an estimation problem:

I have a die, biased. I do not know the chances of the faces appearing in a throw. I roll the die n times and observe that face i appeared n_i times for $1 \leq i \leq 6$. of course, $\sum n_i = n$. let us assume that each $n_i > 0$.

Question: Looking at the data how can I decide the chances for the faces. In other words, let $p_i > 0$ be the chances of face i for $1 \leq i \leq 6$. How do I estimate these numbers?

You say that intuitively n_i/n should be an estimate for p_i. Yes, right, that is what intuition suggests. But does any principle also suggest that. If so, I can use that principle when my intuition fails. Also probably I can develop a theory about such a principle.

The answer is yes, there is a principle which will give you the same answer that your intuition suggested. To begin, remember that in probability theory, you build models; assume some probabilities for outcomes and see the consequences, calculate other interesting probabilities and so on. When you build such models, you have the feeling that an outcome with higher probability is more likely to be seen than an outcome with smaller probability.

So when it comes to estimation, we turn things around and say that since I have seen something, it must have larger probability than others. Of course, you do not reduce it to ridiculous level and say that what I have seen has probability one. Then the numbers (p_i) disappear and there is only one outcome, namely, $(n_i; 1 \leq i \leq 6)$ which has probability one. Our premise is that each $p_i > 0$, that is, the chances for each face are strictly positive. And of course, this entails that the experiment is *not* deterministic. The data could be any thing consistent with multinomial.

In other words we propose the following principle: Our estimate of the parameters (p_i) are those numbers that maximize the probability of our observation. The multinomial probabilities tell you that in case (p_i) are the

unknown probabilities, then chances of coming up with our observation are

$$f(p_1, \cdots, p_6) = \frac{n!}{n_1!n_2!n_3!n_4!n_5!n_6!}\ p_1^{n_1}\ p_2^{n_2}\ p_3^{n_3}\ p_4^{n_4}\ p_5^{n_5}\ p_6^{n_6}. \quad (*)$$

Thus our estimate are those numbers that maximise the above function of the (p_i). Of course, there is a constraint, namely we must have

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1.$$

Thus we are in Lagrange's set up. Since n_i are given numbers, maximizing $(*)$ is same as maximizing the second term there, forgetting the factorials. But again maximizing this is same as maximizing its logarithm. Note that we are assuming that each $p_i > 0$. So logarithm makes sense. Thus the problem reduces to maximizing

$$L(p_1, \cdots, p_n) = \sum n_i \log p_i$$

subject to the constraint $\sum p_i = 1$. Lagrange tells that at an extremum point we must have

$$\frac{n_i}{p_i} - \lambda = 0 \quad 1 \leq i \leq 6.$$

We have used $-\lambda$ for the Lagrange constant instead of $+\lambda$. In other words $p_i = n_i/\lambda$ and now $\sum p_i = 1$ tells us $\lambda = n$ and finally $p_i = n_i/n$.

Of course you need to verify if it is really a point of extremum at all, if so if it is maximum or minimum. In this case you can be sure it is maximum. It can not be minimum and there is a maximum. We shall not detail the argument. it is not difficult. There is no minimum because at the boundary, that is, when one of the $p_i = 0$ this function assumes the value zero. For the same reason combined with compactness, it has a maximum.

Hadamard inequality:

Consider an $n \times n$ matrix (real entries), say, A. Let us denote its rows by $r_1, r_2, \cdots, r_n$. Then Hadamard's inequality says

$$|\det(A)| \leq ||r_1||\ ||r_2||\ \cdots\ ||r_n||.$$

here norm of a vector is the usual Euclidean norm: $\sqrt{\sum v_i^2}$.

If one of the norms is zero, there is nothing to do, that row is zero and done. Let us assume that all are strictly positive. thus the problem reduces

to the following. We are given n strictly positive numbers $d_1, \cdots, d_n$. Let us, as above, denote rows of matrix by r_i.

$$\text{maximize} \quad |A|; \quad \text{subject to} \quad ||r_i|| = d_i; \quad 1 \leq i \leq n.$$

We are in Lagrange set up. We have a function of n^2 variables, instead of denoting the variables with usual linear index, name the variables (x_{ij}). Define

$$f(\{x_{ij}\}) = \det((x_{ij})).$$

name the n constraints

$$\varphi_i(\{x_{ij}\}) = d_i^2 \qquad 1 \leq i \leq n.$$

where

$$\varphi_i(\{x_{ij}\}) = \sum_j x_{ij}^2$$

See that the each constraint depends on variables only in one row.

