$ is $\mathbf{y} = M\mathbf{x}$, so that we get

$$\begin{aligned} y_1 &= -3x_1 + 2x_2 \\ y_2 &= \ \ 2x_1 - x_2 \end{aligned}$$

as the expression for the y_i in terms of the x_i.

Our problem now is: how do we get the matrix M? In practice, you mostly press a key on the calculator, or type a Matlab command. But we need to be able to work abstractly with the matrix — i.e., with symbols, not just numbers, and for this some theoretical ideas are important. The first is that M doesn't always exist.

$$M \text{ exists} \quad \Leftrightarrow \quad |A| \neq 0.$$

The implication $\Rightarrow$ follows immediately from the law **M-5** in section M.1 ($\det(AB) = \det(A)\det(B)$) , since

$$MA = I \quad \Rightarrow \quad |M||A| = |I| = 1 \quad \Rightarrow \quad |A| \neq 0.$$

The implication in the other direction requires more; for the low-dimensional cases, we will produce a formula for M. Let's go to the formal definition first, and give M its proper name, A^{-1}:

Definition. Let A be an $n \times n$ matrix, with $|A| \neq 0$. Then the **inverse** of A is an $n \times n$ matrix, written A^{-1}, such that

$$A^{-1}A = I_n, \qquad AA^{-1} = I_n \tag{11}$$

(It is actually enough to verify either equation; the other follows automatically — see the exercises.)

Using the above notation, our previous reasoning (9) - (10) shows that

$$|A| \neq 0 \quad \Rightarrow \quad \text{the unique solution of } A\,\mathbf{x} = \mathbf{b} \text{ is } \mathbf{x} = A^{-1}\mathbf{b}; \tag{12}$$

$$|A| \neq 0 \quad \Rightarrow \quad \text{the solution of } \mathbf{x} = A\,\mathbf{y} \text{ for the } y_i \text{ is } \mathbf{y} = A^{-1}\mathbf{x}. \tag{12}$$

Calculating the inverse of a 3×3 matrix

Let A be the matrix. The formulas for its **inverse** A^{-1} and for an auxiliary matrix $\operatorname{adj} A$ called the **adjoint** of A (or in some books the **adjugate** of A) are

$$A^{-1} = \frac{1}{|A|} \operatorname{adj} A = \frac{1}{|A|} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}^T . \tag{13}$$

In the formula, A_{ij} is the cofactor of the element a_{ij} in the matrix, i.e., its minor with its sign changed by the checkerboard rule (see section 1 on determinants).

Formula (13) shows that the steps in calculating the inverse matrix are:

1. Calculate the matrix of minors.
2. Change the signs of the entries according to the checkerboard rule.
3. Transpose the resulting matrix; this gives $\operatorname{adj} A$.
4. Divide every entry by $|A|$.

(If inconvenient, for example if it would produce a matrix having fractions for every entry, you can just leave the $1/|A|$ factor outside, as in the formula. Note that step 4 can only be taken if $|A| \neq 0$, so if you haven't checked this before, you'll be reminded of it now.)

The notation A_{ij} for a cofactor makes it look like a matrix, rather than a signed determinant; this isn't good, but we can live with it.

Example 2.2 Find the inverse to $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$.

Solution. We calculate that $|A| = 2$. Then the steps are (T means transpose):

$$\underset{\text{matrix } A}{\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}} \to \underset{\text{cofactor matrix}}{\begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix}} \quad \underset{\text{T}}{\to} \quad \underset{\text{adj } A}{\begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & -1 \\ -1 & 0 & 1 \end{pmatrix}} \to \underset{\text{inverse of } A}{\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}}$$

To get practice in matrix multiplication, check that $A \cdot A^{-1} = I$, or to avoid the fractions, check that $A \cdot \text{adj } (A) = 2I$.

The same procedure works for calculating the inverse of a 2×2 matrix A. We do it for a general matrix, since it will save you time in differential equations if you can learn the resulting formula.

$$\underset{\text{matrix } A}{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \quad \to \quad \underset{\text{cofactors}}{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}} \quad \underset{\text{T}}{\to} \quad \underset{\text{adj } A}{\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}} \quad \to \quad \underset{\text{inverse of } A}{\frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$$

Example 2.3 Find the inverses to: a) $\begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}$ b) $\begin{pmatrix} 1 & 2 & 2 \\ 2 & -1 & 1 \\ 1 & 3 & 2 \end{pmatrix}$

Solution. a) Use the formula: $|A| = 2$, so $A^{-1} = \dfrac{1}{2}\begin{pmatrix} 2 & 0 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{3}{2} & \frac{1}{2} \end{pmatrix}$.

b) Follow the previous scheme:

$$\begin{pmatrix} 1 & 2 & 2 \\ 2 & -1 & 1 \\ 1 & 3 & 2 \end{pmatrix} \to \begin{pmatrix} -5 & -3 & 7 \\ 2 & 0 & -1 \\ 4 & 3 & -5 \end{pmatrix} \to \begin{pmatrix} -5 & 2 & 4 \\ -3 & 0 & 3 \\ 7 & -1 & -5 \end{pmatrix} \to \frac{1}{3}\begin{pmatrix} -5 & 2 & 4 \\ -3 & 0 & 3 \\ 7 & -1 & -5 \end{pmatrix} = A^{-1}.$$

Both solutions should be checked by multiplying the answer by the respective A.

Proof of formula (13) for the inverse matrix.

We want to show $A \cdot A^{-1} = I$, or equivalently, $A \cdot \text{adj } A = |A|I$; when this last is written out using (13) (remembering to transpose the matrix on the right there), it becomes

$$(14) \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} |A| & 0 & 0 \\ 0 & |A| & 0 \\ 0 & 0 & |A| \end{pmatrix}.$$

To prove (14), it will be enough to look at two typical entries in the matrix on the right — say the first two in the top row. According to the rule for multiplying the two matrices on the left, what we have to show is that

$$(15) \qquad a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} = |A|;$$

$$(16) \qquad a_{11}A_{21} + a_{12}A_{22} + a_{13}A_{23} = 0$$

These two equations are both evaluating determinants by Laplace expansions: the first equation (15) evaluates the determinant on the left below by the cofactors of the first row; the second equation (16) evaluates the determinant on the right below by the cofactors of the second row (notice that the cofactors of the second row don't care what's actually in the second row, since to calculate them you only need to know the other two rows).

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \qquad\qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

The two equations (15) and (16) now follow, since the determinant on the left is just $|A|$, while the determinant on the right is 0, since two of its rows are the same. $\square$

The procedure we have given for calculating an inverse works for $n \times n$ matrices, but gets to be too cumbersome if $n > 3$, and other methods are used. The calculation of A^{-1} for reasonable-sized A is a standard package in computer algebra programs and MatLab. Unfortunately, social scientists often want the inverses of very large matrices, and for this special techniques have had to be devised, which produce approximate but acceptable results.

Equations of planes

We have touched on equations of planes previously. Here we will fill in some of the details.

Planes in point-normal form

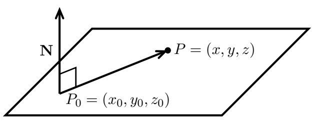
The basic data which determines a plane is a point P_0 in the plane and a vector $\mathbf{N}$ orthogonal to the plane. We call $\mathbf{N}$ a *normal* to the plane and we will sometimes say $\mathbf{N}$ is *normal* to the plane, instead of orthogonal.

Now, suppose we want the equation of a plane and we have a point $P_0 = (x_0, y_0, z_0)$ in the plane and a vector $\overrightarrow{\mathbf{N}} = \langle a, b, c\rangle$ normal to the plane.

Let $P = (x, y, z)$ be an arbitrary point in the plane. Then the vector $\overrightarrow{\mathbf{P_0P}}$ is in the plane and therefore orthogonal to $\mathbf{N}$. This means

$$
\begin{aligned}
& \mathbf{N} \cdot \overrightarrow{\mathbf{P_0P}} = 0 \\
\Leftrightarrow \quad & \langle a, b, c\rangle \cdot \langle x - x_0, y - y_0, z - z_0\rangle = 0 \\
\Leftrightarrow \quad & a(x - x_0) + b(y - y_0) + c(z - z_0) = 0
\end{aligned}
$$

We call this last equation the point-normal form for the plane.



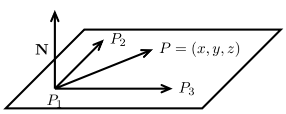
Example 1: Find the plane through the point (1,4,9) with normal $\langle 2, 3, 4\rangle$.

<u>Answer:</u> Point-normal form of the plane is $2(x-1)+3(y-4)+4(z-9) = 0$. We can also write this as $2x + 3y + 4z = 50$.

Example 2: Find the plane containing the points $P_1 = (1, 2, 3)$, $P_2 = (0, 0, 3)$, $P_3 = (2, 5, 5)$.

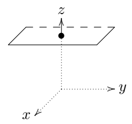
<u>Answer:</u> The goal is to find the basic data, i.e. a point in the plane and a normal to the plane. The point is easy, we already have three of them. To get the normal we note (see figure below) that $\overrightarrow{\mathbf{P_1P_2}}$ and $\overrightarrow{\mathbf{P_1P_3}}$ are vectors in the plane, so their cross product is orthogonal (normal) to the plane. That is,

$$\mathbf{N} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & -2 & 0 \\ 1 & 3 & 2 \end{pmatrix} = -4\mathbf{i} - \mathbf{j}(-2) + \mathbf{k}(-1) = \langle -4, 2, -1\rangle.$$



Using point-normal form (with point P_1) the equation of the plane is

$$-4(x-1)+2(y-2)-(z-3)=0, \text{ or equivalently } \quad -4x+2y-z=-3.$$



Example 3: Find the plane with normal $\mathbf{N}=\widehat{\mathbf{k}}$ containing the point (0,0,3)
Eq. of plane: $\langle 0,0,1\rangle \cdot \langle x,y,z-3\rangle = 0 \Leftrightarrow z=3$.

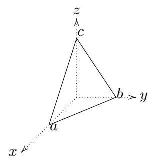
Example 4: Find the plane with x, y and z intercepts a, b and c.

<ins>**Answer:**</ins> We could find this using the method example 1. Instead, we'll use a shortcut that works when all the intercepts are known. In this case, the intercepts are

$$(a,0,0), \quad (0,b,0), \quad (0,0,c)$$

and we simply write the plane as

$$x/a+y/b+z/c=1.$$



You can easily check that each of the given points is on the plane.
For completeness we'll do this using the more general method of example 1.
The 3 points give us 2 vectors in the plane, $\langle -a,b,0\rangle$ and $\langle -a,0,c\rangle$.
$\Rightarrow$ $\mathbf{N}=\langle -a,b,0\rangle \times \langle -a,0,c\rangle = \langle bc,ac,ab\rangle$.
Point-normal form: $bc(x-a)+ac(y-0)+ab(z-0)=0$
$\Leftrightarrow$ $bc\,x+ac\,y+ab\,z=abc \Leftrightarrow x/a+y/b+z/c=1$.

Lines in the plane

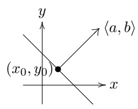
While we're at it, let's look at two ways to write the equation of a line in the xy-plane.

Slope-intercept form: Given the slope m and the y-intercept b the equation of a line can be written $y=mx+b$.

Point-normal form:

We can also use point-normal form to find the equation of a line.

Given a point (x_0,y_0) on the line and a vector $\langle a,b\rangle$ normal to the line the equation of the line can be written $a(x-x_0)+b(y-y_0)=0$.



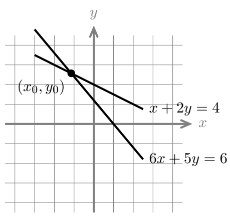
Geometry of linear systems of equations

Very often in math, science and engineering we need to solve a linear system of equations. A simple example of such a system is given by

$$\begin{array}{rcrcl} 6x & + & 5y & = & 6 \\ x & + & 2y & = & 4. \end{array}$$

You have probably already learned algebraic techniques to solve such a system. Later we will also learn to solve such a system using matrix algebra. For now we will focus on the geometric view of this system.

Solving the system means finding a pair (x_0, y_0) which satisfies both equations. Geometrically each of the equations represents a line. That is, each pair (x, y) satisfying the equation is a point on the line. Thus, the solution (x_0, y_0) is the point where the two lines intersect.

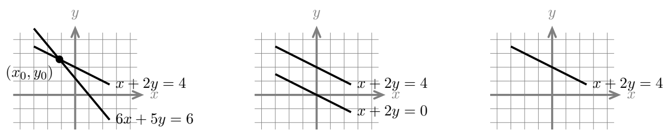


From the graph we can approximate the solution (the exact solution is $(-8/7, 18/7)$), but our interest here is in how many solutions there can be.

The geometric picture makes this obvious. Here are the three possibilities.

1. The two lines intersect in a point, so there is one solution.
2. The two lines are parallel (and not the same), so there are no solutions.
3. The two lines are the same, so there are an infinite number of solutions.

Here are example systems and graphs.



$6x + 5y = 6$	$x + 2y = 4$	$x + 2y = 4$
$x + 2y = 4$	$x + 2y = 0$	$x + 2y = 4$
(intersecting lines: 1 solution)	(parallel lines: no solutions)	(the same line: ∞ solutions)

3 × 3 systems

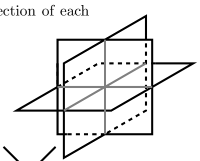
For 3×3 systems there are more possibilities. For example, consider the system

$$\begin{array}{rcrcrcl} 6x & + & 5y & + & 3z & = & 1 \\ x & + & 2y & + & z & = & 4 \\ 2x & - & 2y & - & 2z & = & 8. \end{array}$$

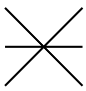
Each equation is the equation of a plane, so, geometrically, solving the system means finding the intersection of three planes, i.e., the point or points which lie on all three planes.

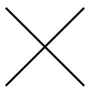
Usually, three planes intersect in a point. You can visualize this by first imagining two of the planes intersecting in a line and then the line intersecting the third plane in a point. Altogether there are four possibilities.

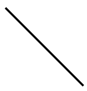
1. Intersect in a point (1 solution to system).
2. Intersect in a line (∞ solutions).
   a) Three different planes, the third plane contains the line of intersection of the first two.
   b) Two planes are the same, the third plane intersects them in a line.
3. Intersect in a plane (∞ solutions)
   a) All three planes are the same.
4. The planes don't all intersect at any point (0 solutions).
   a) The planes are different, but all parallel.
   b) Two planes are parallel, the third crosses them.
   c) The planes are different and none are parallel. but the lines of intersection of each pair are parallel.
   d) Two planes are the same and parallel to the third.

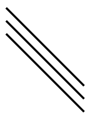


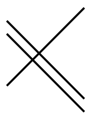
To visualize this we could draw three dimensional figures, for example the figure at the right shows three planes intersecting in a point. Instead, we will visualize the other cases by drawing lines on the page and imagining the planes as extending vertically out of the page.

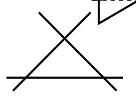

Case (2a)

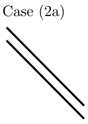

Case (2b)


Case (3a)


Case (4a)


Case (4b)


Case (4c)


Case (4d)

Solutions to linear systems

1. Consider the system of equations

$$\begin{array}{rcrcrcl} x & + & 2y & + & 3z & = & 1 \\ 4x & + & 5y & + & 6z & = & 2 \\ 7x & + & 8y & + & cz & = & 3. \end{array}$$

a) Write the system in matrix form.

b) For which values of c is there exactly one solution?

c) For which values of c are there either 0 or infinitely many solutions?

d) Take the corresponding homogeneous system

$$\begin{array}{rcrcrcl} x & + & 2y & + & 3z & = & 0 \\ 4x & + & 5y & + & 6z & = & 0 \\ 7x & + & 8y & + & cz & = & 0. \end{array}$$

For the value(s) of c found in part (c) give *all* the solutions.

<u>**Answer:**</u> a) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

b) There is exactly one solution when the coefficient matrix has an inverse (i.e., is *invertible*). This happens when the determinant is not zero.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & c \end{vmatrix} = 1(5c - 48) - 2(4c - 42) + 3(32 - 35) = -3c + 27 = 0 \;\Leftrightarrow\; c = 9.$$

There is exactly one solution as long as $c \neq 9$.

c) This is just the complement of part (b): there are zero or infinitely many solutions when $c = 9$.

d) Setting $c = 9$ our coefficient matrix is $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$. Thinking of matrix multiplication as a series of dot products between rows of the left matrix and column(s) of the right one we see that in solving

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

we are looking for vectors $\langle x, y, z \rangle$ that are orthogonal to each of the rows of A. Since $\det(A) = 0$, the rows are all in a plane and we can find orthogonal vectors by taking a cross product of (say) the first two rows.

$$\langle 1, 2, 3 \rangle \times \langle 4, 5, 6 \rangle = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{vmatrix} = \langle -3, 6, -3 \rangle.$$

Since scaling will preserve orthogonality, all the solutions are scalar multiples, i.e., all the solutions are of the form $(x, y, z) = (-3a, 6a, -3a)$. We can make this a little nicer by removing the common factor of three,

$$(x, y, z) = (-a, 2a, -a) = a(-1, 2, -1).$$

Parametric equations of lines

General parametric equations

In this part of the unit we are going to look at parametric curves. This is simply the idea that a point moving in space traces out a path over time. Thus there are four variables to consider, the position of the point (x, y, z) and an independent variable t, which we can think of as time. (If the point is moving in plane there are only three variables, the position of the point (x, y) and the time t.)

Since the position of the point depends on t we write

$$x = x(t), \quad y = y(t), \quad z = z(t)$$

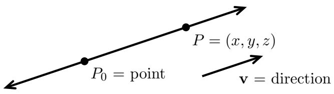
to indicate that x, y and z are functions of t. We call t the parameter and the equations for x, y and z are called *parametric equations*.

In physical examples the parameter often represents time. We will see other cases where the parameter has a different interpretation, or even no interpretation.

Parametric equations of lines

Later we will look at general curves. Right now, let's suppose our point moves on a line.

The basic data we need in order to specify a line are a point on the line and a vector parallel to the line. That is, we need a point and a direction.



Example 1: Write parametric equations for a line through the point $P_0 = (1, 2, 3)$ and parallel to the vector $\mathbf{v} = \langle 1, 3, 5\rangle$.

<u>**Answer:**</u> If $P = (x, y, z)$ is on the line then the vector

$$\overrightarrow{\mathbf{P_0P}} = \langle x - 1, y - 2, z - 3\rangle$$

is parallel to $\langle 1, 3, 5\rangle$. That is, $\overrightarrow{\mathbf{P_0P}}$ is a scalar multiple of $\langle 1, 3, 5\rangle$. We call the scale t and write:

$$\begin{aligned} & \langle x, y, z\rangle = \langle x - 1, y - 2, z - 3\rangle = t\langle 1, 3, 5\rangle \\ \Leftrightarrow \quad & x - 1 = t, \ \ y - 2 = 3t, \quad z - 3 = 5t \\ \Leftrightarrow \quad & x = 1 + t, \ \ y = 2 + 3t, \quad z = 3 + 5t. \end{aligned}$$

Example 2: In example 1, if our direction vector was $\langle 2, 6, 10\rangle = 2\mathbf{v}$ we would get the same line with a different parametrization. That is, the moving point's trajectory would follow the same path as the trajectory in example 1 , but would arrive at each point on the line at a different time.

Example 3: In general, the line through $P_0 = (x_0, y_0, z_0)$ in the direction of (i.e., parallel to) $\mathbf{v} = \langle v_1, v_2, v_3\rangle$ has parametrization

$$\begin{aligned} & \langle x, y, z\rangle = \langle x_0 + tv_1, y_0 + tv_2, z_0 + tv_3\rangle \\ \Leftrightarrow \quad & x = x_0 + tv_1, \quad y = y_0 + tv_2 \quad z = z_0 + tv_3. \end{aligned}$$

Example 4: Find the line through the point $P_0 = (1, 2, 3)$ and $P_1 = (2, 5, 8)$.

Answer: We use the data given to find the basic data (a point and direction vector) for the line.

