Taylor:

The chain rule will now be applied to derive Taylor formula for function of several variables. This will be *exactly* same as the one you learnt last semester. There is absolutely no change. First let us recall the Taylor we know.

Let f be a function on an open interval I which is n times continuously differentiable. Let $a, b \in I$, Then

$$f(b) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(b-a)^k + \int_a^b f^{(n)}(t)\frac{(b-t)^{n-1}}{(n-1)!}dt.$$

If we have functions defined on R^2 and a and b are points in R^2, then the integral term above is a little tricky. so let us reformulate the above equation. We change the variable of integration to

$$t = ub + (1-u)a.$$

The beauty is that as t goes from a to b, the variable u goes from zero to one. It is beautiful because the range of integration no longer depends on the points a and b.

Note that $dt = (b-a)du$ and $(b-t) = (b-a)(1-u)$. Thus

$$f(b) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(b-a)^k + \int_0^1 f^{(n)}(ub+(1-u)a)(b-a)^n\frac{(1-u)^{n-1}}{(n-1)!}du.$$

Let us use convenient notation. We use D for derivative.

Thus $[D]f$ means the function f'. And $[D]f(a)$ means the earlier function evaluated at the point a, namely the number $f'(a)$.

$[4D]f$ means the function $4f'$. And $[4D]f(a)$ means the earlier function evaluated at the point a, namely the number $4f'(a)$.

$[(b-a)D]f$ means the function $(b-a)f'$. And $[(b-a)D]f(13)$ means the earlier function evaluated at the point 13, namely the number $(b-a)f'(13)$.

In particular, $[(b-a)D]f(a)$ means the number $(b-a)f'(a)$.

Symbols can be operated again and again. For example $[(b-a)D]^5 f$ means the function $(b-a)^5 D^5 f = (b-a)^5 f^{(5)}$. Recall $f^{(5)}$ is the fifth derivative of the function f. Thus $[(b-a)D]^5 f(13)$ means you need to evaluate this function at 13, thus you get the number $(b-a)^5 f^{(5)}(13)$.

In general if we say $[(b-a)D]^k f$ means it is the function $(b-a)^k f^{(k)}$; namely, The k-th derivative multiplied by the number $(b-a)^k$.
Thus $[(b-a)D]^k f(a)$ means the number $(b-a)^k f^{(k)}(a)$.

We can rewrite Taylor as follows:

$$f(b) \quad = \quad \sum_{k=0}^{n-1} \frac{[(b-a)D]^k f(a)}{k!} \quad +$$

$$\int_0^1 [(b-a)D]^n f(ub+(1-u)a) \frac{(1-u)^{n-1}}{(n-1)!} du. \quad (\spadesuit)$$

The Taylor formula we shall prove is the following.

Theorem: Let $\Omega \subset R^2$ be an open set. Let $a, b \in \Omega$. Assume that the line joining these points is contained in Ω. Let f be a real valued C^n function defined on Ω. Then

$$f(b) \quad = \quad \sum_{k=0}^{n-1} \frac{[(b-a)\cdot D]^k f(a)}{k!} \quad +$$

$$\int_0^1 [(b-a)\cdot D]^n f(ub+(1-u)a) \frac{(1-u)^{n-1}}{(n-1)!} du. \quad (\clubsuit)$$

You see that this formula is exactly same as the earlier one. We only need to explain the notation. you will see that proof of the theorem is itself not difficult. In fact it is trivial from what you already know in the one variable case and the chain rule.

A function f is C^1 if the partial derivatives f_1 and f_2 are continuous functions on Ω. such a statement means that the partial derivatives at each point exist and the function so obtained on Ω is continuous. We met this notation earlier before defining the concept of derivative.

We say that f is C^2 if these functions f_1 and f_2 are also C^1. In other words, $f_{11}, f_{12}, f_{21}, f_{22}$ are continuous functions. Remember that when this happens we already know that the functions f_{12} and f_{21} are same. Thus there are three second order derivatives.

In general, proceeding to define by induction, we say that f is C^n, if f_1 and f_2 are C^{n-1}. Let us use an earlier notation. $D_1 g$ stands for g_1 and $D_2 g$ stands for g_2, the partial derivatives.

