# Chapter 2

# The Calculus

## 2. 1 Introduction

The **calculus** is the gateway to higher mathematics. Its discovery in the seventeenth century revolutionised the application of mathematics to science. I will choose to introduce the **calculus** through the concept of the **exterior derivative**, which I believe gives a particularly clear intuition as to the real meaning of the subject.

Whether you're aware of it or not, you are familiar with the concept of an **operator**: the symbol "+" is the **addition operator**, "×" is the **multiplication operator**¸ and so on. An **operator** is something that transforms one mathematical entity into another, most commonly into another of the same kind. The *minus sign* "–" is used for two **operators**, first as a **binary operator** between two numbers like "$4 - 3$" or "$x - y$", and secondly as a **unary operator** as in "$-1$". All these, which are called **arithmetic operators**, act on the *value* of the numbers involved and are simply another way of writing our familiar **functions**.

The **calculus** uses an **operator** which on its own I will call the **differential operator**, and which is written "**d**". This **differential operator** is a **unary operator** like the **unary negation** in "$-x$", but it acts slightly differently to the **arithmetic operators** above. It doesn't alter the *value* of the **variable**. Instead, applied to a **variable** like $x$ or $y$, it creates a *new* **variable** which we write as **d**$x$ or **d**$y$, which, although written as two characters are single entities. **d**$x$ is called the **differential** of $x$, and **d**$y$ is called the **differential** of $y$. When I've introduced **vectors** in Chapter 5, I will give a new interpretation to these **differentials**, but for this Chapter you can simply think of them as new number **variables** much like $x$ and $y$ themselves, able to take any assigned numerical value. It needs to be stressed that **d**$x$ and **d**$y$ do *not* represent the **product** of a new variable **d** with either of the old variables. Each **differential d**$x$ or **d**$y$ is a *single new* number, represented by two letters to show its association with the original variable from which it derives.

Applied to a **constant** number like 4 or 3.1415926, the result is rather special, and to stress this I will formulate it as a trivial "axiom 0":

> Axiom 0: The **differential** of a **constant** is identically zero.

So **d**(4) ≡ 0, **d**(−27.253) ≡ 0, and using the standard convention that in algebra letters from the start of the alphabet denote **constants**, as we've already seen in expressions like $ax^2 + bx + c$, we have **d***a* ≡ 0, **d***b* ≡ 0 and **d***c* ≡ 0. I use the "≡" symbol here to stress that these are not **equations**, but **identities**: **d***a* *always* has the value 0.

When applied to an **expression** like $x^2 + y$, the **differential operator** is defined by two further axioms:

Axiom I: $\mathbf{d}(x + y) = \mathbf{d}x + \mathbf{d}y$
Axiom II: $\mathbf{d}(x \times y) = \mathbf{d}x \times y + x \times \mathbf{d}y$

We can rewrite axiom I, using axiom II and axiom 0. By axiom II, we have for the product of a **constant** and a **variable** as in *ax*:

$$\mathbf{d}(ax) = \mathbf{d}(a \times x) = \mathbf{d}a \times x + a \times \mathbf{d}x$$

but since **d***a* = 0 by our "axiom 0" this reduces to just:

$$\mathbf{d}(ax) = a \times \mathbf{d}x.$$

So we can reformulate Axiom I in the more usual form:

Axiom I: $\mathbf{d}(ax + by) = a.\mathbf{d}x + b.\mathbf{d}y.$

This axiom simply says that **d** is a **linear operator**. Axiom II, giving the action of **d** on a product "×" is sometimes called the **derivative property**, sometimes the **Leibniz property**.[1]

Applying just these two axioms over and over on all the terms and factors of an **expression** (or through an **equation**) always leads eventually to an expression that is *linear* in the individual **differentials**. What this means is that however complicated the original expression, the result of applying the **differential operator d** always takes a form like:

$$g(x, y).\mathbf{d}x + h(x, y).\mathbf{d}y$$

if there were just two **variables** *x* and *y*, or like:

$$g(x, y, z).\mathbf{d}x + h(x, y, z).\mathbf{d}y + p(x, y, z).\mathbf{d}z$$

if there are three: *x*, *y* and *z*. Such an expression is called a **linear combination** in the **differentials d***x*, **d***y* and **d***z*.

In effect, the **operator** cascades down through the parts of the **expression** until a strictly linear form spills out.

The resulting **expression** is called the **exterior derivative** of the original **expression**, and the form it takes, which is always that of a **linear combination** in the individual **differentials**, is called a **differential form**. The general definition of a **differential form** at this lowest level is an **expression** which can be written as:

[1] Gottfried Leibniz (1646-1716) was one of the founders of the **calculus**.

$$f_1(x_1, x_2, ..x_n).\mathbf{d}x_1 + f_2(x_1, x_2, ..x_n).\mathbf{d}x_2 + \mathbf{. . .} + f_n(x_1, x_2, ..x_n).\mathbf{d}x_n$$

or in summation notation, $\Sigma_j f_j(x_1, x_2, ..x_n).\mathbf{d}x_j$, where only one **differential** appears in each term, and as a separate factor. In Chapter 6 I will extend the concept of a **differential form** by forming the **exterior derivative** of a **differential form** itself, but for this Chapter we will only need this first level of **differential form**, also called a **1-form**.[2]

Let's look at an example. Suppose we have $z = \varphi(x, y) = ax^2 + 4xy + c$, we can apply the **differential operator** to obtain this sequence:

$$\mathbf{d}z = \mathbf{d}\varphi(x, y) = \mathbf{d}(ax^2 + 4xy + c) = \mathbf{d}(ax^2) + \mathbf{d}(4xy) + \mathbf{d}c$$

by Axiom I in its simple form. So using Axiom 0 and Axiom I in its second form, this gives:

$$\mathbf{d}z = a.\mathbf{d}(x^2) + 4.\mathbf{d}(xy) + 0 = a.(x.\mathbf{d}x + x.\mathbf{d}x) + 4.(y.\mathbf{d}x + x.\mathbf{d}y)$$

using the **derivative property**, Axiom II, twice, expanding $x^2 = x \times x$ on the first term. So, collecting terms:

$$\mathbf{d}z = (2ax + 4y).\mathbf{d}x + 4x.\mathbf{d}y.$$

This is in the form $\mathbf{d}z = \varphi_x(x, y).\mathbf{d}x + \varphi_y(x, y).\mathbf{d}y$, with a **linear combination** in $\mathbf{d}x$ and $\mathbf{d}y$, or a **differential form** or **1-form** on the right.

When expressed like this, with a single variable on the left and a **differential form** on the right with each **differential** appearing just once in the **1-form**, the functions $\varphi_x(x, y)$ and $\varphi_y(x, y)$ appearing as the coefficients of the **differentials**, are called the **partial derivatives** of $z$ with respect to $x$ and $y$ respectively. This is the more common use of the term "**derivative**", but for this text I will emphasize the importance of the **exterior derivative**, and so when I refer simply to a **derivative** without a qualifying adjective, I will mean the **exterior derivative**.

Applying the **differential operator** to any entity is called **differentiating** the **original**. The verb is thus to **differentiate**. I will also widely use **derivative** as an adjective to describe the result of **differentiating**.

I will also refer to the entity **differentiated** as the **original** entity.

So why complicate things by doubling up the number of variables from those we had originally? The immediate usefulness of the new variables is precisely that they abstract the *linear* part of the relationship between the old variables. Suppose we have the equation describing a simple **parabola**:

$$y = x^2.$$

Applying the **differential operator** to this equation gives:

[2] To be consistent with this concept, the ordinary **functions** and **expressions** that have been used throughout up to now are sometimes called **0-forms**. Another term for them is **scalar functions**.

$$\mathbf{d}y = \mathbf{d}(x^2) = \mathbf{d}(x.x) = \mathbf{d}x.x + x.\mathbf{d}x = 2x.\mathbf{d}x.$$

