STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

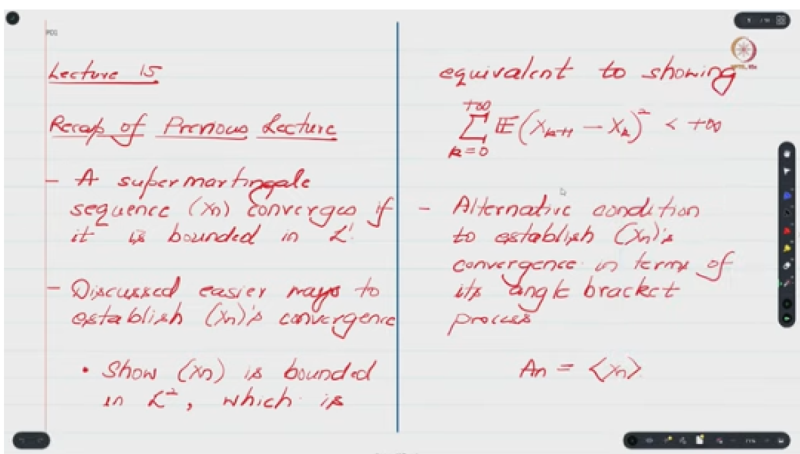
Indian Institute of Science

Week 3

Lecture 15

Revisiting the Martingale Convergence Theorem via Stopped Processes

Hello and Namaste everyone. Welcome to lecture 15 of this NPTEL course on Stochastic Approximation. Let us do a quick recap of what we did in the previous few lectures. So, in the last but one lecture, we discussed conditions for the convergence of a supermartingale, and of course, that implies the convergence of a martingale as well. We said that a supermartingale sequence converges if it is bounded in L1, right? Now, this condition is often challenging to verify, right?



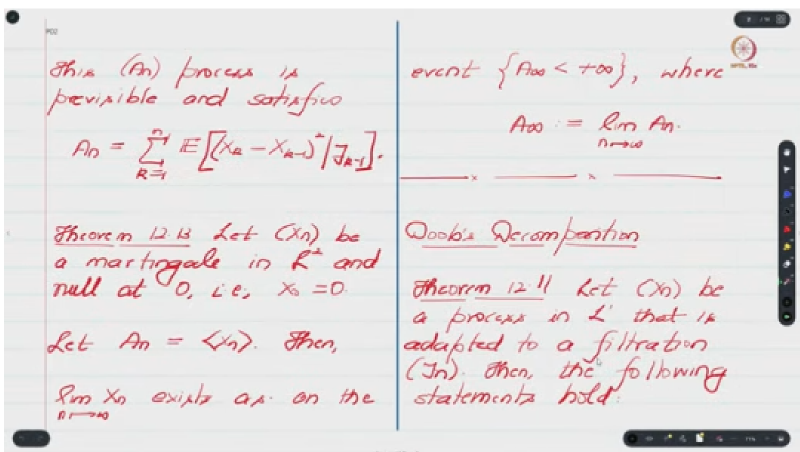
So, we then started discussing easier ways to establish Xn's convergence. One of the ways we discussed was to show that Xn is bounded in L2 instead of L1, right? So, we tried showing it is bounded in L2, and then we said that showing a sequence in L2 is bounded in L2 is equivalent to showing that the sum of the expected value of the square of the difference between successive terms—okay, that infinite sum—if it is less than infinity, then that is a good condition or sufficient condition to establish Xn's convergence. In fact, we will be using this in the discussion later on, right? And we said that this condition by itself is not going to be enough or will put restrictive assumptions during our convergence analysis of stochastic approximation.

And hence, we asked: can we establish a weaker notion or a weaker sufficient condition under which we can establish Xn's convergence? And towards that, we defined this angle bracket process. So, it is referred to as the angle bracket process because of this notation over here.

$$\sum_{k=0}^{\infty} E\left(X_{k+1} - X_k\right)^2 < +\infty$$

$$A_n = < X_n >$$

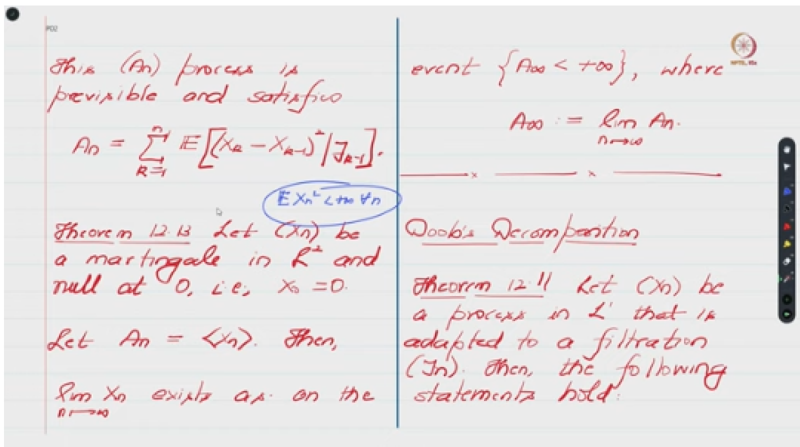
And formally, given a process Xn, the angle bracket process is defined in the following way: it is a sum from k equals 1 to n, and you take the conditional expectation conditional expectation of the square of these successive differences, right? And you can see that since k can at most go to n, the last term will be conditioned on Fn minus 1.



Right, which makes a pre-visible process. Is that okay? So, these are things that we discussed in the previous class, and then we stated one result, which we will be proving today, right.

$$A_n = \sum_{k=1}^{n} E\left[\left(X_k - X_{k-1}\right)^2 |_{F_{k-1}}\right]$$

So, the result went something like this: it said that suppose you have a martingale. So, notice that this result is tailor-made for martingales. Suppose you have a martingale in L2. So, notice that I am not saying it is bounded in L2, but only saying that it is in L2, which means that we only have that the expected value of Xn squared is less than infinity for all n. So, this is the condition; I have not put a sup over here. If I had put a sup over here, that would have been bounded in L2.



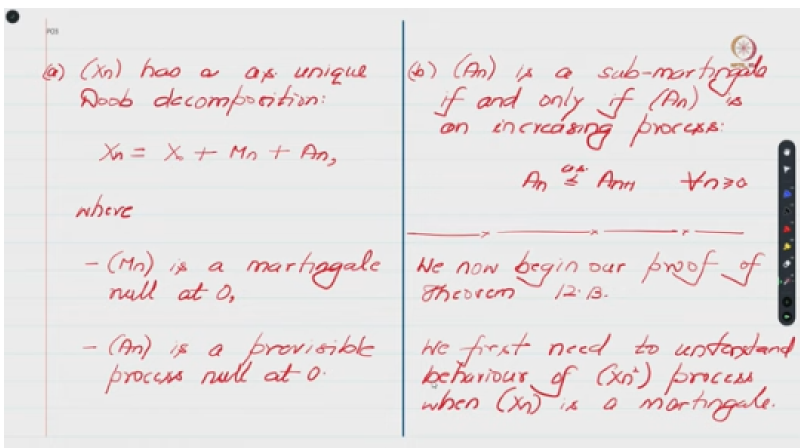
So, that is not what we are presuming here. Right. And, for simplicity, we will presume that the given sequence is null at 0, but this is not a very restrictive assumption. We can take the given sequence, which is not null at 0, and subtract x0 from it to make it null at 0. So, it is not a big deal, but for simplicity, we will presume that xn is null at 0, which means that x0 is 0, and let An be the angle bracket process, okay? So, what this result says is that

your limit Xn exists almost surely on the event A infinity less than infinity. So, An is this sequence, and since Xn is a martingale, one can show that An is an increasing sequence. Because An is an increasing sequence, A infinity exists, which can be defined in the following way. A infinity is a random variable. So, this will be an event, and the result says that on this event, almost surely, the limit exists. Is that okay?

$$A_{\infty} := A_n$$

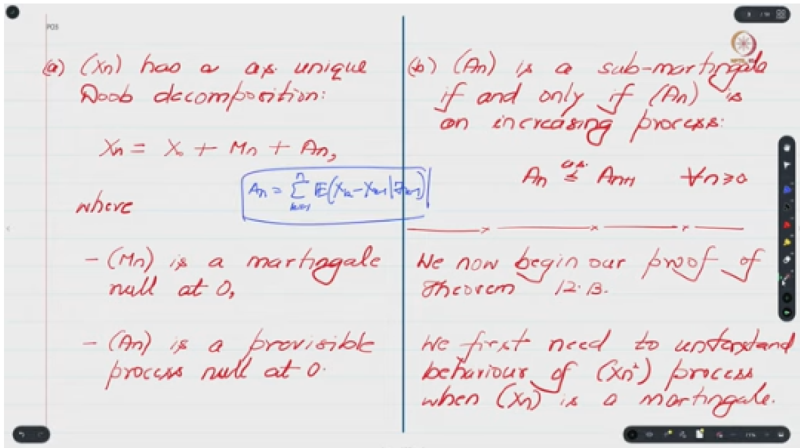
So, you can already see that the sufficient condition over here is very, very weak, right. It does not talk about boundedness in L2 and so on and so forth. Instead, it only says that you define this angle bracket process, and whenever the event holds along all those sample paths, your limit Xn will exist, right. So, it is possible that this event has a very small probability, right, in which case you know this will exist. All we are saying is that, along the set of sample points in this set, except for a subset of measure 0, your limit Xn will exist.

