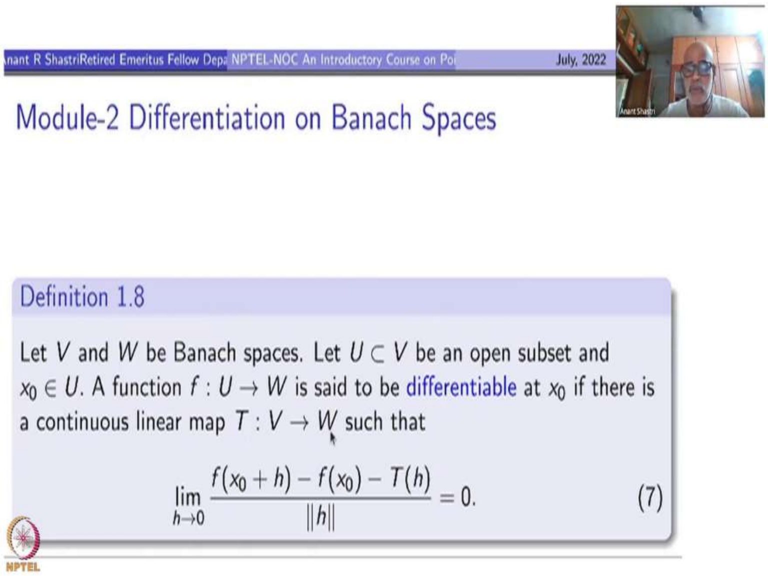
An Introduction to Point Set Topology Part 2
Professor Anant R Shastri
Department of Mathematics
Indian Institute of Technology, Bombay
Lecture 02
Differentiation on Banach Spaces

(Refer Slide Time: 0:16)



Hello, welcome to module 2 of NPTEL NOC an introductory course on point set topology part 2. Today, we will take up the topic of differentiation on Banach spaces. So, many things run parallel to what we do in calculus, one variable calculus or two variable calculus and so on.

Actually, you will see that we copy many ideas from one variable calculus but the things have to be put in a proper perspective. So, start with two Banach spaces one can possibly do many things with just norm linear spaces also but we will concentrate only on Banach spaces. Because our idea of presenting these results is not to do the entire thing in a very general setup but to cover the implicit function theorem and inverse function theorem for Banach spaces as a sample of application of topological methods. So, that is why we will concentrate only on Banach spaces.

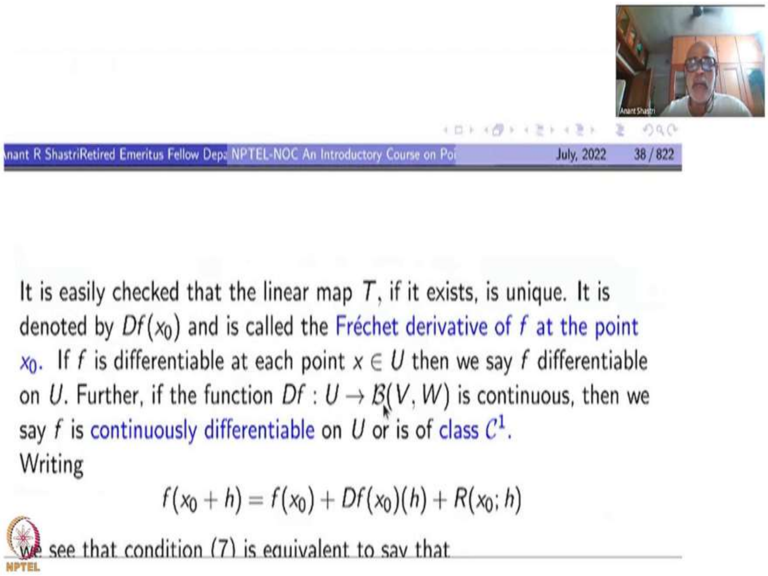
Take a subset U contained inside V, an open subset around a point x_0 inside a Banach space V. Note that the concept of an open subset makes sense here because we are using the norm to induce a metric which in turn induces a topology on V. Start with an open subset U around a point x_0 and a function f defined on this open set into another Banach space. f is said to be differentiable at x_0 if there is a continuous linear map T from V to W such that this limit is equal to 0. What is this stuff? It is $f(x_0 + h) - f(x_0)/\|h\|$. Note that unlike in 1-variable calculus, I cannot divide by h, because h is a vector in V.

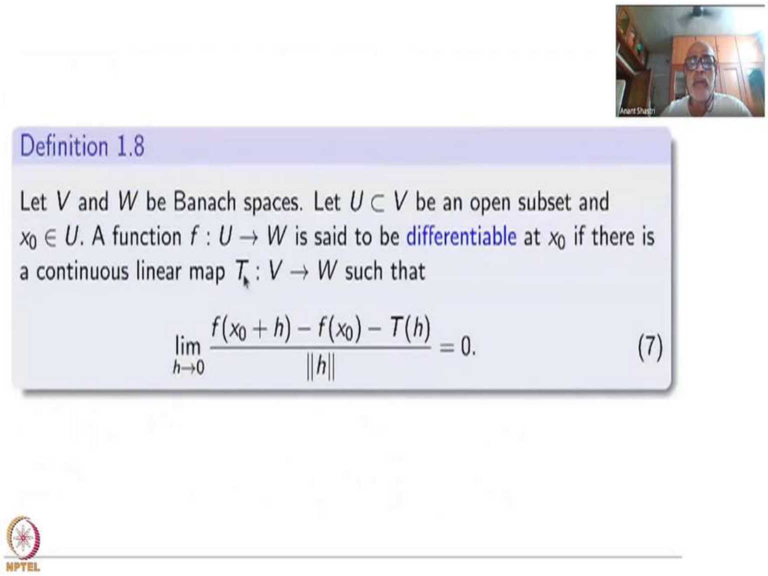
Therefore, to get a real number, I take the norm of h and then divide by it. Note that T is a continuous linear map which is going to be called the derivative of f. Then I can divide by norm h this limit as h tends to 0 which is same thing as saying norms tends to 0. This limit must exist as a vector in W and must be zero. This is the same as saying that the limit of the norm of the difference divided by norm of h as a real number is zero. Note that this numerator is taking value inside W. So, divided by norm h of course that is just a scalar.

So, this is a vector inside W this limit must exist. So, the important point here is that we must have a continuous linear map T. This is the part of the definition. I don't want to question that. Indeed, there are slightly varying definitions, some weaker and some stronger and so on. There are different definitions by different authors, and then they will make under this condition this will be equal to that one that will be called this one and so on. We are not going to study all that in this course. So, let us take this definition namely, there must be a linear map which is bounded, i.e., a continuous linear map T from V to W which satisfies this limit condition.

As soon as such a T exists, it has to be unique. You cannot have two different linear maps T_1 and T_2 satisfying the same property for $f(x_0)$. This follows exactly the same way how you prove the uniqueness in the case of usual calculus of one or several variables. So, uniqueness is not a difficulty.