Thus Lagrange tells us that at an extremum

$$X_{ij} - \lambda_i 2x_{ij} = 0 \qquad 1 \leq i, j \leq n.$$

We denoted by X_{ij} the cofactor of x_{ij}; which is $(-1)^{i+j}$ times the determinant obtained by deleting the row and column containing the entry x_{ij} (thus delete i-th row and j-th column). If you expand the determinant with the help of i-th row, then it is easy to calculate the derivative of f w.r.t. the variables in that row.

Let us recall two facts about the cofactors.

$$\sum_i x_{ij} X_{ij} = \det(A).$$

$$\sum_p x_{pj} X_{ij} = 0 \qquad p \neq i.$$

Combined with the above Lagrange equations we get

$$\det(A) = \lambda_i 2d_i^2; \qquad 1 \leq i \leq n$$

$$\lambda_i \sum_j x_{pj} x_{ij} = 0; \;\; p \neq j.$$

Note that at a maximum $\lambda_i \neq 0$ for all i. This is because, If some $\lambda_i = 0$ then the first equation above tells us that $\det(A) = 0$ but the diagonal matrix

$(d_1, d_2, \cdots, d_n)$ tells us that this can not be maximum.

Thus at a point of maximum, we must have the rows of A must be orthogonal. In other words at a maximum the matrix A is orthogonal matrix. But then for such a matrix A where the maximum is attained we must have

$$|\det(A)|^2 = \det(AA^t) = \text{diag}\{d_1^2, d_2^2, \cdots, d_n^2\}.$$

In other words, at a maximum we have the Hadamard inequality. This proves the inequality.

See how we did not solve the equations and discuss which is max and which is min. We found the equations to be satisfied at an extremum. Then we discarded some and found the equations that should be satisfied at the maximum. This told us that rows must be orthogonal. This was enough to conclude the required inequality. But actually we have found all solutions where the equality is achieved in the Hadamard inequality.

recapitulation of the development:

(1). Points in R are x. Points in R^2 are (x, y); points in R^3 are (x, y, z).

If $f : R \to R$ then $f(x)$ is value of f at x. If $f : R^2 \to R$ then $f(x, y)$ is value of f at (x, y). Similarly we have for $f : R^3 \to R$

If $f : R \to R$ then its graph is a subset of R^2; consists of all points (x, y) such that $f(x) = y$. If $f : R^2 \to R$ then its graph is a subset of R^3; consists of all points (x, y, z) such that $f(x, y) = z$. Similarly for $f : R^3 \to R$, its graph is a subset of R^4.

(2). Derivative at a point x is denoted by $f'(x)$ in case of R.

In case of R^2 it is denoted at a point (x, y) by $f'(x, y)$ or $\nabla f(x, y)$ or $(f_1(x, y), f_2(x, y))$ or sometimes by $(f_x(x, y), f_y(x, y))$. But this last notation is confusing with x appearing as suffix and also argument. Also it has the disadvantage of leading to meaningless things $f_3(3, 4)$ (substitute blindly $x = 3, y = 4$). However if we have a fixed point (a, b) then we could, without fear of confusion, denote $f_x(a, b)$. Similar notation in dimension three holds.

(3) In one dimension we write

$$dy = f'(x)dx; \qquad df = f'(x)dx.$$

In two dimensions

$$dz = f_x dx + f_y dy; \quad df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

Similar notation holds in three dimensions

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

all these notations have their origins in the mean value theorems. df denotes the change in the value of the function for a small change in the value of the arguments. For example in one dimension if x is changed to $x + \epsilon$ then $f(x)$ is changed to $f(x + \epsilon)$ and thus the change in its value is

$$df = f(x + \epsilon) - f(x) = f'(?)\epsilon.$$

where $(?)$ is a point in between, $x + \epsilon$ and x assuming that f is C^1 and noting that ϵ is very small this number $f'(?)$ is approximately $f'(x)$. if you now read the equation above with ϵ replaced by dx you get the interpretation given above.

similarly for functions of two variables

$$f(x + \epsilon_1, y + \epsilon_2) - f(x, y)$$

$$= \{f(x + \epsilon_1, y + \epsilon_2) - f(x, y + \epsilon_2)\} + \{f(x, y + \epsilon_2) - f(x, y)\}$$

$$= f_1(?, y + \epsilon_2)\ \epsilon_1 + f_2(x, ??)\ \epsilon_2.$$

Again assuming C^1 we can argue as above.