We're given a point, $P_0 = (1, 2, 3)$. The direction vector $\mathbf{v} = \overrightarrow{\mathbf{P_0P_1}} = \langle 1, 3, 5\rangle$. So, we get

$$\begin{aligned} & \langle x, y, z\rangle = \overrightarrow{\mathbf{OP_0}} + t\mathbf{v} = \langle 1 + t, 2 + 3t, 3 + 5t\rangle \\ \Leftrightarrow \quad & x = 1 + t, \quad y = 2 + 3t, \quad z = 3 + 5t. \end{aligned}$$

Intersection of a line and a plane

1. Consider the plane $\mathcal{P} = 2x + y - 4z = 4$.

a) Find all points of intersection of $\mathcal{P}$ with the line

$$x = t, \quad y = 2 + 3t, \quad z = t.$$

b) Find all points of intersection of $\mathcal{P}$ with the line

$$x = 1 + t, \quad y = 4 + 2t, \quad z = t.$$

c) Find all points of intersection of $\mathcal{P}$ with the line

$$x = t, \quad y = 4 + 2t, \quad z = t.$$

Answer: a) To find the intersection we substitute the formulas for x, y and z into the equation for $\mathcal{P}$ and solve for t.

$$2(t) + (2 + 3t) - 4(t) = 4 \;\Leftrightarrow\; t = 2.$$

Now use $t = 2$ to find the point of intersection: $(x, y, z) = (2, 8, 2)$.

b) Substituting gives

$$2(1 + t) + (4 + 2t) - 4(t) = 4 \;\Leftrightarrow\; 6 = 4. \;\Leftrightarrow\; \text{no values of } t \text{ satisfy this equation.}$$

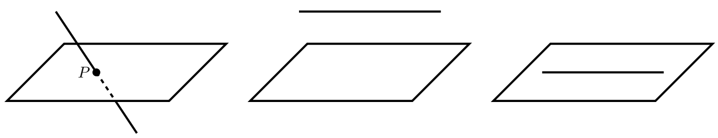
There are no points of intersection.

c) Substituting gives

$$2(t) + (4 + 2t) - 4(t) = 4 \;\Leftrightarrow\; 4 = 4. \;\Leftrightarrow\; \text{all values of } t \text{ satisfy this equation.}$$

The line is contained in the plane, i.e., all points of the line are in its intersection with the plane.

Here are cartoon sketches of each part of this problem.



(a) line intersects the plane in a point

(b) line is parallel to the plane

(c) line is in the plane

Parametric Curves

General parametric equations

We have seen parametric equations for lines. Now we will look at parametric equations of more general trajectories. Repeating what was said earlier, a parametric curve is simply the idea that a point moving in the space traces out a path.

We can use a parameter to describe this motion. Quite often we will use t as the parameter and think of it as time. Since the position of the point depends on t we write

$$x = x(t), \quad y = y(t), \quad z = z(t)$$

to indicate that x, y and z are functions of t. We call t the parameter and the equations for x, y and z are called *parametric equations*.

It is not always necessary to think of the parameter as representing time. We will see cases where it is more convenient to express the position as a function of some other variable.

The position vector

In order to use vector techniques we define the *position vector*

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} = \langle x(t), y(t), z(t) \rangle.$$

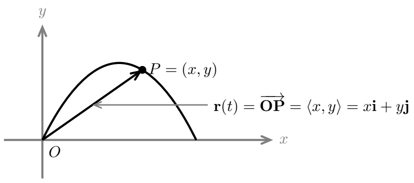
This is just the vector from the origin to the moving point. As the point moves so does the position vector –see the figure with example 1.

Example 1: Thomas Pynchon fires a rocket from the origin. Its initial x-velocity is $v_{0,x}$ and its initial y-velocity is $v_{0,y}$.

You've probably seen this, but in any case, physics tells us that the parametric equations for its parabolic trajectory are

$$x(t) = v_{0,x}t, \quad y(t) = -\frac{1}{2}gt^2 + v_{0,y}t.$$

At time t the rocket is at point $P = (x(t), y(t))$. The position vector can be written in many different ways: $\mathbf{r}(t) = \overrightarrow{\mathbf{OP}} = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} = \langle x, y \rangle$.



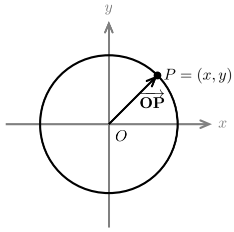
Next we will give a series of examples of parametrized curves. The most important are circles and lines. The last one is the *cycloid*. It is an important example which combines lines and circles.

Circles and ellipses

Consider the parametric curve in the plane

$$x(t) = a\cos t, \quad y(t) = a\sin t.$$

Easily we get the relation $x^2 + y^2 = a^2\cos^2 t + a^2\sin^2 t = a^2$. Therefore the trajectory is on a circle of radius a centered at O.



We will call $x(t) = a\cos t$, $y(t) = a\sin t$ the *parametric form* of the curve and $x^2 + y^2 = a^2$ the *symmetric form*.

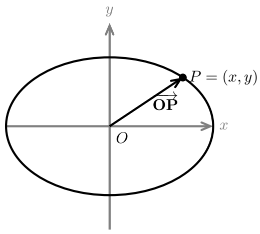
Note, a different parametrization, say

$$x(t) = a\cos(3t), \quad y(t) = a\sin(3t)$$

results in the same path, i.e. the circle $x^2 + y^2 = a^2$, but the two trajectories differ by how fast they travel around the circle.

The circle is easily changed to an ellipse by

$$\text{parametric form:} \quad x(t) = a\cos t, \quad y(t) = b\cos t$$
$$\text{symmetric form:} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$



Lines

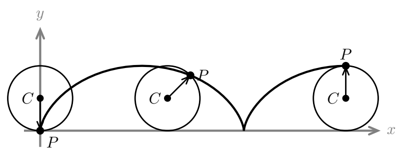
We review parametric equations of lines by writing the the equation of a general line in the plane. We know we can parametrize the line through (x_0, y_0) parallel to $\langle b_1, b_2\rangle$ by

$$x(t) = x_0 + tb_1, \quad y(t) = y_0 + tb_2 \Leftrightarrow \mathbf{r}(t) = \langle x, y\rangle = \langle x_0 + tb_1, y_0 + tb_2\rangle = \langle x_0, y_0\rangle + t\langle b_1, b_2\rangle.$$

The cycloid

The cycloid has a long and storied history and comes up surprisingly often in physical problems. For us it is a curve that has no simple symmetric form, so we will only work with it in its parametric form.

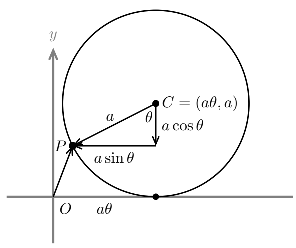
The cycloid is the trajectory of a point on a circle that is rolling without slipping along the x-axis. To be specific, we'll follow the point P that starts at the origin.



The natural parameter to use is the angle θ that the wheel has turned. We'll use vector methods to find the position vector for P as a function of θ.

Our strategy is to break the motion up into translation of the center and rotation about the center. The figure shows the wheel after it has turned through a small θ. We see the position vector

$$\overrightarrow{\mathbf{OP}} = \overrightarrow{\mathbf{OC}} + \overrightarrow{\mathbf{CP}}.$$



We'll compute each piece separately.

After turning θ radians the wheel has rolled a distance $a\theta$, so the center of the circle is at $(a\theta, a)$, i.e.,

$$\overrightarrow{\mathbf{OC}} = \langle a\theta, a\rangle.$$

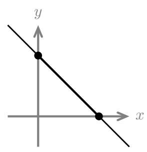
The figure also shows that

$$\overrightarrow{\mathbf{CP}} = \langle -a\sin\theta, -a\cos\theta\rangle.$$

Putting the pieces together we get parametric equations for the cycloid

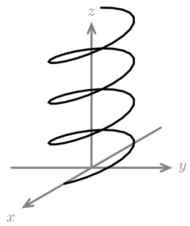
$$\overrightarrow{\mathbf{OP}} = \langle a\theta - a\sin\theta, a - a\cos\theta\rangle$$
$$\Leftrightarrow \quad x(\theta) = a\theta - a\sin\theta, \quad y(\theta) = a - a\cos\theta.$$

Example 2: (Where the symmetric form loses information.)
Find the symmetric form for $x = 3\cos^2 t,\ y = 3\sin^2 t$.



Easily we get: $x + y = 3$, with x, y non-negative.
The symmetric form shows a line, but the parametric trajectory only traces out a part of the line. In fact, it goes back an forth over the part of the line in the first quadrant.

Example 3: The curve $\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + at\,\mathbf{k}$ is a helix winding around the z-axis.



Functions of two variables

Examples: Functions of several variables
$f(x,y) = x^2 + y^2 \;\Rightarrow\; f(1,2) = 5$ etc.
$f(x,y) = xy^2 e^{x+y}$
$f(x,y,z) = xy \log z$
Ideal gas law: $P = kT/V$.

Dependent and independent variables
In $z = f(x,y)$ we say x, y are independent variables and z is a dependent variable. This indicates that x and y are free to take any values and then z depends on these values. For now it will be clear which are which, later we'll have to take more care.

Graphs
For the function $y = f(x)$: there is one independent variable and one dependent variable, which means we need 2 dimensions for its graph.

Graphing technique:

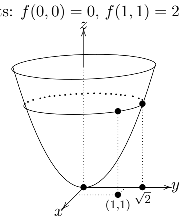
go to x then compute $y = f(x)$ then go up to height y.

For $z = f(x,y)$ we have two independent and one dependent variable, so we need 3 dimensions to graph the function. The technique is the same as before.

Example: Consider $z = f(x,y) = x^2 + y^2$.
To make the graph:

go to (x,y) then compute $z = f(x,y)$ then go up to height z.

We show the plot of three points: $f(0,0) = 0$, $f(1,1) = 2$ and $f(0, \sqrt{2}) = 2$.



The figure above shows more than just the graph of three points. Here are the steps we used to draw the graph. Remember, this is just a sketch, it should suggest the shape of the graph and some of its features.

1. First we draw the axes. The z-axis points up, the y-axis is to the right and the x-axis comes out of the page, so it is drawn at the angle shown. This gives a perspective with the eye somewhere in the first octant.

2. The yz-traces are those curves found by setting $x =$ a constant. We start with the trace when $x = 0$. This is an upward pointing parabola in the yz-plane.

3. Next we sketch the trace with $z = 3$. This is a circle of radius $\sqrt{3}$ at height $z = 3$. Note, the traces where $z =$ constant are generally called *level curves*.

This is enough for this graph. Other graphs take other traces. You should expect to do a certain amount of trial and error before your figure looks right.

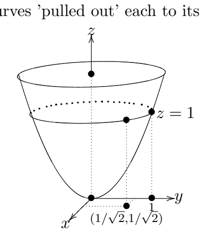
Level Curves and Contour Plots

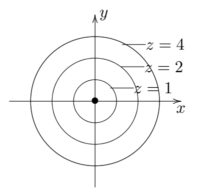
Level curves and *contour plots* are another way of visualizing functions of two variables. If you have seen a topographic map then you have seen a contour plot.

Example: To illustrate this we first draw the graph of $z = x^2 + y^2$. On this graph we draw *contours*, which are curves at a fixed height $z =$ constant.

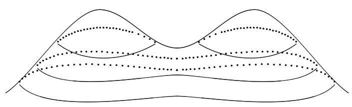
For example the curve at height $z = 1$ is the circle $x^2 + y^2 = 1$. On the graph we have to draw this at the correct height. Another way to show this is to draw the curves in the xy-plane and label them with their z-value. We call these curves *level curves* and the entire plot is called a *contour plot*.

For this example they are shown in the plot on the right. Notice that the 3D graph is simply the level curves 'pulled out' each to its correct height.

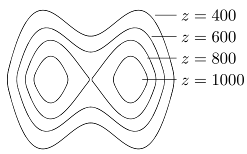




Here is another plot of a 'mountain pass'. Notice that in the contour plot the mountain pass is represented by a level curve that crosses itself. Moving up or down from the cross level curves heights decrease and moving right or left in the other they increase.



Mountain pass

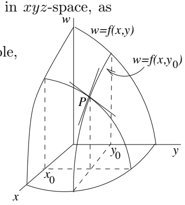


Level curves

Partial derivatives

Partial derivatives

Let $w = f(x, y)$ be a function of two variables. Its graph is a surface in xyz-space, as pictured.



Fix a value $y = y_0$ and just let x vary. You get a function of *one* variable,

(1) $$w = f(x, y_0), \qquad \text{the } \textbf{partial function} \text{ for } y = y_0.$$

Its graph is a curve in the vertical plane $y = y_0$, whose slope at the point P where $x = x_0$ is given by the derivative

(2) $$\frac{d}{dx} f(x, y_0)\bigg|_{x_0}, \quad \text{or} \quad \frac{\partial f}{\partial x}\bigg|_{(x_0,y_0)}.$$

We call (2) the **partial derivative** of f with respect to x at the point (x_0, y_0); the right side of (2) is the standard notation for it. The partial derivative is just the ordinary derivative of the partial function — it is calculated by holding one variable fixed and differentiating with respect to the other variable. Other notations for this partial derivative are

$$f_x(x_0, y_0), \qquad \frac{\partial w}{\partial x}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0;$$

the first is convenient for including the specific point; the second is common in science and engineering, where you are just dealing with relations between variables and don't mention the function explicitly; the third and fourth indicate the point by just using a single subscript.

Analogously, fixing $x = x_0$ and letting y vary, we get the partial function $w = f(x_0, y)$, whose graph lies in the vertical plane $x = x_0$, and whose slope at P is the *partial derivative of f with respect to y*; the notations are

$$\frac{\partial f}{\partial y}\bigg|_{(x_0,y_0)}, \qquad f_y(x_0, y_0), \qquad \frac{\partial w}{\partial y}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial y}\right)_0, \qquad \left(\frac{\partial w}{\partial y}\right)_0.$$

The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ depend on (x_0, y_0) and are therefore functions of x and y.

Written as $\partial w/\partial x$, the partial derivative gives the rate of change of w with respect to x alone, at the point (x_0, y_0): it tells how fast w is increasing as x increases, when y is held constant.

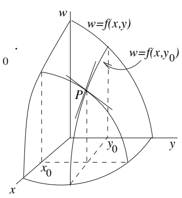
For a function of three or more variables, $w = f(x, y, z, \ldots)$, we cannot draw graphs any more, but the idea behind partial differentiation remains the same: to define the partial derivative with respect to x, for instance, hold all the other variables constant and take the ordinary derivative with respect to x; the notations are the same as above:

$$\frac{d}{dx} f(x, y_0, z_0, \ldots) = f_x(x_0, y_0, z_0, \ldots), \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0.$$

The Tangent Approximation

1. The tangent plane.

For a function of one variable, $w = f(x)$, the tangent line to its graph at a point (x_0, w_0) is the line passing through (x_0, w_0) and having slope $\left(\dfrac{dw}{dx}\right)_0$.



For a function of two variables, $w = f(x, y)$, the natural analogue is the *tangent plane* to the graph, at a point

(x_0, y_0, w_0). What's the equation of this tangent plane? Referring to the picture at right (this figure was also used when we introduced partial derivatives), we see that the tangent plane

(i) must pass through (x_0, y_0, w_0), where $w_0 = f(x_0, y_0)$;

(ii) must contain the tangent lines to the graphs of the two partial functions — this will hold if the plane has the same slopes in the $\mathbf{i}$ and $\mathbf{j}$ directions as the surface does.

Using these two conditions, it is easy to find the equation of the tangent plane. The general equation of a plane through (x_0, y_0, w_0) is

$$A(x - x_0) + B(y - y_0) + C(w - w_0) = 0 .$$

Assume the plane is not vertical; then $C \neq 0$, so we can divide through by C and solve for $w - w_0$, getting

$$w - w_0 = a(x - x_0) + b(y - y_0), \qquad a = A/C,\ b = B/C. \tag{3}$$

The plane passes through (x_0, y_0, w_0); what values of the coefficients a and b will make it also tangent to the graph there? We have

$$\begin{aligned} a &= \text{slope of plane (3) in the } \mathbf{i}\text{-direction} && \text{(by putting } y = y_0 \text{ in (3));} \\ &= \text{slope of graph in the } \mathbf{i}\text{-direction,} && \text{(by (ii) above)} \\ &= \left(\frac{\partial w}{\partial x}\right)_0; && \text{(by the definition of partial derivative);} \quad \text{similarly,} \\ b &= \left(\frac{\partial w}{\partial y}\right)_0 . \end{aligned}$$

Therefore the equation of the **tangent plane** to $w = f(x, y)$ at (x_0, y_0) is

$$w - w_0 = \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0) \tag{4}$$

2. The approximation formula.

The most important use for the tangent plane is to give an approximation that is the basic formula in the study of functions of several variables — almost everything follows in one way or another from it.

The intuitive idea is that if we stay near (x_0, y_0, w_0), the graph of the tangent plane (4) will be a good approximation to the graph of the function $w = f(x, y)$. Therefore if the point (x, y) is close to (x_0, y_0),

$$(5) \qquad \underset{\text{height of graph}}{f(x,y)} \quad \approx \quad \underset{\text{height of tangent plane}}{w_0 + \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)}$$

The function on the right side of (5) whose graph is the tangent plane is often called the **linearization** of $f(x, y)$ at (x_0, y_0): it is the linear function which gives the best approximation to $f(x, y)$ for values of (x, y) close to (x_0, y_0).

An equivalent form of the approximation (5) is obtained by using Δ notation; if we put

$$\Delta x = x - x_0, \qquad \Delta y = y - y_0, \qquad \Delta w = w - w_0,$$

then (5) becomes

$$(6) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad \text{if} \quad \Delta x \approx 0, \ \Delta y \approx 0 .$$

This formula gives the approximate change in w when we make a small change in x and y. We will use it often.

The analogous approximation formula for a function $w = f(x, y, z)$ of three variables would be

$$(7) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y + \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \qquad \text{if} \quad \Delta x, \Delta y \Delta z \approx 0 .$$

Unfortunately, for functions of three or more variables, we can't use a geometric argument for the approximation formula (7); for this reason, it's best to recast the argument for (6) in a form which doesn't use tangent planes and geometry, and therefore can be generalized to several variables. This is done at the end of this Chapter TA; for now let's just assume the truth of (7) and its higher-dimensional analogues.

Here are two typical examples of the use of the approximation formula. Other examples are in the Exercises. In the rest of your study of partial differentiation, you will see how the approximation formula is used to derive the important theorems and formulas.

Example 1. Give a reasonable square, centered at $(1, 1)$, over which the value of $w = x^3 y^4$ will not vary by more than $\pm$ .1 .

Solution. We use (6). We calculate for the two partial derivatives

$$w_x = 3x^2 y^4 \qquad \qquad w_y = 4x^3 y^3$$

and therefore, evaluating the partials at $(1, 1)$ and using (6), we get

$$\Delta w \approx 3\Delta x + 4\Delta y.$$

Thus if $|\Delta x| \le .01$ and $\Delta y| \le .01$, we should have

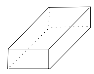
$$|\Delta w| \le 3|\Delta x| + 4|\Delta y| \le .07,$$

which is within the bounds. So the answer is the square with center at (1,1) given by

$$|x - 1| \le .01, \qquad |y - 1| \le .01 .$$

Example 2. The sides a, b, c of a rectangular box have lengths measured to be respectively 1, 2, and 3. To which of these measurements is the volume V most sensitive?

Solution. $V = abc$, and therefore by the approximation formula (7),



$$\begin{aligned} \Delta V &\approx bc\,\Delta a + ac\,\Delta b + ab\,\Delta c \\ &\approx 6\,\Delta a + 3\,\Delta b + 2\,\Delta c, \qquad \text{at } (1,2,3); \end{aligned}$$

thus it is most sensitive to small changes in side a, since Δa occurs with the largest coefficient. (That is, if one at a time the measurement of each side were changed by say .01, it is the change in a which would produce the biggest change in V, namely .06 .)

The result may seem paradoxical — the value of V is most sensitive to the length of the *shortest* side — but it's actually intuitive, as you can see by thinking about how the box looks.

Sensitivity Principle *The numerical value of* $w = f(x, y, \ldots)$, *calculated at some point* $(x_0, y_0, \ldots)$, *will be most sensitive to small changes in that variable for which the corresponding partial derivative* $w_x, w_y, \ldots$ *has the largest absolute value at the point.*

Critical Points

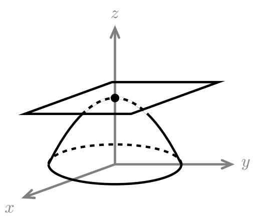
Critical points:
A standard question in calculus, with applications to many fields, is to find the points where a function reaches its relative maxima and minima.

Just as in single variable calculus we will look for maxima and minima (collectively called *extrema*) at points (x_0, y_0) where the first derivatives are 0. Accordingly we define a *critical point* as any point (x_0, y_0) where

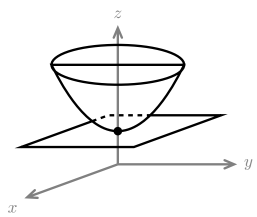
$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \text{ and } \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Often we will abbreviate this as $f_x = 0$ and $f_y = 0$.

Our first job is to verify that relative maxima and minima occur at critical points. The figures below illustrates that they occur at places where the tangent plane is horizontal.



Max. with horizontal tang. plane



Min. with horizontal tang. plane

Since horizontal planes are of the form $z =$ constant. and the equation of the tangent plane at (x_0, y_0, z_0) is

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

we see it is horizontal when

$$f_x(x_0, y_0) = 0 \text{ and } f_y(x_0, y_0) = 0.$$

Thus, extrema occur at critical points. But, just as in single variable calculus, not all critical points are extrema.

Example: Find the critical points of $z = x^2 + y^2 + .5$.

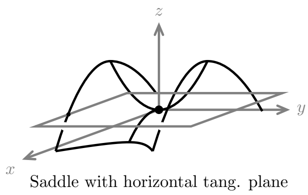
Answer: $\frac{\partial z}{\partial x} = 2x$ and $\frac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a minimum value. (See the figure above on the right.)

Example: Find the critical points of $z = 1 - x^2 - y^2$.