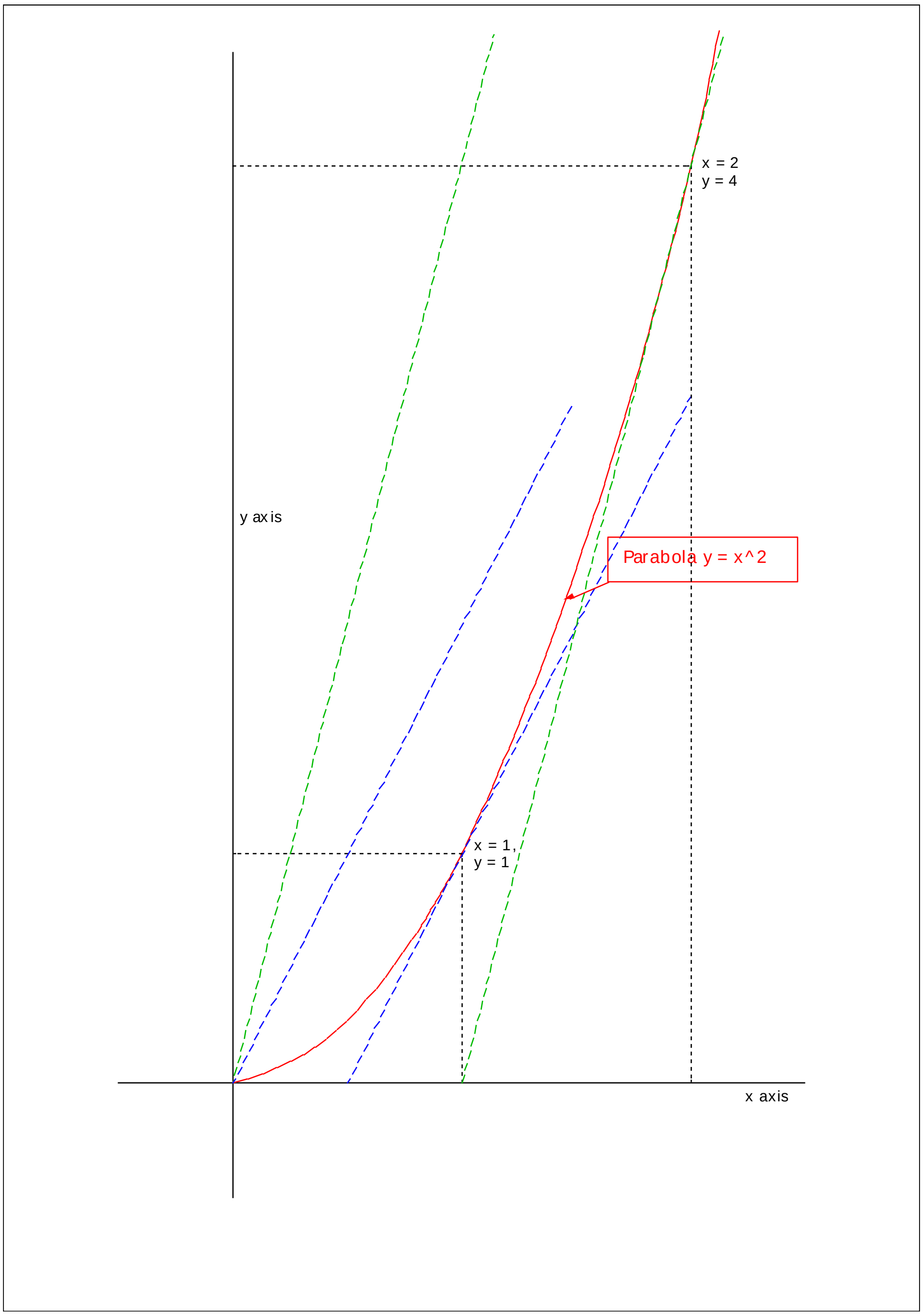


*Figure 2.1.1*

If we now look at the point where $x = 1$, so that $y = x^2 = 1$ also, what $\mathbf{d}y = 2x.\mathbf{d}x = 2.\mathbf{d}x$ shows is that at this point on the parabola the $y$ variable is changing twice as fast as the $x$. The graph in *Figure 2.1.1* shows this: the equation $\mathbf{d}y = 2.\mathbf{d}x$ is the equation of the straight line through the **origin** parallel to the **tangent** to the parabola at the point $x = 1$, $y = 1$. If we move to the point $x = 2$, where $y = x^2 = 4$, we will have $\mathbf{d}y = 2x.\mathbf{d}x = 4.\mathbf{d}x$, so at this point on the parabola, the $y$ value is changing four times as fast as the $x$, and the equation in **differentials** represents a line through the **origin** parallel to the **tangent** at $x = 2$, $y = 4$, again as shown in *Figure 2.1.1*.

Note how at each point on the original graph, there is defined a new equation in the **differentials**, which is always **linear** in them. It's as if we "freeze" or "lock" the **original** variables to see the relationship of the **derivative** ones (the **differentials**) at each point on the original graph. So although the **derivative equation** is in *four* variables – $x$, $y$, **d**$x$ and **d**$y$ – instead of the original *two* $x$ and $y$, we look at the **derivative** form separately at each $(x, y)$ point of the **original**. We can think of the **differential operator** as having generated a *family* of new equations in the **differentials**, one at each point on the original graph.

It is amazing what this apparently trivial concept opens up. I'll begin by working out some examples using the operator, and so start to build up an armoury of easy ready formulae to deal with various algebraic forms.

First, a word of warning: *don't* apply the **differential operator** across an inequality. Applied to an equation the **derivative equation** will also be true. But the independence of the **differentials** means that there is no guarantee that the new *inequality* will hold. So if we had $y > x^2$ we cannot infer that $\mathbf{d}y > 2x.\mathbf{d}x$. This is a more serious danger than the "**inequality trap**" from multiplying by a negative number referred to in Section 1.5. Here there is no connection *at all* between the original inequality and the **derivative** one.

## 2.2 Further Cases

The application to *subtraction* is easy. Imagine $b$ becomes $-b$ in axiom I, so that:

$$\mathbf{d}(a.x - b.y) = \mathbf{d}(a.x + (-b).y) = a.\mathbf{d}x + (-b).\mathbf{d}y = a.\mathbf{d}x - b.\mathbf{d}y$$

much as might be expected.

Division is a little more subtle. To get $\mathbf{d}(x/y)$ put $z = x/y$, so that $x = y.z$. Then

$$\mathbf{d}(x) = \mathbf{d}(y.z) = \mathbf{d}y.z + y.\mathbf{d}z \text{ from axiom II, so}$$

$$\mathbf{d}z = (\mathbf{d}x - \mathbf{d}y.z)/y = (\mathbf{d}x - \mathbf{d}y.(x/y))/y = \mathbf{d}x/y - x.\mathbf{d}y/y^2$$

or, more conveniently,

$$\mathbf{d}(x/y) = (y.\mathbf{d}x - x.\mathbf{d}y)/y^2.$$

We can use this technique for the square root too. Putting $x = \sqrt{x}.\sqrt{x}$,

$$\mathbf{d}x = \sqrt{x}.\mathbf{d}\sqrt{x} + \sqrt{x}.\mathbf{d}\sqrt{x} = 2\sqrt{x}.\mathbf{d}\sqrt{x} \text{ from axiom II,}$$

$$\mathbf{d}\sqrt{x} = \tfrac{1}{2}\mathbf{d}x/\sqrt{x}.$$

This can handle the cube root too. For $x = \sqrt[3]{x}.\sqrt[3]{x}.\sqrt[3]{x}$, so

$$\begin{aligned}\mathbf{d}x &= \mathbf{d}(\sqrt[3]{x}).\sqrt[3]{x}.\sqrt[3]{x} + \sqrt[3]{x}.\mathbf{d}(\sqrt[3]{x}.\sqrt[3]{x})\\ &= \mathbf{d}(\sqrt[3]{x}).\sqrt[3]{x}.\sqrt[3]{x} + \sqrt[3]{x}.(\mathbf{d}(\sqrt[3]{x}).\sqrt[3]{x} + \sqrt[3]{x}.\mathbf{d}(\sqrt[3]{x}))\\ &= 3(\sqrt[3]{x})^2.\mathbf{d}(\sqrt[3]{x})\end{aligned}$$

So $\quad \mathbf{d}(\sqrt[3]{x}) = \tfrac{1}{3}\,\mathbf{d}x/(\sqrt[3]{x})^2.$

A particularly useful result is that for the general power of $x$, $x^n$. This can be demonstrated easily by the technique known as **induction** introduced in Section 1.10, where a result is established for $n = 1$ or $n = 2$, and then it is shown that if it holds for $n$, it must also hold for $n + 1$. We already have the results for $n = 1$ and $n = 2$ as $\mathbf{d}(x^1) = \mathbf{d}x$, and $\mathbf{d}(x^2) = 2.x.\mathbf{d}x$. The $n = 1$ result is special, the definition of the **differential**. The general result follows the $n = 2$ form, and is:

$$\mathbf{d}(x^n) = n.x^{n-1}.\mathbf{d}x.$$

We can show this by **induction** like this. Given the result as above for $n$, we can then put for $n + 1$:

$$\begin{aligned}\mathbf{d}(x^{n+1}) &= \mathbf{d}(x^n.x) = \mathbf{d}(x^n).x + x^n.\mathbf{d}x = n.x^{n-1}\mathbf{d}x.x + x^n.\mathbf{d}x\\ &= n.x^{n-1}.x\,\mathbf{d}x + x^n.\mathbf{d}x = n.x^n\,\mathbf{d}x + x^n.\mathbf{d}x = (n+1).x^n\mathbf{d}x.\end{aligned}$$

If this looks one level out, put $m = n + 1$, so $n = m - 1$, and rewrite the result as:

$$\mathbf{d}(x^m) = m.x^{m-1}.\mathbf{d}x.$$

In the next chapter, we will introduce a formalism that shows that roots can be written as *fractional* **exponents**, so that $\sqrt[3]{x}$ can be written as $x^{1/3}$ and $\sqrt{x}$ as $x^{1/2}$. The same formalism puts the reciprocal of a power like $1/x^n$ as a *negative* **exponent** $x^{-n}$. This leads to the formula above extending to the square and cube roots we developed earlier. So the cube root formula is just

$$\mathbf{d}(\sqrt[3]{x}) = \mathbf{d}(x^{1/3}) = \tfrac{1}{3}.x^{(1/3 - 1)}.\mathbf{d}x = \tfrac{1}{3}.x^{-2/3}\,\mathbf{d}x$$

because $x^{-2/3}$ ($x$ to the *minus two-thirds* ) here is the same thing as $1/(\sqrt[3]{x})^2$.

Another way of showing this is to put $y = \sqrt[3]{x}$, so that $y^3 = x$, when we will have:

$$\mathbf{d}(y^3) = 3y^2.\mathbf{d}y = \mathbf{d}x$$

so $\quad \mathbf{d}y = \mathbf{d}(\sqrt[3]{x}) = \mathbf{d}x/(3y^2) = \tfrac{1}{3}(1/(\sqrt[3]{x})^2).\mathbf{d}x = \tfrac{1}{3}(1/x^{2/3})\mathbf{d}x = \tfrac{1}{3}.x^{-2/3}.\mathbf{d}x.$

If you've followed the ideas so far, you should be comfortable with evaluating the **derivatives** of some reasonably complicated forms such as:

$$\mathbf{d}(y = 4.x^2 + 3x + 17) \text{ gives } \mathbf{d}y = (4.2x + 3).\mathbf{d}x = (8x + 3)\mathbf{d}x$$

where I use the shorthand notation $\mathbf{d}(y = \varphi(x))$ for $(\mathbf{d}y = \mathbf{d}\varphi(x))$, to indicate that applying **d** to both sides of an **equation** will still give an **equation**, so we can think of **d** as applied across the **equation** as a whole. I also re-use Axiom 0, that the derivative of a **constant** is zero, so that, for example, $\mathbf{d}(17) = 0$.

The division formula enables us to tackle:

$$\begin{aligned}\mathbf{d}((7.x + 4)/(x^2 + x)) &= [(x^2 + x).\mathbf{d}(7.x + 4) - (7.x + 4).\mathbf{d}(x^2 + x)]/(x^2 + x)^2 \\ &= [(x^2 + x).7\mathbf{d}x - (7.x + 4).(2x + 1)\mathbf{d}x]/(x^2 + x)^2 \\ &= (7x^2 + 7x - 14x^2 - 7x - 8x - 4)\mathbf{d}x/(x^4 + 2x^3 + x^2) \\ &= [(-7x^2 - 8x - 4)\ /(x^4 + 2x^3 + x^2)]\ \mathbf{d}x.\end{aligned}$$

There need not be any variable actually standing on its own in an equation. For example, the equation:

$x^2 + y = 4x + 7z^2$ can be **differentiated** directly to give:

$$2x\mathbf{d}x + \mathbf{d}y = 4\mathbf{d}x + 14z\mathbf{d}z$$

or

$$(2x - 4)\mathbf{d}x + \mathbf{d}y - 14z\mathbf{d}z = 0.$$

Applying the **operator** to a general equation like this is sometimes called **implicit differentiation**.