(Refer Slide Time: 5:29)





So, that unique T is called the derivative of f at the point x_0 and I am using the same notation $Df(x_0)$ for it, the standard notation. The only thing is you might not have called it a Frechet derivative. We are going to call this T, the Frechet derivative of f at x_0. Frechet started this study of Banach space calculus.

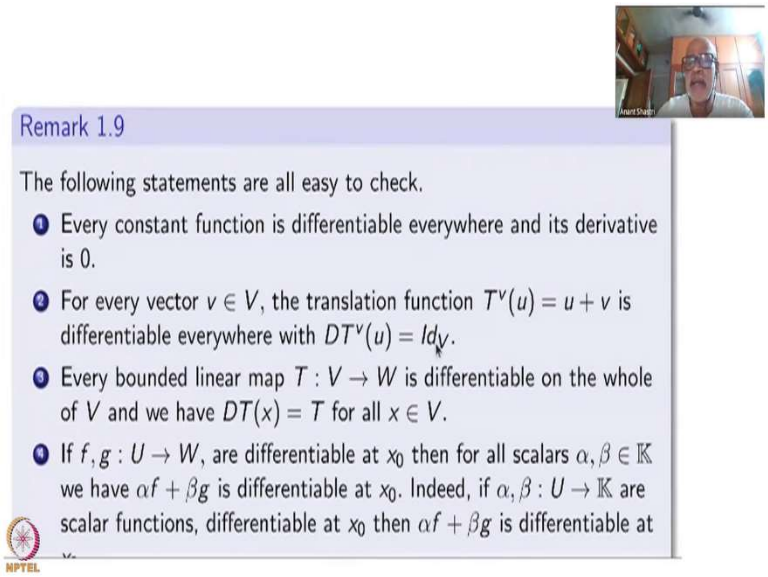
If f is differentiable at each point x inside an open set then we say f is differentiable on the open set U. Further, if the function which assigns to each point x of U, the Frechet derivative of f at x, remember the derivative is a bounded linear map and so, we get a function taking values inside $\mathcal{B}(V, W)$, denoted by Df from U to $\mathcal{B}(V, W)$. If this itself is continuous then we say that f is continuously differentiable on U. Or we can say it is of class C^1 or C^2 so on depending on whether Df is itself is continuously differentiable and so on. So, we will stop here with class C^1 functions here.

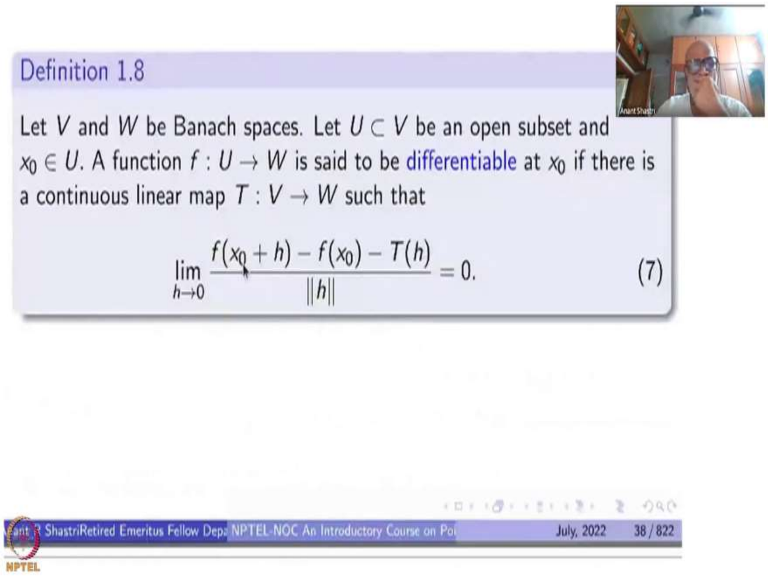
So, if you go back and peep into your calculus course, the very first thing you did now is the so-called increment theorem for differentiable functions or a function which has a derivative at a single point. Rewriting this limit condition, by clearing the denominator and reinterpreting. That is called increment theorem. In other words, we define the quantity $R(x;h)$ by the equation $f(x_0+h) = f(x_0) + D(f)(x_0)(h) + R(x_0;h)$. Then the limit condition can be restated as limit of $R(x_0;h)$ divided by $\|h\|$ is zero as h tends to zero. The term $R(x_0;h)$ which clearly depends on x_0 and h is called the remainder term, etc.

So, this is called the increment theorem or the first approximation, linear approximation to f at the point x_0. If you increase the value of x_0 by h, the increment is roughly $Df(x_0)(h)$. You can ignore this last term $R(x_0;h)$, when h is small. That is the whole idea.

Why? Why should you ignore this one? What I mean is that what allows you to ignore this one? It is not always possible but because of this definition what we have is that if you divide this by $\|h\|$ and take limit then it is 0. So, remainder after the first term here has the property that divided by $\|h\|$ and take the limit as h tends to zero is zero. So, this is called increment theorem exactly as in the case of one variable calculus.

(Refer Slide Time: 9:11)





The following statements are all easy to check exactly as in the case of one variable calculus. Every constant function is differentiable everywhere. Everywhere means what? On the whole of V wherever they are defined. And derivatives of constant functions are always 0, the derivative at all the points is 0, for a constant function.

For every vector $v \in V$, consider the translation function $T^v(u) = u + v$. See, on a vector space we have this function u going to $u + v$, where v is fixed; it is called the translation function; I have written T^v here; maybe I will forget to write this notation every time the translation function is very easy to remember.

It is differentiable everywhere and its derivative is the identity function remember translation is from V to V. So, the derivative will be also from V to V. It will be a continuous linear map. In this case, it is the identity map of V. All that we have to do is to go back to this definition. $f(x_0 + h) - f(x_0)$ will be what? It is $T^v(x_0 + h) - T^v(x_0) = (x_0 + h + v) - (x_0 + v) = h$, $x_0 + v$ cancels out, it is just like $Id(x_0 + h) - Id(x_0)$. So, what should I should take T to be? Take $T = Id_V$, identity map then this numerator itself will be identically 0. So, the limit will be zero as needed. So, the translation maps are differentiable their derivative at every point being the Id_V.

(Refer Slide Time: 11:01)

Remark 1.9

The following statements are all easy to check.

1. Every constant function is differentiable everywhere and its derivative is 0.
2. For every vector $v \in V$, the translation function $T^v(u) = u + v$ is differentiable everywhere with $DT^v(u) = Id_V$.
3. Every bounded linear map $T : V \to W$ is differentiable on the whole of V and we have $DT(x) = T$ for all $x \in V$.
4. If $f, g : U \to W$, are differentiable at x_0 then for all scalars $\alpha, \beta \in \mathbb{K}$ we have $\alpha f + \beta g$ is differentiable at x_0. Indeed, if $\alpha, \beta : U \to \mathbb{K}$ are scalar functions, differentiable at x_0 then $\alpha f + \beta g$ is differentiable at x_0.
5. If f is differentiable at x_0, then it is continuous at x_0.