So, this result does not say how large this probability is. So, that will need to be established externally, right, and then We said, you know, let us try to prove this theorem, and towards that, we established or started discussing a few concepts, which I will quickly recall, right. So, one of the concepts that we discussed was that of Doob's decomposition, okay. So, Doob's decomposition broadly says that let Xn be a process in L1, right.

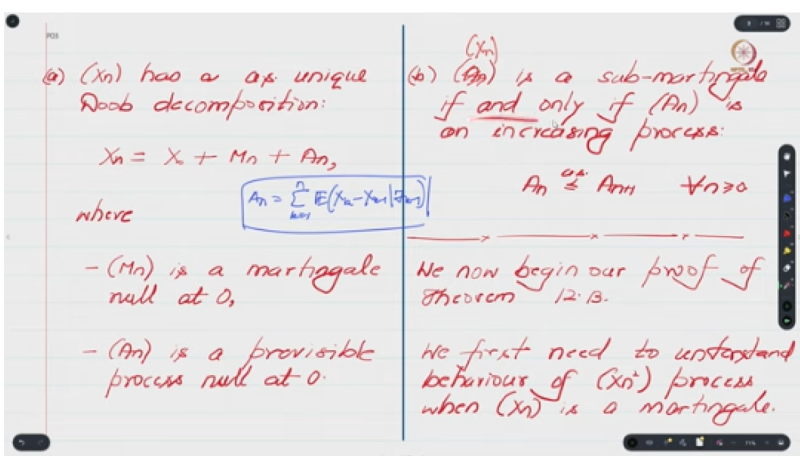


And which is adapted to some filtration FN, then the following statements hold. In particular, XN has a Doob decomposition; that is, XN can be expressed in this way, where your MN is a martingale sequence which is null at 0, and AN is a previsible process which is also null at 0. So, the thing to note here is that for this Doob decomposition to exist, Xn did not have any special property. It is just a process that is in L1 and adapted to this filtration Fn, and as soon as you have such a process, we will have this very nice decomposition. Right, and you know, AN is some previsible process. So, here, please do not misinterpret this AN as the angle bracket process. Here, AN is basically your summation k equals 1 to n expected value of Xk minus Xk minus 1 given

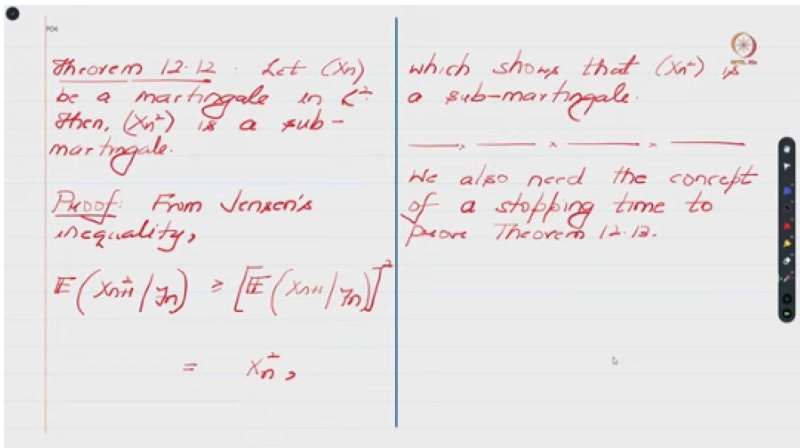
Fk minus 1, okay. So, this is how your AN process was defined. At least, AN will equal this almost surely.



And then we had this additional statement which said that this pre-visible process right is a—sorry, I should make a correction here—this should not be 'an'; rather, it should be 'Xn'. So, the statement said that Xn is a sub-martingale if and only if the pre-visible process is an increasing process, that is, An is less than or equal to A subscript {n plus 1} almost surely. So, this is if and only if Xn is a sub-martingale, then this will hold, and the converse statement is that if your pre-visible process turns out to be increasing, then your Xn will also be a sub-martingale. So, now this is a recap of what we did in the previous class. We will now formally look at the proof of this result that we mentioned before, which is this result. And, you know, this result is a bit challenging, and we will need to introduce several interesting concepts, right.



So, the first of these interesting concepts is the following, right. So, when we have been given a martingale sequence Xn, what can we say about the Xn-squared process, OK? So, this is the question that we need to answer, and the answer is given in this result. Again, I would like to emphasize that this number over here is taken from the 'Probability with Martingales' textbook by David Williams.

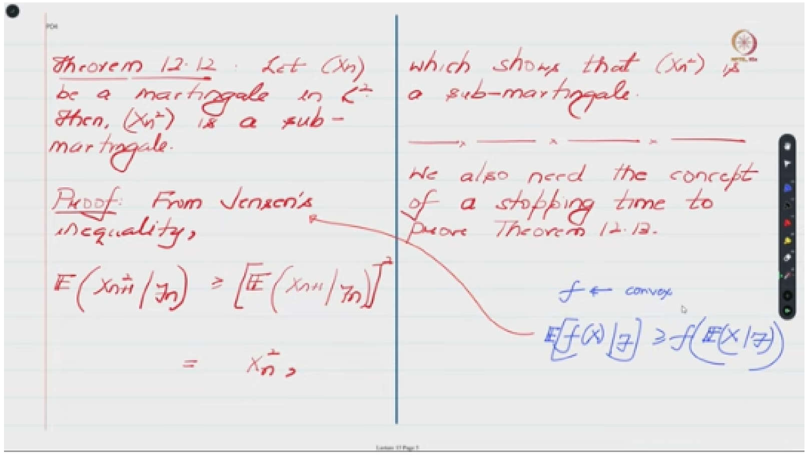


So, if you need more details, you can look it up, right. So, the result goes something as follows. So, let Xn be a martingale in L^2. Right, then Xn-squared is actually a sub-martingale, right. So, Xn is a martingale, and if you look at the square of these individual random variables and consider the resulting process, that resulting process will be a sub-martingale.

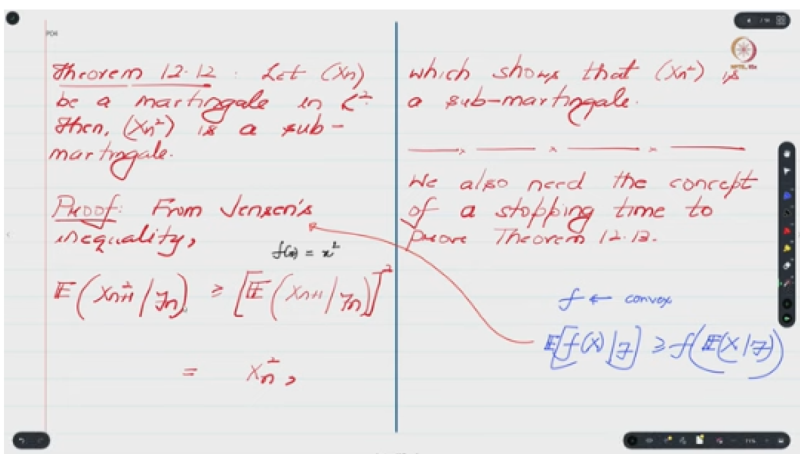
So, let us quickly prove that, okay. So, from Jensen's inequality in the context of conditional expectations, we know that the expected value of f of x given f, right? So, if your f is convex. So, if you look at the conditional expectation of this, then this will be greater than or equal to f of the expected value of x given f, okay.

$$E\left[f(x)|_F\right] \geq f\left(E\left(X|_F\right)\right)$$

So, this was Jensen's inequality for the case of conditional expectations. So, invoking that over here, we get something like this. So, the question is, what is the f that we are using over here? Well, it is straightforward to guess that the f we are looking at is f of little x equals x squared. Now, it is obvious to check that this function is convex in nature.



And hence, this inequality over here applies. So, you can see that you have xn squared here. So, that is f of xn plus 1. So, if you take the conditional expectation of that, that function goes out, and you are left with the conditional expectation of xn plus 1 given fn. Now, since your given xn sequence is a martingale, this expression will turn out to be simply xn, right?

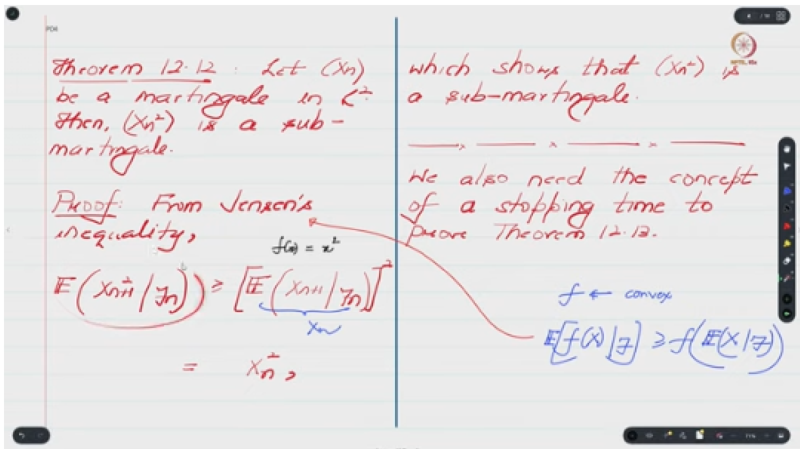


And then you have a square over here. So, together we end up with xn squared. Square, is that OK?