If the explicit form of the **expression** or **equation** is not known, we can apply **functional notation** as described in Section 1.1. Suppose we are given $y$ as a **function** of $x$ in the general form $y = f(x)$. Then we know that whatever the *actual* form of $f$, the (**exterior**) **derivative** must be **linear** in the **differentials**. How do we write it? The most elegant way is to use the beautiful letter $\partial$, which seems to have no name, although it is commonly spoken as “dee” just like the operator itself, and I seem to have a very faint memory of its being sometimes called “del”. Using $\partial$, we write the **derivative equation** as:

$$\mathbf{d}y = \partial f.\mathbf{d}x = \partial f\,(x).\mathbf{d}x$$

where $\partial f\,(x)$ is a new **function** of $x$, called the **ordinary derivative** of $f$. So for example, if the **original function** is $f\,(x) = x^2$, then $\partial f\,(x) = 2x$, because if $y = f\,(x) = x^2$, then $\mathbf{d}y = 2x.\mathbf{d}x$. Another common notation is $\mathbf{d}y = f\,'\mathbf{d}x$, using a superscript tick or “dash” against $f$. The commonest of all however, is the pseudo-division form:

$$\mathbf{d}y = (\mathrm{d}f(x)/\mathrm{d}x).\mathbf{d}x$$

which I will discuss in the next section. If there are two **arguments** to the function, say $z = f\,(x, y)$, the $\partial$ notation lends itself to the elegant form:

$$\mathbf{d}z = \partial_x f.\mathbf{d}x + \partial_y f.\mathbf{d}y$$

which is simply a statement that the (**exterior**) **derivative** must be **linear** in the **differentials** of the **arguments**. Here $\partial_x f$ and $\partial_y f$ constitute a general notation for the **partial derivatives** appearing as the *coefficients* of each **differential**.

These are new **functions** $\partial_x f(x, y)$ and $\partial_y f(x, y)$. You should note that in the conventional terminology, the word **derivative** *without a qualifying adjective* is *only used for these* **functions**, *not* for the **exterior derivative** as I prefer to do in this text. We can without confusion use this same notation for the **ordinary derivative** above as well, so that:

$$\partial f(x) \equiv \partial_x f(x)$$

and I will often write the **ordinary derivative** this way too. The various $\partial_x f(x, y)$, $\partial_y f(x, y)$ forms that appear where there are multiple **arguments** are called **partial derivatives**, as stated above, and since these comprehend the **ordinary derivative** as a special case, I will generally avoid the latter term.

These functions can be **differentiated** again, so for example:

$$\mathbf{d}(\partial_x f(x, y)) = \partial_x\partial_x f(x, y).\mathbf{d}x + \partial_y\partial_x f(x, y).\mathbf{d}y$$

and it is convenient to write this as: $\partial_x^2 f(x, y).\mathbf{d}x + \partial_y\partial_x f(x, y).\mathbf{d}y$, so $\partial_x^2 \equiv \partial_x\partial_x$, another useful shorthand.

If the underlying variables themselves are functions of other variables, say $x$ and $y$ depend on $u$ and $v$, we can proceed by substituting for $\mathbf{d}x$ and $\mathbf{d}y$ using the equations for $x$ and $y$ in terms of $u$ and $v$, which might be $x = g(u, v)$, $y = h(u, v)$ and which will have their own **derivatives:**

$$\mathbf{d}x = \partial_u g.\mathbf{d}u + \partial_v g.\mathbf{d}v$$

and

$$\mathbf{d}y = \partial_u h.\mathbf{d}u + \partial_v h.\mathbf{d}v.$$

If we now substitute these into $\mathbf{d}z = \partial_x f.\mathbf{d}x + \partial_y f.\mathbf{d}y$ we get:

$$\mathbf{d}z = \partial_x f.\,(\partial_u g.\mathbf{d}u + \partial_v g.\mathbf{d}v) + \partial_y f.\,(\partial_u h.\mathbf{d}u + \partial_v h).\mathbf{d}v$$

$$= (\partial_x f.\,\partial_u g + \partial_y f.\,\partial_u h).\mathbf{d}u + (\partial_x f.\partial_v g + \partial_y f.\,\partial_v h).\mathbf{d}v$$

a result known as the **chain rule**.

Care needs to be exercised in using this rule as we must evaluate $\partial_x f$ as $\partial_x f(x(u, v), y(u, v))$, but $\partial_u g$ as $\partial_u g(u, v)$. In other words, we must choose *the same* $(u, v)$ point to evaluate the expression through both levels. As so often, an example may help. Suppose we have:

$$z = 3x^2 + xy$$

$$x = u + 2v$$

$$y = v^3$$

then[3]

$$\mathbf{d}z = 6x.\mathbf{d}x + y.\mathbf{d}x + x.\mathbf{d}y = (6x + y).\mathbf{d}x + x.\mathbf{d}y$$

$$\mathbf{d}x = \mathbf{d}u + 2.\mathbf{d}v$$

$$\mathbf{d}y = 3v^2.\mathbf{d}v.$$

Then $\mathbf{d}z = (6x + y).\mathbf{d}x + x.\mathbf{d}y = (6x(u, v) + y(u, v)).\mathbf{d}x + x(u, v).\mathbf{d}y$

$$= (6x(u, v) + y(u, v)).(\mathbf{d}u + 2.\mathbf{d}v) + x(u, v).(3v^2.\mathbf{d}v)$$

$$= (6(u + 2v) + (v^3)).(\mathbf{d}u + 2.\mathbf{d}v) + (u + 2v).(3v^2.\mathbf{d}v)$$

$$= (6u + 12v + v^3).\mathbf{d}u + (12u + 24v + 2v^3 + 3uv^2 + 6v^3).\mathbf{d}v$$

$$= (6u + 12v + v^3).\mathbf{d}u + (12u + 24v + 8v^3 + 3uv^2).\mathbf{d}v.$$

There's no new maths here. We just need to be consistent in substituting at both the **original** level and at the **derivative** level, respectively for *z*, *x* and *y*, and for **d***z*, **d***x* and **d***y*.

## 2.3 The Infinitesimal Curse

Nowhere have I suggested that the **differentials** have to be small numbers, let alone *infinitely* small. This is in strong contrast to the conventional presentation, which treats **d***x* and **d***y* as being tiny changes in the **original variables** *x* and *y*. I have been at pains to emphasize that the **differential operator** merely creates a **linear differential form** from the original expression or equation.

Nevertheless, the *infinitesimal* idea played a huge role in the founding of the calculus, and we need to look at why it came to have the significance it did.

Let's go back to our parabola $y = x^2$. Imagine a point on this curve $(x, y)$ obeying the relationship $y = x^2$ and imagine another point also on the curve at the *x* value $x + \mathbf{d}x$ where for the moment we *do* assume that **d***x* is a very small value. Call the corresponding *y* value $y^*$, so that $(x + \mathbf{d}x, y^*)$ is also on the curve, and we will have $y^* = (x + \mathbf{d}x)^2$. So

$$y^* = (x + \mathbf{d}x)^2 = x^2 + 2x.\mathbf{d}x + (\mathbf{d}x)^2.$$

Now if **d***x* is very small, say 0.0001, then $(\mathbf{d}x)^2$ will be much smaller again – 0.0000001 – and we can regard this term as negligible, so that

$y^* = x^2 + 2x.\mathbf{d}x$ to a very good approximation. But $y = x^2$, so this becomes:

---

[3] Here I use the functional notation $x = x(u, v)$ to indicate that *x* is a **function** of *u* and *v*. This notation is widely disparaged because "*x*" is being used for both a **variable** and a **function**, but even computer software can be made "intelligent" enough to see the distinction by means of *context-sensitive parsing*, and the suggestive power of the association is very helpful.

$y^* = y + 2x.\mathbf{d}x$

or

$y^* - y = 2x.\mathbf{d}x.$

If we write the small change in *y* corresponding to **d***x* and appearing here as $y^* - y$ as **d***y* or:

$\mathbf{d}y := y^* - y$

using the symbol "**:**=" to mean "is defined as", we can write:

$\mathbf{d}y = 2x.\mathbf{d}x$

so the equation in **differentials** *approximates* the change in the values of the **original** variables very accurately for a small change.

In the early years of the calculus, this **infinitesimal** property was seen as the cardinal defining attribute of the calculus, and the subject actually came to be called the **infinitesimal calculus**. I simply do not see it this way. I have to admit that I am suspicious about any beliefs in infinities or infinitesimals in maths, and have a lot of sympathy with Bishop Berkeley who lambasted the calculus in the early eighteenth century on the basis that its practitioners took **differentials** to be significant or negligible simply as the whim suited them! It seems to me that the cardinal property is the **linearity** of the **derivative** forms, and this can be established purely algebraically without any recourse to infinities or infinitesimals with all their attendant philosophical difficulties. In brief, my thesis is this: *don't resort to infinities or infinitesimals if there is any other interpretation available*.

My picture is that the **differential operator** defines for any **original equation** a family of **linear** equations, which describe for two variables a family of *lines* through the **origin**, and for three variables a family of *planes* through the **origin** – equations in **d***x*, **d***y* and **d***z* – with one unique member of the family for every point on the **original** line if there are two variables, or every point on the **original** *surface* if there are three variables. For more than three variables, the geometrical imagery gets difficult, as we have to think of *hypersurfaces* and *hyperplanes*, but the algebra carries on without a murmur.