NPTEL

Definition 1.8

Let V and W be Banach spaces. Let $U \subset V$ be an open subset and $x_0 \in U$. A function $f : U \to W$ is said to be differentiable at x_0 if there is a continuous linear map $T : V \to W$ such that

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - T(h)}{\|h\|} = 0. \qquad (7)$$

NPTEL

It is easily checked that the linear map T, if it exists, is unique. It is denoted by $Df(x_0)$ and is called the Fréchet derivative of f at the point x_0. If f is differentiable at each point $x \in U$ then we say f differentiable on U. Further, if the function $Df : U \to \mathcal{B}(V, W)$ is continuous, then we say f is continuously differentiable on U or is of class $\mathcal{C}^1$.

Writing

$$f(x_0 + h) = f(x_0) + Df(x_0)(h) + R(x_0; h)$$

we see that condition (7) is equivalent to say that

$$\lim_{h \to 0} \frac{R(x_0; h)}{\|h\|} = 0 \qquad (8)$$

NPTEL

Similarly, every continuous linear map T is differentiable everywhere with its derivative at each point being be the same linear map T.

Here one lucky thing is that we do not have to make further assumptions. Start the linear map you do not know that it is continuous. So, you have to make the assumption you have to put that extra condition continuity. Once it is continuous it is differentiable and its derivative is again the same function T at all the points $x \in V$.

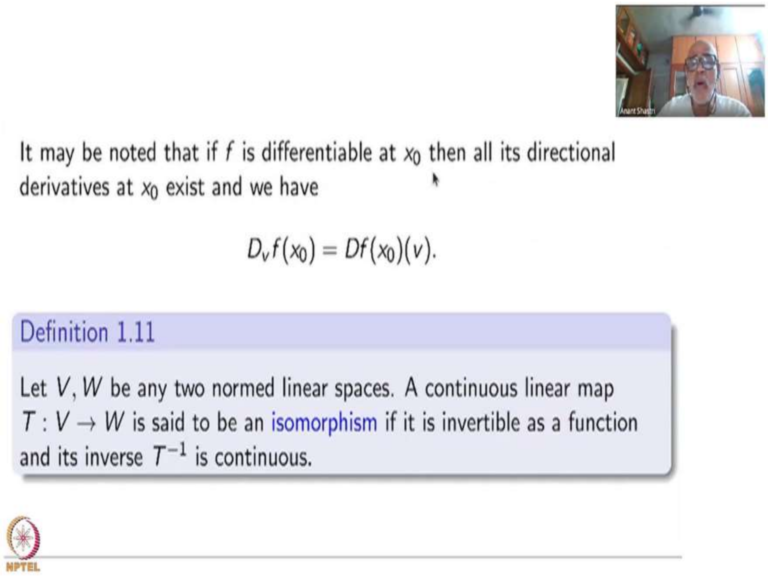
This is again a standard result in multivariable calculus of finite many variable. If you have linear map its derivative it is a linear map itself. You can directly verify it by taking T itself as in the slot in the third slot here. So, $T(x_0 + h) - T(x_0) - T(h)$ is 0 that is all that will give you that T itself is the derivative of T.

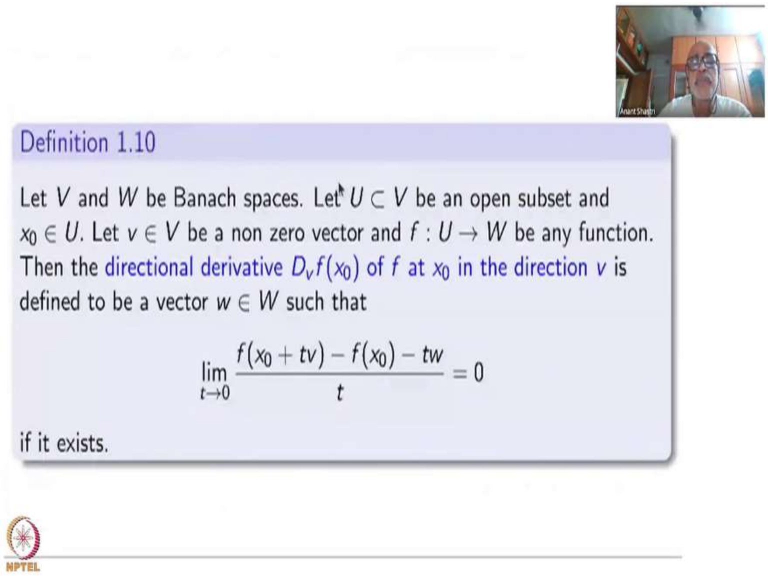
However, in the general case, because in our definition, (we somewhat artificially) demanded that the derivative should be a continuous linear map, we have to also start with a continuous linear map.

And then this standard addition rule and scalar multiplication rule: if f and g are differentiable at x_0 and α, β are scalars, then $\alpha f + \beta g$ is differentiable at x_0. Indeed if α and β are themselves scalar functions functions from U to $\mathbb{K}$ which are differentiable at x_0, then this αf (this is not a composition this is this is just multiplication right with scalar multiplication) is differentiable at x_0.

So, the derivative of αf makes sense. Similarly the derivative of βg makes sense the sum will be also differentiable at x_0. Of course you have to use Leibniz rule here to get the derivative of the sum. Also, if f is differentiable at x_0, then it is continuous at x_0 same proof as in the case of one variable. This is one variable calculus after all. You can just look at the increment theorem here, to show that f is continuous also.

(Refer Slide Time: 14:07)





So far, except that in the definition I have started with continuous linear map, everything is just like one variable calculus. In the one variable calculus, the derivative is just a real number, but if you think carefully (you may have done this already) that real number actually represents a linear map from $\mathbb{R}$ to $\mathbb{R}$, namely, the multiplication by that number. Thus, so far, there is no difference at all.

So, it may be noted that if f is differentiable at x naught then all directional ... very sorry, let me recall this one. I am jumping to say that all directional derivatives also exist. So, let us know what is the meaning of directional derivative in this context. Again this is the same thing as in the case of multivariable calculus.

Starting with two Banach spaces V and W, an open subset U of V, and a point x_0 belonging to U. Now you take any vector preferably a non-zero vector (even the vector 0 is also allowed) take any non-zero vector $v \in V$. Let f from U to W be any function. Then this is the directional derivative of f in the direction of v at the point x_0 is defined as follows:

Namely it is a vector w inside W such that the limit as t tends to zero of $f(x_0 + tv) - f(x_0) - tw$ whole of this divided by t is zero.