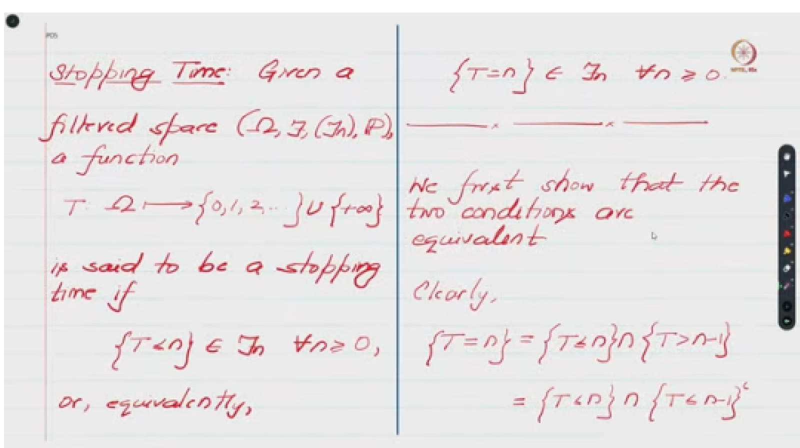
$$E\left(X_{n+1}^2 |_{F_n}\right) \geq \left[E\left(X_{n+1} |_{F_n}\right)\right]^2$$

$$= X_n^2$$

So, what we have managed to show is that the conditional expectation of Xn plus 1 squared given Fn, right, is greater than or equal to Xn squared, which immediately tells us that this Xn squared is a submartingale. Is that OK? So, now, we need one more important concept in order to, you know, have the requisite material to prove this Theorem 12.13 that I stated before, and that concept is that of stopping time, right.



So, in the context of martingales, this stopping time actually plays a very, very important and useful role, and, you know, it is, you know, like, you know, understanding this concept very well will help you, you know, come up with very, very innovative ideas while proving the convergence of martingales or supermartingales. Is that OK? Alright, so the question is, what is a stopping time? So, you know, to define a stopping time, we begin with a filtered space.



So, what is a filtered space? You have the sample space, you have some sigma algebra, and then you have a sequence of sigma algebras which satisfy the property that Fn is a subset of Fn plus 1. So, this property needs to be satisfied, right, and each of them, of course, is a subset of F, and then we have some probability measure P, right, and on this we have a random variable which is defined, right. So, a random variable is a function which goes from this sample space omega, and this time we allow P to take values. either between, you know, either any integer which is larger than or equal to 0, and we will also allow t to take the value plus infinity.

So, either it takes a non-negative integer or the value plus infinity, right? So, you have such a function, and we will refer to this function as the stopping time if it satisfies this condition. So, let us understand this condition. So, t less than or equal to n is So, which basically means you identify all those omega in capital omega such that T of omega is less than n. So, you consider this collection of omega.

So, this will form a set, and this set should belong to Fn for all n greater than or equal to 0. So, notice that the dependence of n comes in over here. And we will now show that equivalently this condition is similar to saying that t equals n. This event belongs to fn for all n greater than 0.

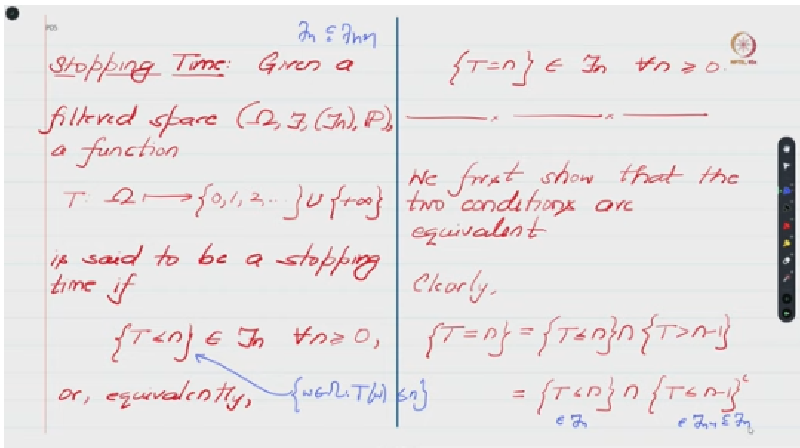
$$T: \Omega \mapsto \{0, 1, 2, ...\} \cup \{+\infty\}$$

$$\{T \leq n\} \in F_n \ \forall n \geq 0$$

$$\{T = n\} \in F_n \ \forall n \geq 0$$

So, I want to give an intuitive explanation of what this means. But before we do that, let us quickly see why this condition and this condition should be the same.

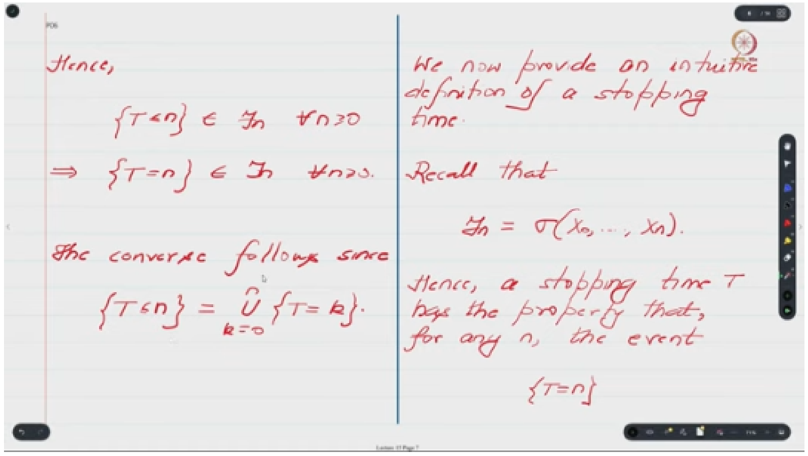
So, first let us presume that this is true, and using this, we will show that this statement holds. And I hope you agree that t equals n is equal to the event t less than n intersection t is strictly bigger than n minus 1. So, I am just carefully writing the event t equals n in this following way: that it should be less than or equal to n, and it should be strictly bigger than n minus 1, and only then t will equal n. Now, this event over here can be written as a complement of t less than or equal to n minus 1. So, now one can ask what is the utility of expressing t equals n in this fashion. So, now we have been told that this belongs to fn, and similarly this belongs to fn minus 1, and since this is a subset of fn, both these events belong to fn. And since fn is a sigma field, if two events belong to fn, then their intersection (finite intersection) will also belong to fn, and from this one can conclude that t equals n also belongs to fn.

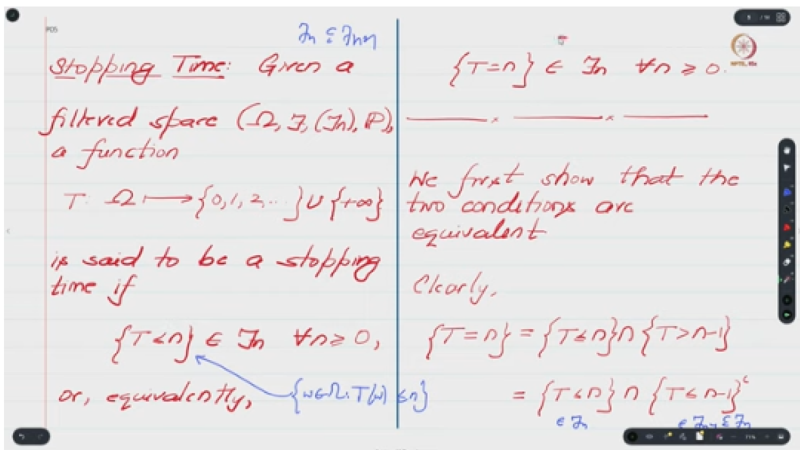


$$\{T = n\} = \{T \le n\} \cap \{T > n - 1\}$$

$$= \{T \le n\} \cap \{T \le n - 1\}^c$$

And one can similarly show the converse by making use of the fact that the event t less than or equal to n can be expressed as this finite union of k equals 0 to n, t equals k. Is that okay? So, it is very easy to see why this equality holds, right? And we have now been told that we are in the converse statement. So, we have now been told that each of these belongs to FK, which is a subset of FN, and hence every event here belongs to FN. This is a finite union of events in FN, and hence this should also belong to FN. So, what we have managed to show is that these two definitions of stopping times are equivalent. So, again, let me repeat.