A major objection to the infinitesimal concept is that it led to the idea that **differentials** couldn't stand on their own in an equation and needed to be "inflated back up" to "proper-size" variables by always using them in a *ratio* form. This in turn led to the "pseudo-division" form for the **coefficients** of the **differentials** in any **linear derivative**, and indeed the normal presentation of the calculus uses this concept almost exclusively, so you have to know it. The idea is that we have a functional form like

$w = f(x, y, z)$

with a **derivative** – in my terminology:

$\mathbf{d}w = \partial_x f.\mathbf{d}x + \partial_y f.\mathbf{d}y + \partial_z f.\mathbf{d}z.$

Now set **d***y* = 0 and **d***z* = 0, and divide through by **d***x* to give:

$(\mathbf{d}w/\mathbf{d}x)_{y,z} = \partial_x f$

where the subscripts indicate that *y* and *z* are being "held constant" – i.e. treated as constants, so $\mathbf{d}y = 0$ and $\mathbf{d}z = 0$. This is commonly written using the ∂ symbol as:

$(\partial w/\partial x)_{y,z}$

and this is the standard notation for a **partial derivative**. The **partial** bit refers to the fact that in this case there are other variables which are "held constant". If we have a relationship between only two variables like $y = f(x)$, so that $\mathbf{d}y = \partial_x f.\mathbf{d}x$ unambiguously, we dispense with the subscripts and use just "d" instead of ∂ to give the **ordinary derivative**:

$\mathrm{d}f(x)/\mathrm{d}x.$

I think this pairwise approach gives quite a wrong picture of what is going on. There isn't normally some paired connection between variables taken two by two, and an equation in **differentials** may often take the more general form:

$\partial_x f.\mathbf{d}x + \partial_y f.\mathbf{d}y + \partial_z f.\mathbf{d}z = 0$

where no individual variable is singled out as having a coefficient of unity (1). We don't treat variables in normal linear equations two by two in this pairwise fashion, and we shouldn't do so with equations in **differentials**.

The auxiliary concept of "holding the other variables constant" permeates the traditional presentation like a canker. It is not there in the actual algebra, and is a quite unnecessary notion.

At this point I should stress again that in the literature the term **derivative** without a qualifying adjective is *only* used for the **ordinary** and **partial derivatives** defined above. Because this usage is so established, I will still use it, referring to forms like $\partial_x f$ as **ordinary** or **partial derivatives** as appropriate, but for brevity I will also use:

**OD** for the **ordinary derivative** $\partial f(x)$, and

**PD** for the **partial derivatives** such as $\partial_y f(x, y, z)$.

## 2.4 Inversion of the Differential Operator

A central question in the calculus is whether we can obtain the **original** form from the **derivative** form. This procedure, the **inverse** of **differentiating**, is commonly called **integrating** or "**indefinite integration**", but the term "**integration**" is also used for a more significant operation called "**definite integration**". The two concepts are quite distinct and I am going to reserve the term **integration** for the concept introduced in Section 2.6. So I will call **inverse differentiation** simply that or else **anti-differentiation**. This suggests that the **original** form from which a **derivative** is obtained may be called the **anti-derivative**. All these terms are rather clumsy but as it happens, we will not need to use any of them very much.

An obvious possibility is to try and find an **inverse operator** to the **differential operator** which we might label something like $\mathbf{d}^{-1}$. Unfortunately, *there is no such inverse* **operator**. The **original** form can only be divined by inspection, although there are some rules. This exercise is the huge field of "Methods of Integration".

It is very important to realize two things: there simply may not *be* an **original** form at all, and even when there is it will not be unique. These two points need some explanation.

First, let's consider some **differential forms**. A **differential form** is a quite specific thing, a **linear** expression or **linear combination**[4] in **differentials**. Examples are:

$$14yz.\mathbf{d}x + x^3y.\mathbf{d}y - 22.3x.\mathbf{d}z$$

$$\partial_x f.\mathbf{d}x + \partial_y f.\mathbf{d}y + \partial_z f.\mathbf{d}z$$

$$2xy.\mathbf{d}x + x^2.\mathbf{d}y$$

$$(x^4 + 2xy - 3pq).\mathbf{d}s + (12p - u^3).\mathbf{d}p - (3p - xu).\mathbf{d}q$$

where in the last I've tried to move away from just $x$ and $y$.

Only the middle two of these are the (**exterior**) **derivatives** of an **original expression**. The other two are simply **differential forms** that could not owe their origin to the application of the **differential operator** to a single **original** expression. The reason is simply that they do not fit the format of the second one. For example, in the first, the coefficient of $\mathbf{d}x$ ($14yz$) would have to be $\partial_x g$ and that of $\mathbf{d}y$ ($x^3y$) would have to be $\partial_y g$ for some function $g(x, y, z)$, and they just don't match that prescription. We can tell that they don't because we can test for this prescription using a reault known as **Young's theorem**, which is:

$$\partial_y\partial_x g(x, y) = \partial_x\partial_y g(x, y) \qquad \text{always.}$$

Applying **Young's theorem** to the expressions $14yz \overset{?}{=} \partial_x g(x, y, z)$ and $x^3y \overset{?}{=} \partial_y g(x, y, z)$ we should have:

$$\partial_y(14yz) = 14z = \partial_x(x^3y) = 3x^2y$$

and clearly we don't. (To see these are the $\partial_y$ and $\partial_x$ **partial derivatives** of $14yz$ and $3x^2y$ respectively, imagine applying **d** to the form $14yz$ to give $14z.\mathbf{d}y + 14y.\mathbf{d}z$ for example. In this expression $14z$ is the $\partial_y$ part of the (**exterior**) **derivative** of an **original** expression $\partial_x g(x, y, z) = 14yz$).

$2xy.\mathbf{d}x + x^2.\mathbf{d}y$, however, does obey this rule, with $\partial_y(2xy) = 2x$, which *does* equal $\partial_x(x^2) = 2x$, and so this one *is* the **derivative** of an **original** form $x^2y$:

$$\mathbf{d}(x^2y) = 2xy.\mathbf{d}x + x^2.\mathbf{d}y.$$

> A **differential form** that *is* the (**exterior**) **derivative** of an **original** expression is called **exact**.

[4] See the Glossary for a brief definition.

The same **derivative** can also come from *more than one* **original**, because “constant” terms drop out. So both $x^2 + 3x + 2$ and $x^2 + 3x – 220$ give rise to the same (**exterior**) **derivative** $(2x + 3)$.**d**$x$. This “unknown constant” or “lost constant” is called the **constant of integration**.[5]

Inversion of **differentiation** is in general nasty, but one result we can get straight away, and this will always be useful. From $\mathbf{d}(x^n) = nx^{n-1}\mathbf{d}x$, putting $m = n - 1$, and so $n = m + 1$, we obtain:

$$(1/(m+1)).\mathbf{d}(x^{m+1}) = x^m.\mathbf{d}x$$

or $$x^m.\mathbf{d}x = \mathbf{d}(x^{m+1}/(m+1))$$

which enables us to invert the terms of any **polynomial** in one variable. So, for example:

$$(ax^2 + bx + c).\mathbf{d}x = ⅓a.\mathbf{d}(x^3) + ½b.\mathbf{d}(x^2) + c.\mathbf{d}x + k$$

$$= \mathbf{d}(⅓a.x^3 + ½b.x^2 + c.x + k)$$

where $k$ is the **constant of integration**. So this is the general form that *any* **original** expression must take to have the (**exterior**) **derivative** $(ax^2 + bx + c)$.**d**$x$.

You may ask whether there is a more general way in which the axiom II of the **d operator** can be used, and indeed there is. Putting it as:

$$\mathbf{d}(u(x).v(x)) = u(x).\mathbf{d}v(x) + \mathbf{d}u(x).v(x) = u(x).\partial_x v(x).\mathbf{d}x + \partial_x u(x).v(x).\mathbf{d}x$$

we have the result:

$$u(x).\partial_x v(x).\mathbf{d}x = \mathbf{d}(u(x).v(x)) - \partial_x u(x).v(x).\mathbf{d}x$$

which seems to have got us no further, as we’ve merely swapped the roles of $u$ and $v$. But by adroit choice of $u$ and $v$, this can be surprisingly useful, and we will use it in Section 2.7 to derive **Taylor’s expansion**. This use of the $u$–$v$ swap using axiom II is called **integration by parts**.[6]

---

5 From the “**indefinite integration**” terminology mentioned at the start of this Section, which I would prefer to avoid.

6 Again using the “**indefinite integration**” terminology.

## 7.2 Elements of Complex Analysis

**Complex analysis** is the theory of the **calculus** applied to **functions** that **map** from the plane *into* (or onto) the plane. We have already encountered several ways of treating points in the plane. Examples are:

- Defining a point **p** in the plane by its Cartesian **coordinates** $x$ and $y$, usually expressed as an **ordered pair** $(x, y)$.
- Expressing $(x, y)$ as a **2-vector** $x\mathbf{e}_1 + y\mathbf{e}_2$.
- Pre-multiplying the **2-vector** by $\mathbf{e}_1$ to give $x\mathbf{e}_1\mathbf{e}_1 + y\mathbf{e}_1\mathbf{e}_2 = x + jy$, where we define $j$ to be the **2-pseudoscalar**: $j := \mathbf{e}_1\mathbf{e}_2$, so that $j^2 = -1$.
- Expressing $x + jy$ as a **magnitude** $r$ and an **angle θ**, which can itself be written either as $\theta_x + j\theta_y$, or as $\mathbf{e}^{j\theta} \equiv \mathbf{e}\uparrow j\theta$, which gives $x + jy$ as $r\mathbf{e}^{j\theta}$.

Both the $x + jy$ and $r\mathbf{e}^{j\theta}$ forms are called **complex numbers**, and because of its built-in ability to handle rotations in the plane simply by multiplying by $\theta_x + j\theta_y$ to rotate a point by an *anticlockwise* **angle θ**, or by the **complex conjugate** of **θ**, which is $\theta_x - j\theta_y$, to turn a point through a *clockwise* **angle −θ**, and for its general power and elegance, the **complex number** formulation is the preferred way to handle points in the plane.