(4). The main and very important change of attitude that put all these definitions on one single platform is to think of these derivatives as linear maps. Think of elements in R^2 as columns and these derivatives are rows. the row operates from left on a column to give a number. This also helped later to give a unified method of thinking of derivatives when we have functions from an R^m to an R^n. Derivative is essentially 'linearizing' a given function at a point.

(5). If you have a point (x_0, y_0) on the graph of a function $f : R \to R$ then tangent at that point to the curve is the affine map on the real line

$$Ax = y_0 + f'(x_0)(x - x_0)$$

Customarily this is written as

$$y - y_0 = f'(x_0)(x - x_0).$$

You think of, not only the analytical expression, but also the graph of this map, namely, straight line.

Of course this line is nothing but translate of the line (subspace) through the origin $y = f'(x_0)x$ by the point (x_0, y_0).

Similarly for function of two variables $f : R^2 \to R$ and a point (x_0, y_0, z_0) on its graph (it is now called surface instead of curve) we define tangent plane as the set of points satisfying

$$z - z_0 = f_1(x_0, y_0)(x - x_0) + f_2(x_0, y_0)(y - y_0)$$

or

$$z - z_0 = \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0).$$

again, this plane is nothing but the translate of the subspace

$$z = \nabla f(x_0, y_0) \cdot (x, y)$$

by the point (x_0, y_0, z_0).

(6). Chain rule is nothing but composition of linear maps.

In one dimensions if we have a map

$$t \mapsto x(t) \mapsto f(x(t))$$

then we have

$$\frac{d}{dt} f(x(t)) = f'(x(t))x'(t); \quad \text{or} \quad \frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

In two dimensions

$$t \mapsto (x(t), y(t)) \mapsto f(x(t), y(t))$$

then

$$\frac{d}{dt} f(x(t), y(t)) = f'(x(t), y(t)) \cdot \langle x'(t), y'(t) \rangle.$$

Remember f' is a row vector and the other is column vector. Sometimes this is written as

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

You see the usage of d and ∂. When you use d you are indicating you have a function of one variable and you are differentiating w.r.t. that one variable.

When you use ∂ you are indicating you have a function of more than one variable and you are differentiating w.r.t. the variable indicated.
Similarly if we have $f(x(t), y(t), z(t))$,

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}.$$

Or if you denote

$$\varphi(t) = (x(t), y(t), z(t)) : R \to R^3$$

and

$$g(x, y, z) : R^3 \to R$$

and

$$f(t) = g(\varphi(t))$$

then

$$f' = g'(\varphi(t))\varphi'(t)$$

When you think of the derivative as linear map, the distinction between column and row disappears. You do not need to put dot (scalar product) etc. It is all composition of linear maps. But of course, you should know what linear map you are talking about and its domain/range.

(7) Level sets for a function are points of constancy. Thus for $f : R \to R$ its level set corresponding to the number a is the set $\{x : f(x) = a\}$. Of course this may be empty. Thus remember level set is not one set. It depends on the number a.

For $f : R^2 \to R$ and a number $a \in R$ the level set of f corresponding to the value a is the set $\{(x, y) : f(x, y) = a\}$. Thus if f is the temperature defined on a geographical region $\Omega \subset R^2$ then level sets are nothing but isothermal lines. This will give an idea of the function.

what next:

We have developed the basic concepts and ideas to understand functions of several variables. There are several directions one can take.

We could develop improper integrals completely. We have started once and I gave up after explaining some ideas. I felt that it was not going well.

You asked about the fundamental theorem of calculus. Yes, this is one topic that could be developed. yes, there is an analogue of fundamental theorem. In one dimensions it said the following. If $F' = f$ then

$$\int_{[a,b]} f = F(b) - F(a).$$

Do not think of this theorem as explaining full evaluation of the integral. A space of dimension less than one (whatever it may mean) is a set of two points $\{a < b\}$, the boundary of $[a, b]$ If we interpret the right side above as integral of F over this set, then the fundamental theorem is giving you a reduction of the complicated integral to a lower dimension. To see the beauty, let us denote derivative by not prime, but ∂. Thus instead of $F' = f$ we write $\partial F = f$. Thus the above equality can be written as follows. Let us abbreviate the interval to I

$$\int_I \partial F = \int_{\partial I} F.$$

Did you see how beautifully the ∂ travelled from top to bottom. Left side is integral of a probably simple function on a complicated set, interval. The right side is integral of a probably complicated function but on a simpler set.