Answer: $\frac{\partial z}{\partial x} = -2x$ and $\frac{\partial z}{\partial y} = -2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a maximum value. (See the figure above on the left.)

Example: Find the critical points of $z = -x^2 + y^2$.

Answer: $\dfrac{\partial z}{\partial x} = -2x$ and $\dfrac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0,0)$. Thus, there is a single critical point at $(0,0)$. The figure shows it is neither a minimum or a maximum.



Saddle with horizontal tang. plane

Example: Making a box with minimum material.

A box is made of cardboard with double thick sides, a triple thick bottom, single thick front and back and no top. It's volume = 3.

What dimensions use the least amount of cardboard?

Answer: The box shown has dimensions x, y, and z.

The area of one side $= yz$. There are two double thick sides $\Rightarrow$ cardboard used $= 4yz$.

The area of the front (and back) $= xz$. It is single thick $\Rightarrow$ cardboard used $= 2xz$.

The area of the bottom $= xy$. It is triple thick $\Rightarrow$ cardboard used $= 3xy$.

Thus, the total cardboard used is

$$w = 4yz + 2xz + 3xy.$$

The volume $= 3 = xyz \Rightarrow z = \frac{3}{xy}$. Substituting this in the formula for w gives

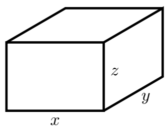
$$w = \frac{12}{x} + \frac{6}{y} + 3xy.$$

We find the critical points of w.

$$w_x = -\frac{12}{x^2} + 3y = 0, \quad w_y = -\frac{6}{y^2} + 3x = 0.$$

The first equation implies $y = \dfrac{4}{x^2}$. Substituting this in the second equation gives $-\dfrac{6}{16}x^4 + 3x = 0$. Thus, $x = 0$ or 2. We reject 0 since then y is undefined. Using $x = 2$ we find $y = 1$. Thus, there there is one critical point at (2,1). and at this point we have $z = 3/2$.

This point gives the box with minimum cardboard used because physically we know it must have a minimum somewhere. Later we will learn to check this with the second derivative test.

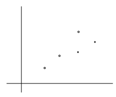


Least Squares Interpolation

1. The least-squares line.

Suppose you have a large number n of experimentally determined points, through which you want to pass a curve. There is a formula (the Lagrange interpolation formula) producing a polynomial curve of degree $n-1$ which goes through the points exactly. But normally one wants to find a simple curve, like a line, parabola, or exponential, which goes approximately through the points, rather than a high-degree polynomial which goes exactly through them. The reason is that the location of the points is to some extent determined by experimental error, so one wants a smooth-looking curve which averages out these errors, not a wiggly polynomial which takes them seriously.

In this section, we consider the most common case — finding a line which goes approximately through a set of data points.



Suppose the data points are

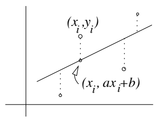
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

and we want to find the line

$$y = ax + b \tag{1}$$

which "best" passes through them. Assuming our errors in measurement are distributed randomly according to the usual bell-shaped curve (the so-called "Gaussian distribution"), it can be shown that the right choice of a and b is the one for which the sum D of the squares of the deviations

$$D \;=\; \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2 \tag{2}$$



is a *minimum*. In the formula (2), the quantities in parentheses (shown by dotted lines in the picture) are the **deviations** between the observed values y_i and the ones $ax_i + b$ that would be predicted using the line (1).

The deviations are squared for theoretical reasons connected with the assumed Gaussian error distribution; note however that the effect is to ensure that we sum only positive quantities; this is important, since we do not want deviations of opposite sign to cancel each other out. It also weights more heavily the larger deviations, keeping experimenters honest, since they tend to ignore large deviations ("I had a headache that day").

This prescription for finding the line (1) is called the **method of least squares**, and the resulting line (1) is called the **least-squares** line or the **regression** line.

To calculate the values of a and b which make D a minimum, we see where the two partial derivatives are zero:

$$\begin{aligned} \frac{\partial D}{\partial a} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) \;=\; 0 \\ \frac{\partial D}{\partial b} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) \;=\; 0\;. \end{aligned} \tag{3}$$

These give us a pair of *linear* equations for determining a and b, as we see by collecting terms and cancelling the 2's:

$$\begin{aligned} \left(\sum x_i^2\right) a \ + \ \left(\sum x_i\right) b \ &= \ \sum x_i y_i \\ \left(\sum x_i\right) a \ + \ n\, b \ &= \ \sum y_i \ . \end{aligned} \tag{4}$$

(Notice that it saves a lot of work to differentiate (2) using the chain rule, rather than first expanding out the squares.)

The equations (4) are usually divided by n to make them more expressive:

$$\begin{aligned} \bar{s}\, a \ + \ \bar{x}\, b \ &= \ \frac{1}{n}\sum x_i y_i \\ \bar{x}\, a \ + \ b \ &= \ \bar{y} \ , \end{aligned} \tag{5}$$

where $\bar{x}$ and $\bar{y}$ are the average of the x_i and y_i, and $\bar{s} = \sum x_i^2/n$ is the average of the squares.

From this point on use linear algebra to determine a and b. It is a good exercise to see that the equations are always solvable unless all the x_i are the same (in which case the best line is vertical and can't be written in the form (1)).

In practice, least-squares lines are found by pressing a calculator button, or giving a MatLab command. Examples of calculating a least-squares line are in the exercises accompanying the course. Do them from scratch, starting from (2), since the purpose here is to get practice with max-min problems in several variables; don't plug into the equations (5). Remember to differentiate (2) using the chain rule; don't expand out the squares, which leads to messy algebra and highly probable error.

2. Fitting curves by least squares.

If the experimental points seem to follow a curve rather than a line, it might make more sense to try to fit a second-degree polynomial

$$y = a_0 + a_1 x + a_2 x^2 \tag{6}$$

to them. If there are only three points, we can do this exactly (by the Lagrange interpolation formula). For more points, however, we once again seek the values of a_0, a_1, a_2 for which the sum of the squares of the deviations

$$D \ = \ \sum_1^n \big(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\big)^2 \tag{7}$$

is a minimum. Now there are three unknowns, a_0, a_1, a_2. Calculating (remember to use the chain rule!) the three partial derivatives $\partial D/\partial a_i,\ i = 0, 1, 2$, and setting them equal to zero leads to a square system of three linear equations; the a_i are the three unknowns, and the coefficients depend on the data points (x_i, y_i). They can be solved by finding the inverse matrix, elimination, or using a calculator or MatLab.

If the points seem to lie more and more along a line as $x \to \infty$, but lie on one side of the line for low values of x, it might be reasonable to try a function which has similar behavior, like

$$y \ = \ a_0 + a_1 x + a_2 \frac{1}{x} \tag{8}$$

and again minimize the sum of the squares of the deviations, as in (7). In general, this method of least squares applies to a trial expression of the form

$$y = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_r f_r(x), \tag{9}$$

where the $f_i(x)$ are given functions (usually simple ones like $1, x, x^2, 1/x, e^{kx}$, etc. Such an expression (9) is called a **linear combination** of the functions $f_i(x)$. The method produces a square inhomogeneous system of linear equations in the unknowns $a_0, \ldots, a_r$ which can be solved by finding the inverse matrix to the system, or by elimination.

The method also applies to finding a linear function

$$z = a_1 + a_2 x + a_3 y \tag{10}$$

to fit a set of data points

$$(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n)\ . \tag{11}$$

where there are two independent variables x and y and a dependent variable z (this is the quantity being experimentally measured, for different values of (x, y)). This time after differentiation we get a 3×3 system of linear equations for determining a_1, a_2, a_3 .

The essential point in all this is that the unknown coefficients a_i should occur *linearly* in the trial function. Try fitting a function like ce^{kx} to data points by using least squares, and you'll see the difficulty right away. (Since this is an important problem — fitting an exponential to data points — one of the Exercises explains how to adapt the method to this type of problem.)

Least squares interpolation

1. Use the method of least squares to fit a line to the three data points

$$(0,0), \quad (1,2), \quad (2,1).$$

<u>Answer:</u> We are looking for the line $y = ax + b$ that best models the data. The deviation of a data point (x_i, y_i) from the model is

$$y_i - (ax_i + b).$$

By best we mean the line that minimizes the sum of the squares of the deviation. That is we want to minimize

$$\begin{aligned} D &= (0 - (a \cdot 0 + b))^2 + (2 - (a \cdot 1 + b))^2 + (1 - (a \cdot 2 + b))^2 \\ &= b^2 + (2 - a - b)^2 + (1 - 2a - b)^2. \end{aligned}$$

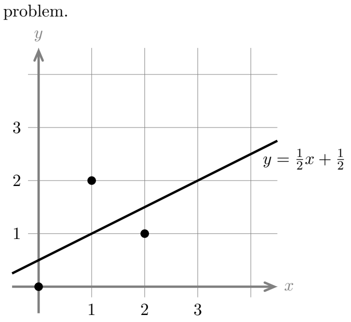
(Remember, the variables whose values are to be found are a and b.) We do not expand out the squares, rather we take the derivatives first. Setting the derivatives equal to 0 gives

$$\begin{aligned} &\tfrac{\partial D}{\partial a} = -2(2 - a - b) - 4(1 - 2a - b) = 0 \;\Rightarrow\; 10a + 6b = 8 \;\Rightarrow\; 5a + 3b = 4 \\ &\tfrac{\partial D}{\partial b} = 2b - 2(2 - a - b) - 2(1 - 2a - b) = 0 \;\Rightarrow\; 6a + 6b = 6 \;\Rightarrow\; 3a + 3b = 3. \end{aligned}$$

This linear system of two equations in two unknowns is easy to solve. We get

$$a = \frac{1}{2}, \qquad b = \frac{1}{2}.$$

Here is a plot of the problem.



Second Derivative Test

1. The Second Derivative Test

We begin by recalling the situation for twice differentiable functions $f(x)$ of one variable. To find their local (or "relative") maxima and minima, we

1. find the critical points, i.e., the solutions of $f'(x) = 0$;
2. apply the second derivative test to each critical point x_0:

$$f''(x_0) > 0 \;\Rightarrow\; x_0 \text{ is a local minimum point;}$$
$$f''(x_0) < 0 \;\Rightarrow\; x_0 \text{ is a local maximum point.}$$

The idea behind it is: at x_0 the slope $f'(x_0) = 0$; if $f''(x_0) > 0$, then $f'(x)$ is strictly increasing for x near x_0, so that the slope is negative to the left of x_0 and positive to the right, which shows that x_0 is a minimum point. The reasoning for the maximum point is similar.

If $f''(x_0) = 0$, the test fails and one has to investigate further, by taking more derivatives, or getting more information about the graph. Besides being a maximum or minimum, such a point could also be a horizontal point of inflection.

The analogous test for maxima and minima of functions of two variables $f(x, y)$ is a little more complicated, since there are several equations to satisfy, several derivatives to be taken into account, and another important geometric possibility for a critical point, namely a **saddle point**. This is a local minimax point; around such a point the graph of $f(x, y)$ looks like the central part of a saddle, or the region around the highest point of a mountain pass. In the neighborhood of a saddle point, the graph of the function lies both above and below its horizontal tangent plane at the point.

The second-derivative test for maxima, minima, and saddle points has two steps.

1. Find the critical points by solving the simultaneous equations $\begin{cases} f_x(x, y) = 0, \\ f_y(x, y) = 0. \end{cases}$

Since a critical point (x_0, y_0) is a solution to both equations, both partial derivatives are zero there, so that the tangent plane to the graph of $f(x, y)$ is horizontal.

2. To test such a point to see if it is a local maximum or minimum point, we calculate the three second derivatives at the point (we use subscript 0 to denote evaluation at (x_0, y_0), so for example $(f)_0 = f(x_0, y_0)$), and denote the values by A, B, and C:

$$A = (f_{xx})_0, \qquad B = (f_{xy})_0 = (f_{yx})_0, \qquad C = (f_{yy})_0, \tag{1}$$

(we are assuming the derivatives exist and are continuous).

Second-derivative test. *Let (x_0, y_0) be a critical point of $f(x, y)$, and A, B, and C be as in (1). Then*

$$AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \;\Rightarrow\; (x_0, y_0) \text{ is a minimum point;}$$
$$AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \;\Rightarrow\; (x_0, y_0) \text{ is a maximum point;}$$
$$AC - B^2 < 0 \;\Rightarrow\; (x_0, y_0) \text{ is a saddle point.}$$

If $AC - B^2 = 0$, the test fails and more investigation is needed.

Note that if $AC - B^2 > 0$, then $AC > 0$, so that A and C must have the same sign.

Example 1. Find the critical points of $w = 12x^2 + y^3 - 12xy$ and determine their type.

Solution. We calculate the partial derivatives easily:

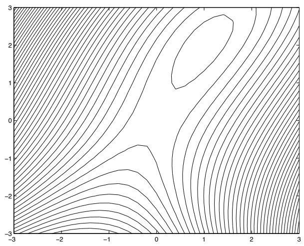
$$\begin{array}{ll} w_x = 24x - 12y & A = w_{xx} = 24 \\ & B = w_{xy} = -12 \\ w_y = 3y^2 - 12x & C = w_{yy} = 6y \end{array} \tag{2}$$

To find the critical points we solve simultaneously the equations $w_x = 0$ and $w_y = 0$; we get

$$\begin{array}{l} w_x = 0 \\ w_y = 0 \end{array} \;\Rightarrow\; \begin{array}{l} y = 2x \\ y^2 = 4x \end{array} \;\Rightarrow\; 4x^2 = 4x \;\Rightarrow\; x = 0, 1 \;\Rightarrow\; \begin{array}{l} (x,y) = (0,0) \\ (x,y) = (1,2) \end{array}.$$

Thus there are two critical points: $(0,0)$ and $(1,2)$. To determine their type, we use the second derivative test: we have $AC - B^2 = 144y - 144$, so that

at $(0,0)$, we have $AC - B^2 = -144$, so it is a saddle point;
at $(1,2)$, we have $AC - B^2 = 144$ and $A > 0$, so it is a a minimum point.



A plot of the level curves is given at the right, which confirms the above. Note that the behavior of the level curves near the origin can be determined by using the approximation $w \approx 12x^2 - 12xy$; this shows the level curves near $(0,0)$ look like those of the function $x(x-y)$: the family of hyperbolas $x(x-y) = c$, with asymptotes given by the degenerate hyperbola $x(x-y) = 0$, i.e., the pair of lines $x = 0$ (the y-axis) and $x - y = 0$ (the diagonal line $y = x$).

2. Justification for the Second-derivative Test.

The test involves the quantity $AC - B^2$. In general, whenever we see the expressions $B^2 - 4AC$ or $B^2 - AC$ or their negatives, it means the quadratic formula is involved, in one of its two forms (the second is often used to get rid of the excess two's):

$$Ax^2 + Bx + C = 0 \quad\Rightarrow\quad x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \tag{3}$$

$$Ax^2 + 2Bx + C = 0 \quad\Rightarrow\quad x = \frac{-B \pm \sqrt{B^2 - AC}}{A} \tag{4}$$

This is what is happening here. We want to know whether, near a critical point P_0, the graph of our function $w = f(x,y)$ always stays on one side of its horizontal tangent plane (P_0 is then a maximum or minimum point), or whether it lies partly above and partly below the tangent plane (P_0 is then a saddle point). As we will see, this is determined by how the graph of a quadratic function $f(x)$ lies with respect to the x-axis. Here is the basic lemma.

Lemma. For the quadratic function $Ax^2 + 2Bx + C$,

$$AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \quad\Rightarrow\quad Ax^2 + 2Bx + C > 0 \quad \text{for all } x; \tag{5}$$

$$AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \quad\Rightarrow\quad Ax^2 + 2Bx + C < 0 \quad \text{for all } x; \tag{6}$$

$$AC - B^2 < 0 \quad\Rightarrow\quad \begin{cases} Ax^2 + 2Bx + C > 0, & \text{for some } x; \\ Ax^2 + 2Bx + C < 0, & \text{for some } x. \end{cases} \tag{7}$$

Proof of the Lemma. To prove (5), we note that the quadratic formula in the form (4) shows that the zeros of $Ax^2 + 2Bx + C$ are imaginary, i.e., it has no real zeros. Therefore its graph must lie entirely on one side of the x-axis; which side can be determined from either A or C, since

$$A > 0 \ \Rightarrow \ \lim_{x\to\infty} Ax^2 + 2Bx + C = \infty; \qquad C > 0 \ \Rightarrow \ Ax^2 + 2Bx + C > 0 \text{ when } x = 0.$$

If $A < 0$ or $C < 0$, the reasoning is analogous and proves (6).

If on the other hand $AC - B^2 < 0$, formula (4) shows the quadratic function has two real roots, so that its parabolic graph crosses the x-axis twice, and hence lies partly above and partly below it. This proves (7). □

Proof of the Second-derivative Test in a special case.

The simplest function is a linear function, $w = w_0 + ax + by$, but it does not in general have maximum or minimum points and its second derivatives are all zero. The simplest functions to have interesting critical points are the quadratic functions, which we write in the form (the 2's will be explained momentarily):

$$w = w_0 + ax + by + \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{8}$$

Such a function has in general a unique critical point, which we will assume is $(0,0)$; this gives the function a special form, which we can determine by evaluating its partial derivatives at $(0,0)$:

$$\begin{array}{ll} (w_x)_0 = a & w_{xx} = A \\ (w_y)_0 = b & w_{xy} = B \\ & w_{yy} = C \end{array} \tag{9}$$

(The neat look of the above explains the $\frac{1}{2}$ and $2B$ in (8).) Since $(0,0)$ is a critical point, (9) shows that $a = 0$ and $b = 0$, so our quadratic function has the form

$$w - w_0 = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{10}$$

We moved w_0 to the left side since the tangent plane at $(0,0)$ is the horizontal plane $w = w_0$, and we are interested in whether the graph of the quadratic function lies above or below this tangent plane, i.e., whether $w - w_0 > 0$ or $w - w_0 < 0$ at points other than the origin.

If $(x,y) \neq (0,0)$, then either $x \neq 0$ or $y \neq 0$; say $y \neq 0$. Then we write (10) as

$$w - w_0 \ = \ \frac{y^2}{2}\left[A\left(\frac{x}{y}\right)^2 + 2B\left(\frac{x}{y}\right) + C\right] \tag{11}$$

We know that $y^2 > 0$ if $y \neq 0$; applying our previous lemma to the factor on the right of (11), (or if $y = 0$, switching the roles of x and y in (11) and applying the lemma), we get

$$\begin{array}{rcl} AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 & \Rightarrow & w - w_0 > 0 \quad \text{for all } (x,y) \neq (0,0); \\ & \Rightarrow & (0,0) \text{ is a minimum point;} \\ AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 & \Rightarrow & w - w_0 < 0 \quad \text{for all } (x,y) \neq (0,0); \\ & \Rightarrow & (0,0) \text{ is a maximum point;} \\ AC - B^2 < 0 & \Rightarrow & \begin{cases} w - w_0 > 0, & \text{for some } (x,y); \\ w - w_0 < 0, & \text{for some } (x,y); \end{cases} \\ & \Rightarrow & (0,0) \text{ is a saddle point.} \end{array}$$

Argument for the Second-derivative Test for a general function.

This part won't be rigorous, only suggestive, but it will give the right idea.

We consider a general function $w = f(x, y)$, and assume it has a critical point at (x_0, y_0), and continuous second derivatives in the neighborhood of the critical point. Then by a generalization of Taylor's formula to functions of several variables, the function has a best quadratic approximation at the critical point. To simplify the notation, we will move the critical point to the origin by making the change of variables

$$u = x - x_0, \qquad v = y - y_0.$$

Then the best quadratic approximation is (if the x, y on the left and u, v on the right is upsetting, just imagine u and v replaced everywhere by $x - x_0$ and $y - y_0$):

$$w \;=\; f(x,y) \;\approx\; w_0 + \frac{1}{2}\big(Au^2 + 2Buv + Cv^2\big); \tag{13}$$

here the coefficients A, B, C are given as in (1) by the second partial derivatives with respect to u and v at $(0, 0)$, or what is the same (according to the chain rule—see the footnote below), by the second partial derivatives with respect to x and y at (x_0, y_0).

> (Intuitively, one can see the coefficients have these values by differentiating both sides of (13) and pretending the approximation is an equality. There are no linear terms in u and v on the right since $(0, 0)$ is a critical point.)