Since $x$ and $y$ are commonly used for the **coordinates** of a point in the plane, the letter $z$ is preferred for the points (or **complex numbers**) themselves, and it is used in various guises, as plain $z$, as $z_1$, $z_2$, and so on, and as Greek $\zeta$. So we will often write $z = x + jy$, and therefore **functions** defined with a point in the plane as **argument** appear as, for example, $\varphi(z)$ or $\varphi(x + jy)$ rather than $\varphi(x, y)$. Because I cannot enter overscores − the usual notation for **complex conjugates** − on this word processor, I will use the alternative, and slightly more rare, asterisk notation, so:

$$z = x + jy \rightarrow z^* := x - jy.$$

For **functions** that **map** from the plane *into* (or onto) the plane, the *resulting value* of φ is *also* a **complex number**, so that φ takes the form:

$$\varphi(x, y) = \varphi(x + jy) = u(x + jy) + jv(x + jy)$$

so that φ *itself* has both an $x$ and a $y$ component, with

$$\varphi_x(x, y) = u(x + jy)$$
$$\varphi_y(x, y) = v(x + jy).$$

The $u$ and $v$ here are simply the parts of the *double* **function** φ, which takes a point in the *plane* $(x, y)$ and **maps** it into another point in the plane $(u(x, y), v(x, y))$. Put another way, now treating the plane as the space of **2-vectors**, this is:

$$\varphi(x\mathbf{e}_x + y\mathbf{e}_y) = u(x, y)\mathbf{e}_x + v(x, y)\mathbf{e}_y.$$

Bringing in the $z$ notation, the **function** φ might be written as:

$$\zeta = \varphi(z) = u(z) + jv(z) \equiv u(z)\mathbf{e}_x + v(z)\mathbf{e}_y = u(x, y)\mathbf{e}_x + v(x, y)\mathbf{e}_y.$$

I may refer to these as **complex-to-complex functions**, or to emphasize that the **complex number** aspect is simply an *interpretation*, as **plane-to-plane functions**.

The key theorem that gives the foundation of **complex analysis** is **Green's theorem** from Section 6.8, which I will repeat here:

$$(\partial_x g(x, y) - \partial_y f(x, y)).\mathbf{d}x\mathbf{d}y \mid \Sigma = (f(x, y).\mathbf{d}x + g(x, y).\mathbf{d}y) \mid \mathbf{b}\Sigma$$

where $\Sigma$ is a bounded area in the plane, with **boundary b**$\Sigma$. In the **complex** formulation, the theorem is given a rather elegant form by the introduction of two **derivative operators** rather than using the **differential operator** directly. These two **operators** are defined as:

$$\partial_z := \tfrac{1}{2}(\partial_x - j\partial_y) \qquad \text{and} \qquad \partial_z{}^* := \tfrac{1}{2}(\partial_x + j\partial_y).$$

These **operators** are not actually numbers, but are treated as obeying the regular algebra of **complex numbers** and distributing across their arguments accordingly. They have a *pseudo-***complex conjugate** form, and *note which way round they are*: the one with the **complex conjugate** symbol "*" is now the one with the *plus sign*. Their constituent **derivatives** $\partial_x$ and $\partial_y$ just mean what they always mean − they define the **partial derivatives** or **PD**'s with respect to $x$ and $y$. These **operators** really only affect $\partial_z z$, $\partial_z z^*$, $\partial_z{}^*z$, and $\partial_z{}^*z^*$ because, like all **derivatives**, they simply cascade down through functions of $z$ and $z^*$ by the **chain rule** of Section 2.2, remembering that $z = z(x, y) = x + jy$, and $z^* = z^*(x, y) = x - jy$. So, for example:

$$\partial_z(z^3) = 3z^2.\partial_z z.$$

They are written the "wrong" way round with good reason, as we see if we expand the four basic forms:

$$\partial_z z = \tfrac{1}{2}(\partial_x - j\partial_y)(x + jy) = \tfrac{1}{2}(\partial_x x + j\partial_x y - j\partial_y x - j^2\partial_y y) = \tfrac{1}{2}(\partial_x x + \partial_y y) = \tfrac{1}{2}(1 + 1) = 1$$

$$\partial_z z^* = \tfrac{1}{2}(\partial_x - j\partial_y)(x - jy) = \tfrac{1}{2}(\partial_x x - j\partial_x y - j\partial_y x + j^2\partial_y y) = \tfrac{1}{2}(\partial_x x - \partial_y y) = \tfrac{1}{2}(1 - 1) = 0$$

$$\partial_z{}^*z = \tfrac{1}{2}(\partial_x + j\partial_y)(x + jy) = \tfrac{1}{2}(\partial_x x + j\partial_x y + j\partial_y x + j^2\partial_y y) = \tfrac{1}{2}(\partial_x x - \partial_y y) = \tfrac{1}{2}(1 - 1) = 0$$

$$\partial_z{}^*z^* = \tfrac{1}{2}(\partial_x + j\partial_y)(x - jy) = \tfrac{1}{2}(\partial_x x - j\partial_x y + j\partial_y x - j^2\partial_y y) = \tfrac{1}{2}(\partial_x x + \partial_y y) = \tfrac{1}{2}(1 + 1) = 1$$

where we remember that $\partial_x y = \partial_y x = 0$, and $\partial_x x = \partial_y y = 1$. So likewise, these **complex derivative operators** give $\partial_z z^* = \partial_z{}^*z = 0$ and $\partial_z z = \partial_z{}^*z^* = 1$.

Just to show that these **operators** are consistent, I'll give another example:

$$\partial_z(z^2) = \tfrac{1}{2}(\partial_x - j\partial_y)(x + jy)^2 = \tfrac{1}{2}(\partial_x - j\partial_y)(x^2 + 2jxy + j^2y^2)$$

$$= \tfrac{1}{2}(2x + 2jy - 2j^2x - 2j^3y) = x + jy + x + jy$$

$$= 2(x + jy) = 2z.$$

Applying $\partial_z{}^*$ to a **complex-to-complex function** $\varphi(z) = u(z) + jv(z)$, we find:

$\partial_z{*}\varphi(z) = ½(\partial_x + j\partial_y)(u(z) + jv(z)) = ½[\partial_x u(z) + j^2.\partial_y v(z) + j\partial_y u(z) + j\partial_x v(z)]$

$= ½[\partial_x u(z) - \partial_y v(z) + j.[\partial_y u(z) + \partial_x v(z)]].$

So $\partial_z{*}\varphi(z) = 0$ if and only if $\partial_x u(z) = \partial_y v(z)$ and $\partial_y u(z) = -\partial_x v(z)$.

These subsidiary equations in terms of $\partial_x$ and $\partial_y$ on $u(z)$ and $v(z)$ are called the **Cauchy-Riemann equations**. What makes them important is that they give a critical condition for the application of **Green's theorem**. For if we take a **1-form** in the plane of the form:

$\varphi(z).\mathbf{d}z = (u(z) + jv(z))(\mathbf{d}x + j\mathbf{d}y)$

where we have simply applied the **differential operator** in the usual way to $z = x + jy$ to get:

$\mathbf{d}z = \mathbf{d}x + j\mathbf{d}y,$

the expression for $\varphi(z).\mathbf{d}z$ expands to give:

$\varphi(z).\mathbf{d}z = (u(z)\mathbf{d}x + j^2 v(z)\mathbf{d}y) + j(u(z)\mathbf{d}y + v(z)\mathbf{d}x)$

$= (u(z)\mathbf{d}x - v(z)\mathbf{d}y) + j.[u(z)\mathbf{d}y + v(z)\mathbf{d}x].$

Now, replacing $z$ with $x$, $y$, and applying the **1-form** to a **boundary** of a closed region $\Sigma$, the *second* term becomes:

$j.(v(x, y).\mathbf{d}x + u(x, y).\mathbf{d}y) \mid \mathbf{b}\Sigma = j.(\partial_x u(x, y) - \partial_y v(x, y)).\mathbf{d}x\mathbf{d}y \mid \Sigma.$

Applying **Green's theorem**, if $\partial_z{*}\varphi(z) = 0$, then $\partial_x u(z) = \partial_y v(z)$ by the first **Cauchy-Riemann equation**, and so the **integral** is zero.

Likewise for the *first* term, we obtain:

$(u(x, y).\mathbf{d}x - v(x, y).\mathbf{d}y) \mid \mathbf{b}\Sigma = (-\partial_x v(x, y) - \partial_y u(x, y)).\mathbf{d}x\mathbf{d}y \mid \Sigma,$

and again, if $\partial_z{*}\varphi(z) = 0$, then $\partial_y u(z) = -\partial_x v(z)$ by the second **Cauchy-Riemann equation**, and so the **integral** is zero.

This proves **Cauchy's integral theorem**: that if $\partial_z{*}\varphi(z) = 0$, then $\varphi(z).\mathbf{d}z \mid \mathbf{b}\Sigma = 0$.

**Complex-to-complex functions** which obey $½(\partial_x + j\partial_y)\varphi(z) = \partial_z{*}\varphi(z) = 0$ or the **Cauchy-Riemann equations**, are said to be **analytic** or **holomorphic**[7] because they *do* have a valid $\partial_z\varphi(z)$ **derivative**. *These terms are very important.*

We will also need a second result, known as **Cauchy's integral formula**. For this, we need first to evaluate the **line integral**:

$\varphi(z).\mathbf{d}z \mid \mathbf{C} = (z - z_0)^m\mathbf{d}z \mid \mathbf{C}$

---

[7] As e.g. Nickerson, Spencer and Steenrod p.510.

where **C** is a small circle of radius ρ about the point $z_0$. To evaluate this, we **parametrise C** by $t$ on the **interval** [0, 2π], giving:

$$z(t) = z_0 + \rho\mathbf{e}^{jt}.$$

This is of course using the convenient **parametrisation** of a circle introduced in Sections 4.4 and 4.6, this time around a point $z_0$ away from the **origin**. Now:

$$(z - z_0)^m = \rho^m\mathbf{e}^{jmt} \qquad \text{and} \qquad \mathbf{d}z = j\rho\mathbf{e}^{jt}.\mathbf{d}t,$$

so the **line integral** around the circle, [0, 2π] being the **pullback** of **C**, becomes:

$$(z(t) - z_0)^m.\mathbf{d}z \mid \mathbf{C} = (\rho^m\mathbf{e}^{\,jmt} \times j\rho\mathbf{e}^{\,jt}).\mathbf{d}t \mid [0, 2\pi] = j\rho^{(m+1)} \times \mathbf{e}^{\,j\,(m+1)t}\mathbf{d}t \mid [0, 2\pi]\ .$$

When $m = -1$, this gives:

$$j\rho^0 \times \mathbf{e}^0\mathbf{d}t \mid [0, 2\pi] = j.\mathbf{d}t \mid [0, 2\pi] = j.[2\pi - 0] = 2\pi j$$

because $x^0 = 1$ always, and for $m \neq -1$,

$$\mathbf{e}^{\,j\,(m+1)t}\mathbf{d}t \mid [0, 2\pi] = \mathbf{e}^{\,j\,(m+1)t}/(j(m+1)) \mid \mathbf{b}[0, 2\pi] = 0,$$

because $\mathbf{e}^0 = \mathbf{e}^{n2\pi} = 0$.