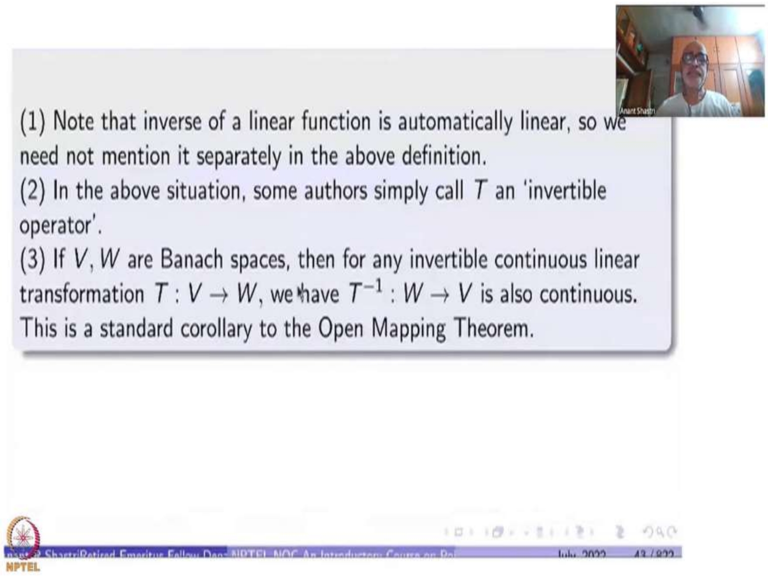
So in the definition of differentiability, you have replaced the bounded linear T by a very specific one viz, t going to tw; the limit is taken not over all vectors h but restricted to multiples tv of v. That is the differennce. The numerator is just a function the real variable t and takes values in W. You are dividing by t itself, no norm. x_0, v and w is fixed. So, it is function of real variable, one variable. Then this limit must be 0. In other words, if you just look at the function t maps to $f(x_0 + tv)$, then this function must be differentiable as a function of t and its derivative is w.

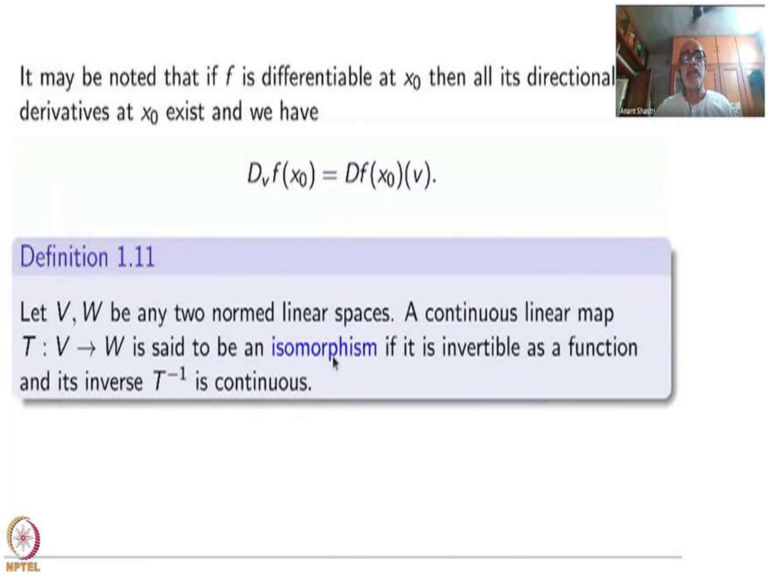
That vector w is called the directional derivative of f at x_0 in the direction of v. Exactly same definition as in the case of usual multivariable calculus.

And you can immediately verify that all the directional derivatives will exist as soon as the derivative at x_0 exists. So, often one calls the other derivative which you have defined as the total derivative. You can talk about partial derivatives but then you have to fix up coordinates. In Banach spaces coordinate fixing is something very fishy. You do not want to do that.

So, let V and W be any two normed linear spaces. A continuous linear map T from V to W is said to be an isomorphism, if it is invertible as a function and the inverse is also continuous. So, here I have taken this definition for all norm linear spaces.

(Refer Slide Time: 18:36)





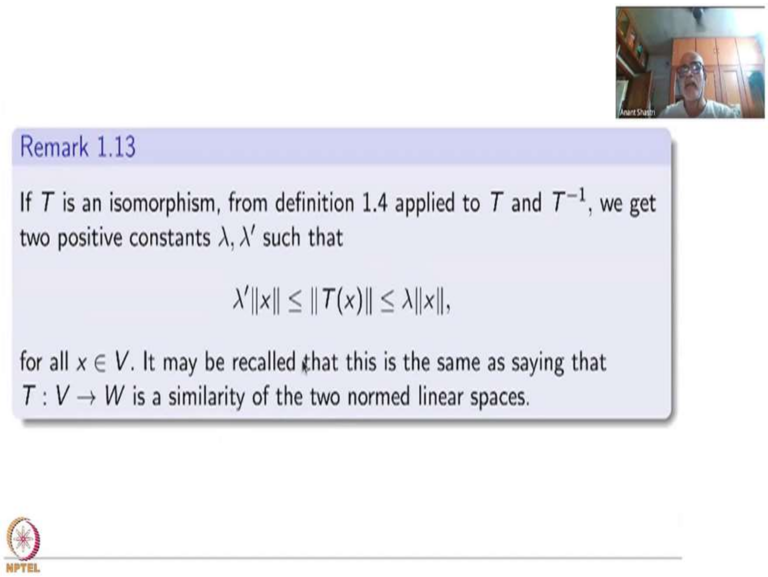
The remark here is that the inverse of a linear function is automatically linear. This is elementary algebra on any vector space. If you have just a bijection which is linear then the inverse is automatically linear that is not a problem.

But inverse may not be continuous this is what we have been telling. Even if it is a bijection and continuous, its inverse may not be continuous. So, in the above definition, we have to mention it separately. In the above situation some authors simply call T an invertible operator. They do not separately mention that inverse is also continuous etc, they will just say invertible operator. But the continuity of the inverse is also assumed. So, I will also use that terminology.

If V and W are Banach spaces then for any invertible (bijective) continuous linear transformation V from V to W automatically T^{-1} will be continuous, i.e., T and T^{-1} are both invertible operators. See remember invertible just for me does not mean that it is continuous. So, I want to be very careful invertible operator is by convention inverse is also continuous.

A linear map may be invertible it has an inverse but its inverse may not be continuous even if it is continuous. That is why I am making this extra caution. However, if V and W are Banach spaces then there is no problem. Automatically the inverse will be also continuous. But this one needs a deeper theorem there, namely, what is called as open mapping theorem. We are not going that deep into function analysis here. This is not a course on function analysis. But I am just mentioning this. I will never use this property because we are not going to prove this here. I am just mentioning it as an information; that is all.