With this notation, $f \in C^n$ is same as saying the following: whenever you take a sequence $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ consisting of ones and twos, the function

$$D_{\epsilon_n} D_{\epsilon_{n-1}} \cdots D_2 D_1 f$$

should exist and is continuous.

Using equality of mixed derivatives that we proved, it is a joy to show the following. When the above happens, the function so obtained depends not on the exact sequence, but only on the number of ones and twos in the sequence!

To continue with the notation, we use D for the vector (D_1, D_2). Of course, you should not get worried because, so far, vectors are something like $(4,3)$ and so on consisting of numbers. Here this vector we have introduced has symbols. Let it be. We shall do more symbolic operations.

Observe $(b-a) \in R^2$, is a vector. By $(b-a) \cdot D$ we mean inner product between these vectors. Again, you should not get worried. We only mean that $(b-a) \cdot D$ stands for the symbol

$$(b_1 - a_1)D_1 + (b_2 - a_2)D_2.$$

Of course, what is the meaning of this symbol? Just as the symbol D_1 applied to a function f gives you a new function $D_1 f = f_1$ so is this symbol.

$$[(b-a) \cdot D]f = [(b_1 - a_1)D_1 + (b_2 - a_2)D_2]f$$

$$= (b_1 - a_1)f_1 + (b_2 - a_2)f_2.$$

Thus if you want to calculate this function at a point, say, $(4,3)$ you get

$$[(b-a) \cdot D]f(4,3) == (b_1 - a_1)f_1(4,3) + (b_2 - a_2)f_2(4,3).$$

More symbolic things would follow. When we say $[(b-a) \cdot D]^2$ we understand exactly the same thing as in the one dimensional case above.

$$[(b-a) \cdot D]^2 == [(b_1 - a_1)D_1 + (b_2 - a_2)D_2]^2$$

$$= (b_1 - a_1)^2 D_1^2 + 2(b_1 - a_1)(b_2 - a_2) D_2 D_1 + (b_2 - a_2)^2 D_2^2.$$

Thus

$$[(b-a)\cdot D]^2 f = (b_1 - a_1)^2 f_{11} + 2(b_1 - a_1)(b_2 - a_2) f_{12} + (b_2 - a_2)^2 f_{22}.$$

More specifically, if you want to evaluate the function at any point you can do so

$$[(b-a)\cdot D]^2 f(a) =$$
$$(b_1 - a_1)^2 f_{11}(a) + 2(b_1 - a_1)(b_2 - a_2) f_{12}(a) + (b_2 - a_2)^2 f_{22}(a).$$

If you have got a feeling, then

$$[(b-a)\cdot D]^k = \sum_{j=0}^{k} \binom{k}{j} (b_1 - a_1)^j (b_2 - a_2)^{k-j} D_2^{k-j} D_1^j.$$

Or

$$[(b-a)\cdot D]^k f = \sum_{j=0}^{k} \binom{k}{j} (b_1 - a_1)^j (b_2 - a_2)^{k-j} D f_{1^j 2^{k-j}}.$$

For example $f_{1^3 2^4}$ means $f_{1112222}$. Still more specifically, if you want to calculate the function above at the point a,

$$[(b-a)\cdot D]^k f(a) = \sum_{j=0}^{k} \binom{k}{j} (b_1 - a_1)^j (b_2 - a_2)^{k-j} D f_{1^j 2^{k-j}}(a).$$

Now let us go back to the simple chain rule and calculate higher order derivatives of composed function in a special case. The special case is the following. We have an open set $\Omega \subset R^2$ and a real valued c^2 function on Ω and two points $a, b \in \Omega$. We have an open interval $I \subset R$. Define

$$\Phi(u) = ub + (1-u)a.$$

We consider the real valued function $F(u) = f(\Phi(u))$ defined on I. By earlier chain rule

$$F'(u) = f'(\Phi(u)) \cdot \Phi'(u) = [(b-a)\cdot D] f(\Phi(u)). \qquad (*)$$

Recall that to compare with earlier notation, here we have

$$\Phi(u) = (\varphi_1(u), \varphi_2(u))$$

$$\varphi_1(u) = ub_1 + (1-u)a_1; \quad \varphi_2(u) = ub_2 + (1-u)a_2.$$
$$\Phi' = (b_1 - a_1, b_2 - a_2)$$

Thus $(*)$ is a consequence of earlier chain rule. Now let us consider this function F' as a sum of two composed function.