Now we can proceed straight to **Cauchy's integral formula**, which states that

> **Cauchy's integral formula**: if φ(z) is **analytic** in a region including a closed path **C**, (i.e. $\partial_z{}^*\varphi(z) = 0$ in this region), then for any point $z_0$ enclosed by the path **C**:
>
> $$[\varphi(z)/(z - z_0)].\mathbf{d}z \mid \mathbf{C} = 2\pi j.\varphi(z_0).$$

First we evaluate:

$$\partial_z{}^*[\varphi(z)/(z - z_0)] = \partial_z{}^*[\varphi(z)(z - z_0)^{-1}] = (\partial_z{}^*\varphi(z))(z - z_0)^{-1} - \varphi(z)(z - z_0)^{-2}\partial_z{}^*(z - z_0)$$

$$= 0.(z - z_0)^{-1} - \varphi(z)(z - z_0)^{-2}\partial_z{}^*z = 0 - \varphi(z)(z - z_0)^{-2}.0 = 0$$

so, defining $\psi(z) := \varphi(z)/(z - z_0)$, we can say that ψ(z) is also **analytic** *wherever* $z \neq z_0$, so **Cauchy's integral theorem** holds for it.

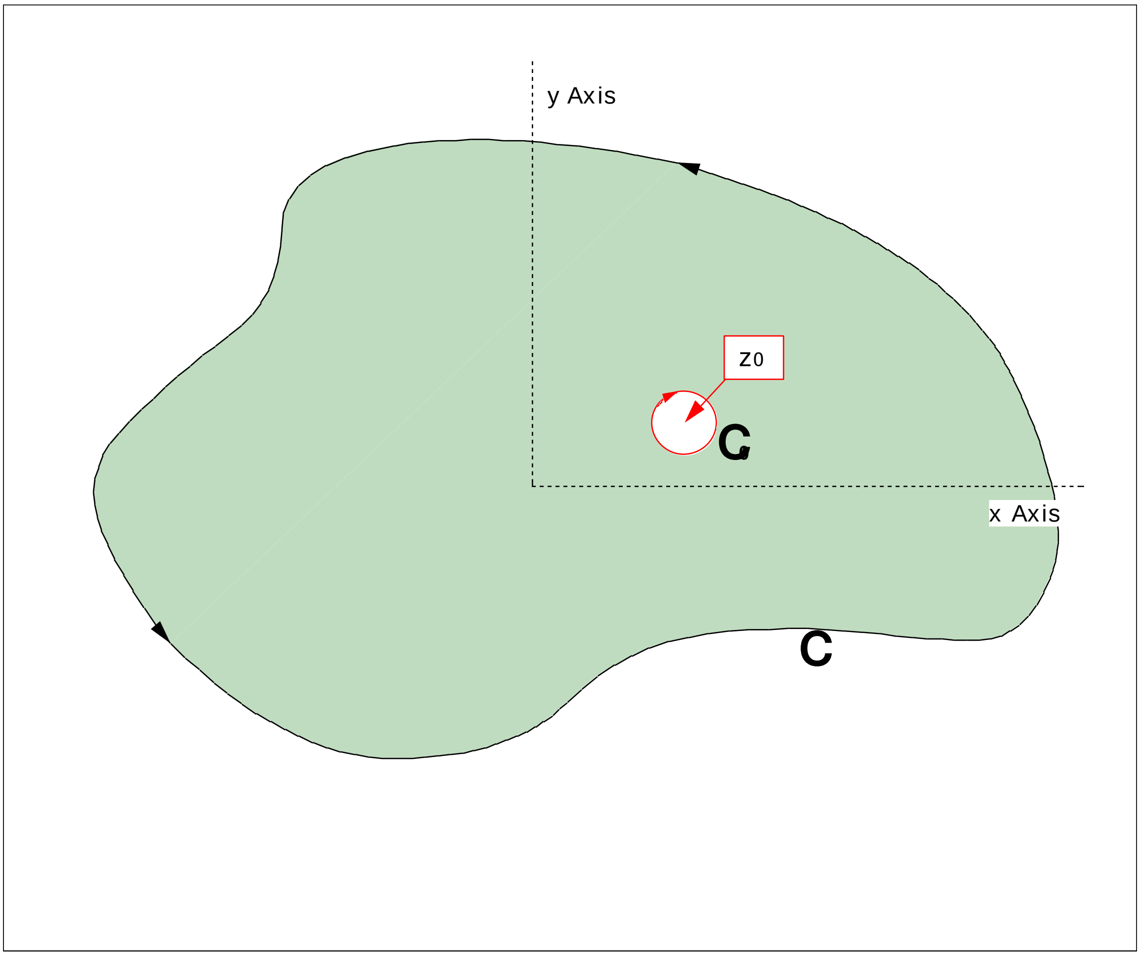


*Figure 7.2.1*

Now we use the device indicated in *Figure 7.2.1*, putting an arbitrarily small ring $\mathbf{C}_0$ around $z_0$ and in the annular space − shown shaded in the *Figure* − between **C** and $\mathbf{C}_0$,

$\psi(z) = \varphi(z_0)/(z - z_0)$ is **analytic** there and so, by **Cauchy's integral theorem**, the combined **integral**:

$$[\varphi(z)/(z - z_0)].\mathbf{d}z \mid (\mathbf{C} + \mathbf{C}_0)$$

over the *two* curves is zero. Note that here we go round **C** *anticlockwise* and $\mathbf{C}_0$ *clockwise* to give a consistently orientated **boundary** to the annulus. So it must be that:

$$[\varphi(z)/(z - z_0)].\mathbf{d}z \mid \mathbf{C} = -[\varphi(z)/(z - z_0)].\mathbf{d}z \mid \mathbf{C}_0.$$

If $\mathbf{C}_0$ is small enough − and we can make it as small as we choose − then $\varphi(z) \approx \varphi(z_0)$ throughout the area of $\mathbf{C}_0$, and so:

$$-[\varphi(z)/(z - z_0)].\mathbf{d}z \mid \mathbf{C}_{0(clockwise)} = [\varphi(z)/(z - z_0)].\mathbf{d}z \mid \mathbf{C}_{0(anticlockwise)}$$

$$\approx [\varphi(z_0)/(z - z_0)].\mathbf{d}z \mid \mathbf{C}_{0(anticlockwise)} = \varphi(z_0).[(z - z_0)^{-1}.\mathbf{d}z \mid \mathbf{C}_0] = 2\pi j.\varphi(z_0).$$

This **Cauchy integral formula** opens up tremendous possibilities, because these **line integrals** in the plane come up again and again, and they usually involve **functions** that are **analytic** *everywhere except at a few singular points*. If such is the case, we can “hive off” the **singularities** as shown in *Figure 7.2.2* and evaluate the **line integrals** around them by **Cauchy’s integral formula**. Because the **function** is **analytic**, the **line integral** *overall* must be zero, so that over the enveloping curve − **C** in the *Figure* − it must be *equal and opposite* to the sum of the **line integrals** around the **singularities**. This trick is called **integration by the method of residues**.

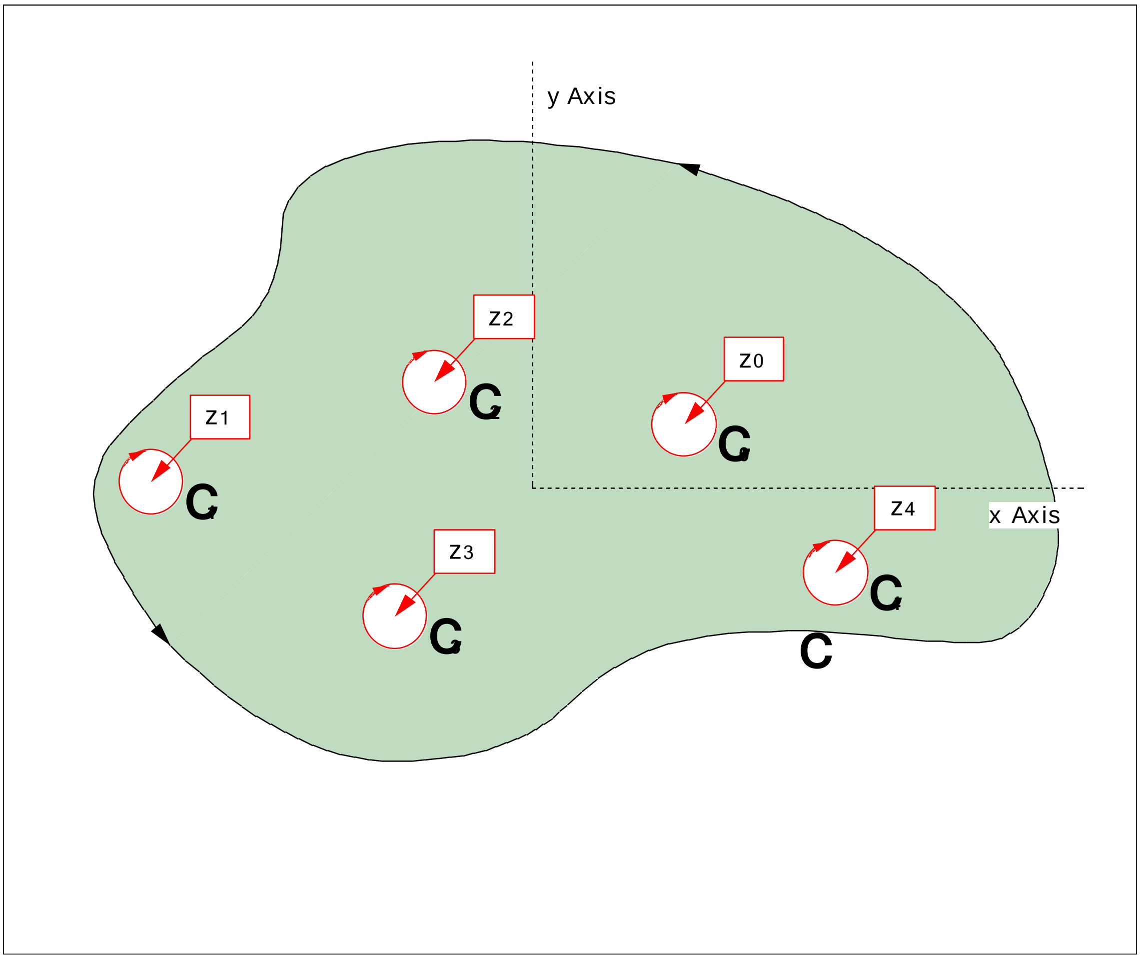


*Figure 7.2.2*

As an example, suppose we have $\varphi(z) = (4 - 3z)/(z^2 - z)$. This can be expanded by **partial fractions** as introduced in Section 1.9:

$$(4 - 3z)/z(z - 1) = A/z + B/(z - 1) = (A(z - 1) + Bz)/z(z - 1).$$

Equating like powers of $z$, $A + B = -3$, $-A = 4$, so $A = -4$, $B = 1$, and

$$(4 - 3z)/(z^2 - z) = -4/z + 1/(z - 1).$$

**Partial fractions**, by the way, are much used in evaluating **integrals** (in the sense of finding **antiderivatives**).