If you stare at the above equation it makes perfect sense even in R^2. For example you could take a rectangle for I. You know its boundary. You could take a nice function f on the rectangle. Explain what could be the meaning of ∂F to make the equation above correct and discover that F. In other words, the idea is to make a good definition so that the beauty seen above is preserved. This can be done. This you will learn in the next course on vector calculus.

One can go to develop some concepts of geometry of curves and surfaces. We may not be able to do much, but we use integration and develop notions of lengths of curves. We shall do that.

We can also develop some vector calculus terminology that is useful in physics. We shall do that. One can develop some physics that explains these concepts, but we may not do that.

There is a profound symphony, complex analysis, which you will learn later. We can make a beginning and tie it up with what we learnt so far. We shall do that.

lengths of curves:

A curve is a continuous function f defined on a closed bounded interval $[a, b] \subset R$. Sometimes we have curves defined on an open interval too. But we do not need this generality. If the function takes values in R^2 then it is a curve in plane; if it takes values in R^3 then it is a curve in space. We do not discuss other curves.

The first question is whether the curve is the function or the set of points consisting the image. Thus, for example if you consider the curve

$$f(t) = (\cos t, \sin t); \quad 0 \leq t \leq 2\pi.$$

is the curve this function or is it the circle that you see in the plane described this set of points? Actually, it is both. Suppose you consider the function

$$g(t) = (\cos t, \sin t); \quad 0 \leq t \leq 4\pi.$$

Then this also gives the same circle, but the curve goes along the circle twice. So thinking of the set of points alone does not completely help. Or consider the function

$$h(t) = (\cos t/2, \sin t/2); \quad 0 \leq t \leq 4\pi.$$

This is also the same circle, also goes around exactly once but rather slowly. so you should keep both in mind; the function and the geometric picture. But in case of a dispute, it is the definition that wins, the curve is the function.

How shall we define length of a curve? To make a beginning, let us recall how we defined area under a graph. We all agreed upon, without any dispute, the area of a rectangle, namely, it is the product of lengths of its sides. So we approximated the area by rectangles. We shall do the same thing for length. after all, we have agreement on length of a straight line. The line joining two points (x_1, y_1) and (x_2, y_2) in the plane has length

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Given a curve $\varphi : [a, b] \to R^2$, we take points

$$t_0 = a < t_1 < t_2 < \cdots < t_k = b.$$

Consider the points on the curve

$$\varphi(t_0) = P_0; \varphi(t_1) = P_1; \varphi(t_2) = P_2; \cdots, \varphi(t_k) = P_k.$$

Length of the curve is at least as much as the sum

$$\sum_{i=0}^{k-1} ||P_{i+1} - P_i||.$$

We agree to this because if you have two points, the smallest distance between them is the length of the line joining them. Any other curve joining them must have a larger length.

Given the curve φ, and a partition π of the interval, we define

$$L(\varphi, \pi) = \sum ||\varphi(t_{i+1}) - \varphi(t_i)||$$

Just now we agreed that length of the curve is at least as much as the above. We define

$$L(\varphi) = \sup\{L(\varphi, \pi) : \pi \text{ partition of } [a, b]\}.$$

Thus length of the curve must be at least the above quantity. As the partition becomes finer, the lines joining the partition points on the curve move closer and closer to the curve; and finally trace the curve (of course, there is nothing like 'finally'). Thus this sup is length of the curve. However this sup may be ∞. We say that the curve has length if this is finite and then we define length of the curve to be this number.

Technical name for curves which have length is 'rectifiable'. Of course even if the above quantity is infinity, we could have said that the curve has length and its length is infinity. it makes no difference, but we shall not do it.

For example consider the function

$$f(t) = t \sin\left\{\frac{\pi}{t}\right\}; \quad 0 < t \leq 1; \quad f(0) = 0;$$

and the curve is its graph, that is,

$$\varphi(t) = (t, f(t)); \quad 0 \leq t \leq 1.$$

This curve is not rectifiable. That is, the sup will be infinity. Observe that this is a continuous function. This is easy to see. if you take the partition

$$\{0, 1/100, 1/99, \cdots, 1/2, 1\}.$$

Note that at successive points of the partition the points on the curve can be easily calculated and you get partial sum of the series $\sum(1/n)$.