Since the quadratic function on the right of (13) is the best approximation to $w = f(x, y)$ for (x, y) close to (x_0, y_0), it is reasonable to suppose that their graphs are essentially the same near (x_0, y_0), so that if the quadratic function has a maximum, minimum or saddle point there, so will $f(x, y)$. Thus our results for the special case of a quadratic function having the origin as critical point carry over to the general function $f(x, y)$ at a critical point (x_0, y_0), if we interpret $A,\ B,\ C$ as the second partial derivatives at (x_0, y_0).
This is what the second derivative test says. □

Footnote: Using $u = x - x_0$ and $v = y - y_0$, we can apply the chain rule for partial derivatives, which tells us that for all x, y and the corresponding u, v, we have

$$w_x = w_u \frac{\partial u}{\partial x} + w_v \frac{\partial v}{\partial x} = w_u, \ \text{ since } u_x = 1 \text{ and } v_x = 0,$$

and similarly, $w_y = w_v$. Therefore at the corresponding points,

$$(w_x)_{(x_0,y_0)} = (w_u)_{(0,0)}, \qquad\qquad (w_y)_{(x_0,y_0)} = (w_v)_{(0,0)},$$

and differentiating once more and using the same reasoning,

$$(w_{xx})_{(x_0,y_0)} = (w_{uu})_{(0,0)}, \qquad (w_{xy})_{(x_0,y_0)} = (w_{uv})_{(0,0)}, \qquad (w_{yy})_{(x_0,y_0)} = (w_{vv})_{(0,0)}.$$

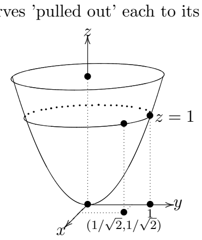
Level Curves and Contour Plots

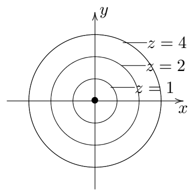
Level curves and *contour plots* are another way of visualizing functions of two variables. If you have seen a topographic map then you have seen a contour plot.

Example: To illustrate this we first draw the graph of $z = x^2 + y^2$. On this graph we draw *contours*, which are curves at a fixed height $z =$ constant.

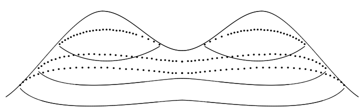
For example the curve at height $z = 1$ is the circle $x^2 + y^2 = 1$. On the graph we have to draw this at the correct height. Another way to show this is to draw the curves in the xy-plane and label them with their z-value. We call these curves *level curves* and the entire plot is called a *contour plot.*

For this example they are shown in the plot on the right. Notice that the 3D graph is simply the level curves 'pulled out' each to its correct height.

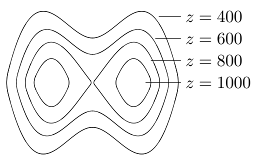




Here is another plot of a 'mountain pass'. Notice that in the contour plot the mountain pass is represented by a level curve that crosses itself. Moving up or down from the cross level curves heights decrease and moving right or left in the other they increase.



Mountain pass

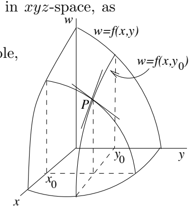


Level curves

Partial derivatives

Partial derivatives

Let $w = f(x, y)$ be a function of two variables. Its graph is a surface in xyz-space, as pictured.



Fix a value $y = y_0$ and just let x vary. You get a function of *one* variable,

(1) $$w = f(x, y_0), \qquad \text{the } \textbf{partial function} \text{ for } y = y_0.$$

Its graph is a curve in the vertical plane $y = y_0$, whose slope at the point P where $x = x_0$ is given by the derivative

(2) $$\frac{d}{dx} f(x, y_0)\bigg|_{x_0}, \quad \text{or} \quad \frac{\partial f}{\partial x}\bigg|_{(x_0,y_0)}.$$

We call (2) the **partial derivative** of f with respect to x at the point (x_0, y_0); the right side of (2) is the standard notation for it. The partial derivative is just the ordinary derivative of the partial function — it is calculated by holding one variable fixed and differentiating with respect to the other variable. Other notations for this partial derivative are

$$f_x(x_0, y_0), \qquad \frac{\partial w}{\partial x}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0;$$

the first is convenient for including the specific point; the second is common in science and engineering, where you are just dealing with relations between variables and don't mention the function explicitly; the third and fourth indicate the point by just using a single subscript.

Analogously, fixing $x = x_0$ and letting y vary, we get the partial function $w = f(x_0, y)$, whose graph lies in the vertical plane $x = x_0$, and whose slope at P is the *partial derivative of f with respect to y*; the notations are

$$\frac{\partial f}{\partial y}\bigg|_{(x_0,y_0)}, \qquad f_y(x_0, y_0), \qquad \frac{\partial w}{\partial y}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial y}\right)_0, \qquad \left(\frac{\partial w}{\partial y}\right)_0.$$

The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ depend on (x_0, y_0) and are therefore functions of x and y.

Written as $\partial w/\partial x$, the partial derivative gives the rate of change of w with respect to x alone, at the point (x_0, y_0): it tells how fast w is increasing as x increases, when y is held constant.

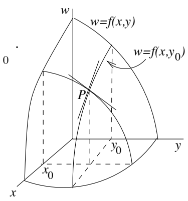
For a function of three or more variables, $w = f(x, y, z, \dots)$, we cannot draw graphs any more, but the idea behind partial differentiation remains the same: to define the partial derivative with respect to x, for instance, hold all the other variables constant and take the ordinary derivative with respect to x; the notations are the same as above:

$$\frac{d}{dx} f(x, y_0, z_0, \dots) \;=\; f_x(x_0, y_0, z_0, \dots), \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0.$$

The Tangent Approximation

1. The tangent plane.

For a function of one variable, $w = f(x)$, the tangent line to its graph at a point (x_0, w_0) is the line passing through (x_0, w_0) and having slope $\left(\dfrac{dw}{dx}\right)_0$.



For a function of two variables, $w = f(x, y)$, the natural analogue is the *tangent plane* to the graph, at a point (x_0, y_0, w_0). What's the equation of this tangent plane? Referring to the picture at right (this figure was also used when we introduced partial derivatives), we see that the tangent plane

(i) must pass through (x_0, y_0, w_0), where $w_0 = f(x_0, y_0)$;

(ii) must contain the tangent lines to the graphs of the two partial functions — this will hold if the plane has the same slopes in the $\mathbf{i}$ and $\mathbf{j}$ directions as the surface does.

Using these two conditions, it is easy to find the equation of the tangent plane. The general equation of a plane through (x_0, y_0, w_0) is

$$A(x - x_0) + B(y - y_0) + C(w - w_0) \;=\; 0 \;.$$

Assume the plane is not vertical; then $C \neq 0$, so we can divide through by C and solve for $w - w_0$, getting

$$(3) \qquad w - w_0 \;=\; a(x - x_0) + b(y - y_0), \qquad a = A/C,\; b = B/C.$$

The plane passes through (x_0, y_0, w_0); what values of the coefficients a and b will make it also tangent to the graph there? We have

$$\begin{aligned} a \;&=\; \text{slope of plane (3) in the } \mathbf{i}\text{-direction} && \text{(by putting } y = y_0 \text{ in (3));} \\ &=\; \text{slope of graph in the } \mathbf{i}\text{-direction,} && \text{(by (ii) above)} \\ &=\; \left(\frac{\partial w}{\partial x}\right)_0 ; && \text{(by the definition of partial derivative);} \quad \text{similarly,} \\ b \;&=\; \left(\frac{\partial w}{\partial y}\right)_0 . \end{aligned}$$

Therefore the equation of the **tangent plane** to $w = f(x, y)$ at (x_0, y_0) is

$$(4) \qquad w - w_0 \;=\; \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)$$

2. The approximation formula.

The most important use for the tangent plane is to give an approximation that is the basic formula in the study of functions of several variables — almost everything follows in one way or another from it.

The intuitive idea is that if we stay near (x_0, y_0, w_0), the graph of the tangent plane (4) will be a good approximation to the graph of the function $w = f(x, y)$. Therefore if the point (x, y) is close to (x_0, y_0),

$$(5) \qquad \underset{\text{height of graph}}{f(x,y)} \quad \approx \quad \underbrace{w_0 + \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)}_{\text{height of tangent plane}}$$

The function on the right side of (5) whose graph is the tangent plane is often called the **linearization** of $f(x, y)$ at (x_0, y_0): it is the linear function which gives the best approximation to $f(x, y)$ for values of (x, y) close to (x_0, y_0).

An equivalent form of the approximation (5) is obtained by using Δ notation; if we put

$$\Delta x \;=\; x - x_0, \qquad \Delta y \;=\; y - y_0, \qquad \Delta w \;=\; w - w_0,$$

then (5) becomes

$$(6) \qquad \Delta w \;\approx\; \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad \text{if} \quad \Delta x \approx 0,\ \Delta y \approx 0\ .$$

This formula gives the approximate change in w when we make a small change in x and y. We will use it often.

The analogous approximation formula for a function $w = f(x, y, z)$ of three variables would be

$$(7) \qquad \Delta w \;\approx\; \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y + \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \qquad \text{if} \quad \Delta x, \Delta y \Delta z \approx 0\ .$$

Unfortunately, for functions of three or more variables, we can't use a geometric argument for the approximation formula (7); for this reason, it's best to recast the argument for (6) in a form which doesn't use tangent planes and geometry, and therefore can be generalized to several variables. This is done at the end of this Chapter TA; for now let's just assume the truth of (7) and its higher-dimensional analogues.

Here are two typical examples of the use of the approximation formula. Other examples are in the Exercises. In the rest of your study of partial differentiation, you will see how the approximation formula is used to derive the important theorems and formulas.

Example 1. Give a reasonable square, centered at $(1, 1)$, over which the value of $w = x^3 y^4$ will not vary by more than $\pm\ .1$.

Solution. We use (6). We calculate for the two partial derivatives

$$w_x = 3x^2 y^4 \qquad\qquad w_y = 4x^3 y^3$$

and therefore, evaluating the partials at $(1, 1)$ and using (6), we get

$$\Delta w \;\approx\; 3\Delta x + 4\Delta y.$$

Thus if $|\Delta x| \leq .01$ and $\Delta y| \leq .01$, we should have

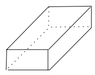
$$|\Delta w| \;\leq\; 3|\Delta x| + 4|\Delta y| \;\leq\; .07,$$

which is within the bounds. So the answer is the square with center at (1,1) given by

$$|x - 1| \leq .01, \qquad |y - 1| \leq .01\ .$$

Example 2. The sides a, b, c of a rectangular box have lengths measured to be respectively 1, 2, and 3. To which of these measurements is the volume V most sensitive?

Solution. $V = abc$, and therefore by the approximation formula (7),



$$\begin{aligned} \Delta V &\approx bc\,\Delta a + ac\,\Delta b + ab\,\Delta c \\ &\approx 6\,\Delta a + 3\,\Delta b + 2\,\Delta c, \qquad \text{at } (1,2,3); \end{aligned}$$

thus it is most sensitive to small changes in side a, since Δa occurs with the largest coefficient. (That is, if one at a time the measurement of each side were changed by say .01, it is the change in a which would produce the biggest change in V, namely .06 .)

The result may seem paradoxical — the value of V is most sensitive to the length of the *shortest* side — but it's actually intuitive, as you can see by thinking about how the box looks.

Sensitivity Principle *The numerical value of $w = f(x, y, \dots)$, calculated at some point $(x_0, y_0, \dots)$, will be most sensitive to small changes in that variable for which the corresponding partial derivative $w_x, w_y, \dots$ has the largest absolute value at the point.*

Critical Points

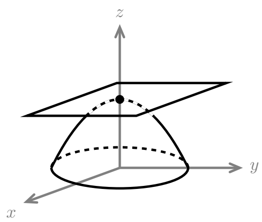
Critical points:
A standard question in calculus, with applications to many fields, is to find the points where a function reaches its relative maxima and minima.

Just as in single variable calculus we will look for maxima and minima (collectively called *extrema*) at points (x_0, y_0) where the first derivatives are 0. Accordingly we define a *critical point* as any point (x_0, y_0) where

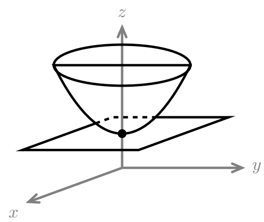
$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \text{ and } \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Often we will abbreviate this as $f_x = 0$ and $f_y = 0$.

Our first job is to verify that relative maxima and minima occur at critical points. The figures below illustrates that they occur at places where the tangent plane is horizontal.



Max. with horizontal tang. plane



Min. with horizontal tang. plane

Since horizontal planes are of the form $z =$ constant. and the equation of the tangent plane at (x_0, y_0, z_0) is

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

we see it is horizontal when

$$f_x(x_0, y_0) = 0 \text{ and } f_y(x_0, y_0) = 0.$$

Thus, extrema occur at critical points. But, just as in single variable calculus, not all critical points are extrema.

Example: Find the critical points of $z = x^2 + y^2 + .5$.

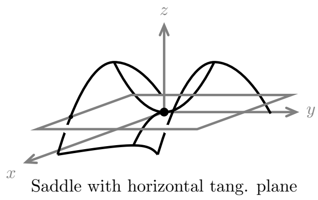
Answer: $\dfrac{\partial z}{\partial x} = 2x$ and $\dfrac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a minimum value. (See the figure above on the right.)

Example: Find the critical points of $z = 1 - x^2 - y^2$.

Answer: $\dfrac{\partial z}{\partial x} = -2x$ and $\dfrac{\partial z}{\partial y} = -2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a maximum value. (See the figure above on the left.)

Example: Find the critical points of $z = -x^2 + y^2$.

<u>Answer:</u> $\dfrac{\partial z}{\partial x} = -2x$ and $\dfrac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0,0)$. Thus, there is a single critical point at $(0,0)$. The figure shows it is neither a minimum or a maximum.



Saddle with horizontal tang. plane

Example: Making a box with minimum material.

A box is made of cardboard with double thick sides, a triple thick bottom, single thick front and back and no top. It's volume = 3.

What dimensions use the least amount of cardboard?

<u>Answer:</u> The box shown has dimensions x, y, and z.

The area of one side $= yz$. There are two double thick sides $\Rightarrow$ cardboard used $= 4yz$.

The area of the front (and back) $= xz$. It is single thick $\Rightarrow$ cardboard used $= 2xz$.

The area of the bottom $= xy$. It is triple thick $\Rightarrow$ cardboard used $= 3xy$.

Thus, the total cardboard used is

$$w = 4yz + 2xz + 3xy.$$

The volume $= 3 = xyz \Rightarrow z = \frac{3}{xy}$. Substituting this in the formula for w gives

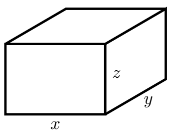
$$w = \frac{12}{x} + \frac{6}{y} + 3xy.$$

We find the critical points of w.

$$w_x = -\frac{12}{x^2} + 3y = 0, \quad w_y = -\frac{6}{y^2} + 3x = 0.$$

The first equation implies $y = \dfrac{4}{x^2}$. Substituting this in the second equation gives $-\dfrac{6}{16}x^4 + 3x = 0$. Thus, $x = 0$ or 2. We reject 0 since then y is undefined. Using $x = 2$ we find $y = 1$. Thus, there there is one critical point at (2,1). and at this point we have $z = 3/2$.

This point gives the box with minimum cardboard used because physically we know it must have a minimum somewhere. Later we will learn to check this with the second derivative test.

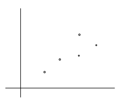


Least Squares Interpolation

1. The least-squares line.

Suppose you have a large number n of experimentally determined points, through which you want to pass a curve. There is a formula (the Lagrange interpolation formula) producing a polynomial curve of degree $n-1$ which goes through the points exactly. But normally one wants to find a simple curve, like a line, parabola, or exponential, which goes approximately through the points, rather than a high-degree polynomial which goes exactly through them. The reason is that the location of the points is to some extent determined by experimental error, so one wants a smooth-looking curve which averages out these errors, not a wiggly polynomial which takes them seriously.

In this section, we consider the most common case — finding a line which goes approximately through a set of data points.



Suppose the data points are

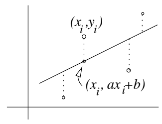
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

and we want to find the line

(1) $$y = ax + b$$

which "best" passes through them. Assuming our errors in measurement are distributed randomly according to the usual bell-shaped curve (the so-called "Gaussian distribution"), it can be shown that the right choice of a and b is the one for which the sum D of the squares of the deviations

(2) $$D \;=\; \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2$$



is a *minimum*. In the formula (2), the quantities in parentheses (shown by dotted lines in the picture) are the **deviations** between the observed values y_i and the ones $ax_i + b$ that would be predicted using the line (1).

The deviations are squared for theoretical reasons connected with the assumed Gaussian error distribution; note however that the effect is to ensure that we sum only positive quantities; this is important, since we do not want deviations of opposite sign to cancel each other out. It also weights more heavily the larger deviations, keeping experimenters honest, since they tend to ignore large deviations ("I had a headache that day").

This prescription for finding the line (1) is called the **method of least squares**, and the resulting line (1) is called the **least-squares** line or the **regression** line.

To calculate the values of a and b which make D a minimum, we see where the two partial derivatives are zero:

(3) $$\begin{aligned} \frac{\partial D}{\partial a} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) \;=\; 0 \\ \frac{\partial D}{\partial b} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) \;=\; 0\,. \end{aligned}$$

These give us a pair of *linear* equations for determining a and b, as we see by collecting terms and cancelling the 2's:

$$
\begin{aligned}
\left(\sum x_i^2\right) a \;+\; \left(\sum x_i\right) b \;&=\; \sum x_i y_i \\
\left(\sum x_i\right) a \;+\; n\, b \;&=\; \sum y_i \;.
\end{aligned}
\tag{4}
$$

(Notice that it saves a lot of work to differentiate (2) using the chain rule, rather than first expanding out the squares.)

The equations (4) are usually divided by n to make them more expressive:

$$
\begin{aligned}
\bar{s}\, a \;+\; \bar{x}\, b \;&=\; \frac{1}{n}\sum x_i y_i \\
\bar{x}\, a \;+\; b \;&=\; \bar{y}\;,
\end{aligned}
\tag{5}
$$

where $\bar{x}$ and $\bar{y}$ are the average of the x_i and y_i, and $\bar{s} = \sum x_i^2/n$ is the average of the squares.

From this point on use linear algebra to determine a and b. It is a good exercise to see that the equations are always solvable unless all the x_i are the same (in which case the best line is vertical and can't be written in the form (1)).

In practice, least-squares lines are found by pressing a calculator button, or giving a MatLab command. Examples of calculating a least-squares line are in the exercises accompanying the course. Do them from scratch, starting from (2), since the purpose here is to get practice with max-min problems in several variables; don't plug into the equations (5). Remember to differentiate (2) using the chain rule; don't expand out the squares, which leads to messy algebra and highly probable error.

2. Fitting curves by least squares.

If the experimental points seem to follow a curve rather than a line, it might make more sense to try to fit a second-degree polynomial

$$y = a_0 + a_1 x + a_2 x^2 \tag{6}$$

to them. If there are only three points, we can do this exactly (by the Lagrange interpolation formula). For more points, however, we once again seek the values of a_0, a_1, a_2 for which the sum of the squares of the deviations

$$D \;=\; \sum_1^n \big(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\big)^2 \tag{7}$$

is a minimum. Now there are three unknowns, a_0, a_1, a_2. Calculating (remember to use the chain rule!) the three partial derivatives $\partial D/\partial a_i,\ i = 0, 1, 2$, and setting them equal to zero leads to a square system of three linear equations; the a_i are the three unknowns, and the coefficients depend on the data points (x_i, y_i). They can be solved by finding the inverse matrix, elimination, or using a calculator or MatLab.

If the points seem to lie more and more along a line as $x \to \infty$, but lie on one side of the line for low values of x, it might be reasonable to try a function which has similar behavior, like

$$y \;=\; a_0 + a_1 x + a_2\, \frac{1}{x} \tag{8}$$

and again minimize the sum of the squares of the deviations, as in (7). In general, this method of least squares applies to a trial expression of the form

$$y = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_r f_r(x), \tag{9}$$

where the $f_i(x)$ are given functions (usually simple ones like $1, x, x^2, 1/x, e^{kx}$, etc. Such an expression (9) is called a **linear combination** of the functions $f_i(x)$. The method produces a square inhomogeneous system of linear equations in the unknowns $a_0, \ldots, a_r$ which can be solved by finding the inverse matrix to the system, or by elimination.

The method also applies to finding a linear function

$$z = a_1 + a_2 x + a_3 y \tag{10}$$

to fit a set of data points

$$(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n)\ . \tag{11}$$

where there are two independent variables x and y and a dependent variable z (this is the quantity being experimentally measured, for different values of (x, y)). This time after differentiation we get a 3×3 system of linear equations for determining a_1, a_2, a_3 .

The essential point in all this is that the unknown coefficients a_i should occur *linearly* in the trial function. Try fitting a function like ce^{kx} to data points by using least squares, and you'll see the difficulty right away. (Since this is an important problem — fitting an exponential to data points — one of the Exercises explains how to adapt the method to this type of problem.)