Now $\varphi(z)$ appears as the sum of two **functions** of the type in **Cauchy's integral formula** with **singularities** − points where we get a division by zero − at $z = 0$ and at $z = 1$, and so a **line integral** in $\varphi(z)$ can be expanded as:

$$\varphi(z)\mathbf{d}z \mid \mathbf{C} = [-4/(z-0)]\mathbf{d}z \mid \mathbf{C} + [1/(z-1)]\mathbf{d}z \mid \mathbf{C}$$

$$= [\nu_0(z_0)/(z-0)]\mathbf{d}z \mid \mathbf{C} + [\nu_1(z_1)/(z-1)]\mathbf{d}z \mid \mathbf{C}.$$

Here I write 4 as $\nu_0(z_0)$ with $z_0 = 0$, and similarly 1 as $\nu_1(z_1)$ with $z_1 = 1$ to emphasize the analogy with the **Cauchy formula**. So if **C** encloses only the **origin** ($z = 0$), the **integral** will be $2\pi j.\nu_0(z_0) = 2\pi j.(-4) = -8\pi j$; if it encloses only the point $z = 1$, the **integral** will be $2\pi j.\nu_1(z_1) = 2\pi j.(1) = 2\pi j$; if it encloses *both* it will be $\pi j(2 - 8) = -6\pi j$.

In practice, **integration by residues** is a bit more subtle than this, but this gives the general idea.

Before leaving this Section, I'll mention an interesting historical note, which highlights the influence of fashion even in advanced mathematical work.

Because **complex numbers** are *numbers* and constitute an extension of the normal number system of **real numbers**, it occurred to the hugely influential German mathematician David Hilbert that one could set up **vector spaces** with **complex numbers** acting as the **scalars** of the space. So, in a way, we'd have **n-vector spaces** with **2-vector scalars**! Such spaces are called **Hilbert spaces**, and they are widely used. In particular, because Hilbert made this suggestion at just about the time that **quantum mechanics** was starting to be developed, they came to be used as a foundation for the mathematics of **quantum** physics. This led to the notion that there was something *unavoidably* "complex numberish" about **quantum** theory. It could be argued that this notion has recently been challenged by the Geometric Algebra school, as GA shows that one of the central quantum algebras, the Pauli algebra, described in Section 5.9, appears naturally in 3-space when the **geometric product** is used. As it happens, this was apparently known to Pauli himself, who knew of Clifford's work.

# Chapter 9

# Projective Geometry

## 9.1 Introduction

**Projective Geometry** has its origins in the theory of **perspective**. To put this into context, I'll give a brief résumé of general ideas involved in the **projection** of a three-dimensional scene or object onto a two-dimensional surface.

In architecture and engineering, a rather limited set of geometrical projections are used, which fall into two main classes: **parallel** and **perspective**. This may seem to be in marked contrast to geographers, who select from a very large range of projections to accommodate the problem of mapping a spherical earth onto plane paper. But there's a big difference: **projections** in engineering are mapping a three-dimensional object onto a plane two-dimensional surface; the **projections** used in cartography are mapping a curved two-dimensional surface onto a plane two-dimensional surface.

As there's a certain amount of confusion about terminology here, I will follow the terminology used in F. D. K. Ching's *Architectural Graphics*, a standard U.S. textbook, which gives the terms familiar to most American architects and which are those used in the almost universally American software employed in CAD work.

The **projections** are most easily distinguished by the use of two auxiliary concepts. The first is that of **projector lines** or **projectors**, which are simply straight lines so defined that precisely one alone passes through each point of the three-dimensional object being depicted. The second is that of the plane on which the two-dimensional image will appear, which is called the **picture plane**. Exactly how the **projectors** are defined defines the **projection**: for where the unique **projector** through any point in space P with **coordinates** $(x, y, z)$ passes through the **picture plane** defines the **image** of P. Prof. Ching distinguishes three main classes:

- **Orthographic Projection**

  The **projectors** are parallel to each other and perpendicular to the **picture plane**.

- **Oblique Projection**

  The **projectors** are parallel to each other and at an oblique angle (i.e. an angle other than a right angle) to the **picture plane**.

- **Perspective Projection**

  The **projectors** all pass through a single point in space SP unique to the **projection** that represents a single eye of the observer. This unique point is called the **eye point** or **station point**. I will use both terms interchangeably.

The first two are both **parallel** projections, which is why I refer to only two main classes, **parallel** and **perspective**. Of these projections, **perspective** is much the most interesting, both mathematically and artistically, and its understanding by Brunelleschi in 1425 was one of the turning points of the modern era. There's a widespread belief today that perspective is some sort of Western convention, but it isn't: it's a true scientific discovery, because in perspective the **image** in the **picture plane** of any point in space lies in exactly the direction, as seen by the observer's eye, in which the point itself does. So a **perspective image** is a true realization of what the observer sees.

**Parallel projections** are sometimes referred to simply as "**paraline**" projections in the USA, but this is an informal term. **Orthographic projections** are sometimes mistakenly called **orthogonal**, but because this term has established usage in mathematics it should be avoided.

The three **projections** can be defined very easily algebraically. To do this, we need to choose **coordinates** specific to each of the three **projections,** and if we also define **picture plane coordinates** $\xi$ and $\psi$ measuring distances in the $x$ and $y$ directions *within the* **picture plane**, we can look at the three main **projection** types in detail, using diagrams showing just the $xz$-plane. The logic for the $yz$-plane is similar in each case.

The **picture plane**, which we assume lies at right angles to the $xz$-plane, and indeed, except for the **oblique** case, actually is parallel to the $xy$-plane, now appears as a line, since it's seen edge on. That line is marked PP in the three diagrams which follow, and for clarity, I've chosen to put each on a separate page.

• For **Orthographic Projection**

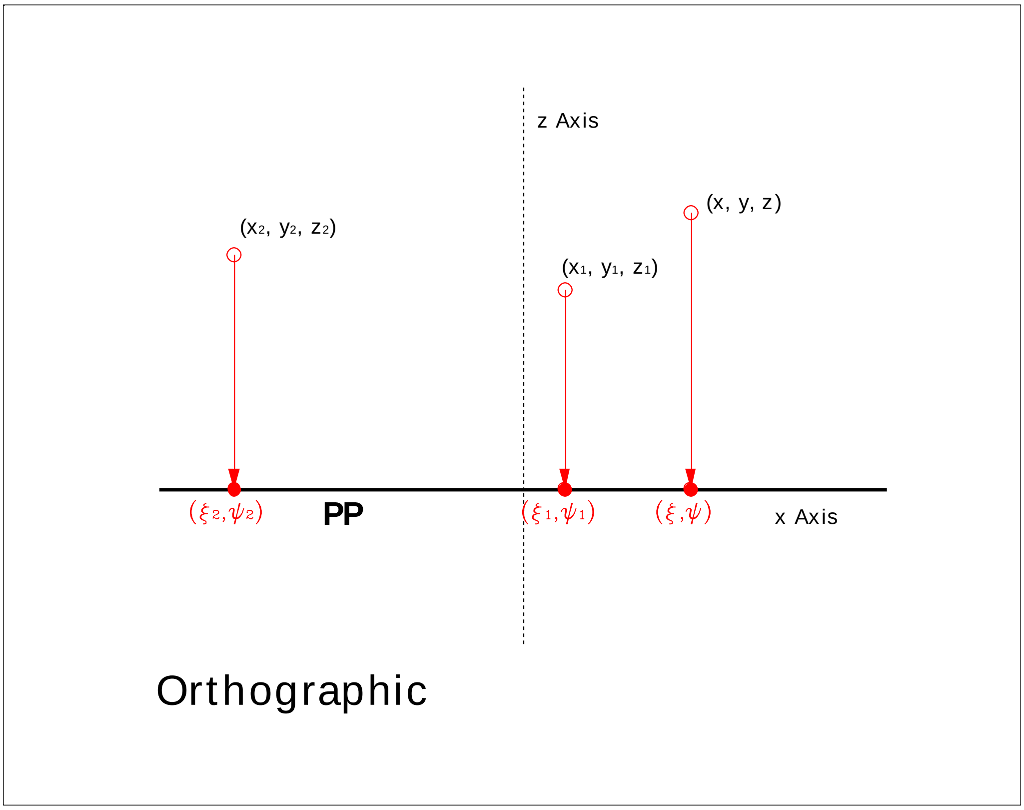


*Figure 9.1.1*

Choose **coordinates** such that the **projectors** are parallel to the *z* axis and the *x* and *y* axes are parallel to the **picture plane**. Then the **image** of (*x*, *y*, *z*) in the **picture plane** is at the point (ξ, ψ) = (*x*, *y*), independent of the value of *z*. In other words, all points with the same value of *z* map into the same **image point**.