(Refer Slide Time: 21:13)

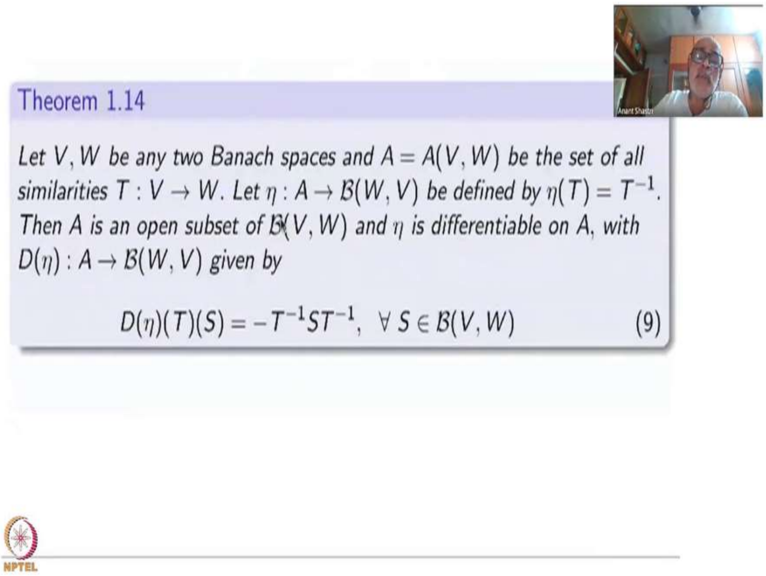


If T is an isomorphism from definition 1.4 (of boundedness) applied to both T and T^{-1}, we get two constants λ and λ' such that $\lambda'\|x\|$ is less than or equal to etc.

This right-hand side constant λ says T is continuous: $\|T(x)\| \leq \lambda\|x\|$. Similarly I must have other way around also for T^{-1} which will give you a λ' on this side. viz., $\|x\| \leq \|T(x)\|$ which you can rewrite it as $\|T^{-1}(x)\| \leq \|x\|/\lambda'$. So, both sides you get a constant. This may remind you the concept of two linear transformations being similar.

This is a special case in some sense. Here only one function T is involved. You may say T itself is a similarity, because it is similar to the identity map. It is not the identity map but it is similar to the identity map. So, a linear map T from V to W is called a similarity of the two norm linear spaces if it it satisfies this condition. We have studied similarities in the part 1.

(Refer Slide Time: 22:44)



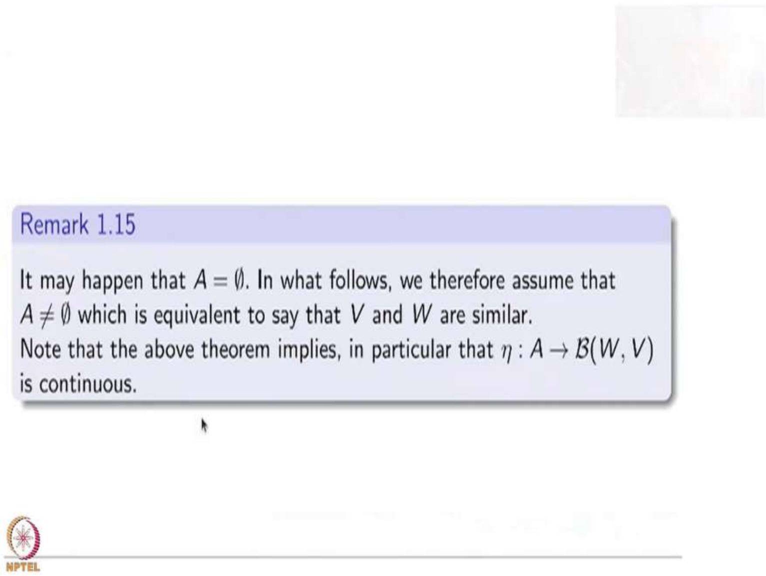
Now here is a theorem that I need to use. So, go through this carefully. Start with two Banach spaces put $A := A(V, W)$ (A is a temporary short form when V and W are understood) the set subset of $\mathcal{B}(V, W)$, of all similarities T from V to W. Not all bounded linear transformation but only similarities. Consider the function eta from A to $\mathcal{B}(V, W)$, defined by $\eta(T) = T^{-1}$.

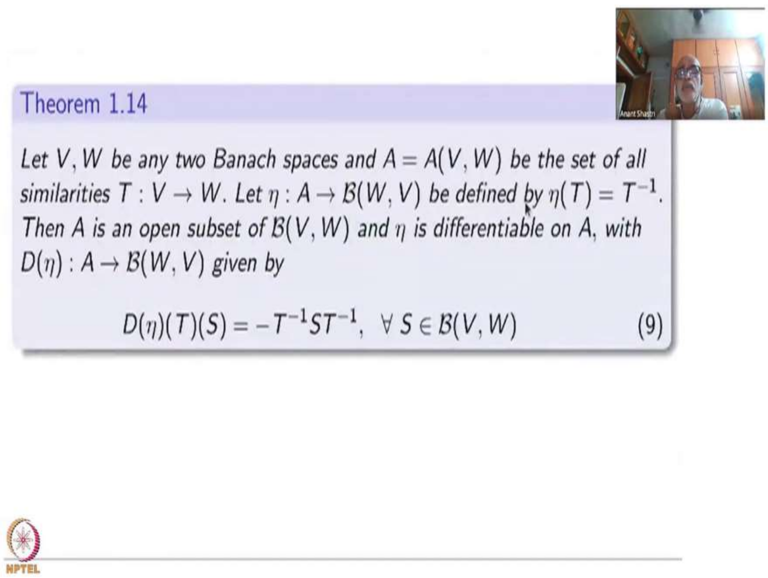
So, each element in A is inevitable. So, I can take their inverse. I am getting inside $A(W, V)$. Everything is happening $\mathcal{B}(V, W)$ and $\mathcal{B}(W, V)$ which are Banach spaces. Then A is an open subset of this $\mathcal{B}(V, W)$ and η is differentiable on the entire of A.

So, I am not stating A non-empty. (If it is empty the statement is trivial.) First of all I say that this subset A is an open subset of $\mathcal{B}(V, W)$ (remember $\mathcal{B}(V, W)$ itself is a Banach space) and η is differentiable on A. Further the derivative $D(\eta)$ at any $T \in A$ is an element of $\mathcal{B}(\mathcal{B}(V, W), \mathcal{B}(W, V))$, because η itself is a map of A in into $\mathcal{B}(W, V)$. For each point T of A you have a bounded linear map from $\mathcal{B}(V, W)$ to $\mathcal{B}(W, V)$.

It is obtained by pre-composing with T^{-1} first then again post-composing with T^{-1} and finally putting a minus sign. So, it looks a bit complicated. For every S belong to $\mathcal{B}(V, W)$, (not necessarily invertible) $D(\eta)(T)(S)$ is $-T^{-1}ST^{-1}$. A is a subset of $\mathcal{B}(V, W)$. I am claiming that A is an open subset. On this open subset you have a function. You can talk about whether it is differentiable or not. The statement is that it is differentiable and its derivative is given by this formula. So, that is statement. So, let us see the proof which is not all that difficult.