We shall show that for a C^1 curve, length exists (and so the curve is rectifiable). We shall also give a formula for its length. Let us denote the curve by

$$\varphi(t) = (x(t), y(t)); \quad a \leq t \leq b.$$

The curve is C^1 means that these functions x and y are C^1 functions. If you took a partition π and calculate $L(\varphi, \pi)$ then mean value theorem tells you

$$L(\varphi, \pi) = \sum \sqrt{[x'(\xi_i)]^2 + [y'(\eta_i)]^2} \ \ (t_{i=1} - t_i).$$

Here (ξ_i) and (η_i) are selections for the partition. This should remind you of

$$\int_a^b ||\varphi'(t)|| \ \ dt.$$

We shall show that this is actually true. Where is the problem? We have two selections, if there is only one selection (ξ_i) for both x' and y' then the above sum is exactly the Riemann sum for the later integral.

Riemann sums revisited:

We need a fact. suppose we have continuous functions as follows:

$$x : [a, b] \to R; \quad y : [a, b] \to R$$

$$g : R^2 \to R$$

We define

$$f(t) = g(x(t), y(t)) : [a, b] \to R$$

From last semester we know the following. Suppose we have a sequence of partitions π_n of $[a, b]$ with $||\pi_n|| \to 0$ and for each n, a selection ξ_n for the partition π_n. Then

$$R(f, \pi_n, \xi_n) = \sum g(x(\xi_i), y(\xi_i))(t_{i+1} - t_i) \to \int_a^b f(t)dt.$$

Of course, here the sum is over the partition points of π_n and also what we denoted by ξ_i is actually $\xi_n(i)$, from the selection ξ_n. We did not burden the notation. But if you have confusion, you should write for yourself completely and clearly.

What we shall now claim is the following. Suppose that we have the above situation and two selections ξ_n and η_n for each n. Then

$$R(f, \pi_n, \xi_n, \eta_n) = \sum g(x(\xi_i), y(\eta_i))(t_{i+1} - t_i) \to \int_a^b f(t)dt.$$

Integral is robust, it will accommodate you if you deviate little bit. after all, since the norms of the partitions are getting smaller, if you take two selections, then they must be close and it should make no difference. Here is the proof.

We only need to show

$$\sum \{g(x(\xi_i), y(\eta_i)) - g(x(\xi_i), y(\xi_i))\} (t_{i+1} - t_i) \to 0. \quad (*)$$

Please note that the ξ and η appearing above are the n-th selections ξ_n and η_n.

Let $\epsilon > 0$ be given. Choose δ_1 such that

$$|a_1 - a_2| < \delta_1; |b_1 - b_2| < \delta_1 \Rightarrow |g(a_1, b_1) - g(a_2, b_2)| < \epsilon/(b - a).$$

choose $\delta > 0$ so that

$$|s - t| < \delta \Rightarrow |y(s) - y(t)| < \delta_1$$

Both are possible by uniform continuity of the functions involved. Now

$$||\pi_n|| < \delta \Rightarrow | \{g(x(\xi_i), y(\eta_i)) - g(x(\xi_i), y(\xi_i))\} | < \epsilon/(b - a)$$

so that the sum in $(*)$ is at most ϵ.

This completes the proof of the claim.

return to length:

the result proved just now shows

$$||\pi_n|| \to 0 \quad \Rightarrow \quad L(\varphi, \pi_n) \to \int_a^b ||\varphi'(t)||dt.$$

Thus every C^1 curve has length and is given by

$$L = \int_a^b ||\varphi'(t)||dt.$$

You might be wondering what happened to the sup we took of the $L(\varphi, \pi)$. There are several ways to see this. Here is a way.

Take a sequence of partitions π_n such that $L(\varphi, \pi_n)$ converges to the sup. Remember sup of any set is actually limit of a carefully chosen sequence of points from the set. Note that if π_1 has one extra point than π then triangle inequality immediately gives you $L(\varphi, \pi) \leq L(\varphi, \pi_1)$. Thus by considering the partition

$$\Pi_n = \pi_1 \vee \pi_2 \vee \cdots \vee \pi_n$$

you see that

$$||\Pi_n|| \quad \downarrow; \quad L(\varphi, \Pi_n) \to L.$$

If $||\Pi_n|| \not\to 0$, add extra points to make it converge to zero. Thus you have a sequence of partitions Π_n with norm converging to zero and the corresponding chord length sums converge to L. But these chord lengths are precisely the Riemann sums converging to the integral. This completes the proof.