$$F'(u) = f_1(\Phi(u))(b_1 - a_1) + f_2(\Phi(u))(b_2 - a_2).$$

Exactly the same argument applied to f_1 and f_2 in place of $f$gives

$$[f_1(\Phi(u)]' = f_{11}(\Phi(u))(b_1 - a_1) + f_{12}(b_2 - a_2).$$

$$[f_2(\Phi(u)]' = f_{21}(\Phi(u))(b_1 - a_1) + f_{22}(b_2 - a_2).$$

Substituting above and simplifying we get

$$F''(u) = [(b - a) \cdot D]^2 f(u). \qquad (**)$$

If you have understood the argument of arriving at $(**)$ from $(*)$, you should have no problem in arriving at the following, assuming that f is C^3.

$$F^{(3)}(u) = [(b - a) \cdot D]^3 f(\Phi(u)). \qquad (***)$$

Thus, by induction, one has

$$F^{(k)}(u) = [(b - a) \cdot D]^k f(\Phi(u)).$$

<u>Proof of Taylor:</u>

Finally, we return to proof of (♣). Since the line joining a and b is contained in Ω and Omega is open, you can easily see that there is a an $\epsilon > 0$ so that

$$\Phi(u) = ub + (1 - u)a$$

defined on $(-\epsilon, 1 + \epsilon)$ has range contained in Ω. Now consider the function

$$F(u) = f(\Phi(u))$$

on this interval. all the above argument shows that when f is C^n then so is F and gives a formula to calculate its derivatives.

Expand $F(1)$ around zero. This means apply the usual one variable Taylor, namely (♠), to F with $b = 1$ and $a = 0$.

$$f(1) = \sum_{k=0}^{n-1} \frac{[D]^k F(0)}{k!} + \int_0^1 [D]^n F(u) \frac{(1 - u)^{n-1}}{(n - 1)!} du.$$

That is,

$$F(1) = \sum_{k=0}^{n-1} \frac{F^{(k)}(0)}{k!} + \int_0^1 F^{(n)}(u)\frac{(1-u)^{n-1}}{(n-1)!}du.$$

Observe

$$F(1) = f(b); \quad F^{(k)}(0) = [(b-a)\cdot D]^k f(a).$$

This completes proof.

Again, this can be stated differently

$$f(b) = \sum_{k=0}^{n-1} \frac{[(b-a)\cdot D]^k f(a)}{k!} + \frac{[(b-a)\cdot D]^n f(\theta)}{n!};$$

where θ is a point on the line joining a to b. This follows from the version proved above by noting that the integral is between $c/n!$ and $C/n!$ where c and C are bounds for $[(b-a)\cdot D]^n f(\theta)$ as θ runs on the line.

It is also customary to state the Taylor with x instead of b

$$f(x) \quad = \quad \sum_{k=0}^{n-1} \frac{[(x-a)\cdot D]^k f(a)}{k!} \quad +$$

$$\int_0^1 [(x-a)\cdot D]^n f(ux+(1-u)a)\frac{(1-u)^{n-1}}{(n-1)!}du.$$

But there is a subtle scope for confusion in understanding this and so I did not state it this way.

Exact the same formula holds in dimensions more than two as well. Now $(b-a)$ and D are of that dimension. But formula is the same and so is the proof.

Though we do not have much use of this formula, we shall see one important consequence of this formula. But it is satisfying to realise that the ideas of last semester are indeed very powerful to let the same formulae to hold even in higher dimensions.

Extrema:

Let $\Omega \subset R^2$ be open and f be a real valued C^1 function on Ω. Let $a \in \Omega$. We say a is a point of local maximum if there is an $r > 0$ such that $f(x) \leq f(a)$ for all $x \in B(a,r)$. call it strict local maximum if the inequality is strict for every x in the ball different from a.

Similarly we say that a is a point of local minimum if there is an $r > 0$ such that $f(x) \geq f(a)$ for all $x \in B(a, r)$. A point which is a local minimum or local maximum is called a local extremum.

Let a be a local maximum. If $a = (a_1, a_2)$ then clearly, $x \mapsto f(x, a_2)$ has local maximum at $x = a_1$ and $y \mapsto f(a_1, y)$ has local maximum at $y = a_2$. Thus

$$f_1(a) = 0 = f_2(a).$$

Same holds even if a is a local minimum.