Least squares interpolation

1. Use the method of least squares to fit a line to the three data points

$$(0,0), \quad (1,2), \quad (2,1).$$

<u>Answer:</u> We are looking for the line $y = ax + b$ that best models the data. The deviation of a data point (x_i, y_i) from the model is

$$y_i - (ax_i + b).$$

By best we mean the line that minimizes the sum of the squares of the deviation. That is we want to minimize

$$\begin{aligned} D &= (0-(a\cdot 0+b))^2 + (2-(a\cdot 1+b))^2 + (1-(a\cdot 2+b))^2 \\ &= b^2 + (2-a-b)^2 + (1-2a-b)^2. \end{aligned}$$

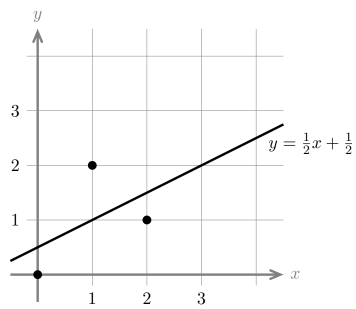
(Remember, the variables whose values are to be found are a and b.) We do not expand out the squares, rather we take the derivatives first. Setting the derivatives equal to 0 gives

$$\begin{aligned} &\tfrac{\partial D}{\partial a} = -2(2-a-b) - 4(1-2a-b) = 0 \;\Rightarrow\; 10a + 6b = 8 \;\Rightarrow\; 5a + 3b = 4 \\ &\tfrac{\partial D}{\partial b} = 2b - 2(2-a-b) - 2(1-2a-b) = 0 \;\Rightarrow\; 6a + 6b = 6 \;\Rightarrow\; 3a + 3b = 3. \end{aligned}$$

This linear system of two equations in two unknowns is easy to solve. We get

$$a = \frac{1}{2}, \qquad b = \frac{1}{2}.$$

Here is a plot of the problem.



Second Derivative Test

1. The Second Derivative Test

We begin by recalling the situation for twice differentiable functions $f(x)$ of one variable. To find their local (or "relative") maxima and minima, we

1. find the critical points, i.e., the solutions of $f'(x) = 0$;
2. apply the second derivative test to each critical point x_0:

$$f''(x_0) > 0 \;\Rightarrow\; x_0 \text{ is a local minimum point;}$$
$$f''(x_0) < 0 \;\Rightarrow\; x_0 \text{ is a local maximum point.}$$

The idea behind it is: at x_0 the slope $f'(x_0) = 0$; if $f''(x_0) > 0$, then $f'(x)$ is strictly increasing for x near x_0, so that the slope is negative to the left of x_0 and positive to the right, which shows that x_0 is a minimum point. The reasoning for the maximum point is similar.

If $f''(x_0) = 0$, the test fails and one has to investigate further, by taking more derivatives, or getting more information about the graph. Besides being a maximum or minimum, such a point could also be a horizontal point of inflection.

The analogous test for maxima and minima of functions of two variables $f(x, y)$ is a little more complicated, since there are several equations to satisfy, several derivatives to be taken into account, and another important geometric possibility for a critical point, namely a **saddle point**. This is a local minimax point; around such a point the graph of $f(x, y)$ looks like the central part of a saddle, or the region around the highest point of a mountain pass. In the neighborhood of a saddle point, the graph of the function lies both above and below its horizontal tangent plane at the point.

The second-derivative test for maxima, minima, and saddle points has two steps.

1. Find the critical points by solving the simultaneous equations $\begin{cases} f_x(x,y) = 0, \\ f_y(x,y) = 0. \end{cases}$

Since a critical point (x_0, y_0) is a solution to both equations, both partial derivatives are zero there, so that the tangent plane to the graph of $f(x, y)$ is horizontal.

2. To test such a point to see if it is a local maximum or minimum point, we calculate the three second derivatives at the point (we use subscript 0 to denote evaluation at (x_0, y_0), so for example $(f)_0 = f(x_0, y_0)$), and denote the values by A, B, and C:

$$A = (f_{xx})_0, \qquad B = (f_{xy})_0 = (f_{yx})_0, \qquad C = (f_{yy})_0, \tag{1}$$

(we are assuming the derivatives exist and are continuous).

Second-derivative test. *Let (x_0, y_0) be a critical point of $f(x, y)$, and A, B, and C be as in (1). Then*

$$\begin{aligned} AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \;&\Rightarrow\; (x_0, y_0) \textit{ is a minimum point;} \\ AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \;&\Rightarrow\; (x_0, y_0) \textit{ is a maximum point;} \\ AC - B^2 < 0 \;&\Rightarrow\; (x_0, y_0) \textit{ is a saddle point.} \end{aligned}$$

If $AC - B^2 = 0$, the test fails and more investigation is needed.

Note that if $AC - B^2 > 0$, then $AC > 0$, so that A and C must have the same sign.

Example 1. Find the critical points of $w = 12x^2 + y^3 - 12xy$ and determine their type.

Solution. We calculate the partial derivatives easily:

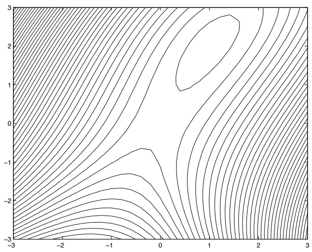
$$\begin{aligned} w_x &= 24x - 12y \\ w_y &= 3y^2 - 12x \end{aligned} \qquad \begin{aligned} A &= w_{xx} = 24 \\ B &= w_{xy} = -12 \\ C &= w_{yy} = 6y \end{aligned} \tag{2}$$

To find the critical points we solve simultaneously the equations $w_x = 0$ and $w_y = 0$; we get

$$\begin{matrix} w_x = 0 \\ w_y = 0 \end{matrix} \quad \Rightarrow \quad \begin{matrix} y = 2x \\ y^2 = 4x \end{matrix} \quad \Rightarrow \quad 4x^2 = 4x \quad \Rightarrow \quad x = 0, 1 \quad \Rightarrow \quad \begin{matrix} (x, y) = (0, 0) \\ (x, y) = (1, 2) \end{matrix}.$$

Thus there are two critical points: $(0, 0)$ and $(1, 2)$. To determine their type, we use the second derivative test: we have $AC - B^2 = 144y - 144$, so that

at $(0, 0)$, we have $AC - B^2 = -144$, so it is a saddle point;
at $(1, 2)$, we have $AC - B^2 = 144$ and $A > 0$, so it is a a minimum point.



A plot of the level curves is given at the right, which confirms the above. Note that the behavior of the level curves near the origin can be determined by using the approximation $w \approx 12x^2 - 12xy$; this shows the level curves near $(0, 0)$ look like those of the function $x(x - y)$: the family of hyperbolas $x(x - y) = c$, with asymptotes given by the degenerate hyperbola $x(x - y) = 0$, i.e., the pair of lines $x = 0$ (the y-axis) and $x - y = 0$ (the diagonal line $y = x$).

2. Justification for the Second-derivative Test.

The test involves the quantity $AC - B^2$. In general, whenever we see the expressions $B^2 - 4AC$ or $B^2 - AC$ or their negatives, it means the quadratic formula is involved, in one of its two forms (the second is often used to get rid of the excess two's):

$$Ax^2 + Bx + C = 0 \quad \Rightarrow \quad x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \tag{3}$$

$$Ax^2 + 2Bx + C = 0 \quad \Rightarrow \quad x = \frac{-B \pm \sqrt{B^2 - AC}}{A} \tag{4}$$

This is what is happening here. We want to know whether, near a critical point P_0, the graph of our function $w = f(x, y)$ always stays on one side of its horizontal tangent plane (P_0 is then a maximum or minimum point), or whether it lies partly above and partly below the tangent plane (P_0 is then a saddle point). As we will see, this is determined by how the graph of a quadratic function $f(x)$ lies with respect to the x-axis. Here is the basic lemma.

Lemma. For the quadratic function $Ax^2 + 2Bx + C$,

$$AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \quad \Rightarrow \quad Ax^2 + 2Bx + C > 0 \quad \text{for all } x; \tag{5}$$

$$AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \quad \Rightarrow \quad Ax^2 + 2Bx + C < 0 \quad \text{for all } x; \tag{6}$$

$$AC - B^2 < 0 \quad \Rightarrow \quad \begin{cases} Ax^2 + 2Bx + C > 0, & \text{for some } x; \\ Ax^2 + 2Bx + C < 0, & \text{for some } x. \end{cases} \tag{7}$$

Proof of the Lemma. To prove (5), we note that the quadratic formula in the form (4) shows that the zeros of $Ax^2 + 2Bx + C$ are imaginary, i.e., it has no real zeros. Therefore its graph must lie entirely on one side of the x-axis; which side can be determined from either A or C, since

$$A > 0 \;\Rightarrow\; \lim_{x\to\infty} Ax^2 + 2Bx + C = \infty; \qquad C > 0 \;\Rightarrow\; Ax^2 + 2Bx + C > 0 \text{ when } x = 0.$$

If $A < 0$ or $C < 0$, the reasoning is analogous and proves (6).

If on the other hand $AC - B^2 < 0$, formula (4) shows the quadratic function has two real roots, so that its parabolic graph crosses the x-axis twice, and hence lies partly above and partly below it. This proves (7). □

Proof of the Second-derivative Test in a special case.

The simplest function is a linear function, $w = w_0 + ax + by$, but it does not in general have maximum or minimum points and its second derivatives are all zero. The simplest functions to have interesting critical points are the quadratic functions, which we write in the form (the 2's will be explained momentarily):

$$w = w_0 + ax + by + \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{8}$$

Such a function has in general a unique critical point, which we will assume is $(0,0)$; this gives the function a special form, which we can determine by evaluating its partial derivatives at $(0,0)$:

$$\begin{array}{ll} (w_x)_0 = a & w_{xx} = A \\ (w_y)_0 = b & w_{xy} = B \\ & w_{yy} = C \end{array} \tag{9}$$

(The neat look of the above explains the $\frac{1}{2}$ and $2B$ in (8).) Since $(0,0)$ is a critical point, (9) shows that $a = 0$ and $b = 0$, so our quadratic function has the form

$$w - w_0 = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{10}$$

We moved w_0 to the left side since the tangent plane at $(0,0)$ is the horizontal plane $w = w_0$, and we are interested in whether the graph of the quadratic function lies above or below this tangent plane, i.e., whether $w - w_0 > 0$ or $w - w_0 < 0$ at points other than the origin.

If $(x,y) \neq (0,0)$, then either $x \neq 0$ or $y \neq 0$; say $y \neq 0$. Then we write (10) as

$$w - w_0 \;=\; \frac{y^2}{2}\left[A\left(\frac{x}{y}\right)^2 + 2B\left(\frac{x}{y}\right) + C\right] \tag{11}$$

We know that $y^2 > 0$ if $y \neq 0$; applying our previous lemma to the factor on the right of (11), (or if $y = 0$, switching the roles of x and y in (11) and applying the lemma), we get

$$\begin{array}{rcl}
AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 & \Rightarrow & w - w_0 > 0 \quad \text{for all } (x,y) \neq (0,0); \\
& \Rightarrow & (0,0) \text{ is a minimum point;} \\
AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 & \Rightarrow & w - w_0 < 0 \quad \text{for all } (x,y) \neq (0,0); \\
& \Rightarrow & (0,0) \text{ is a maximum point;} \\
AC - B^2 < 0 & \Rightarrow & \begin{cases} w - w_0 > 0, & \text{for some } (x,y); \\ w - w_0 < 0, & \text{for some } (x,y); \end{cases} \\
& \Rightarrow & (0,0) \text{ is a saddle point.}
\end{array}$$

Argument for the Second-derivative Test for a general function.

This part won't be rigorous, only suggestive, but it will give the right idea.

We consider a general function $w = f(x, y)$, and assume it has a critical point at (x_0, y_0), and continuous second derivatives in the neighborhood of the critical point. Then by a generalization of Taylor's formula to functions of several variables, the function has a best quadratic approximation at the critical point. To simplify the notation, we will move the critical point to the origin by making the change of variables

$$u = x - x_0, \qquad v = y - y_0.$$

Then the best quadratic approximation is (if the x, y on the left and u, v on the right is upsetting, just imagine u and v replaced everywhere by $x - x_0$ and $y - y_0$):

$$w \;=\; f(x,y) \;\approx\; w_0 + \frac{1}{2}\big(Au^2 + 2Buv + Cv^2\big); \tag{13}$$

here the coefficients A, B, C are given as in (1) by the second partial derivatives with respect to u and v at $(0, 0)$, or what is the same (according to the chain rule—see the footnote below), by the second partial derivatives with respect to x and y at (x_0, y_0).

(Intuitively, one can see the coefficients have these values by differentiating both sides of (13) and pretending the approximation is an equality. There are no linear terms in u and v on the right since $(0, 0)$ is a critical point.)

Since the quadratic function on the right of (13) is the best approximation to $w = f(x, y)$ for (x, y) close to (x_0, y_0), it is reasonable to suppose that their graphs are essentially the same near (x_0, y_0), so that if the quadratic function has a maximum, minimum or saddle point there, so will $f(x, y)$. Thus our results for the special case of a quadratic function having the origin as critical point carry over to the general function $f(x, y)$ at a critical point (x_0, y_0), if we interpret $A,\ B,\ C$ as the second partial derivatives at (x_0, y_0).
This is what the second derivative test says. □

Footnote: Using $u = x - x_0$ and $v = y - y_0$, we can apply the chain rule for partial derivatives, which tells us that for all x, y and the corresponding u, v, we have

$$w_x = w_u \frac{\partial u}{\partial x} + w_v \frac{\partial v}{\partial x} = w_u, \ \text{ since } u_x = 1 \text{ and } v_x = 0,$$

and similarly, $w_y = w_v$. Therefore at the corresponding points,

$$(w_x)_{(x_0,y_0)} = (w_u)_{(0,0)}, \qquad\qquad (w_y)_{(x_0,y_0)} = (w_v)_{(0,0)},$$

and differentiating once more and using the same reasoning,

$$(w_{xx})_{(x_0,y_0)} = (w_{uu})_{(0,0)}, \qquad (w_{xy})_{(x_0,y_0)} = (w_{uv})_{(0,0)}, \qquad (w_{yy})_{(x_0,y_0)} = (w_{vv})_{(0,0)}.$$

Second derivative test

1. Find and classify all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

Answer: Taking the first partials and setting them to 0:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 3y^2 - 12 = 0.$$

The first equation implies $x = -1$ and the second implies $y = \pm 2$. Thus, the critical points are $(-1, 2)$ and $(-1, -2)$.

Taking second partials:

$$\frac{\partial^2 z}{\partial x^2} = 30x^4, \quad \frac{\partial^2 z}{\partial xy} = 0, \quad \frac{\partial^2 z}{\partial y^2} = 6y.$$

We analyze each critical point in turn.
At (-1,-2): $A = z_{xx}(-1, -2) = 30, \quad B = z_{xy}(-1, -2) = 0, \quad C = z_{yy}(-1, -2) = -12.$
Therefore $AC - B^2 = -360 < 0$, which implies the critical point is a saddle.

At (-1,2): $A = z_{xx}(-1, 2) = 30, \quad B = z_{xy}(-1, 2) = 0, \quad C = z_{yy}(-1, 2) = 12.$
Therefore $AC - B^2 = 360 > 0$ and $A > 0$, which implies the critical point is a minimum.

Chain rule

Now we will formulate the chain rule when there is more than one independent variable.

We suppose w is a function of x, y and that x, y are functions of u, v. That is,

$$w = f(x, y) \text{ and } x = x(u, v),\ y = y(u, v).$$

The use of the term chain comes because to compute w we need to do a chain of computations

$$(u, v) \to (x, y) \to w.$$

We will say w is a *dependent* variable, u and v are *independent* variables and x and y are *intermediate* variables.

Since w is a function of x and y it has partial derivatives $\dfrac{\partial w}{\partial x}$ and $\dfrac{\partial w}{\partial y}$.

Since, ultimately, w is a function of u and v we can also compute the partial derivatives $\dfrac{\partial w}{\partial u}$ and $\dfrac{\partial w}{\partial v}$. The chain rule relates these derivatives by the following formulas.

$$\begin{aligned}
\frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} \\
\frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v}.
\end{aligned}$$

Example: Given $w = x^2y + y^2 + x$, $x = u^2v$, $y = uv^2$ find $\dfrac{\partial w}{\partial u}$.

Answer: First we compute

$$\frac{\partial w}{\partial x} = 2xy + 1,\ \frac{\partial w}{\partial y} = x^2 + 2y,\ \frac{\partial x}{\partial u} = 2uv,\ \frac{\partial y}{\partial u} = v^2,\ \frac{\partial x}{\partial v} = u^2,\ \frac{\partial y}{\partial v} = 2uv.$$

The chain rule then implies

$$\begin{aligned}
\frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} \\
&= (2xy + 1)2uv + (x^2 + 2y)v^2 \\
\frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v} \\
&= (2xy + 1)u^2 + (x^2 + 2y)2uv.
\end{aligned}$$

Often, it is okay to leave the variables mixed together. If, for example, you wanted to compute $\frac{\partial w}{\partial u}$ when $(u, v) = (1, 2)$ all you have to do is compute x and y and use these values, along with u, v, in the formula for $\dfrac{\partial w}{\partial u}$.

$$x = 2,\ y = 4 \ \Rightarrow\ \frac{\partial w}{\partial u} = (5)(4) + (12)(4) = 68.$$

If you actually need the derivatives expressed in just the variables u and v then you would have to substitute for x, y and z.

Proof of the chain rule:
Just as before our argument starts with the tangent approximation at the point (x_0, y_0).

$$\Delta w \approx \left.\frac{\partial w}{\partial x}\right|_o \Delta x + \left.\frac{\partial w}{\partial y}\right|_o \Delta y.$$

Now hold v constant and divide by Δu to get

$$\frac{\Delta w}{\Delta u} \approx \left.\frac{\partial w}{\partial x}\right|_o \frac{\Delta x}{\Delta u} + \left.\frac{\partial w}{\partial y}\right|_o \frac{\Delta y}{\Delta u}.$$

Finally, letting $\Delta u \to 0$ gives the chain rule for $\dfrac{\partial w}{\partial u}$.

Ambiguous notation

Often you have to figure out the dependent and independent variables from context.

Thermodynamics is a big player here. It has, for example, the variables P, T, V, U, S. and *any* two can be taken to be independent and the others are functions of those two.

We will do more with this topic in the future.

Gradient: definition and properties

Definition of the gradient

If $w = f(x,y)$, then $\dfrac{\partial w}{\partial x}$ and $\dfrac{\partial w}{\partial y}$ are the rates of change of w in the $\mathbf{i}$ and $\mathbf{j}$ directions.

It will be quite useful to put these two derivatives together in a vector called the *gradient* of w.

$$\operatorname{grad} w = \left\langle \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y} \right\rangle.$$

We will also use the symbol $\boldsymbol{\nabla} w$ to denote the gradient. (You read this as 'gradient of w' or 'grad w'.)

Of course, if we specify a point $P_0 = (x_0, y_0)$, we can evaluate the gradient at that point. We will use several notations for this

$$\operatorname{grad} w(x_0, y_0) = \boldsymbol{\nabla}\, w|_{P_0} = \boldsymbol{\nabla}\, w|_o = \left\langle \left.\frac{\partial w}{\partial x}\right|_o, \left.\frac{\partial w}{\partial y}\right|_o \right\rangle.$$

Note well the following: (as we look more deeply into properties of the gradient these can be points of confusion).

1. The gradient takes a scalar function $f(x,y)$ and produces a vector $\boldsymbol{\nabla} f$.
2. The vector $\boldsymbol{\nabla} f(x,y)$ lies in the plane.

For functions $w = f(x,y,z)$ we have the gradient

$$\operatorname{grad} w = \boldsymbol{\nabla} w = \left\langle \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial w}{\partial z} \right\rangle.$$

That is, the gradient takes a scalar function of three variables and produces a three dimensional vector.

The gradient has many geometric properties. In the next session we will prove that for $w = f(x,y)$ the gradient is perpendicular to the level curves $f(x,y) = c$. We can show this by direct computation in the following example.

Example 1: Compute the gradient of $w = (x^2 + y^2)/3$ and show that the gradient at $(x_0, y_0) = (1,2)$ is perpendicular to the level curve through that point.

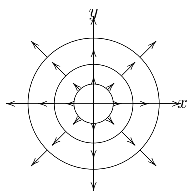
<u>Answer:</u> The gradient is easily computed

$$\boldsymbol{\nabla} w = \langle 2x/3, 2y/3 \rangle = \frac{2}{3}\langle x, y \rangle.$$

At $(1,2)$ we get $\boldsymbol{\nabla} w(1,2) = \frac{2}{3}\langle 1, 2\rangle$. The level curve through $(1,2)$ is

$$(x^2 + y^2)/3 = 5/3,$$

which is identical to $x^2 + y^2 = 5$. That is, it is a circle of radius $\sqrt{5}$ centered at the origin. Since the gradient at (1,2) is a multiple of $\langle 1, 2\rangle$, it points radially outward and hence is perpendicular to the circle. Below is a figure showing the gradient field and the level curves.



$z = (x^2 + y^2)/3$

Example 2: Consider the graph of $y = \mathrm{e}^x$. Find a vector perpendicular to the tangent to $y = \mathrm{e}^x$ at the point $(1, \mathrm{e})$.

Old method: Find the slope take the negative reciprocal and make the vector.

New method: This graph is the level curve of $w = y - \mathrm{e}^x = 0$.

$\boldsymbol{\nabla} w = \langle -\mathrm{e}^x, 1\rangle \;\Rightarrow\;$ (at $x = 1$) $\boldsymbol{\nabla} w(1, e) = \langle -\mathrm{e}, 1\rangle$ is perpendicular to the tangent vector to the graph, $\mathbf{v} = \langle 1, \mathrm{e}\rangle$.

Higher dimensions

Similarly, for $w = f(x, y, z)$ we get level surfaces $f(x, y, z) = c$. The gradient is perpendicular to the level surfaces.