You can calculate length of circle and see that it coincides with what you felt.

You saw two examples of 'parametrizing' the circle, both go through exactly once but with different speeds. It is possible to bring in uniform code for curves so that the parametrisation traces the curve at uniform speed. This is how we make it precise.

Let $\varphi : [a, b] \to R^2$ be a C^1 curve with length L. Let us assume that φ' does not completely vanish during any interval. In other words, in any given interval, $||\varphi'(t)|| \neq 0$ for at least one t. For each $a \leq t \leq b$, let $L(t)$ be the length of the curve upto t. That is, it is length of the curve

$$\gamma(s) = \varphi(s); \quad a \leq s \leq t.$$

you will see that its length is nothing but

$$L(t) = \int_0^t ||\varphi'(t)||dt.$$

Thus L is a strictly increasing C^1 function on $[a, b]$ onto $[0, L]$. Let its inverse be denoted by L^{-1}. Define the curve

$$\Psi(s) = \varphi(L^{-1}(s)); \quad 0 \leq s \leq L$$

Then it is not difficult to see that Ψ describes the same curve, has the same length L. The interesting point is the following. At any time s if you ask: what is the length of the curve Ψ traced so far; the answer is s. This is called

parametrisation of the curve by arc length.

vector calculus:

Actually R^n is a vector space and what we have been doing is indeed calculus on (finite dimensional real) vector spaces. However, the word vector calculus is usually referred to certain notions that are found useful in physics. We have discussed some of these in a homework.

Physicists have a good nomenclature for functions to distinguish: real valued or vector valued. if f is real valued, they say it is a scalar function. If f takes values in R^2 or $R^3$and so on, they say it is a vector field. Of course, from a mathematical point real numbers are also vectors, one dimensional vectors. (And when you study vector spaces, you have a underlying field and R also plays the role of that field and you refer to elements of the field as scalars.) Usually small letters f, g etc.are used for scalar functions and capital letters F, G etc. are used for vector fields.

Let $f : R^3 \to R$ be a scalar function then ∇f is a vector field. for every point P it associates the vector $(f_x(P), f_y(P), f_z(P))$. If you think of ∇ as the symbolic operator

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).$$

Thus you can think of ∇ operating on f and giving ∇f.

suppose that $F : R^3 \to R^3$ is a vector field. Thus

$$F(P) = (F_1(P), F_2(P), F_3(P)).$$

Since now ∇ and F are vectors of size three it makes sense to talk about their scalar product (inner product) and their vector product. They are very important and are defined s follows.

$$\text{div}F(P) = \nabla \cdot F(P) = \frac{\partial F_1}{\partial x}(P) + \frac{\partial F_2}{\partial y}(P) + \frac{\partial F_3}{\partial z}(P)$$

This is called divergence of the vector field F. This is a scalar function.

$$\text{curl } F = \nabla \times F = (D_2F_3 - D_3F_2, D_3F_1 - D_1F_3, D_1F_2 - D_2F_1).$$

This is again a vector field. For every point P it associates a vector, namely the right side where the derivatives are evaluated at the point P.

There is another important operator

$$\Delta = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

This is called Laplacian. For a scalar C^2 function f, its Laplacian is Δf another scalar function.

There are several important relations among them. Here are some. Most of them are routine to verify.

(1) Curl(grad f) =0; that is, $\nabla \times \nabla f = 0$

(2) div (Curl F) =0; that is $\nabla \cdot (\nabla \times F) = 0$.

(3) $\text{div}(fF) = F \cdot \nabla f + f \cdot \text{div}F$

(4) $\text{Curl}(fF) = \nabla f \times F + f \text{ curl}(F)$.

(5) $\text{div}(F \times G) = G \cdot \text{curl}(F) - F \cdot \text{Curl}(G)$.

(6) $\Delta(F) = \nabla(\text{div}(F)) - \text{Curl}(\text{Curl}(F))$.
Here Laplacian for the field F is coordinate-wise.

complex derivative:

Recall that we have made R^2 into a field by defining addition as usual coordinate-wise. Multiplication is defined as follows:

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

to be clear whether we are talking about R^2 as a vector space or as a field with the above multiplication, we denote R^2 by C when we think of it as a field. We also write $z = x + iy$ instead of (x, y). Thus as a set they are same but the way we think about them depends on what we are talking about. For example when you are thinking of it as a vector space, it makes no sense to divide by a vector. When we are thinking of it as a field, division definitely makes sense.