• For **Oblique Projection**

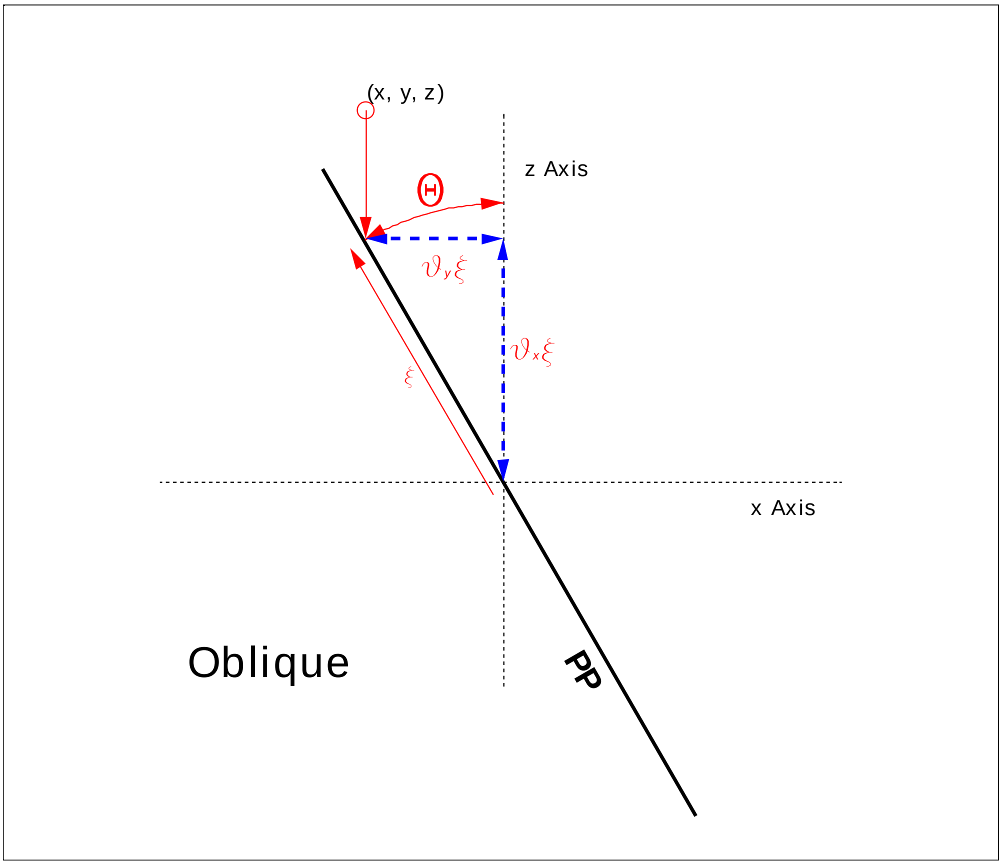


*Figure 9.1.2*

Choose **coordinates** such that the **projectors** are parallel to the $z$ axis and at an angle $\boldsymbol{\theta} = (\theta_x, \theta_z)$ to the **picture plane** in the $xz$-plane, but at right angles to the **picture plane** in the $yz$-plane. In other words, rotate the $x$ and $y$ axes about the $z$ until the obliquity falls in the $xz$-plane. Then the **image** of $(x, y, z)$ in the **picture plane** will fall at the point $(\xi, \psi) = (x/\theta_y, y)$ as can be seen in *Figure 9.1.2*. In this diagram, $\theta_x$, the **cosine** part of $\boldsymbol{\theta}$, runs along the $z$ axis, and $\theta_y$, the **sine** part, runs parallel to the $x$ axis. This is because the angle here is being measured anticlockwise from the $z$ axis in this diagram. Note that if $\boldsymbol{\theta} = \mathbf{j} \equiv 0 + j.1$, so the angle between the **picture plane** PP and the $z$ axis is 90°, then $(\xi, \psi) = (x/1, y) = (x, y)$ and the projection reduces to the ordinary **orthographic projection**.

- For **Perspective Projection**

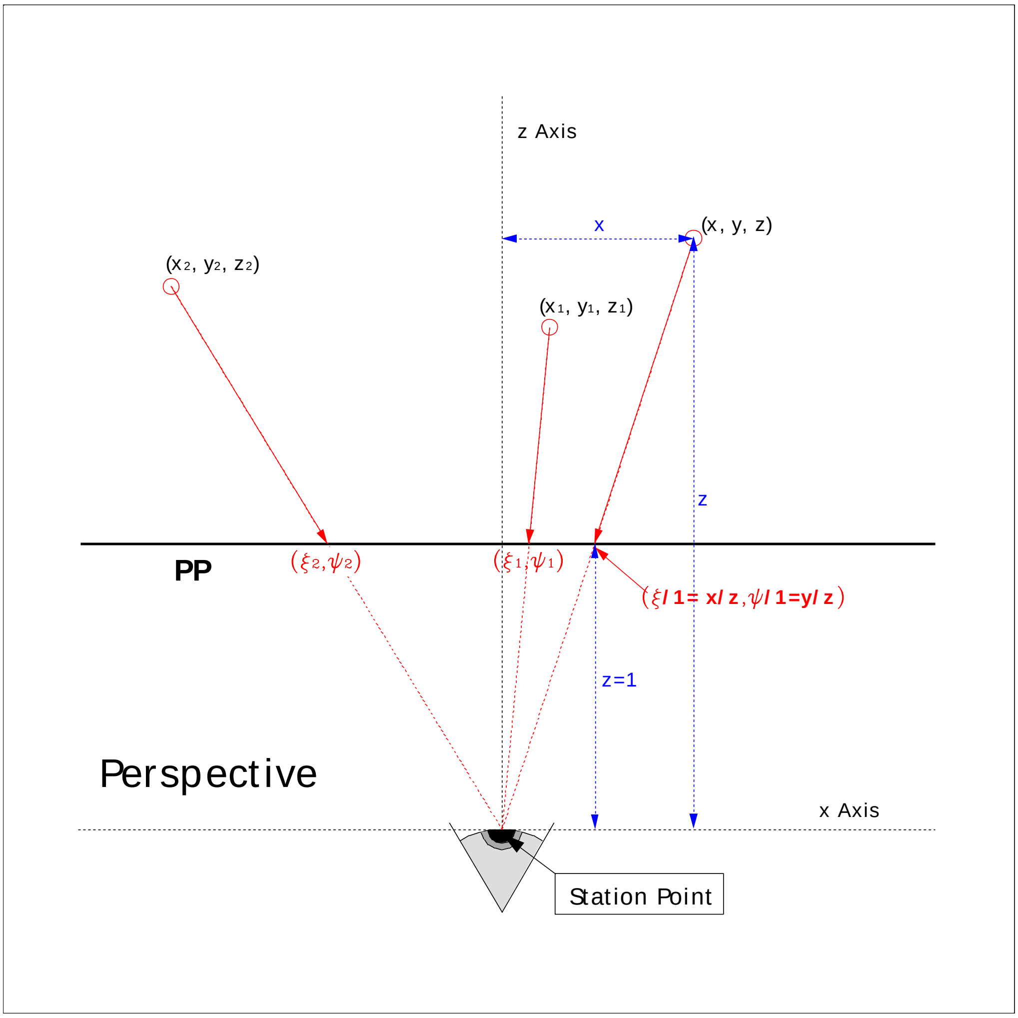


*Figure 9.1.3*

Choose **coordinates** so that the *z* axis is at right angles to the **picture plane**, and so the *x* and *y* axes are parallel to the **picture plane**. Let the unique **eye point** or **station point** be at the **origin** (0, 0, 0). Now scale the **coordinates** so that the **picture plane** coincides with the plane $z = 1$. Then, by similar triangles, as in *Figure 9.1.3*, we can see that the **image** of $(x, y, z)$ in the **picture plane** will be at the point $(\xi, \psi) = (x/z, y/z)$. This division by the third **coordinate** will be hugely important in what follows.

**Perspective** is unique in that all points in a given *direction* have **images** closer and closer to a specific **limiting point** in the **picture plane** as the points are selected further and further from the **eye point**. This **limiting point** is unique to that direction. We can see this easily enough.

Take any point in the $z = 0$ plane $(x_0, y_0, 0)$, and take any direction defined as $(\theta_x, \theta_y, 1)$. Then all points along the line through $(x_0, y_0, 0)$ running in the direction $(\theta_x, \theta_y, 1)$ will obey the equation:

$$P = (x, y, z) = (x_0, y_0, 0) + t.(\theta_x, \theta_y, 1),$$

where we're assuming ordinary **vector algebra**. Then as $t \to \infty$, the **image** of P is:

$$p = (x/z, y/z) = ((x_0 + t\theta_x)/t, (y_0 + t\theta_y)/t) = (x_0/t + \theta_x, y_0/t + \theta_y) \to (\theta_x, \theta_y).$$

So the **limiting point** of the **images** is *independent of* $(x_0, y_0)$. *All* points lying far enough away in the direction defined by $(\theta_x, \theta_y)$ have their **images** in the neighbourhood of *the same* **image point** given by precisely that direction itself $(\theta_x, \theta_y)$.

This unique **image point** for any given direction from the **eye point** is called the **vanishing point** for that direction. By taking $(x_0, y_0) = (0, 0)$, or the **eye point** itself (we had assumed $z = 0$ here) we can see that this **image point** is precisely the point where the unique line through the **eye point** in the direction $(\theta_x, \theta_y, 1)$ intersects the **picture plane**.

In this I've assumed that "$\theta_z$" = 1 always to avoid a notorious case. If we take "$\theta_z$" = 0, so we run in a direction $(\theta_x, \theta_y, 0)$ from the **eye point**, our line *never* intersects the **picture plane** but simply runs off to infinity *parallel* to the **picture plane**. So *directions parallel to the* **picture plane** do not have **vanishing points** *lying in the* **picture plane**. This anomaly was the entire reason for the introduction of **projective geometry**.

Algebraically we now have:

$$P = (x, y, z) = (x_0, y_0, 0) + t.(\theta_x, \theta_y, 0),$$

so as $t \to \infty$, the **image point** would be undefined as the $z$ **coordinate** of P is consistently zero:

$$p = (x/z, y/z) = ((x_0 + t\theta_x)/0, (y_0 + t\theta_y)/0) \to (\infty, \infty).$$

So, in the sense that $x/0 = \infty$, *any* point in a direction parallel to the **picture plane**, *however near to the* **eye point**, has its **image** at infinity.

This issue is so important that I will devote the next Section to the consideration of **vanishing points**.