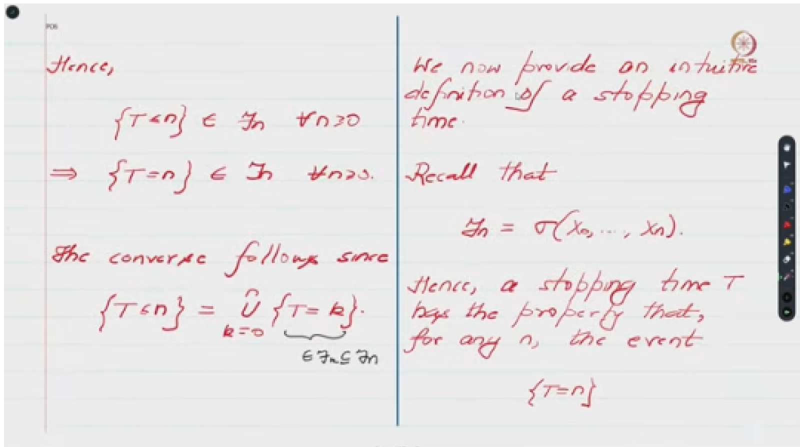
So, a random variable t, which takes values in the set of non-negative integers or plus infinity, right? You know, is said to be a stopping time if either t less than or equal to n belongs to Fn for all n greater than 0 or equivalently t equals n belongs to Fn for all n greater than or equal to 0. So, we have now verified the equivalence between these two conditions.

$$\{T \le n\} \in F_n \quad \forall\ n \ge 0$$

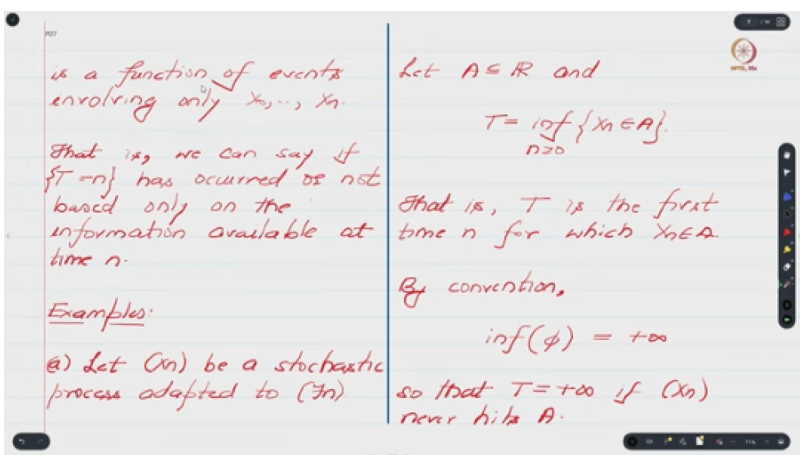
$$\{T = n\} \in F_n \quad \forall\ n \ge 0$$

$$\{T \le n\} = \cup_{k=0}^{n} \{T = k\}$$

So, now what we will do is we will look at an intuitive description of a stopping time. So, recall that typically we define Fn to be the sigma field that is generated by X0 to Xn. This could also be some given filtration, which need not be this, but it could be either different or more elaborate than this. But this is, you know, the typical sigma field that we consider, right?



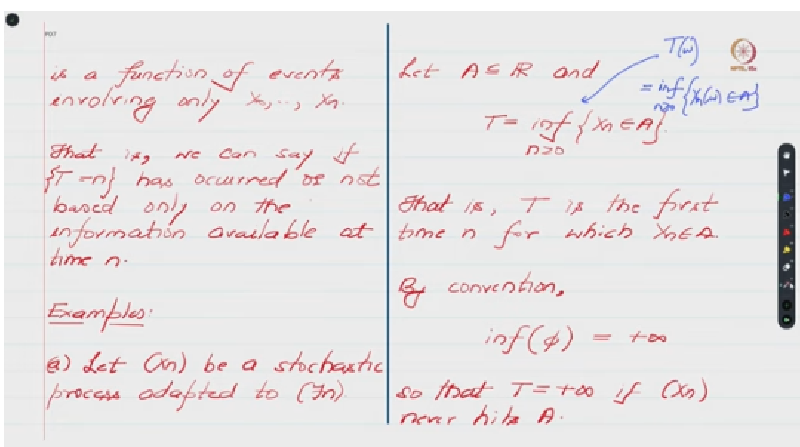
So, in this context. When we say t has the property that t equals n belongs to Fn for all n, it is equivalent to saying that this event over here, t equals n, is in some sense expressible in terms of events that only involve X0 to Xn. In other words, t equals n belongs to Fn is equivalent to saying that the event t equals n is expressible in terms of events involving only X0 to Xn. Is that okay? And, you know, in other words, we can also interpret the notion of a stopping time in the following way: by just looking at the information that is available at time n, which is like the values of X0 to Xn, one can conclude whether the event t equals n has occurred or not.



So, this is very important. I am not saying that one should be able to say t equals n has occurred or, you know, it has not occurred. One of the two—if you are able to say it unambiguously—then we will say that your t is a stopping time. So, let us look at some examples, and that will help clarify, you know, the details. So, let us say Xn is some stochastic process adapted to Fn. I mean, this Fn need not be the sigma-field that I introduced previously.

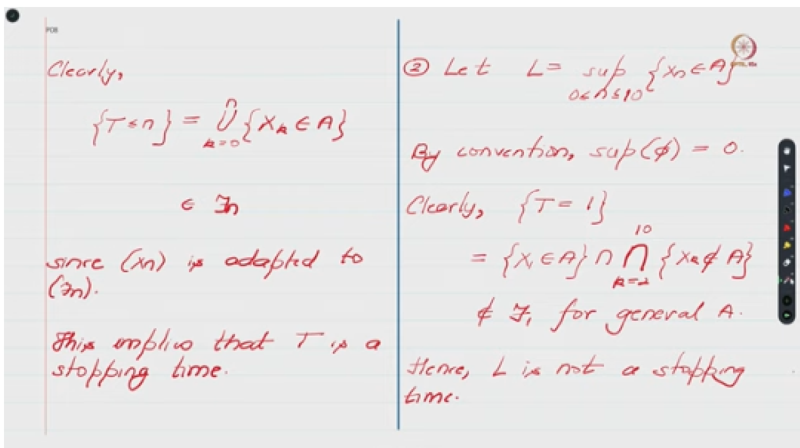
This could be some arbitrary Fn. And let us say A is some subset of R. Like, for example, this could be the set of non-negative real numbers and so on and so forth. And let T be defined to be the inf over n of the event Xn belongs to A. So, what this is saying is you are looking for the first time at which the process Xn belongs to A. That is, it is the first time Xn takes a value that lies in A. And this is also referred to as the time to hit A. So, that is what I have said over here.

T is the first time for which Xn belongs to A. Now, it is possible that, you know, along the sample path, Xn, you know, never takes a value that belongs to A, right? In that case, this set will be empty, and in that, you know, scenario, we will, you know, use this convention that inf of an empty set is plus infinity to conclude that on that sample point, you know, t equals plus infinity, OK? So, I have said that t equals the value of plus infinity if Xn never hits A. So, again, I would like to, you know, elaborate what this means a bit. So, we have to, you know, interpret this definition on a sample point-by-sample point basis, which means that T of omega equals inf n greater than or equal to 0 xn of omega belongs to A. Is that OK?



Now, the question is, you know, is this time that we have defined—the first time to hit A—is it a stopping time or not? So, you can see that for this notion of T, right, the event

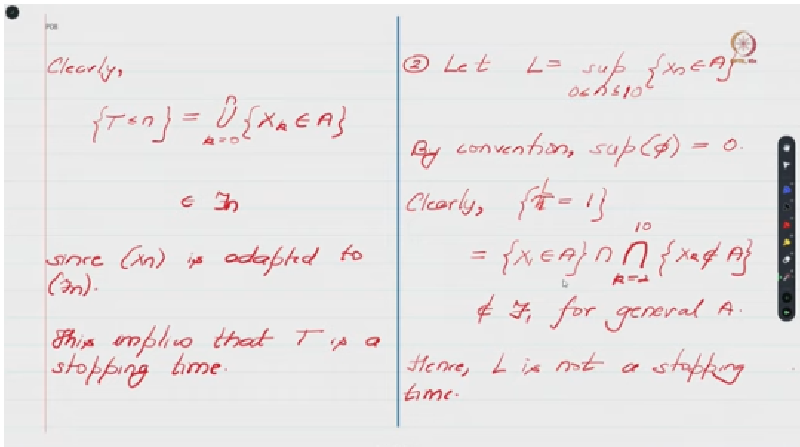
T less than or equal to n is equivalent to the union that Xk belongs to A. So, let us just understand: T less than or equal to n means that at some point less than or equal to n, your process was taking a value in A, which is exactly what is written over here, right? And because you can express T less than or equal to n in this fashion. And we know that your Xn process is adapted, and in particular, Xk belongs to A must belong to Fk, and Fk is a subset of Fn. Hence, every event here belongs to Fn, and hence their union must also belong to Fn, which can help us conclude that



So, you know, this belongs to Fn for all n, and we can finally conclude that T is a stopping time. Is that okay? So, I have now given you a very, very simple example of a stopping time. Let us now look at an example which is an example of something that is not a stopping time, okay? So, here is that example, right? So, let L be the random variable which is defined in the following way.