For a complex number $z = x + iy$ we refer to x as the real part of z and y as the imaginary part of z. We also use $Re(z)$ and $Im(z)$ for the real and imaginary parts of z. Thus, remember the real and imaginary parts are real

numbers. Conjugate of a complex number $z = x + iy$ is the number $x - iy$ denoted by $\overline{z}$.

When the imaginary part is zero, we would not take the trouble of writing $x + i0$ but just write x. similarly, when the real part is zero we shall not write $0 + iy$, but simply as iy.

The real number system is a subset of C, identified as complex numbers with imaginary part zero. The multiplication and additions agree. By the way multiplication is defined we have $i \times i = -1$. Thus sometimes we also write $\sqrt{-1}$ or $+\sqrt{-1}$ for i.

Since you are familiar and also we discussed in an homework, we shall not spend time on routine matters.,

The definition of convergence is same as that of R^2. Thus a sequence of complex numbers z_n converge to a complex number z if the real parts converge to the real part of z and similarly the imaginary parts converge. Functions from C to C are thus essentially functions from R^2 to R^2. Continuity is just the same. Thus $f : C \to C$ is continuous if $f(z_n) \to f(z)$ whenever $z_n \to z$.

Let now $\Omega \subset C$ be an open set. We say that f is (complex) differentiable at a point $z_0 \in \Omega$ if there is a complex number α such that

$$z_n \to z_0; (\forall n) z_n \neq z_0 \quad \Rightarrow \quad \frac{f(z_n) - f(z_0)}{z_n - z_0} \to \alpha$$

or equivalently

$$h_n \in C; h_n \to 0, (\forall n) h_n \neq 0 \quad \Rightarrow \quad \frac{f(z + h_n) - f(z)}{h_n} \to \alpha.$$

The important question is the following. Set theoreticlly C is same as R^2. How does this derivative relate to what we have learnt? if you think of f as $f(x, y)$ (from R^2 to R^2) then does it have derivative? Is that all?

We shall now see that the real and imaginary parts of f do have partial derivatives and are related. Complex analysis has intrinsic beauty, if f is C^1 in a region then it is actually C^∞! You will have a course in complex analysis later when you will learn these. Right now our interest is only the relation to what we have learnt so far.

so let $f : C \to C$ be differentiable at a point $z_0 = x_0 + iy_0$. Let the derivative be $a + ib$. Let us denote

$$f(x + iy) = u(x + iy) + iv(x + iy).$$

Here u and v are the real and imaginary parts. Let us take a sequence of real numbers (h_n) all different from zero. Thus we have

$$\frac{u(x_0 + h_n, y_0) + iv(x_0 + h_n, y_0) - u(x_0, y_0) - iv(x_0, y_0)}{h_n} \to a + ib.$$

Since the real parts and imaginary parts converge, this implies u and v have partial derivatives w.r.t. x and

$$u_x(x_0, y_0) = a; \quad v_x(x_0, y_0) = b. \qquad (*)$$

since $ih_n \to 0$ whenever $h_n \to 0$ we also have

$$\frac{u(x_0, y_0 + h_n) + iv(x_0, y_0 + h_n) - u(x_0, y_0) - iv(x_0, y_0)}{ih_n} \to a + ib.$$

This means,

$$\frac{u(x_0, y_0 + h_n) + iv(x_0, y_0 + h_n) - u(x_0, y_0) - iv(x_0, y_0)}{h_n} \to ia - b.$$

This means that u and v have partial derivatives w.r.t. y and

$$v_y(x_0, y_0) = a; \qquad u_y(x_0.y_0) = -b. \quad (**)$$

comparing $(*)$ and $(**)$ we see

$$u_x = v_y; \quad u_y = -v_x \qquad (\spadesuit)$$

at the point (x_0, y_0). The equations above are called **Cauchy-Riemann equations**.

Thus any complex differentiable function satisfies the Cauchy Riemann equations; this means the real and imaginary parts of the complex function satisfy the above equations. Interestingly enough, if u and v are real valued C^1 functions on Ω and satisfy the above equations, then the function $f(z) = u(z) + iv(z)$ is a complex differentiable function.