Example 3: Find the tangent plane to the surface $x^2 + 2y^2 + 3z^2 = 6$ at the point $P = (1, 1, 1)$.

Answer: Introduce a new variable

$$w = x^2 + 2y^2 + 3z^2.$$

Our surface is the level surface $w = 6$. Saying the gradient is perpendicular to the surface means exactly the same thing as saying it is normal to the tangent plane. Computing

$$\boldsymbol{\nabla} w = \langle 2x, 4y, 6z\rangle \;\Rightarrow\; \boldsymbol{\nabla} w|_P = \langle 2, 4, 6\rangle.$$

Using point normal form we get the equation of the tangent plane is

$$2(x - 1) + 4(y - 1) + 6(z - 1) = 0, \quad \text{or} \quad 2x + 4y + 6z = 12.$$

Gradient: proof that it is perpendicular to level curves and surfaces

Let $w = f(x, y, z)$ be a function of 3 variables. We will show that at any point $P = (x_0, y_0, z_0)$ on the level surface $f(x, y, z) = c$ (so $f(x_0, y_0, z_0) = c$) the gradient $\nabla f|_P$ is perpendicular to the surface.

By this we mean it is perpendicular to the tangent to any curve that lies on the surface and goes through P. (See figure.)

This follows easily from the chain rule: Let

$$\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle$$

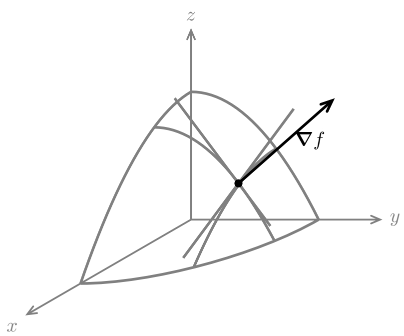
be a curve on the level surface with $\mathbf{r}(t_0) = \langle x_0, y_0, z_0 \rangle$. We let $g(t) = f(x(t), y(t), z(t))$. Since the curve is on the level surface we have $g(t) = f(x(t), y(t), z(t)) = c$. Differentiating this equation with respect to t gives

$$\frac{dg}{dt} = \left.\frac{\partial f}{\partial x}\right|_P \left.\frac{dx}{dt}\right|_{t_0} + \left.\frac{\partial f}{\partial y}\right|_P \left.\frac{dy}{dt}\right|_{t_0} + \left.\frac{\partial f}{\partial z}\right|_P \left.\frac{dz}{dt}\right|_{t_0} = 0.$$

In vector form this is

$$\left\langle \left.\frac{\partial f}{\partial x}\right|_P, \left.\frac{\partial f}{\partial y}\right|_P, \left.\frac{\partial f}{\partial z}\right|_P \right\rangle \cdot \left\langle \left.\frac{dx}{dt}\right|_{t_0}, \left.\frac{dy}{dt}\right|_{t_0}, \left.\frac{dz}{dt}\right|_{t_0} \right\rangle = 0$$
$$\Leftrightarrow \quad \nabla f|_P \cdot \mathbf{r}'(t_0) = 0.$$

Since the dot product is 0, we have shown that the gradient is perpendicular to the tangent to any curve that lies on the level surface, which is exactly what we needed to show.



Tangent Plane to a Level Surface

1. Find the tangent plane to the surface $x^2 + 2y^2 + 3z^2 = 36$ at the point $P = (1, 2, 3)$.

Answer: In order to use gradients we introduce a new variable

$$w = x^2 + 2y^2 + 3z^2.$$

Our surface is then the the level surface $w = 36$. Therefore the normal to surface is

$$\nabla w = \langle 2x, 4y, 6z \rangle.$$

At the point P we have $\nabla w|_P = \langle 2, 8, 18 \rangle$. Using point normal form, the equation of the tangent plane is

$$2(x-1) + 8(y-2) + 18(z-3) = 0, \text{ or equivalently } 2x + 8y + 18z = 72.$$

2. Use gradients and level surfaces to find the normal to the tangent plane of the graph of $z = f(x, y)$ at $P = (x_0, y_0, z_0)$.

Answer: Introduce the new variable

$$w = f(x, y) - z.$$

The graph of $z = f(x, y)$ is just the level surface $w = 0$. We compute the normal to the surface to be

$$\nabla w = \langle f_x, f_y, -1 \rangle.$$

At the the point P the normal is $\langle f_x(x_0, y_0), f_y(x_0, y_0), -1 \rangle$, so the equation of the tangent plane is

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) - (z - z_0) = 0.$$

We can write this in a more compact form as

$$(z - z_0) = \left.\frac{\partial f}{\partial x}\right|_0 (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_0 (y - y_0),$$

which is exactly the formula we saw earlier for the tangent plane to a graph.

Directional Derivatives

Directional derivative

Like all derivatives the *directional derivative* can be thought of as a ratio. Fix a unit vector $\mathbf{u}$ and a point P_0 in the *plane*. The **directional derivative** of w at P_0 in the direction $\mathbf{u}$ is defined as

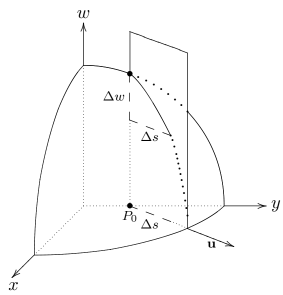
$$\left.\frac{dw}{ds}\right|_{P_0,\mathbf{u}} = \lim_{\Delta s \to 0} \frac{\Delta w}{\Delta s}.$$

Here Δw is the change in w caused by a step of length Δs in the direction of $\mathbf{u}$ (all in the xy-plane).

Below we will show that

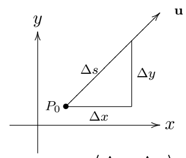
$$\left.\frac{dw}{ds}\right|_{P_0,\mathbf{u}} = \boldsymbol{\nabla} w(P_0) \cdot \mathbf{u}. \tag{1}$$

We illustrate this with a figure showing the graph of $w = f(x, y)$. Notice that Δs is measured in the plane and Δw is the change of w on the graph.



Proof of equation 1

The figure below represents the change in position from P_0 resulting from taking a step of size Δs in the $\mathbf{u}$ direction.



Since $(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2$ we have that $\left\langle \frac{\Delta x}{\Delta s}, \frac{\Delta y}{\Delta s} \right\rangle$. is a unit vector, so

$$\mathbf{u} = \left\langle \frac{\Delta x}{\Delta s}, \frac{\Delta y}{\Delta s} \right\rangle.$$

The tangent plane approximation at P_0 is

$$\Delta w \approx \left.\frac{\partial w}{\partial x}\right|_{P_0} \Delta x + \left.\frac{\partial w}{\partial y}\right|_{P_0} \Delta y$$

Dividing this approximation by Δs gives

$$\frac{\Delta w}{\Delta s} \approx \left.\frac{\partial w}{\partial x}\right|_{P_0} \frac{\Delta x}{\Delta s} + \left.\frac{\partial w}{\partial y}\right|_{P_0} \frac{\Delta y}{\Delta s}.$$

We can rewrite this as a dot product

$$\frac{\Delta w}{\Delta s} \approx \left\langle \left.\frac{\partial w}{\partial x}\right|_{P_0}, \left.\frac{\partial w}{\partial y}\right|_{P_0} \right\rangle \cdot \left\langle \frac{\Delta x}{\Delta s}, \frac{\Delta y}{\Delta s} \right\rangle.$$

In the dot product the first term is $\nabla w|_{P_0}$ and the second is just $\mathbf{u}$, so,

$$\frac{\Delta w}{\Delta s} \approx \nabla w|_{P_0} \cdot \mathbf{u}.$$

Now taking the limit we get equation (1).

Example: (Algebraic example) Let $w = x^3 + 3y^2$.
Compute $\frac{dw}{ds}$ at $P_0 = (1, 2)$ in the direction of $\mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$.
<u>Answer:</u> We compute all the necessary pieces:
i) $\nabla w = \langle 3x^2, 6y\rangle \;\Rightarrow\; \nabla w|_{(1,2)} = \langle 3, 12\rangle$.
ii) $\mathbf{u}$ must be a unit vector, so $\mathbf{u} = \frac{\mathbf{v}}{|\mathbf{v}|} = \langle \frac{3}{5}, \frac{4}{5}\rangle$.
iii) $\left.\frac{dw}{ds}\right|_{P_0,\mathbf{u}} = \nabla w|_{(1,2)} \cdot \mathbf{u} = \langle 3, 12\rangle \cdot \langle \frac{3}{5}, \frac{4}{5}\rangle = \boxed{\frac{57}{5}.}$

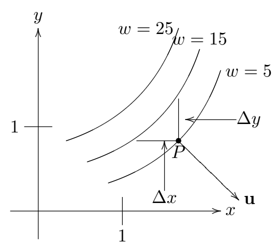
Example: (Geometric example) Let $\mathbf{u}$ be the direction of $\langle 1, -1\rangle$.
Using the picture at right estimate $\left.\frac{\partial w}{\partial x}\right|_P$, $\left.\frac{\partial w}{\partial y}\right|_p$, and $\left.\frac{dw}{ds}\right|_{P,\mathbf{u}}$.
By measuring from P to the next in level curve in the x direction we see that $\Delta x \approx -.5$.

$\Rightarrow \boxed{\left.\frac{\partial w}{\partial x}\right|_P \approx \frac{\Delta w}{\Delta x} \approx \frac{10}{-.5} = -20.}$

Similarly, we get $\boxed{\left.\frac{\partial w}{\partial y}\right|_P \approx 20.}$

Measuring in the $\mathbf{u}$ direction we get $\Delta s \approx -.3$

$\Rightarrow \boxed{\left.\frac{dw}{ds}\right|_{P,\mathbf{u}} \approx \frac{\Delta w}{\Delta s} \approx \frac{10}{.3} = -33.3.}$



Direction of maximum change:
The direction that gives the maximum rate of change is in the same direction as ∇w. The proof of this uses equation (1). Let θ be the angle between ∇w and $\mathbf{u}$. Then the geometric form of the dot product says

$$\left.\frac{dw}{ds}\right|_{\mathbf{u}} = \nabla w \cdot \mathbf{u} = |\nabla w||\mathbf{u}|\cos\theta = |\nabla w|\cos\theta.$$

Lagrange Multipliers

We will give the argument for why Lagrange multipliers work later. Here, we'll look at where and how to use them. Lagrange multipliers are used to solve constrained optimization problems. That is, suppose you have a function, say $f(x, y)$, for which you want to find the maximum or minimum value. But, you are not allowed to consider all (x, y) while you look for this value. Instead, the (x, y) you can consider are constrained to lie on some curve or surface. There are lots of examples of this in science, engineering and economics, for example, optimizing some utility function under budget constraints.

Lagrange multipliers problem:
Minimize (or maximize) $w = f(x, y, z)$ constrained by $g(x, y, z) = c$.

Lagrange multipliers solution:
Local minima (or maxima) must occur at a *critical point*. This is a point where $\nabla f = \lambda \nabla g$, and $g(x, y, z) = c$.

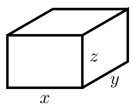
Example: Making a box using a minimum amount of material.

A box is made of cardboard with double thick sides, a triple thick bottom, single thick front and back and no top. It's volume is fixed at 3.

What dimensions use the least amount of cardboard?

Answer: We did this problem once before by solving for z in terms of x and y and substituting for it. That led to an unconstrained optimization problem in x and y. Here we will do it as a constrained problem. It is important to be able to do this because eliminating one variable is not always easy.

The box shown has dimensions x, y, and z.



The area of one side $= yz$. There are two double thick sides $\Rightarrow$ cardboard used $= 4yz$.

The area of the front (and back) $= xz$. It is single thick $\Rightarrow$ cardboard used $= 2xz$.

The area of the bottom $= xy$. It is triple thick $\Rightarrow$ cardboard used $= 3xy$.

Thus, the total cardboard used is

$$w = f(x, y, z) = 4yz + 2xz + 3xy.$$

The fixed volume acts as the constraint. It forces a relation between x, y and z so they can't all be varied independently. The constraint is

$$V = xyz = 3.$$

Our first job is to set up the equations to look for critical points. $\nabla f = \langle 2z + 3y, 4z + 3x, 4y + 2x \rangle$ and $\nabla V = \langle yz, xz, xy \rangle$.

The Lagrange multiplier equations are then

$$\begin{aligned} & \nabla f = \lambda \nabla V, \text{ and } V = 3 \\ \Leftrightarrow \quad & \langle 2z + 3y,\, 4z + 3x,\, 4y + 2x \rangle = \lambda \langle yz,\, xz,\, xy \rangle, \quad xyz = 3 \end{aligned}$$

Next we solve these equations for critical points. We do this by solving for λ in each equation (we call this *solving symmetrically*).

$\frac{2z+3y}{yz} = \lambda \quad \frac{4z+3x}{xz} = \lambda, \quad \frac{4y+2x}{xy} = \lambda, \quad xyz = 3 \qquad \Rightarrow \frac{2}{y} + \frac{3}{z} = \frac{4}{x} + \frac{3}{z} = \frac{4}{x} + \frac{2}{y}$

$\Rightarrow \frac{2}{y} = \frac{4}{x} \Rightarrow x = 2y$ and $\frac{3}{z} = \frac{2}{y} \Rightarrow z = \frac{3}{2}y$

Now, $xyz = 3 \Rightarrow 3y^3 = 3 \Rightarrow y = 1$

Answer: $x = 2, \quad y = 1, \quad z = \frac{3}{2}, \quad w = 18.$

Sphere example:

Minimize $w = y$ constrained to $x^2 + y^2 + z^2 = 1$.

Answer: $\nabla f = \langle 0, 1, 0\rangle$, $\nabla g = \langle 2x, 2y, 2z\rangle$

$\nabla f = \lambda \nabla g \Rightarrow \langle 0, 1, 0\rangle = \lambda\langle 2x, 2y, 2z\rangle \Rightarrow x = z = 0.$

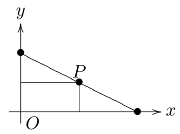
Constraint $\Rightarrow y = \pm 1$. (Gives the minimum and maximum respectively).

Example: (checking the boundary)

A rectangle in the plane is placed in the first quadrant so that one corner O is at the origin and the two sides adjacent to O are on the axes. The corner P opposite O is on the curve $x + 2y = 1$. Using Lagrange multipliers find for which point P the rectangle has maximum area. Say how you know this point gives the maximum.

Answer: We need some names

$g(x, y) = x + 2y = 1 =$ the constraint and $f(x, y) = xy =$ the area.



The gradients are: $\nabla g = \hat{\mathbf{i}} + 2\hat{\mathbf{j}}, \quad \nabla f = y\,\hat{\mathbf{i}} + x\,\hat{\mathbf{j}}$.

Lagrange multipliers: $\Rightarrow y = \lambda, \quad x = 2\lambda, \quad x + 2y = 1.$

The first two equations $\Rightarrow x = 2y$;

Combine this with the third equation $\Rightarrow 4y = 1$.

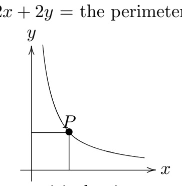
$\Rightarrow y = 1/4, \quad x = 1/2 \Rightarrow P = (1/2, 1/4).$

We know this is a maximum because the maximum occurs either at a critical point or on the boundary. In this case, the boundary points are on the axes at (1,0) and (0,1/2), which gives a rectangle with area = 0.

Example: (boundary at ∞)

A rectangle in the plane is placed in the first quadrant so that one corner O is at the origin and the two sides adjacent to O are on the axes. The corner P opposite O is on the curve $xy = 1$. Using Lagrange multipliers find for which point P the rectangle has minimum perimeter. Say how you know this point gives the minimum.

Answer: Let $g(x, y) = xy = 1 =$ the constraint and $f(x, y) = 2x + 2y =$ the perimeter.



Gradients: $\nabla g = y\,\hat{\mathbf{i}} + x\,\hat{\mathbf{j}}, \quad \nabla f = 2\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$.

Lagrange multipliers: $\Rightarrow$

$$2 = \lambda y$$
$$2 = \lambda x$$
$$xy = 1$$

The first two equations $\Rightarrow x = y$;

Combine this with the third equation $\Rightarrow x^2 = 1$.

$\Rightarrow x = 1, \quad x = 1 \Rightarrow P = (1, 1).$

We know this is a minimum because the minimum occurs either at a critical point or on the boundary. In this case the boundary points are infinitely far out on the axes which gives a rectangle with perimeter $= \infty$.

Proof of Lagrange Multipliers

Here we will give two arguments, one geometric and one analytic for why Lagrange multipliers work.

Critical points
For the function $w = f(x, y, z)$ constrained by $g(x, y, z) = c$ (c a constant) the critical points are defined as those points, which satisfy the constraint and where ∇f is parallel to ∇g. In equations:

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z) \quad \textit{and} \quad g(x, y, z) = c.$$

Statement of Lagrange multipliers
For the constrained system local maxima and minima (collectively extrema) occur at the critical points.

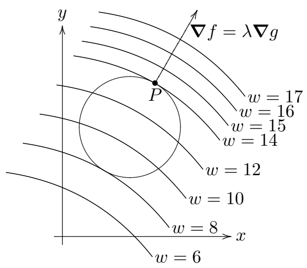
Geometric proof for Lagrange
(We only consider the two dimensional case, $w = f(x, y)$ with constraint $g(x, y) = c$.)

For concreteness, we've drawn the constraint curve, $g(x, y) = c$, as a circle and some level curves for $w = f(x, y) = c$ with explicit (made up) values. Geometrically, we are looking for the point on the circle where w takes its maximum or minimum values.

Now, start at the level curve with $w = 17$, which has no points on the circle. So, clearly, the maximum value of w on the constraint circle is less than 17. Move down the level curves until they first touch the circle when $w = 14$. Call the point where the first touch P. It is clear that P gives a local maximum for w on $g = c$, because if you move away from P in either direction on the circle you'll be on a level curve with a smaller value.

Since the circle is a level curve for g, we know ∇g is perpendicular to it. We also know ∇f is perpendicular to the level curve $w = 14$, since the curves themselves are tangent, these two gradients must be parallel.

Likewise, if you keep moving down the level curves, the last one to touch the circle will give a local minimum and the same argument will apply.



Analytic proof for Lagrange (in three dimensions)

Suppose f has a local maximum at P on the constraint surface.

Let $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ be an arbitrary parametrized curve which lies on the constraint surface and has $(x(0), y(0), z(0)) = P$. Finally, let $h(t) = f(x(t), y(t), z(t))$. The setup guarantees that $h(t)$ has a maximum at $t = 0$.

Taking a derivative using the chain rule in vector form gives

$$h'(t) = \boldsymbol{\nabla} f|_{\mathbf{r}(t)} \cdot \mathbf{r}'(t).$$

Since $t = 0$ is a local maximum, we have

$$h'(0) = \boldsymbol{\nabla} f|_P \cdot \mathbf{r}'(0) = 0.$$

Thus, $\boldsymbol{\nabla} f|_P$ is perpendicular to any curve on the constraint surface through P.

This implies $\boldsymbol{\nabla} f|_P$ is perpendicular to the surface. Since $\boldsymbol{\nabla} g|_P$ is also perpendicular to the surface we have proved $\boldsymbol{\nabla} f|_P$ is parallel to $\boldsymbol{\nabla} g|_P$. QED

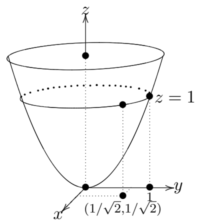
Level Curves and Contour Plots

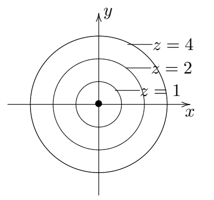
Level curves and *contour plots* are another way of visualizing functions of two variables. If you have seen a topographic map then you have seen a contour plot.

Example: To illustrate this we first draw the graph of $z = x^2 + y^2$. On this graph we draw *contours*, which are curves at a fixed height $z =$ constant.

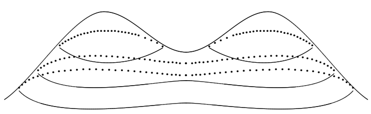
For example the curve at height $z = 1$ is the circle $x^2 + y^2 = 1$. On the graph we have to draw this at the correct height. Another way to show this is to draw the curves in the xy-plane and label them with their z-value. We call these curves *level curves* and the entire plot is called a *contour plot*.

For this example they are shown in the plot on the right. Notice that the 3D graph is simply the level curves 'pulled out' each to its correct height.

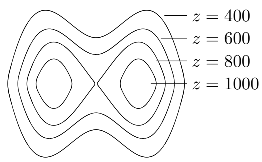




Here is another plot of a 'mountain pass'. Notice that in the contour plot the mountain pass is represented by a level curve that crosses itself. Moving up or down from the cross level curves heights decrease and moving right or left in the other they increase.



Mountain pass

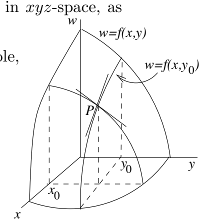


Level curves

Partial derivatives

Partial derivatives

Let $w = f(x, y)$ be a function of two variables. Its graph is a surface in xyz-space, as pictured.