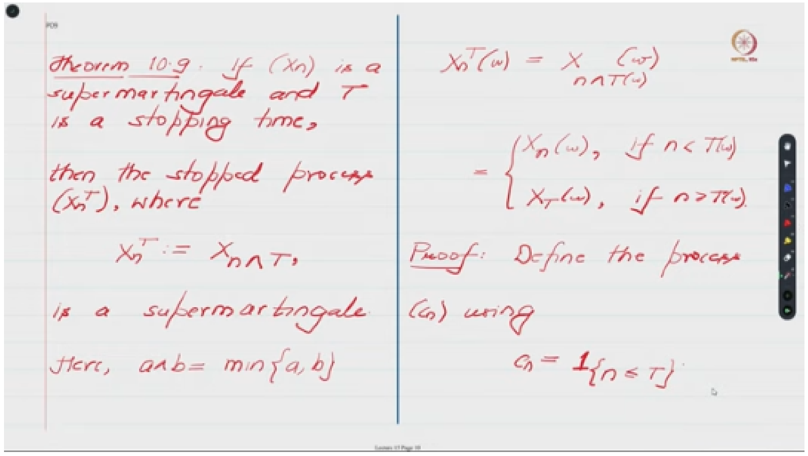
It is the supremum—and let us say the supremum of n within the indices 0 to 10—such that Xn belongs to A. So, this says, what is the last time that Xn was in A? So, that is what, you know, describes L. And again, it is possible that, you know, So, this set could be empty—in that, you know, Xn was never in A—in which case this will be empty, and we will use the convention that the supremum of an empty set is 0. So, this is the convention that we will use. So, let us try to understand whether this L is a stopping time or not, and one can check very easily that the event T equals 1 equals the fact that X1 belongs to A, right?

So, this is the supremum. Sorry, I think I made a mistake over here. This should be L, okay? So, the event L equals 1 corresponds to X1 belonging to A, and this should be the last time you should be in A, which implies that you know, for every k which is from 2 to 10, your Xk should not belong to A, right?



So, we want this also to happen and this also to happen, right? And, you know, for a general choice of A, So, you know this belongs to F1, but you know these events need not belong to F1, and hence in general this intersection need not belong to F1 for a generic choice of A, right? And since this does not belong to F1, one can conclude that L is not a stopping time. So, you know, I hope after looking at these two examples, you have some sense of what is a stopping time and what is not a stopping time.

The loose summary is that t is a stopping time. If the event t equals n, you know, you can say whether it has occurred or not based only on the information that is available from time 0 to n. All right. Very good. So, as I told you, this Theorem 12.13 is a bit challenging, and we require a lot of concepts. So, we have now understood this Doob decomposition, and then we have understood this property of Xn square when Xn is a martingale, and now we also introduced this concept of a stopping time, and now we will see the power of stopping time via this.



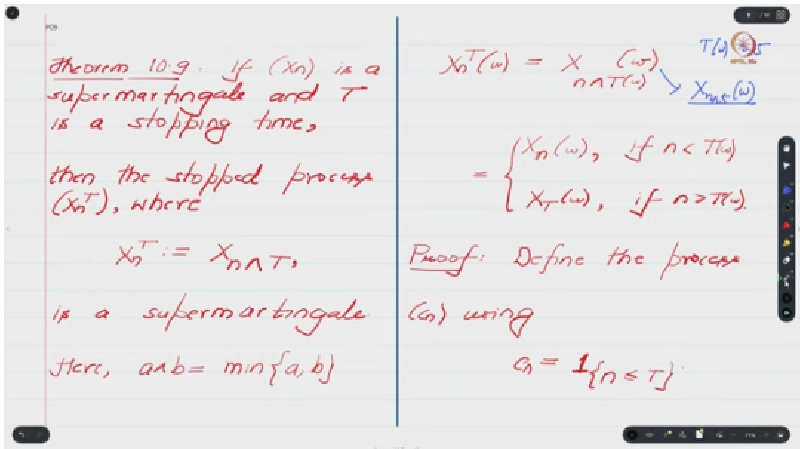
So, again, this number over here is taken from this probability with, you know, martingales textbook by David Williams, right. So, what does this result say? It says that suppose Xn is a supermartingale, right. Suppose Xn is a supermartingale and T is some stopping time, right. Then the process XnT, which is referred to as the stopped process, right, is also a supermartingale. Okay, so the result says that if you start with a supermartingale, then the stopped process is also a supermartingale. So, what do we mean by stopped process? I will now try to explain. So, XnT is defined to be X{min(n, T)}.

So, this notation over here is defined here, right. So, n is an integer, T is a random variable which, on different sample points, will take either a finite integer or plus infinity, right. So, in either case, we can look at the min of these two values, right, and this is what this notation over here means.

$$X_n^T := X_{n \wedge T}$$

So, let us spend a few seconds understanding why this process is referred to as a stopped process. So, you know, let us try to understand or try to make sense of this definition for some sample point omega.

So, for little omega, you know, you would have some T of omega, so, you know, then, you know. X of Xn min T of omega of omega, right. So, this now makes sense, right? Like, for example, if your T of omega is 5, right, this will be like Xn, sorry. So, we can understand this value for different values of n now, and one can immediately see that if n is strictly less than or equal to T of omega. Then the min will basically be n, hence this value will equal this value.



On the other hand, if n is greater than or equal to t of omega, then this value will equal this. Now, if t omega is 5, so for every n which is bigger than or equal to 5, your xn t of omega will equal x5 of omega. So, even if you keep increasing n beyond 5, 6, 7, 8, whatever you want, the value of this will stay at X5 of omega. So, you can see that for all large enough n, the value of this process will be Xt of omega. And that is why we refer to this process as a stopped process.

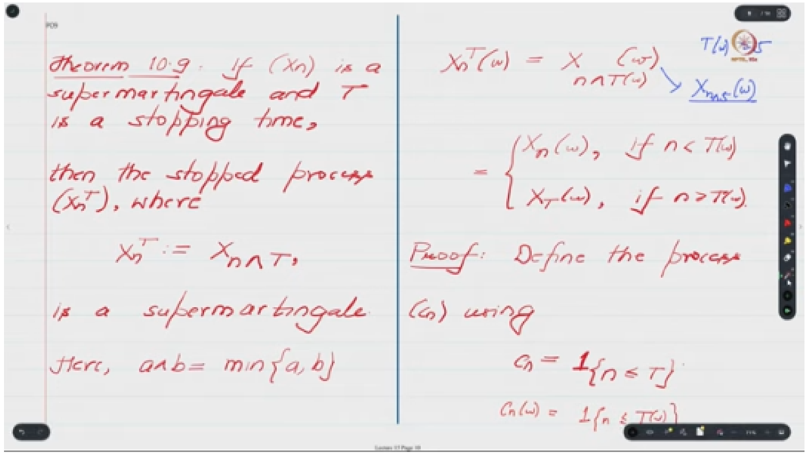
Is that okay? Alright.

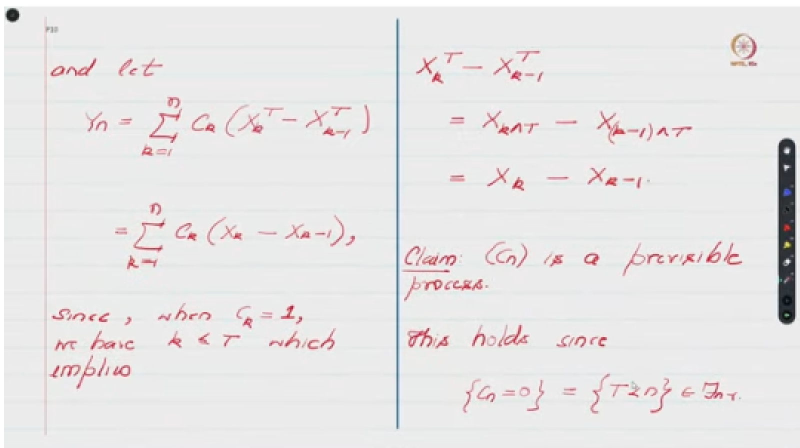
$$X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega)$$

$$= \{X_n(\omega), \; if \; n < T(\omega) \; X_T(\omega), \quad if \; n > T(\omega)$$

So, now our goal is to show that if Xn is a supermartingale, then the stopped process XnT is also a supermartingale. And this is proved in a very clever and easy fashion, and we will go over it now. So, you know the idea over here is to define a new process Cn.

given by C n equals indicator n less than or equal to t. Is this okay? So, again this is a random variable since C n of omega equals indicator n less than or equal to t of omega. So, this is what we have. Now, what we will do is we will let Yn be this process over here. So, it is like the sum of Ck times XkT minus Xk minus 1T.





And I am claiming that this sum is equal to this sum. So, the difference between this sum and this sum is that here you have T, whereas here you do not have T, which means that this sum over here is expressed in terms of your stop process, whereas this sum over here is expressed in terms of the original process, right? So, let us see why this is helpful—why this equivalence holds, right? Now, notice that whenever ck is 1, right.