(Refer Slide Time: 26:06)



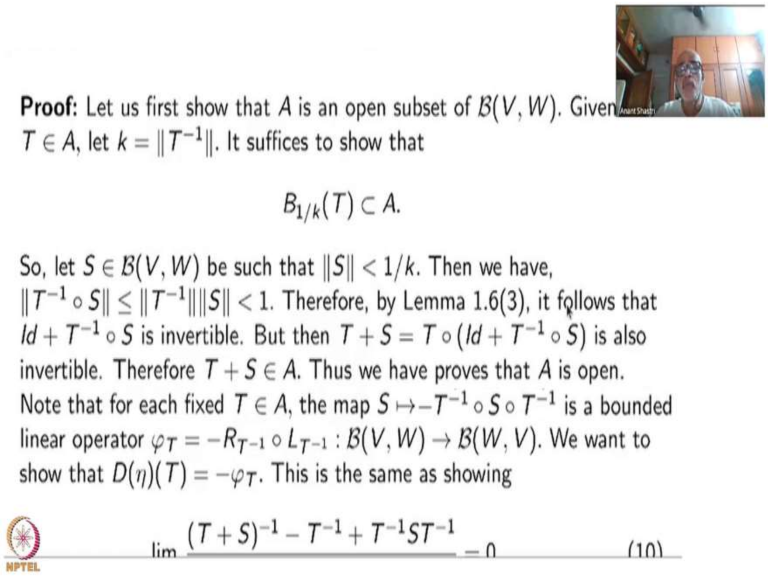


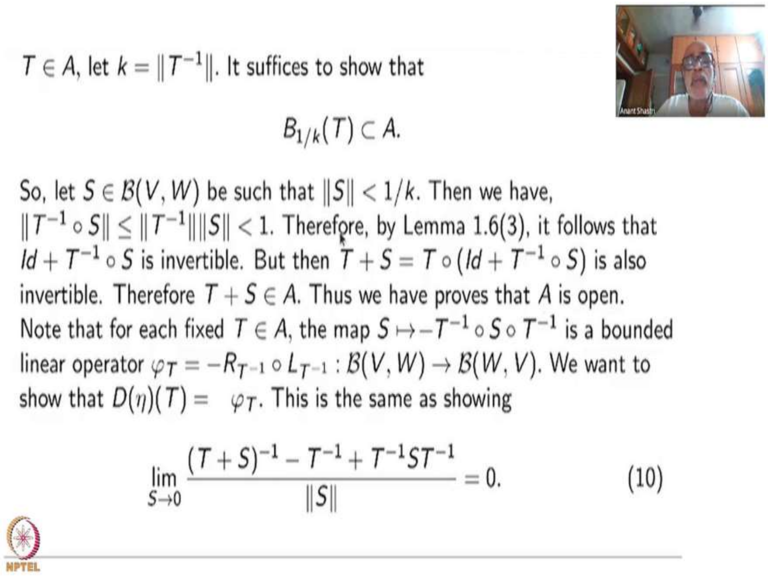
First of all, what may happen if this A is empty. What is the meaning of that? There may not be any similarities between V and W. So, if you want to say anything there is no statement about this being an empty set. Whatever I have stated is vacuously true. So, we should assume that A is non-empty that is all otherwise you do not have to prove anything all.

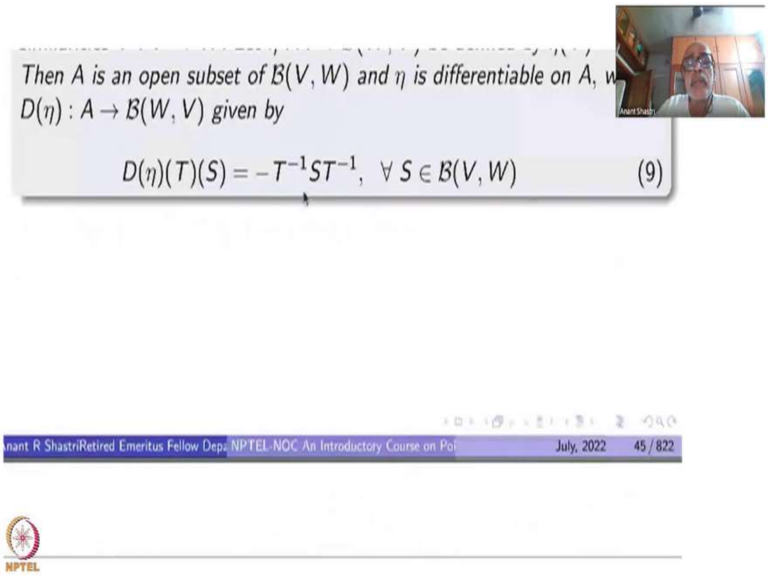
So, assume A is non-empty. What is the meaning of A is non-empty there is some similarity which means V and W are similar. Already that is a non-trivial assumption.

The above theorem implies in particular that η is continuous. Because we have already remarked that any function which is differentiable at a point is continuous at that point. We are going to prove that this eta is differentiable on the whole of A. Therefore it is continuous on A which is not stated here but it is an easy consequence of that. So, we will use that also at an appropriate place.

(Refer Slide Time: 27:40)







First, toward the proof of that A is an open subset, in the last lecture, we have already made the preparation for this. So, let us see how? Take a point T inside A. What is it? It is a similarity from V to W. I am taking k equal to $\|T^{-1}\|$. Then I am claiming that the ball of radius $1/k$ around T is contained inside A. T is an arbitrary point of A, a ball of radius $1/k$ is contained inside A. This $1/k$ is obviously is non-zero. For each point in A, you have an open ball contained inside A. So, A is open.

Remember we are working in $\mathcal{B}(V, W)$; A subset of $\mathcal{B}(V, W)$. Let S belonging to $\mathcal{B}(V, W)$ be such that $\|S\| < 1/k$. Then we know that $\|T^{-1}S\| \leq \|T^{-1}\|\|S\|$, which in turn is less than 1. That is the whole idea why I took $k = \|T^{-1}\|$. So, it is less than 1 therefore by lemma 1.6(3) that we have proved, it follows that $(Id + T^{-1}S)$ is invertible.

There it is identity minus you can take $-S$ here and you can put this will become $Id + T^{-1}S$ is invertible because norm of this one is less than 1. But then you can write $T + S$ as T composed with $Id + T^{-1}S$, which mean $T + S$ is invertible.