Fix a value $y = y_0$ and just let x vary. You get a function of *one* variable,

(1) $$w = f(x, y_0), \qquad \text{the } \textbf{partial function} \text{ for } y = y_0.$$

Its graph is a curve in the vertical plane $y = y_0$, whose slope at the point P where $x = x_0$ is given by the derivative

(2) $$\frac{d}{dx} f(x, y_0)\bigg|_{x_0}, \quad \text{or} \quad \frac{\partial f}{\partial x}\bigg|_{(x_0, y_0)}.$$

We call (2) the **partial derivative** of f with respect to x at the point (x_0, y_0); the right side of (2) is the standard notation for it. The partial derivative is just the ordinary derivative of the partial function — it is calculated by holding one variable fixed and differentiating with respect to the other variable. Other notations for this partial derivative are

$$f_x(x_0, y_0), \qquad \frac{\partial w}{\partial x}\bigg|_{(x_0, y_0)}, \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0;$$

the first is convenient for including the specific point; the second is common in science and engineering, where you are just dealing with relations between variables and don't mention the function explicitly; the third and fourth indicate the point by just using a single subscript.

Analogously, fixing $x = x_0$ and letting y vary, we get the partial function $w = f(x_0, y)$, whose graph lies in the vertical plane $x = x_0$, and whose slope at P is the *partial derivative of f with respect to y*; the notations are

$$\frac{\partial f}{\partial y}\bigg|_{(x_0, y_0)}, \qquad f_y(x_0, y_0), \qquad \frac{\partial w}{\partial y}\bigg|_{(x_0, y_0)}, \qquad \left(\frac{\partial f}{\partial y}\right)_0, \qquad \left(\frac{\partial w}{\partial y}\right)_0.$$

The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ depend on (x_0, y_0) and are therefore functions of x and y.

Written as $\partial w/\partial x$, the partial derivative gives the rate of change of w with respect to x alone, at the point (x_0, y_0): it tells how fast w is increasing as x increases, when y is held constant.

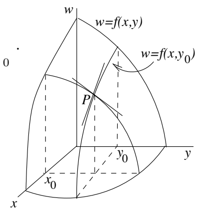
For a function of three or more variables, $w = f(x, y, z, \ldots)$, we cannot draw graphs any more, but the idea behind partial differentiation remains the same: to define the partial derivative with respect to x, for instance, hold all the other variables constant and take the ordinary derivative with respect to x; the notations are the same as above:

$$\frac{d}{dx} f(x, y_0, z_0, \ldots) = f_x(x_0, y_0, z_0, \ldots), \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0.$$

The Tangent Approximation

1. The tangent plane.

For a function of one variable, $w = f(x)$, the tangent line to its graph at a point (x_0, w_0) is the line passing through (x_0, w_0) and having slope $\left(\dfrac{dw}{dx}\right)_0$.



For a function of two variables, $w = f(x, y)$, the natural analogue is the *tangent plane* to the graph, at a point

(x_0, y_0, w_0). What's the equation of this tangent plane? Referring to the picture at right (this figure was also used when we introduced partial derivatives), we see that the tangent plane

(i) must pass through (x_0, y_0, w_0), where $w_0 = f(x_0, y_0)$;

(ii) must contain the tangent lines to the graphs of the two partial functions — this will hold if the plane has the same slopes in the $\mathbf{i}$ and $\mathbf{j}$ directions as the surface does.

Using these two conditions, it is easy to find the equation of the tangent plane. The general equation of a plane through (x_0, y_0, w_0) is

$$A(x - x_0) + B(y - y_0) + C(w - w_0) \;=\; 0\ .$$

Assume the plane is not vertical; then $C \neq 0$, so we can divide through by C and solve for $w - w_0$, getting

$$w - w_0 \;=\; a(x - x_0) + b(y - y_0), \qquad a = A/C,\ b = B/C. \tag{3}$$

The plane passes through (x_0, y_0, w_0); what values of the coefficients a and b will make it also tangent to the graph there? We have

$$\begin{aligned}
a \;&=\; \text{slope of plane (3) in the } \mathbf{i}\text{-direction} \qquad &&\text{(by putting } y = y_0 \text{ in (3));}\\
&=\; \text{slope of graph in the } \mathbf{i}\text{-direction,} &&\text{(by (ii) above)}\\
&=\; \left(\frac{\partial w}{\partial x}\right)_0; &&\text{(by the definition of partial derivative);} \quad \text{similarly,}\\
b \;&=\; \left(\frac{\partial w}{\partial y}\right)_0 .
\end{aligned}$$

Therefore the equation of the **tangent plane** to $w = f(x, y)$ at (x_0, y_0) is

$$w - w_0 \;=\; \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0) \tag{4}$$

2. The approximation formula.

The most important use for the tangent plane is to give an approximation that is the basic formula in the study of functions of several variables — almost everything follows in one way or another from it.

The intuitive idea is that if we stay near (x_0, y_0, w_0), the graph of the tangent plane (4) will be a good approximation to the graph of the function $w = f(x,y)$. Therefore if the point (x,y) is close to (x_0, y_0),

$$(5) \qquad \underset{\text{height of graph}}{f(x,y)} \quad \approx \quad \underset{\text{height of tangent plane}}{w_0 + \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)}$$

The function on the right side of (5) whose graph is the tangent plane is often called the **linearization** of $f(x,y)$ at (x_0, y_0): it is the linear function which gives the best approximation to $f(x,y)$ for values of (x,y) close to (x_0, y_0).

An equivalent form of the approximation (5) is obtained by using Δ notation; if we put

$$\Delta x = x - x_0, \qquad \Delta y = y - y_0, \qquad \Delta w = w - w_0,$$

then (5) becomes

$$(6) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad \text{if} \quad \Delta x \approx 0,\ \Delta y \approx 0\ .$$

This formula gives the approximate change in w when we make a small change in x and y. We will use it often.

The analogous approximation formula for a function $w = f(x,y,z)$ of three variables would be

$$(7) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y + \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \qquad \text{if} \quad \Delta x, \Delta y \Delta z \approx 0\ .$$

Unfortunately, for functions of three or more variables, we can't use a geometric argument for the approximation formula (7); for this reason, it's best to recast the argument for (6) in a form which doesn't use tangent planes and geometry, and therefore can be generalized to several variables. This is done at the end of this Chapter TA; for now let's just assume the truth of (7) and its higher-dimensional analogues.

Here are two typical examples of the use of the approximation formula. Other examples are in the Exercises. In the rest of your study of partial differentiation, you will see how the approximation formula is used to derive the important theorems and formulas.

Example 1. Give a reasonable square, centered at $(1,1)$, over which the value of $w = x^3 y^4$ will not vary by more than $\pm .1$.

Solution. We use (6). We calculate for the two partial derivatives

$$w_x = 3x^2 y^4 \qquad\qquad w_y = 4x^3 y^3$$

and therefore, evaluating the partials at $(1,1)$ and using (6), we get

$$\Delta w \approx 3\Delta x + 4\Delta y.$$

Thus if $|\Delta x| \le .01$ and $\Delta y| \le .01$, we should have

$$|\Delta w| \le 3|\Delta x| + 4|\Delta y| \le .07,$$

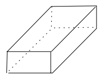
which is within the bounds. So the answer is the square with center at (1,1) given by

$$|x - 1| \le .01, \qquad |y - 1| \le .01\ .$$

Example 2. The sides a, b, c of a rectangular box have lengths measured to be respectively 1, 2, and 3. To which of these measurements is the volume V most sensitive?

Solution. $V = abc$, and therefore by the approximation formula (7),

$$\begin{aligned} \Delta V &\approx bc\,\Delta a + ac\,\Delta b + ab\,\Delta c \\ &\approx 6\,\Delta a + 3\,\Delta b + 2\,\Delta c, \qquad \text{at } (1,2,3); \end{aligned}$$



thus it is most sensitive to small changes in side a, since Δa occurs with the largest coefficient. (That is, if one at a time the measurement of each side were changed by say .01, it is the change in a which would produce the biggest change in V, namely .06 .)

The result may seem paradoxical — the value of V is most sensitive to the length of the *shortest* side — but it's actually intuitive, as you can see by thinking about how the box looks.

Sensitivity Principle *The numerical value of $w = f(x, y, \dots)$, calculated at some point $(x_0, y_0, \dots)$, will be most sensitive to small changes in that variable for which the corresponding partial derivative $w_x, w_y, \dots$ has the largest absolute value at the point.*

Non-independent Variables

1. We give a worked example here. A fuller explanation will be given in the next session.

Let

$$w = x^3y^2 + x^2y^3 + y$$

and assume x and y satisfy the relation

$$x^2 + y^2 = 1.$$

We consider x to be the independent variable, then, because y depends on x we have w is ultimately a function of the single variable x.

a) Compute $\dfrac{dw}{dx}$ using implicit differentiation.

b) Compute $\dfrac{dw}{dx}$ using total differentials.

<u>**Answer:**</u>

a) Implicit differentiation means remembering that y is a function of x, e.g., $\dfrac{dy^2}{dx} = 2y\dfrac{dy}{dx}$.

Thus,

$$\frac{dw}{dx} = 3x^2y^2 + 2x^3y\frac{dy}{dx} + 2xy^3 + 3x^2y^2\frac{dy}{dx} + \frac{dy}{dx}.$$

Now we differentiate the constraint to find $\dfrac{dy}{dx}$.

$$x^2 + y^2 = 1 \;\Rightarrow\; 2x + 2y\frac{dy}{dx} = 0 \;\Rightarrow\; \frac{dy}{dx} = -\frac{x}{y}.$$

Substituting this in the equation for $\dfrac{dw}{dx}$ gives

$$\frac{dw}{dx} = 3x^2y^2 - 2x^3y\frac{x}{y} + 2xy^3 - 3x^2y^2\frac{x}{y} - \frac{x}{y} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

b) Taking total differentials of both w and the constraint equation gives

$$\begin{aligned} dw &= 3x^2y^2\,dx + 2x^3y\,dy + 2xy^3\,dx + 3x^2y^2\,dy + dy \\ &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\,dy \end{aligned}$$

$$2x\,dx + 2y\,dy = 0.$$

We can solve the second equation for dy and substitute in the equation for dw.

$$\begin{aligned} dy &= -\frac{x}{y}\,dx \;\Rightarrow \\ dw &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\left(-\frac{x}{y}\right)dx \\ &= (3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y})\,dx \end{aligned}$$

Thus,

$$\frac{dw}{dx} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

Second derivative test

1. Find and classify all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

<u>Answer:</u> Taking the first partials and setting them to 0:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 3y^2 - 12 = 0.$$

The first equation implies $x = -1$ and the second implies $y = \pm 2$. Thus, the critical points are $(-1, 2)$ and $(-1, -2)$.

Taking second partials:

$$\frac{\partial^2 z}{\partial x^2} = 30x^4, \quad \frac{\partial^2 z}{\partial xy} = 0, \quad \frac{\partial^2 z}{\partial y^2} = 6y.$$

We analyze each critical point in turn.
At (-1,-2): $A = z_{xx}(-1, -2) = 30, \quad B = z_{xy}(-1, -2) = 0, \quad C = z_{yy}(-1, -2) = -12.$
Therefore $AC - B^2 = -360 < 0$, which implies the critical point is a saddle.

At (-1,2): $A = z_{xx}(-1, 2) = 30, \quad B = z_{xy}(-1, 2) = 0, \quad C = z_{yy}(-1, 2) = 12.$
Therefore $AC - B^2 = 360 > 0$ and $A > 0$, which implies the critical point is a minimum.

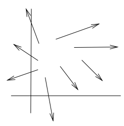
V1. Plane Vector Fields

1. Vector fields in the plane; gradient fields.

We consider a function of the type

$$\mathbf{F}(x,y) \;=\; M(x,y)\,\mathbf{i} + N(x,y)\,\mathbf{j}\;. \tag{1}$$

where M and N are both functions of two variables. To each pair of values (x_0, y_0) for which both M and N are defined, such a function assigns a vector $\mathbf{F}(x_0, y_0)$ in the plane. $\mathbf{F}$ is therefore called a **vector function of two variables**. The set of points (x, y) for which $\mathbf{F}$ is defined is called the *domain* of $\mathbf{F}$.



To visualize the function $\mathbf{F}(x, y)$, at each point (x_0, y_0) in the domain we place the corresponding vector $\mathbf{F}(x_0, y_0)$ so that its tail is at (x_0, y_0). Thus each point of the domain is the tail end of a vector, and what we get is called a **vector field**. This vector field gives a picture of the vector function $\mathbf{F}(x, y)$.

Conversely, given a vector field in a region of the xy-plane, it determines a vector function of the type (1), by expressing each vector of the field in terms of its $\mathbf{i}$ and $\mathbf{j}$ components. Thus there is no real distinction between "vector function" and "vector field". Mindful of the applications to physics, in these notes we will mostly use "vector field". We will use the same symbol $\mathbf{F}$ to denote both the field and the function, saying "the vector field $\mathbf{F}$", rather than "the vector field corresponding to the vector function $\mathbf{F}$".

We say the vector field $\mathbf{F}$ is *continuous* in some region of the plane if both $M(x, y)$ and $N(x, y)$ are continuous functions in that region. The intuitive picture of a continuous vector field is that the vectors associated to points sufficiently near (x_0, y_0) should have direction and magnitude very close to that of $\mathbf{F}(x_0, y_0)$ — in other words, as you move around the field, the vectors should change direction and magnitude smoothly, without sudden jumps in size or direction.

In the same way, we say $\mathbf{F}$ is *differentiable* in some region if M and N are differentiable, that is, if all the partial derivatives

$$\frac{\partial M}{\partial x}\,, \qquad \frac{\partial M}{\partial y}\,, \qquad \frac{\partial N}{\partial x}\,, \qquad \frac{\partial N}{\partial y}$$

exist in the region. We say $\mathbf{F}$ is *continuously differentiable* in the region if all these partial derivatives are themselves continuous there. In general, all the commonly used vector fields are continuously differentiable, except perhaps at isolated points, or along certain curves. But as you will see, these points or curves affect the properties of the field in very important ways.

Where do vector fields arise in science and engineering?

One important way is as **gradient vector fields**. If

$$w = f(x,y) \tag{2}$$

is a differentiable function of two variables, then its *gradient*

$$\nabla w = \frac{\partial w}{\partial x}\,\mathbf{i} + \frac{\partial w}{\partial y}\,\mathbf{j} \tag{3}$$

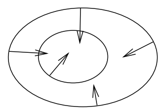
is a vector field, since both partial derivatives are functions of x and y. We recall the geometric interpretation of the gradient:

$$\begin{aligned} \text{dir } \nabla w &= \text{the direction } \mathbf{u} \text{ in which } \left.\frac{dw}{ds}\right|_{\mathbf{u}} \text{ is greatest;} \\ |\nabla w| &= \text{this greatest value of } \left.\frac{dw}{ds}\right|_{\mathbf{u}}, \end{aligned} \tag{4}$$

where $\left.\dfrac{dw}{ds}\right|_{\mathbf{u}} = \nabla w \cdot \mathbf{u}$ is the directional derivative of w in the direction $\mathbf{u}$.

Another important fact about the gradient is that if one draws the contour curves of $f(x,y)$, which by definition are the curves

$$f(x,y) = c, \qquad c \text{ constant},$$

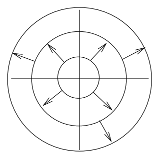


then at every point (x_0, y_0), the gradient vector ∇w at this point is perpendicular to the contour line passing through this point, i.e.,

(5) *the gradient field of f is perpendicular to the contour curves of f .*

Example 1. Let $w = \sqrt{x^2+y^2} = r$. Using the definition (3) of gradient, we find

$$\nabla w = \frac{x}{r}\,\mathbf{i} + \frac{y}{r}\,\mathbf{j} = \frac{x\,\mathbf{i} + y\,\mathbf{j}}{r}.$$



The domain of ∇w is the xy-plane with $(0,0)$ deleted, and it is continuously differentiable in this region. Since $|x\,i + y\,j| = r$, we see that $|\nabla w| = 1$. Thus all the vectors of the vector field ∇w are unit vectors, and they point radially outward from the origin. This makes sense by (4), since the definition of w shows that dw/ds should be greatest in the radially outward direction, and have the value 1 in that direction.

Finally, the contour curves for w are circles centered at $(0,0)$, which are perpendicular to the vectors ∇w everywhere, as (5) predicts.

2. Force and velocity fields.

Continuing our search for ways in which vector fields arise, here are two physical situations which are described mathematically by vector fields. We shall refer to them often in the sequel, using our physical intuition to suggest the sort of mathematical properties that vector fields ought to have.

Force fields.

From physics, we have the two-dimensional electrostatic force fields arising from a distribution of static (i.e., not moving) charges in the plane. At each point (x_0, y_0) of the plane, we put a vector representing the force which would act on a unit positive charge placed at that point.

In the same way, we get vector fields arising from a distribution of masses in the xy-plane, representing the gravitational force acting at each point on a unit mass. There are also the electromagnetic fields arising from moving electric charges and/or a distribution of magnets, representing the magnetic force at each point.

Any of these we shall simply refer to as a **force field**.

Example 2. Express in $\mathbf{i} - \mathbf{j}$ form the electrostatic force field $\mathbf{F}$ in the xy-plane arising from a unit positive charge placed at the origin, given that the force vector at (x, y) is directed radially away from the origin and that it has magnitude c/r^2, c constant.

Solution. Since the vector $x\,\mathbf{i} + y\,\mathbf{j}$ with tail at (x, y) is directed radially outward and has magnitude r, it has the right direction, and we need only change its magnitude to c/r^2. We do this by multiplying it by c/r^3, which gives

$$\mathbf{F} \;=\; \frac{cx}{r^3}\,\mathbf{i} + \frac{cy}{r^3}\,\mathbf{j} \;=\; c\,\frac{x\,\mathbf{i} + y\,\mathbf{j}}{(x^2+y^2)^{3/2}}\ .$$

Flow fields and velocity fields

A second way vector fields arise is as the steady-state *flow fields* and *velocity fields.*

Imagine a fluid flowing in a horizontal shallow tank of uniform depth, and assume that the flow pattern at any point is purely horizontal and not changing with time. We will call this a *two-dimensional steady-state flow* or for short, simply a *flow.* The fluid can either be compressible (like a gas), or incompressible (like water). We also allow for the possibility that at various points, fluid is beiong added to or subtracted from the flow; for instance, someone could be standing over the tank pouring in water at a certain point, or over a certain area. We also allow the density to vary from point to point, as it would for an unevenly heated gas.

With such a flow we can associate two vector fields.

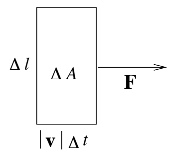
There is the **velocity field** $\mathbf{v}(x, y)$ where the vector $\mathbf{v}(x, y)$ at the point (x, y) represents the velocity vector of the flow at that point — that is, its direction gives the direction of flow, and its magnitude gives the speed of the flow.

Then there is the **flow field**, defined by

$$\mathbf{F} \;=\; \delta(x, y)\,\mathbf{v}(x, y) \tag{6}$$

where $\delta(x, y)$ gives the density of the fluid at the point (x, y), in terms of mass per unit area. Assuming it is not 0 at a point (x, y), we can interpret $\mathbf{F}(x, y)$ as follows:

$$\begin{aligned} \text{dir}\,\mathbf{F} \;&=\; \text{direction of fluid flow at } (x,y); \\ |\mathbf{F}| \;&=\; \begin{cases} \text{rate (per unit length per second) of mass transport} \\ \text{across a line perpendicular to the flow direction at } (x,y). \end{cases} \end{aligned} \tag{7}$$



Namely, we see that first by (6) and then by the picture,

$$|\mathbf{F}|\,\Delta l\,\Delta t \;=\; \delta\,|\mathbf{v}|\,\Delta t\,\Delta l \;=\; \text{mass in } \Delta A,$$

from which (7) follows by dividing by $\Delta l\,\Delta t$ and letting Δl and $\Delta t \to 0$.

If the density is a constant δ_0, as it would be for an incompressible fluid at a uniform temperature, then the flow field and velocity field are essentially the same, by (6) — the vectors of one are just a constant scalar multiple of the vectors of the other.

Example 3. Describe and interpret $\mathbf{F} \;=\; \dfrac{x\,\mathbf{i} + y\,\mathbf{j}}{x^2+y^2}$ as a flow field and a force field.

Solution. As in Example 2, the field $\mathbf{F}$ is defined everywhere except $(0,0)$ and its direction is radially outward; now, however, its magnitude is r/r^2, i.e., $|\mathbf{F}| = 1/r$.

$\mathbf{F}$ is the *flow field* for a source of magnitude 2π at the origin. To see this, look at a circle of radius a centered at the origin. At each point P on this circle, the flow is radially outward and by (7),

$$\text{mass transport rate at P} \;=\; \frac{1}{a}, \qquad \text{so that}$$
$$\text{mass transport rate across circle} \;=\; \frac{1}{a}\cdot 2\pi a \;=\; 2\pi\;.$$

This shows that in one second, 2π mass flows out through every circle centered at the origin. This is the flow field for a source of magnitude 2π at the origin — for example, one could imagine a narrow pipe placed over the tank, introducting 2π mass units per second at the point $(0,0)$.