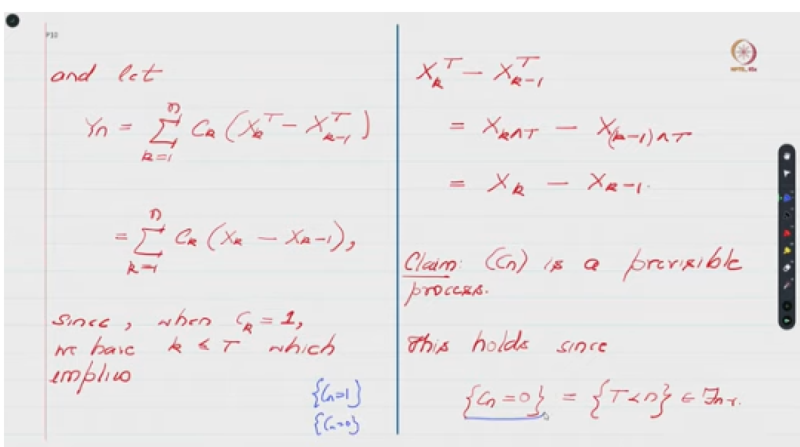
So, from the definition of cn over here, one can see that whenever ck is 1, we have that k is less than or equal to t. So, whenever c_k is 1, k is less than or equal to T. Hence, this difference must equal Xk min T and Xk minus 1 T. And since k is less than or equal to T, this expression must equal Xk. And this expression again, since k minus 1 is less than k and k is less than T, this must also equal Xk minus 1. So, in other words, this difference must equal this difference over here. Is that okay?

So, what we have managed to show is that whenever ck is 1, this expression equals this expression. Now, it is also possible that ck is 0, but if ck is 0, you know, it does not matter what I have over here because this product will be 0. Is that okay? So, hence we do not have to worry about what this value is when ck is 0. We only need to worry about what this value is when ck is 1.

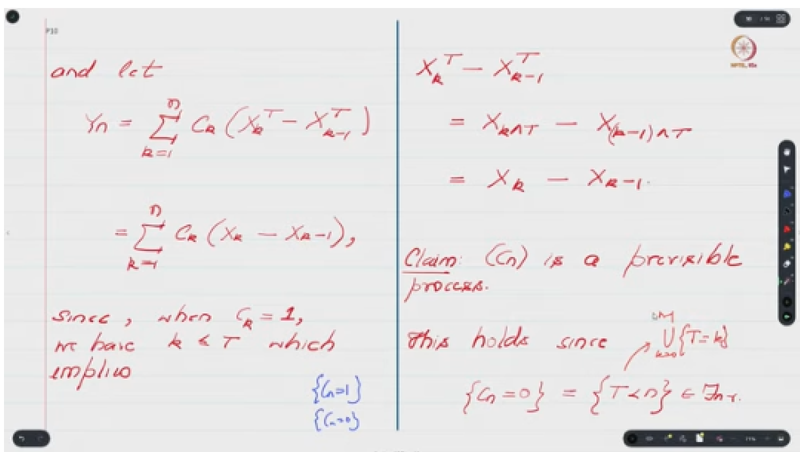
And we have shown that when CK is 1, these two things are the same. And hence, one can conclude that this sum equals this sum over here. However, once we have this relation, we can easily conclude something. First, we can see that Cn is a pre-visible process, and I will show you how that can be shown. But if Cn is a pre-visible process, then one can use one of the results that we have proved earlier to show that Yn, using this fact,

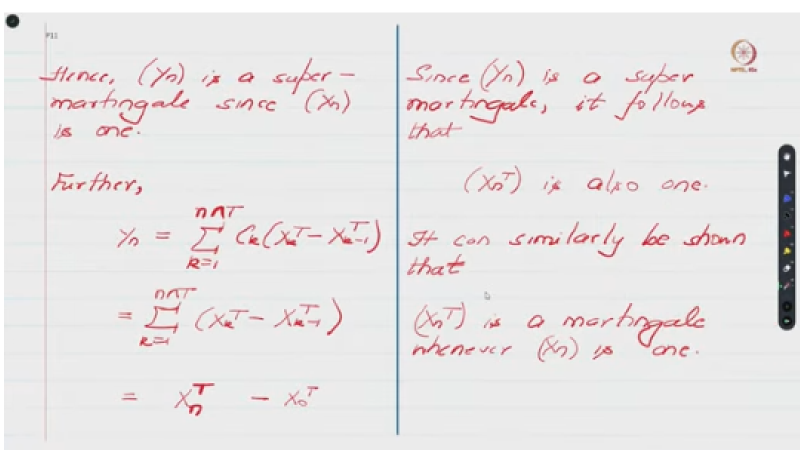
must be a supermartingale. Since Xn is a supermartingale, Yn also will be a supermartingale. So, that fact we can immediately conclude. So, let us quickly check why Cn is a pre-visible process. So, you know, for this, what we will do is we will look at the two events that can be defined in relation to Cn.

So, Cn is an indicator random variable. So, either you have Cn equals 1 or Cn equals 0, and they are complements of each other. So, let us focus on the event Cn equals 0. So, what does Cn equals 0 mean? Let us recall the definition of Cn. So, Cn equals 0 means that, you know,



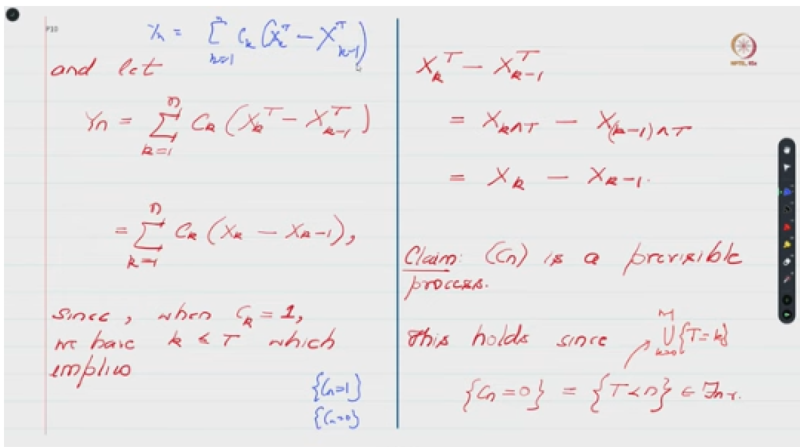
So, here we have n less than or equal to t, or t must be greater than or equal to n for Cn to be 1. So, Cn equals 0 means that t is strictly less than n. So, t is strictly less than n means that this is the union k equals 0 till n minus 1. t is k. So, this event equals this, and because this is n minus 1 over here, we know that this union must belong to Fn minus 1. So, since you know Cn equals 0 belongs to Fn minus 1, and since Fn minus 1 is a sigma-algebra, one can immediately conclude that Cn equals 1 must also belong to Fn minus 1, which in turn implies that Cn is measurable with respect to Fn minus 1, and hence Cn is a pre-visible process. So, and as I said, since Cn is a pre-visible process, by using one of the previous results we had proved, one can conclude that since Xn is a supermartingale, we immediately have that Yn is also a supermartingale.



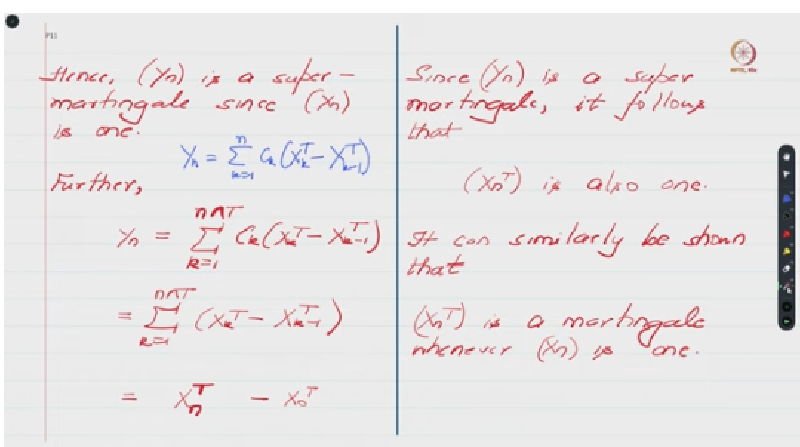


But, you know, one can now use an alternative description of Yn. So, Yn, you know, was defined in the following way. Now, as soon as Ck becomes 0, right, whatever you add over here, you know, you are only adding 0 terms over here, right. So, from that property, from that observation, one can see that Yn has this equivalent description.

Is that okay? So, let us just understand this right. So, you know, for some value, ok, let me just see if I can explain it in a better way, right. So, your Yn sequence, sorry, your Yn sequence was summation k equals 1 till n Ck Xk t minus Xk minus 1 t, is that ok. So, I should say Yn equals summation k equals 1 to n Ck Xk t minus Xk minus 1 t. Is that okay?

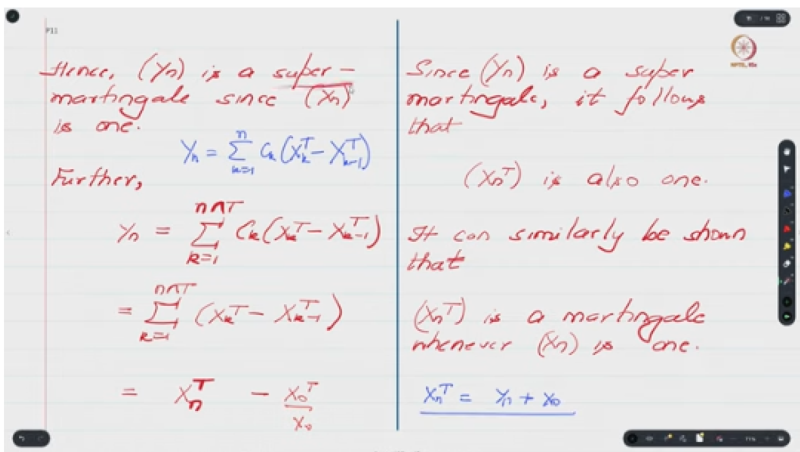


One can immediately see that for any omega, if n is larger than t of omega, then for all those values of k, right, your ck will be 0. Hence, you know, these terms will be 0. Hence, it suffices to only look at, you know, this sum over here. Now, for k which is between 1 and n min t, the value of ck is 1. Hence, we can replace this ck with 1 over here, and this expression will become this.