Just like in one dimensions here too, the above equations would not guarantee either a maximum or minimum. For example $f(x, y) = x^3$ has both derivatives at $(0, 0)$ but f has neither a max nor min at that point. This is not surprising.

However, now something spectacular may also happen. The point may be maximum in several directions and minimum in several other directions at that point! For example,

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

has derivatives zero at the point $(0, 0)$. There are several lines passing through the origin such that if you restrict f to that line it has a max at this point and for several other lines it is a point of minimum.

In one dimensions the strict positivity of the second derivative at a would ensure that a is a local minimum. Same is true here too.

Assume that f is C^2. Recall there are now four (actually three distinct) second derivatives. Let us temporarily, only for this section, denote by f'' the following 2×2 matrix.

$$f''(a) = \begin{pmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{pmatrix}.$$

Just to draw your attention to the fact that the matrix is symmetric, we have written, in the first entry of second row, f_{12} instead of the expected $f_{21} = f_{12}$.

Let A be a 2×2 symmetric matrix.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Let us say that A is positive definite or simply positive if $v^t A v > 0$ for every non-zero vector v. That is

$$a_{11}v_1^2 + 2a_{12}v_1v_2 + a_{22}v_2^2 > 0.$$

We write this as $A \gg 0$. We shall need matrix theory, but no deep results. Here is a fact to which we shall return later.

$$A \gg 0 \quad \leftrightarrow \quad a_{11} > 0; \;\; a_{11}a_{22} - a_{12}^2 > 0.$$

Here is a criterion for local minimum.

Theorem: Let f be a C^2 function on an open set $\Omega \subset R^2$ and $a \in \Omega$. If $f'(a) = 0$ and $f'' \gg 0$ the a is a point of strict local minimum.

Proof is simple. Consider the two functions: f_{11} and $f_{11}f_{22} - f_{12}^2$. These are continuous and are strictly positive at the point a. So fix $r > 0$ so that they are strictly positive for all $x \in B(a, r)$. Now take any point b in this ball, use Taylor

$$f(b) = f(a) + [(b-a)\cdot D]f(a) + \frac{[(b-a)\cdot D]^2 f(\theta)}{2!};$$

where θ is a point on the line joining a to b. But whatever it be, the positive definiteness of f'' says that the last term on the right side is positive. The second term is zero. Thus $f(b) > f(a)$. This completes the proof.

Similarly, we have a criterion for local minimum too. Say that A is negative definite if for all non-zero vectors $v^t A v < 0$. We write $A \ll 0$.

$$A \ll 0 \quad \leftrightarrow \quad a_{11} < 0; \;\; a_{11}a_{22} - a_{12}^2 > 0.$$

The same proof as earlier shows that If $f'' \ll 0$ at a point a, then it is so in a ball around the point a.

functions without formula:

So far what we have been doing is just imitation of the development of calculus we learnt last semester. of course, this statement does not mean it is a trivial job. It has taken us rather too far. We can now differentiate functions defined on any Euclidean space (R^k) and taking values in any Euclidean space (R^n).

If you think about it, the achievement is really spectacular. Imagine, $n \times n$ matrix is nothing but a point in n^2-dimensional euclidean space. Since determinant is a continuous function of the entries of the matrix (it is a polynomial), the set of non-singular matrices is an open set (determinant non zero). Thus you can define a function on this open set of n^2 dimensional space to itself, namely, the matrix inverse map. We know how to differentiate this function!

This is all fine. But is there anything we can do now about functions of one variables (we learnt last semester) which we could not do last semester? Yes, there are several things that we can do now even for functions of one variable. Just because this course is functions of several variables, you should not be under any wrong impression.