We know that $|\mathbf{F}| \;=\; \delta\,|\mathbf{v}| \;=\; 1/r$. Two important cases are:

• if the fluid is incompressible, like water, then its density is constant, so the flow speed must decrease like $1/r$ — the flow outward gets slower the further you are from the origin;

• if it is compressible like a gas, and its flow speed stays constant, then the density must decrease like $1/r$.

We now interpret the same field as a *force field.*

Suppose we think of the z-axis in space as a long straight wire, bearing a uniform positive electrostatic charge. This gives us a vector field in space, representing the electrostatic force field.

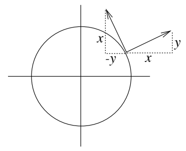
Since one part of the wire looks just like any other part, radial symmetry shows first that the vectors in the force field have 0 as their $\mathbf{k}$-component, i.e., they are pointed radially outward from the wire, and second that their magnitude depends only on their distance r from the wire. It can be shown in fact that the resulting force field is $\mathbf{F}$, up to a constant factor.

Such a field is called "two-dimensional", even though it is a vector field in space, because z and $\mathbf{k}$ don't enter into its description — once you know how it looks in the xy-plane, you know how it looks all through space.

The important thing to notice is that the magnitude of the force field in the xy-plane decreases like $1/r$, *not* like $1/r^2$, as it would if the charge were all at a point.

In the same way, the gravitational field of a uniform mass distribution along the z-axis would be $-\mathbf{F}$, up to a constant factor, and would be called a "two-dimensional gravitational field". Naturally, we don't have infinite long straight wires, but if you have a long straight wire, and stay away from its ends, or have only a short straight wire, but stay close to it, the force field will look like $\mathbf{F}$ near the wire.

Example 4. Find the velocity field of a fluid with density 1 in a shallow tank, rotating with constant angular velocity ω counterclockwise around the origin.



Solution. First we find the field direction at each point (x, y).

We know the vector $x\,\mathbf{i} + y\,\mathbf{j}$ is directed radially outward. Therefore a vector perpendicular to it in the counterclockwise direction (see picture) will be $-y\,\mathbf{i} + x\,\mathbf{j}$ (since its scalar product with $x\,\mathbf{i} + y\,\mathbf{j}$ is 0 and the signs are correct).

The preceding vector has magnitude r. If the angular velocity is ω, then the linear velocity is given by

$$|\mathbf{v}| \ = \ \omega r,$$

so to get the velocity field, we should multiply the above field by ω :

$$\mathbf{v} \ = \ -\omega y\,\mathbf{i} + \omega x\,\mathbf{j} \ .$$

Non-independent Variables

1. We give a worked example here. A fuller explanation will be given in the next session.

Let

$$w = x^3y^2 + x^2y^3 + y$$

and assume x and y satisfy the relation

$$x^2 + y^2 = 1.$$

We consider x to be the independent variable, then, because y depends on x we have w is ultimately a function of the single variable x.

a) Compute $\frac{dw}{dx}$ using implicit differentiation.

b) Compute $\frac{dw}{dx}$ using total differentials.

Answer:

a) Implicit differentiation means remembering that y is a function of x, e.g., $\frac{dy^2}{dx} = 2y\frac{dy}{dx}$.

Thus,

$$\frac{dw}{dx} = 3x^2y^2 + 2x^3y\frac{dy}{dx} + 2xy^3 + 3x^2y^2\frac{dy}{dx} + \frac{dy}{dx}.$$

Now we differentiate the constraint to find $\frac{dy}{dx}$.

$$x^2 + y^2 = 1 \;\Rightarrow\; 2x + 2y\frac{dy}{dx} = 0 \;\Rightarrow\; \frac{dy}{dx} = -\frac{x}{y}.$$

Substituting this in the equation for $\frac{dw}{dx}$ gives

$$\frac{dw}{dx} = 3x^2y^2 - 2x^3y\frac{x}{y} + 2xy^3 - 3x^2y^2\frac{x}{y} - \frac{x}{y} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

b) Taking total differentials of both w and the constraint equation gives

$$\begin{aligned} dw &= 3x^2y^2\,dx + 2x^3y\,dy + 2xy^3\,dx + 3x^2y^2\,dy + dy \\ &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\,dy \end{aligned}$$

$$2x\,dx + 2y\,dy = 0.$$

We can solve the second equation for dy and substitute in the equation for dw.

$$\begin{aligned} dy &= -\frac{x}{y}\,dx \;\Rightarrow \\ dw &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\left(-\frac{x}{y}\right)\,dx \\ &= (3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y})\,dx \end{aligned}$$

Thus,

$$\frac{dw}{dx} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

Second derivative test

1. Find and classify all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

Answer: Taking the first partials and setting them to 0:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 3y^2 - 12 = 0.$$

The first equation implies $x = -1$ and the second implies $y = \pm 2$. Thus, the critical points are $(-1, 2)$ and $(-1, -2)$.

Taking second partials:

$$\frac{\partial^2 z}{\partial x^2} = 30x^4, \quad \frac{\partial^2 z}{\partial xy} = 0, \quad \frac{\partial^2 z}{\partial y^2} = 6y.$$

We analyze each critical point in turn.
At (-1,-2): $A = z_{xx}(-1, -2) = 30, \quad B = z_{xy}(-1, -2) = 0, \quad C = z_{yy}(-1, -2) = -12.$
Therefore $AC - B^2 = -360 < 0$, which implies the critical point is a saddle.

At (-1,2): $A = z_{xx}(-1, 2) = 30, \quad B = z_{xy}(-1, 2) = 0, \quad C = z_{yy}(-1, 2) = 12.$
Therefore $AC - B^2 = 360 > 0$ and $A > 0$, which implies the critical point is a minimum.

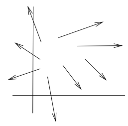
V1. Plane Vector Fields

1. Vector fields in the plane; gradient fields.

We consider a function of the type

$$\mathbf{F}(x,y) \;=\; M(x,y)\,\mathbf{i} + N(x,y)\,\mathbf{j}\ . \tag{1}$$

where M and N are both functions of two variables. To each pair of values (x_0, y_0) for which both M and N are defined, such a function assigns a vector $\mathbf{F}(x_0, y_0)$ in the plane. $\mathbf{F}$ is therefore called a **vector function of two variables**. The set of points (x, y) for which $\mathbf{F}$ is defined is called the *domain* of $\mathbf{F}$.



To visualize the function $\mathbf{F}(x, y)$, at each point (x_0, y_0) in the domain we place the corresponding vector $\mathbf{F}(x_0, y_0)$ so that its tail is at (x_0, y_0). Thus each point of the domain is the tail end of a vector, and what we get is called a **vector field**. This vector field gives a picture of the vector function $\mathbf{F}(x, y)$.

Conversely, given a vector field in a region of the xy-plane, it determines a vector function of the type (1), by expressing each vector of the field in terms of its $\mathbf{i}$ and $\mathbf{j}$ components. Thus there is no real distinction between "vector function" and "vector field". Mindful of the applications to physics, in these notes we will mostly use "vector field". We will use the same symbol $\mathbf{F}$ to denote both the field and the function, saying "the vector field $\mathbf{F}$", rather than "the vector field corresponding to the vector function $\mathbf{F}$".

We say the vector field $\mathbf{F}$ is *continuous* in some region of the plane if both $M(x, y)$ and $N(x, y)$ are continuous functions in that region. The intuitive picture of a continuous vector field is that the vectors associated to points sufficiently near (x_0, y_0) should have direction and magnitude very close to that of $\mathbf{F}(x_0, y_0)$ — in other words, as you move around the field, the vectors should change direction and magnitude smoothly, without sudden jumps in size or direction.

In the same way, we say $\mathbf{F}$ is *differentiable* in some region if M and N are differentiable, that is, if all the partial derivatives

$$\frac{\partial M}{\partial x}\,, \qquad \frac{\partial M}{\partial y}\,, \qquad \frac{\partial N}{\partial x}\,, \qquad \frac{\partial N}{\partial y}$$

exist in the region. We say $\mathbf{F}$ is *continuously differentiable* in the region if all these partial derivatives are themselves continuous there. In general, all the commonly used vector fields are continuously differentiable, except perhaps at isolated points, or along certain curves. But as you will see, these points or curves affect the properties of the field in very important ways.

Where do vector fields arise in science and engineering?

One important way is as **gradient vector fields**. If

$$w = f(x,y) \tag{2}$$

is a differentiable function of two variables, then its *gradient*

$$\nabla w = \frac{\partial w}{\partial x}\,\mathbf{i} + \frac{\partial w}{\partial y}\,\mathbf{j} \tag{3}$$

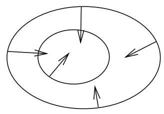
is a vector field, since both partial derivatives are functions of x and y. We recall the geometric interpretation of the gradient:

$$\begin{aligned} \text{dir } \nabla w &= \text{the direction } \mathbf{u} \text{ in which } \left.\frac{dw}{ds}\right|_{\mathbf{u}} \text{ is greatest;} \\ |\nabla w| &= \text{this greatest value of } \left.\frac{dw}{ds}\right|_{\mathbf{u}}, \end{aligned} \tag{4}$$

where $\left.\frac{dw}{ds}\right|_{\mathbf{u}} = \nabla w \cdot \mathbf{u}$ is the directional derivative of w in the direction $\mathbf{u}$.

Another important fact about the gradient is that if one draws the contour curves of $f(x,y)$, which by definition are the curves

$$f(x,y) = c, \qquad c \text{ constant},$$

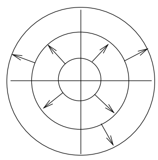


then at every point (x_0, y_0), the gradient vector ∇w at this point is perpendicular to the contour line passing through this point, i.e.,

(5) *the gradient field of f is perpendicular to the contour curves of f .*

Example 1. Let $w = \sqrt{x^2+y^2} = r$. Using the definition (3) of gradient, we find

$$\nabla w = \frac{x}{r}\,\mathbf{i} + \frac{y}{r}\,\mathbf{j} = \frac{x\,\mathbf{i} + y\,\mathbf{j}}{r}\,.$$



The domain of ∇w is the xy-plane with $(0,0)$ deleted, and it is continuously differentiable in this region. Since $|x\,i + y\,j| = r$, we see that $|\nabla w| = 1$. Thus all the vectors of the vector field ∇w are unit vectors, and they point radially outward from the origin. This makes sense by (4), since the definition of w shows that dw/ds should be greatest in the radially outward direction, and have the value 1 in that direction.

Finally, the contour curves for w are circles centered at $(0,0)$, which are perpendicular to the vectors ∇w everywhere, as (5) predicts.

2. Force and velocity fields.

Continuing our search for ways in which vector fields arise, here are two physical situations which are described mathematically by vector fields. We shall refer to them often in the sequel, using our physical intuition to suggest the sort of mathematical properties that vector fields ought to have.

Force fields.

From physics, we have the two-dimensional electrostatic force fields arising from a distribution of static (i.e., not moving) charges in the plane. At each point (x_0, y_0) of the plane, we put a vector representing the force which would act on a unit positive charge placed at that point.

In the same way, we get vector fields arising from a distribution of masses in the xy-plane, representing the gravitational force acting at each point on a unit mass. There are also the electromagnetic fields arising from moving electric charges and/or a distribution of magnets, representing the magnetic force at each point.

Any of these we shall simply refer to as a **force field**.

Example 2. Express in $\mathbf{i}-\mathbf{j}$ form the electrostatic force field $\mathbf{F}$ in the xy-plane arising from a unit positive charge placed at the origin, given that the force vector at (x, y) is directed radially away from the origin and that it has magnitude c/r^2, c constant.

Solution. Since the vector $x\,\mathbf{i} + y\,\mathbf{j}$ with tail at (x, y) is directed radially outward and has magnitude r, it has the right direction, and we need only change its magnitude to c/r^2. We do this by multiplying it by c/r^3, which gives

$$\mathbf{F} \;=\; \frac{cx}{r^3}\,\mathbf{i} + \frac{cy}{r^3}\,\mathbf{j} \;=\; c\,\frac{x\,\mathbf{i} + y\,\mathbf{j}}{(x^2+y^2)^{3/2}}\ .$$

Flow fields and velocity fields

A second way vector fields arise is as the steady-state *flow fields* and *velocity fields*.

Imagine a fluid flowing in a horizontal shallow tank of uniform depth, and assume that the flow pattern at any point is purely horizontal and not changing with time. We will call this a *two-dimensional steady-state flow* or for short, simply a *flow*. The fluid can either be compressible (like a gas), or incompressible (like water). We also allow for the possibility that at various points, fluid is beiong added to or subtracted from the flow; for instance, someone could be standing over the tank pouring in water at a certain point, or over a certain area. We also allow the density to vary from point to point, as it would for an unevenly heated gas.

With such a flow we can associate two vector fields.

There is the **velocity field** $\mathbf{v}(x, y)$ where the vector $\mathbf{v}(x, y)$ at the point (x, y) represents the velocity vector of the flow at that point — that is, its direction gives the direction of flow, and its magnitude gives the speed of the flow.

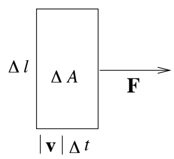
Then there is the **flow field**, defined by

$$\mathbf{F} \;=\; \delta(x, y)\,\mathbf{v}(x, y) \tag{6}$$

where $\delta(x, y)$ gives the density of the fluid at the point (x, y), in terms of mass per unit area. Assuming it is not 0 at a point (x, y), we can interpret $\mathbf{F}(x, y)$ as follows:

$$\text{dir}\,\mathbf{F} \;=\; \text{direction of fluid flow at } (x, y); \tag{7}$$

$$|\mathbf{F}| \;=\; \begin{cases} \text{rate (per unit length per second) of mass transport} \\ \text{across a line perpendicular to the flow direction at } (x, y). \end{cases}$$



Namely, we see that first by (6) and then by the picture,

$$|\mathbf{F}|\,\Delta l\,\Delta t \;=\; \delta\,|\mathbf{v}|\,\Delta t\,\Delta l \;=\; \text{mass in } \Delta A,$$

from which (7) follows by dividing by $\Delta l\,\Delta t$ and letting Δl and $\Delta t \to 0$.

If the density is a constant δ_0, as it would be for an incompressible fluid at a uniform temperature, then the flow field and velocity field are essentially the same, by (6) — the vectors of one are just a constant scalar multiple of the vectors of the other.

Example 3. Describe and interpret $\mathbf{F} \;=\; \dfrac{x\,\mathbf{i} + y\,\mathbf{j}}{x^2+y^2}$ as a flow field and a force field.

Solution. As in Example 2, the field $\mathbf{F}$ is defined everywhere except $(0,0)$ and its direction is radially outward; now, however, its magnitude is r/r^2, i.e., $|\mathbf{F}| = 1/r$.

$\mathbf{F}$ is the *flow field* for a source of magnitude 2π at the origin. To see this, look at a circle of radius a centered at the origin. At each point P on this circle, the flow is radially outward and by (7),

$$\text{mass transport rate at P} \;=\; \frac{1}{a}, \qquad \text{so that}$$
$$\text{mass transport rate across circle} \;=\; \frac{1}{a}\cdot 2\pi a \;=\; 2\pi\,.$$

This shows that in one second, 2π mass flows out through every circle centered at the origin. This is the flow field for a source of magnitude 2π at the origin — for example, one could imagine a narrow pipe placed over the tank, introducting 2π mass units per second at the point $(0,0)$.

We know that $|\mathbf{F}| \;=\; \delta\,|\mathbf{v}| \;=\; 1/r$. Two important cases are:

• if the fluid is incompressible, like water, then its density is constant, so the flow speed must decrease like $1/r$ — the flow outward gets slower the further you are from the origin;

• if it is compressible like a gas, and its flow speed stays constant, then the density must decrease like $1/r$.

We now interpret the same field as a *force field*.

Suppose we think of the z-axis in space as a long straight wire, bearing a uniform positive electrostatic charge. This gives us a vector field in space, representing the electrostatic force field.

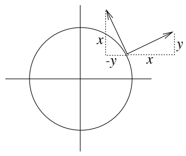
Since one part of the wire looks just like any other part, radial symmetry shows first that the vectors in the force field have 0 as their $\mathbf{k}$-component, i.e., they are pointed radially outward from the wire, and second that their magnitude depends only on their distance r from the wire. It can be shown in fact that the resulting force field is $\mathbf{F}$, up to a constant factor.

Such a field is called “two-dimensional”, even though it is a vector field in space, because z and $\mathbf{k}$ don’t enter into its description — once you know how it looks in the xy-plane, you know how it looks all through space.

The important thing to notice is that the magnitude of the force field in the xy-plane decreases like $1/r$, *not* like $1/r^2$, as it would if the charge were all at a point.

In the same way, the gravitational field of a uniform mass distribution along the z-axis would be $-\mathbf{F}$, up to a constant factor, and would be called a “two-dimensional gravitational field". Naturally, we don't have infinite long straight wires, but if you have a long straight wire, and stay away from its ends, or have only a short straight wire, but stay close to it, the force field will look like $\mathbf{F}$ near the wire.

Example 4. Find the velocity field of a fluid with density 1 in a shallow tank, rotating with constant angular velocity ω counterclockwise around the origin.



Solution. First we find the field direction at each point (x, y).

We know the vector $x\,\mathbf{i} + y\,\mathbf{j}$ is directed radially outward. Therefore a vector perpendicular to it in the counterclockwise direction (see picture) will be $-y\,\mathbf{i} + x\,\mathbf{j}$ (since its scalar product with $x\,\mathbf{i} + y\,\mathbf{j}$ is 0 and the signs are correct).

The preceding vector has magnitude r. If the angular velocity is ω, then the linear velocity is given by

$$|\mathbf{v}| = \omega r,$$

so to get the velocity field, we should multiply the above field by ω :

$$\mathbf{v} = -\omega y\,\mathbf{i} + \omega x\,\mathbf{j} .$$

Non-independent Variables

1. We give a worked example here. A fuller explanation will be given in the next session. Let

$$w = x^3y^2 + x^2y^3 + y$$

and assume x and y satisfy the relation

$$x^2 + y^2 = 1.$$

We consider x to be the independent variable, then, because y depends on x we have w is ultimately a function of the single variable x.

a) Compute $\dfrac{dw}{dx}$ using implicit differentiation.

b) Compute $\dfrac{dw}{dx}$ using total differentials.

<u>Answer:</u>

a) Implicit differentiation means remembering that y is a function of x, e.g., $\dfrac{dy^2}{dx} = 2y\dfrac{dy}{dx}$. Thus,

$$\frac{dw}{dx} = 3x^2y^2 + 2x^3y\frac{dy}{dx} + 2xy^3 + 3x^2y^2\frac{dy}{dx} + \frac{dy}{dx}.$$

Now we differentiate the constraint to find $\dfrac{dy}{dx}$.

$$x^2 + y^2 = 1 \;\Rightarrow\; 2x + 2y\frac{dy}{dx} = 0 \;\Rightarrow\; \frac{dy}{dx} = -\frac{x}{y}.$$

Substituting this in the equation for $\dfrac{dw}{dx}$ gives

$$\frac{dw}{dx} = 3x^2y^2 - 2x^3y\frac{x}{y} + 2xy^3 - 3x^2y^2\frac{x}{y} - \frac{x}{y} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

b) Taking total differentials of both w and the constraint equation gives

$$\begin{aligned} dw &= 3x^2y^2\,dx + 2x^3y\,dy + 2xy^3\,dx + 3x^2y^2\,dy + dy \\ &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\,dy \end{aligned}$$

$$2x\,dx + 2y\,dy = 0.$$

We can solve the second equation for dy and substitute in the equation for dw.

$$\begin{aligned} dy &= -\frac{x}{y}\,dx \;\Rightarrow \\ dw &= (3x^2y^2 + 2xy^3)\,dx + (2x^3y + 3x^2y^2 + 1)\left(-\frac{x}{y}\right)\,dx \\ &= (3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y})\,dx \end{aligned}$$

Thus,

$$\frac{dw}{dx} = 3x^2y^2 - 2x^4 + 2xy^3 - 3x^3y - \frac{x}{y}.$$

Second derivative test

1. Find and classify all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

Answer: Taking the first partials and setting them to 0:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 3y^2 - 12 = 0.$$

The first equation implies $x = -1$ and the second implies $y = \pm 2$. Thus, the critical points are $(-1, 2)$ and $(-1, -2)$.

Taking second partials:

$$\frac{\partial^2 z}{\partial x^2} = 30x^4, \quad \frac{\partial^2 z}{\partial xy} = 0, \quad \frac{\partial^2 z}{\partial y^2} = 6y.$$

We analyze each critical point in turn.
At (-1,-2): $A = z_{xx}(-1, -2) = 30$, $B = z_{xy}(-1, -2) = 0$, $C = z_{yy}(-1, -2) = -12$.
Therefore $AC - B^2 = -360 < 0$, which implies the critical point is a saddle.

At (-1,2): $A = z_{xx}(-1, 2) = 30$, $B = z_{xy}(-1, 2) = 0$, $C = z_{yy}(-1, 2) = 12$.
Therefore $AC - B^2 = 360 > 0$ and $A > 0$, which implies the critical point is a minimum.