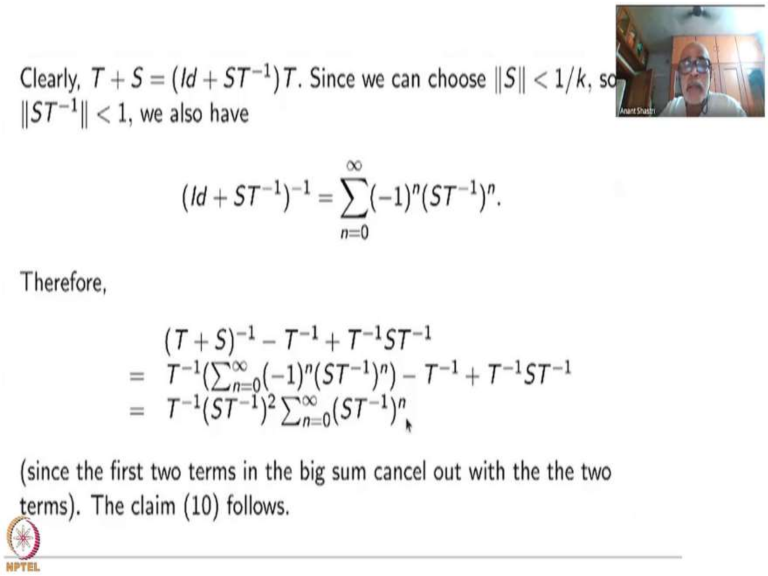
So, $T + S$ is an element of A. So, these are the points inside the open ball. Every element in this open ball looks like $T + S$ where $\|S\| < 1/k$, that is the ball. So, the whole ball is contained inside A, that is all we have proved that A is open. Now, we want to show the differentiability and the formula for the derivative. Fix a T. I want to show that η is differentiable at T.

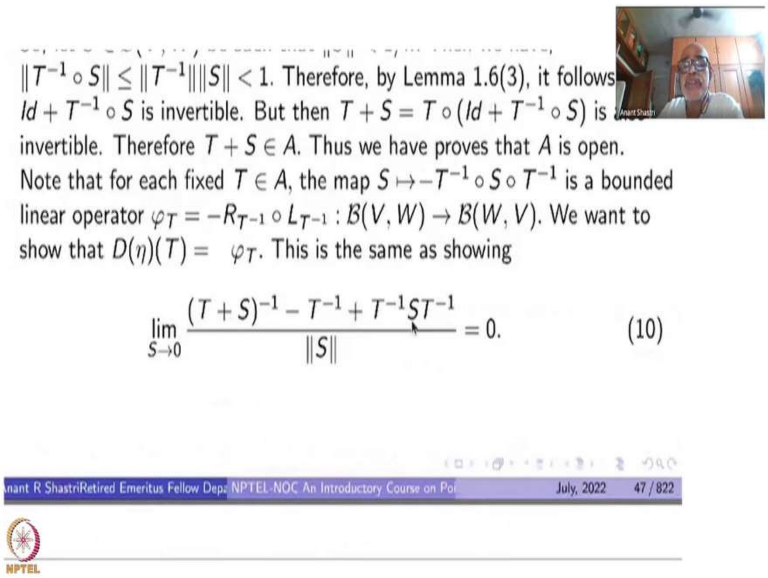
Clearly, the map S going to $T^{-1}ST^{-1}$ is a bounded linear operator. Why? S is the variable here I am taking the right composition and then the left composition by some other bounded linear operators viz., T^{-1} on both sides now. So, we have seen that composing left or right is again a bounded linear transformation. What are they? actually $L_{T^{-1}}$ and $R_{T^{-1}}$.

So, this map S going to $T^{-1}ST^{-1}$ is nothing but I have a short notation is $\phi(T) = -R_{T^{-1}}L_{T^{-1}}$, that is its minus sign coming here you see in the statement is minus sign. So, $\phi(T)(S) = -T^{-1}ST^{-1}$.

This is a linear map from $\mathcal{B}(V, W)$ to $\mathcal{B}(W, V)$. We want to show that it is the derivative of η at T. This is the same thing we are showing that the limit of $\eta(T+S) - \eta(T) - \phi(T)(S)/\|S\|$ is zero. The numerator is $(T+S)^{-1} - T^{-1} + T^{-1}ST^{-1}$.

(Refer Slide Time: 33:44)





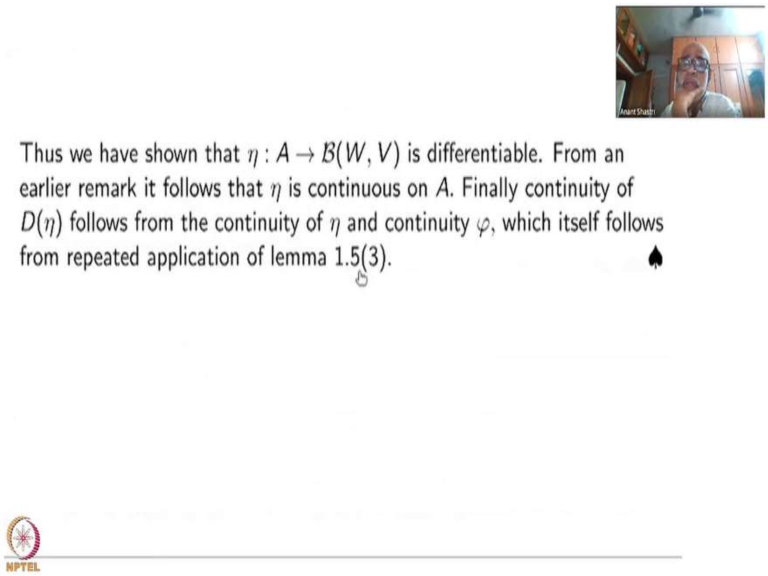
So, first of all $T + S$ can be written as $(Id + ST^{-1})T$. (I have stopped writing composition and using the simplified multiplicative notation: ST directly for S composite T etc. So, this is just for convenience everywhere composite I have already stopped writing compositions here. But now when I am doing computation I am just using multiplicative notation. You must understand that these are compositions that is all.

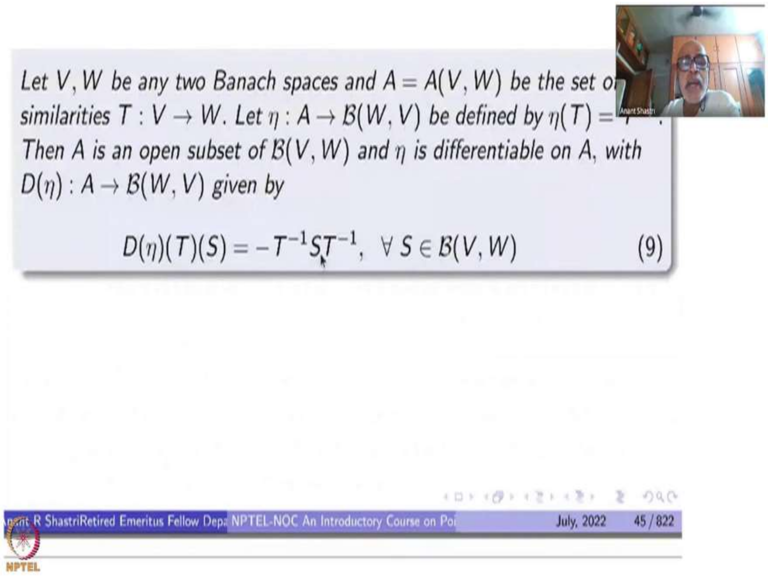
So, $T + S = (Id + ST^{-1})T$. So, since we can choose $\|S\| < 1/k$, while taking the limit. Then $\|ST^{-1}\| < 1$ and hence we have this $Id + ST^{-1}$ can be inverted $Id + ST^{-1}$ will be nothing but the summation from 0 to infinity of $(-1)^n(ST^{-1})^n$.