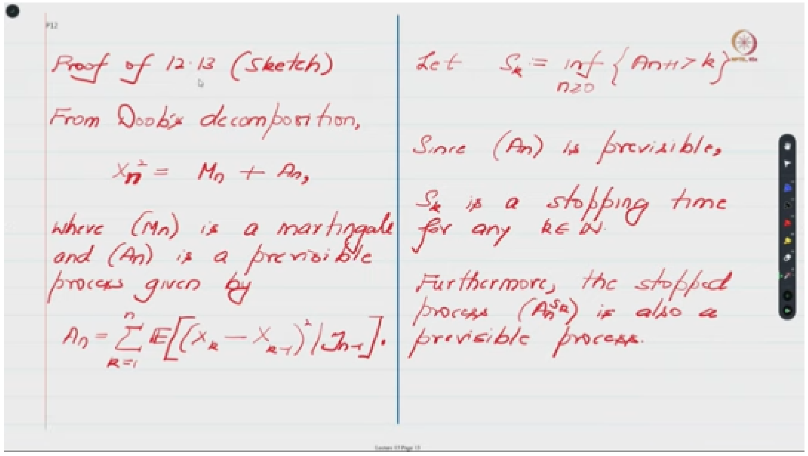


But once we write it in this way, we have a telescopic sum, and hence we will only end up with the last term in this sequence, which is x and t, and the first term over here, which is for the value of k equals 1, which is x0 t. And it is easy to see that X0 t is actually X0. So, we have Xn t minus X0. In other words, your Xn t is actually equal to Yn plus X0.

Is that okay? And since we have proven earlier that Yn is a supermartingale, it follows that XnT is also a supermartingale. And by repeating, you know, the proof argument, one can also show the following: if Xn is a martingale, not just a supermartingale, if Xn is a martingale, then your stopped process will also be a martingale. So, let me summarize quickly what we have shown. We have shown that, you know, if you start with a supermartingale and you stop it, then that stopped process will also be a supermartingale.



And similarly, if you start with a martingale, then the stopped process will also be a martingale. Is that okay? So, now, you know, I… You know, I proved Theorem 12.13 as promised. Is that okay? So, you know, I want to keep this as an exercise so that those who are listening to this video can try proving this on their own.



So, I am going to basically build upon all the results that we have seen so far, right, and prove Theorem 12.13. Is that okay? So, recall what we had to show in Theorem 12.13: we had to show that your limit Xn exists almost surely on the event A infinity strictly less than infinity. Is that okay? And we have not been told that Xn is bounded in L2. So, we have just been given some Xn process in L2, not bounded in L2, right? And we need to now show that it converges almost surely on the event A infinity less than infinity.

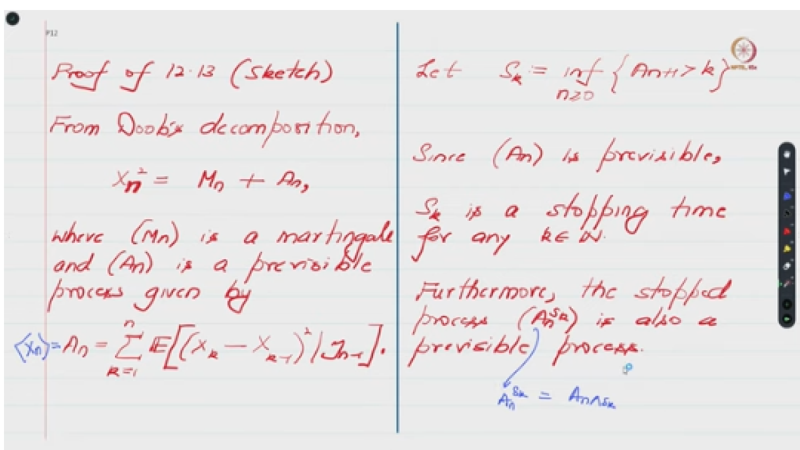
So, what we will do is we will first work with this squared process, that is Xn squared, and we know that since Xn is a martingale, this will be a submartingale. And now, since Xn squared is a submartingale and hence some process. We can invoke the Doob decomposition to conclude that Xn squared can be broken down into two Mn plus An. Is that okay?

Where Mn is a martingale and An is a previsible process. So, if this was an arbitrary process, right, then An has some structure. But since, you know, you have Xn square over here, one can show that An has this precise description, right? I encourage you to figure out why this description holds, right? So, the previsible process has this precise description, and this process, if you recall, corresponds to the angle bracket process associated with Xn. Is that okay?

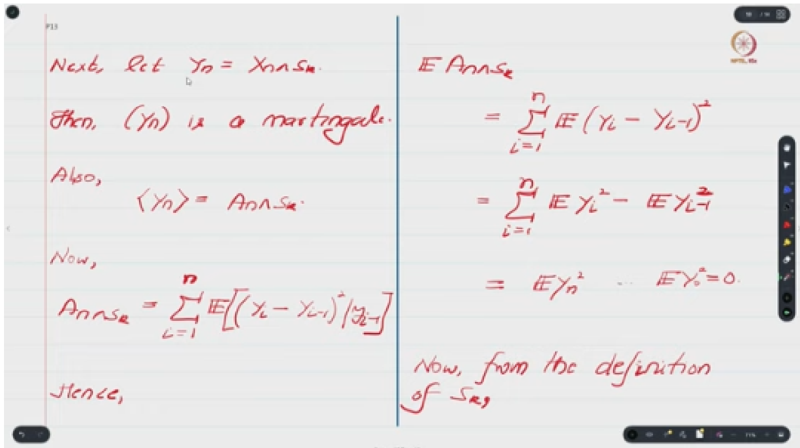
So, let me summarize: from the Doob decomposition, we know that Xn square can be expressed as Mn plus An, right? And An has this description, which corresponds to the angle bracket process corresponding to Xn. So, in the proof of 12.13, the clever thing that is used is this definition of the stopping time. This is—you know, once you read this proof and understand it, you will see that this is a very, very clever definition of the stopping time. So, what we do is let Sk be the first time where An plus 1 is strictly bigger than k. So, notice this plus 1 over here, and that actually plays a very, very crucial role throughout this proof. So, Sk is the first time where An plus 1 is strictly bigger than k, and one can check that because An is previsible, even though you have n plus 1 over here.

One can show that your Sk is actually a stopping time for any k. So, one can actually show that. And I encourage you to prove this. I am not proving this. And separately, one can show that the stopped process AnSk. So, what is AnSk?

So, AnSk. Is basically a min SK. So, this is what an SK is. So, one can show that the stop process an SK is also a pre-visible process. So, again, I encourage the listener to actually verify this fact on their own.

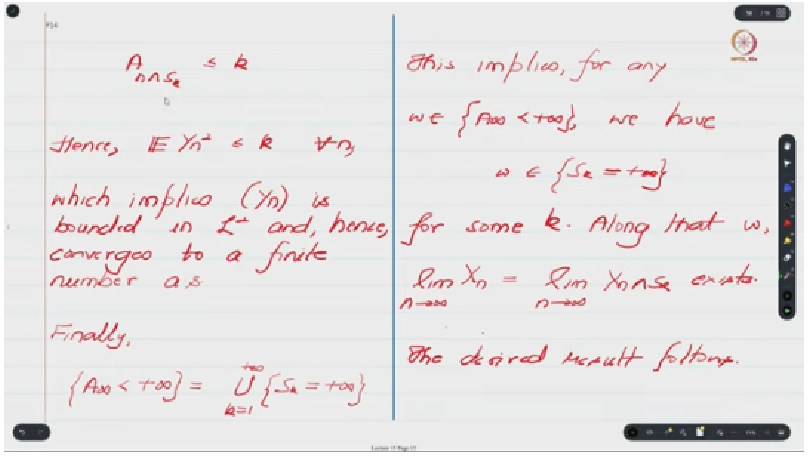


And let us now define Yn to be Xn stopped at Sk, and since Xn is a martingale and Sk is a stopping time, then one can immediately conclude that Yn is a martingale and moreover, If you look at the, you know, the angle bracket process associated with Yn, then one can show that the angle bracket process associated with Yn is the, you know, the process An which is stopped at Sk. So, let us, you know, summarize what we have done. So, An is the angle bracket process associated with Xn, and if we define Sk in this way and, you know, you stop Xn at Sk and define Yn, then Yn is a martingale, and the angle bracket process of Yn is related to the angle bracket process of Xn in that. The angle bracket process associated with Yn is the, you know, angle bracket, the stopped angle bracket process associated with Xn, right?