Let us start with a simple example. Consider the function $y = \varphi(x)$ defined by the formula $x^2 + y^2 - 1 = 0$. Of course you might ask if there is such a function at all. This is a happy situation, you can explicitly solve the equation. There are two functions defined on the open interval $(-1, = 1)$ by the formula

$$\varphi(x) = +\sqrt{1-x^2}; \qquad \psi(x) = -\sqrt{1-x^2}.$$

Just a word about notation: I do not have to put + sign in describing φ because by convention a square root is always taken positive. However we did so to draw your attention.

$$\varphi'(x) = -\frac{x}{\sqrt{1-x^2}}.$$

Here is another way of arriving at the answer. Consider the function

$$f(x,y) = x^2 + y^2 - 1$$

on R^2 and $F(x) = f(x, \varphi(x))$. Then we know that $F \equiv 0$ and hence $F'(x) \equiv 0$. But by chain rule

$$F'(x) = (1, \varphi'(x)) \cdot (2x, 2y) = 2x + 2y\varphi'(x) \equiv 0.$$

This gives

$$\varphi'(x) = -\frac{x}{y} = -\frac{x}{\sqrt{1-x^2}}.$$

Of course what is the purpose of using the chain rule when you can get explicit formula for your function and calculate using expertise of last semester. As far as this example is concerned this is just another way of doing it. But sometimes this may be the only way of doing it! This happens especially when you are not lucky to get a formula in your hand.

Let us now consider another example.

Consider the function $y = \varphi(x)$ on the interval $(1, \infty)$ defined by the formula

$$y = \log(x + y).$$

is there such a function? Yes, consider the equation

$$x = e^y - y.$$

The right side as a function of y starts off at 1 when $y = 0$; derivative being positive it strictly increases towards infinity as y becomes large. Thus it assumes all values between one and infinity exactly once as y travels from zero to infinity. Thus given any number $x > 1$ there is exactly one $y > 0$ satisfying the above equation. This is the number $\varphi(x)$.

Is this function differentiable? We do not have explicit formula. suppose it is differentiable. What is φ'? Again as above consider the function

$$f(x, y) = y - \log(x + y)$$

so that

$$F(x) = f(x, \varphi(x)) \equiv 0.$$

Thus

$$F'(x) = (1, \varphi'(x)) \cdot (\frac{1}{x+y}, 1 - \frac{1}{x+y}) = 0.$$

giving

$$\varphi'(x) = \frac{1}{x+y} \; \frac{x+y}{x+y-1} = \frac{1}{x+y-1}.$$

Thus without any explicit formula for our function, we have been able to differntiate.

Of course, you can still avoid two variable calculus. You can say, consider

$$\psi(y) = e^y - y : (0, \infty) \to (1, \infty)$$

Thus $\varphi(x)$ is precisely inverse of this map and we know

$$\varphi'(x) = \frac{1}{\psi'(\varphi(x))} = \frac{1}{e^y - 1} = \frac{1}{x + y - 1}.$$

But many times even such a recognition is not possible. Then we need to use the method outlined above.

Such functions are called 'implicit functions'. That is, functions that *are there already* in the relation you want to be satisfied. They may have explicit formula or may not.

But as soon as you know that you have a differentiable function $y = \varphi(x)$ satisfying the relation $f(x, y) \equiv 0$ then the above argument tells us that

$$\varphi'(x) = -\frac{f_1(x, \varphi(x))}{f_2(x, \varphi(x))}.$$

Thus we need to understand the problem: given a relation to be satisfied between x and y is there a function $\varphi(x)$ so that when you take $y = \varphi(x)$ then (x, y) satisfies your relation. Is such a function differentiable? This is precisely the question answered by the 'implicit function theorem'.

Imagine starting with the relation $x^2 + y^2 + 1 = 0$, so that,

$$f(x, y) = x^2 + y^2 + 1$$

and the above formula gives

$$\varphi'(x) = -\frac{x}{y}.$$

But this is non-sense because there is no function at all!

So let us take a point (x_0, y_0) satisfying the given relation. You can see from the formula for the derivative of $y = \varphi$, we need $f_2(x_0, y_0) \neq 0$). Surprisingly this condition is enough.

Theorem (Implicit function theorem)

Let $f : \Omega \subset R^2 \to R$ be a C^1 function. Suppose $(a, b) \in \Omega$ Suppose $f(a, b) = 0$ and $f_2(a, b)] \neq 0$. Then

(i) there is a rectangle $Q = (a - \delta, a + \delta) \times (b - \eta, b + \eta) \subset \Omega$ and a unique function φ defined on the interval $(a - \delta, a + \delta)$ whose graph is contained in the rectangle Q and such that $f(x, \varphi(x)) \equiv 0$.

(ii) This function is differentiable.
(iii)

$$\varphi'(x) = -\frac{f_1(x, \varphi(x))}{f_2(x, \varphi(x))}.$$