From the formula above $T + S$, we get $(T + S)^{-1}$ is equal to T^-1 composite with the above geometric series. So, the first and the second term cancel out, we can pull out $(ST^{-1})^2$. So, the numerator becomes $T^{-1}(ST^{-1})^2$ composited with the alternate summation from 0 to infinity of ST^{-1} powers.

Now what we have to do? We have to divide it by $\|S\|$ and then show that the limit as $\|S\|$ tends to zero is zero. Instead, it is enough to show that the entire norm tends to zero. So, when you take the norm of the numerator, it will be less than or equal to square of $\|S\|$ times some constant. One power cancels out with the denominator. Therefore, as $\|S\|$ tends to 0, the norm of the entire thing also tends to zero.

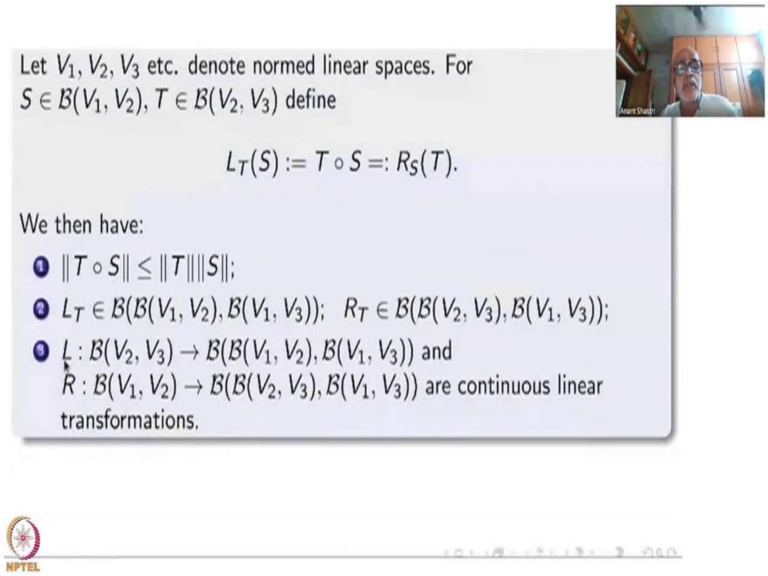
(Refer Slide Time: 38:18)

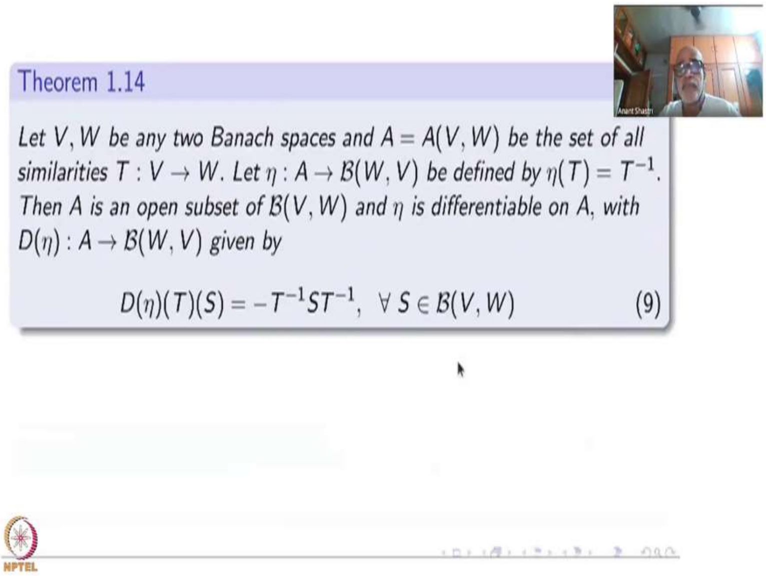




So, we have shown that η is differentiable. From an earlier remark it is continuous. I have already told you I am repeating it. η is continuous. But how to show that $D(\eta)$ is continuous. We use the formula for $D(\eta)$. Look at the formula. Formula says that $D(\eta)(T)$ is $-L_{T^{-1}}R_{T^{-1}}$.

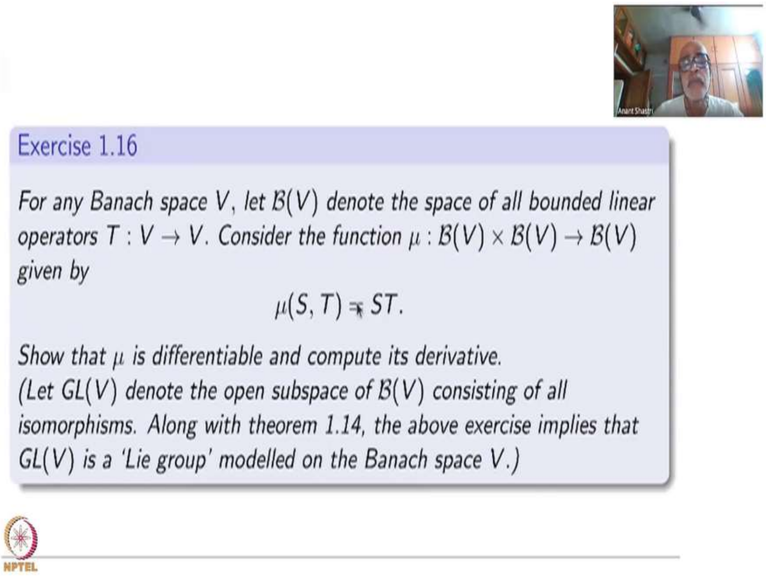
(Refer Slide Time: 39:49)





From part I, we know that the left composition L and right composition R are continuous. Since η is continuous, it follows that $T \mapsto L_{T^{-1}}$ and $T \mapsto R_{T^{-1}}$ are continuous. Therefore their composite is also continuous. Therefore $D(\eta)$ is continuous.

(Refer Slide Time: 40:24)



So, here is an exercise I am not going to use it in the course but this is something which is very very useful for people who want to study Lie groups and so on. So, that is precisely what it is take any Banach space that $\mathcal{B}(V)$ will denote the space of all bounded linear operators from V to V.

Obviously you can compose two elements of $\mathcal{B}(V)$. This composition is differentiable. So if you take the set of invertible elements, then it is something more that a topological group namely what is called a Lie group modelled on the Banach space $\mathcal{B}(V)$. So, if you want to study Banach Lie groups this is the starting point.

So, I am giving you this an exercise. Show that mu is differentiable compute its derivative. Let $GL(V)$ denote this open subspace of $\mathcal{B}(V)$ consisting of all isomorphisms. They form a group Along with our theorem above, this exercise implies that $GL(V)$ is a Lie group modeled on Banach space. So, that is just an exercise. The only thing is you will see how to differentiate this one, where the derivative taking values if you figure it out then you will know what the derivative. So, thank you very much this is all for today.