And that stopping has to be with Sk, is that okay? And again, you know, An, you know, because it is the angle bracket process, one can see that this has the following description over here, right? So, because it has this description, we can take expectation on both sides to conclude that the expected value of A min Sk is the sum of these expectations. So, we are invoking the linearity property and the iterated expectation property to make this conclusion. And then, you know, one can see that because your yi's are martingales, you know, this expectation actually turns out to be the difference between the squares, that is, you can check what happens with the cross term, and because we have the difference of squares, we have a telescopic sum.

And we can cancel off corresponding terms and we will end up with the expected value of Yn squared minus the expected value of Y0 squared. And if you recall in the statement of this result, we have assumed that your xn's start at 0, that is, they are null at 0. And hence, one can show that from this definition, your y0 must also be 0, and hence y0 squared must be 0, and hence the expected value of y0 squared must be 0. In other words, this expectation must equal this expectation. So, now from the definition of your Sk, recall that Sk was the smallest time such that a {n plus 1} is strictly bigger than or equal to k. And notice that here you have n.



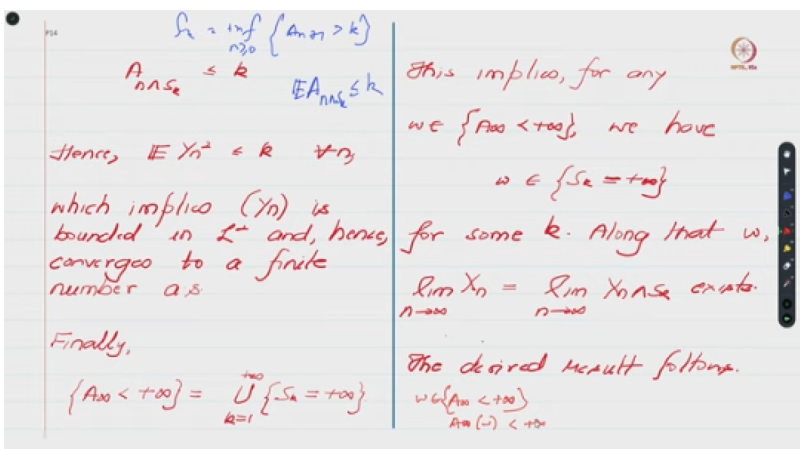
Right. And we are taking n min Sk. Right. And S_k is defined in terms of a {n plus 1}. Right.

So we can show that until you achieve this or you hit this set. So, a {n min Sk} is until you hit that set, and since you have not hit this set—in other words, your a_n values are still bigger than or equal to k—one can conclude that a {n min Sk} is actually less than or equal to k for all n. So, this is a very important conclusion; you will soon see why, and it is a very interesting conclusion which just follows from this definition of Sk and the fact that you have plus 1 over here. So, once we know that this is less than k for every n, one can conclude that your expected value of an greater than Sk will also be less than or equal to k, and since in the previous slide we showed that this expectation equals this expectation, one can conclude that the expected value of yn squared is less than or equal to k for all n. So, the left-hand side right, you know, is varying with n, and on the right-hand side you have no n, and hence one can conclude that your Yn process is bounded in L2, right. So, while the given Xn process was not, you know, told to us to be bounded in L2, one can show that the stopped process is actually bounded in L2, and now, you know, if you remember one of the first results that I discussed in today's class, we can invoke that to conclude that Yn

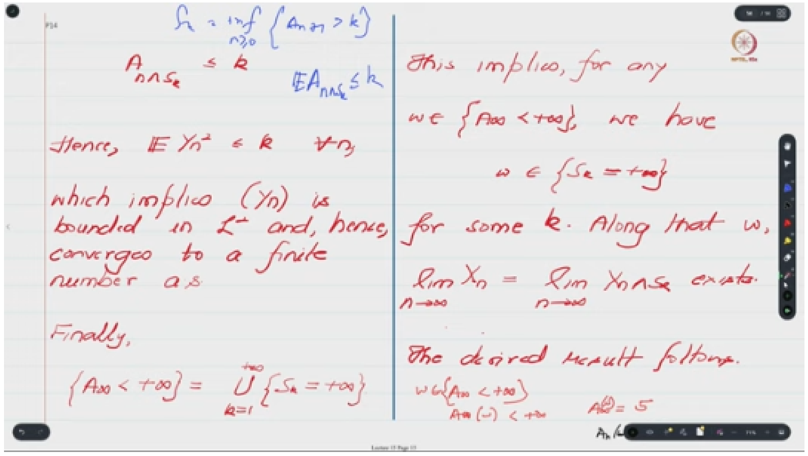
You know, it converges and converges to a finite number almost surely, right? So, what we have managed to show is we wanted to show Xn converges. So, towards that, what we did was we defined a stopped process Yn, right? And we are saying that the stopped process converges, right? And you know, we can define, you know, Yn, the stopped process with respect to Sk for any k. So, we can define that, and what we are saying is whatever Sk you pick over here, for every k, Sk is a stopping time. So, whatever Sk you pick, your Yn will be a stopped process, and the conclusion holds.

Now, let us look at this event: A infinity is less than infinity, right? So, A infinity less than infinity is, recall, the event of interest for us. In particular, we want to show that A infinity less than infinity is the event on which, almost surely, the limit Xn exists, right? And one can very easily see that, you know, from the definition of Sk that is given over here, right? From the definition of Sk that is given over here,

from the definition of Sk that is given over here, one can conclude that if A infinity is less than infinity, right. So, let us, you know, elaborate a bit. So, A infinity is less than infinity means that A infinity of, like, if omega belongs over here, then that means that A infinity of omega is less than infinity, right? So, less than infinity means it is going to take some finite value, right?

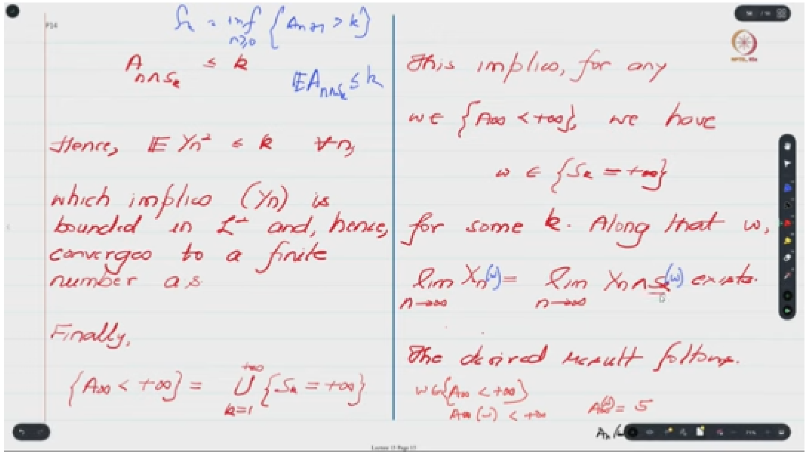


Let us say A infinity is, let us say, A infinity of omega is, let us say, 5 as an example, right? So, if it is 5 and because you know your, you know, this an process is defined with respect to your xn square, and since xn square is a submartingale, your an is monotonically increasing, right? One can conclude that your an of omega. This is less than 5 for all n, right? So, an of omega is less than 5 for all n, right? And since it is less than or equal to 5, right?



So, if you pick a k which is larger than you know, 5 or bigger than 5, then one can conclude that this event will never happen, right? On this omega, right. So, on this omega, you're a n plus 1 omega will never be bigger than k, where k is chosen to be 5, in which case this will be an empty set, right? And since this is an empty set, and you know, by convention, we have that the corresponding sk will be infinity. So, from that observation, one can see that A infinity is less than infinity will only happen if Sk equals infinity for some k. A infinity is less than or equal to infinity if Sk equals plus infinity for some k. So, if this happens. We have that if you have an omega in this event, right, then omega must belong to Sk equals infinity for some k, right.

Now, if you take such an omega, right, you know, which belongs to this, then for this omega, your limit xn of omega actually equals limit xn min Sk of omega, right? So, why does that happen? Because for this omega, your Sk of omega is actually infinity. So, n min infinity is n. Hence, this xn and this xn min infinity is going to be xn only, and hence their limits have to match.



However, we know that the stopped process limit always exists or almost surely exists, right? And from that observation, one can conclude that on such an omega, your limit Xn exists, okay. And if you think a bit about what this implies, it is that almost surely on this event, your limit Xn actually exists by using this fact over here, and we can now conclude the desired result. So, this brings me to the end of the class. Let us just quickly summarize what we have done in this class and, in particular, throughout this week. In this class, we have established a weak condition for the convergence of martingales. In particular, we defined this, you know, angle bracket process, and we said that whenever the angle bracket process is less than infinity almost surely on that event, your limit Xn exists, right? And for that, we made use of this stopping time concept and proved that result, okay.

So, this summarizes what we did in this class. Now, let me summarize what we have done throughout this week. So, throughout this week, we have studied martingales, the convergence of martingales. In particular, we made use of Doob's upcrossing lemma and so on and so forth. And finally, we have established one condition, you know, based on this angle bracket process, that corresponds to the condition under which we can establish the convergence of a martingale. So, this property, or this last theorem, 12.13, we will be using, you know, during our convergence analysis of stochastic approximation. So, next week, we will actually be discussing another topic, another fundamental topic, which is that of ordinary differential equations.

So, martingales is one topic that we have discussed so far. Then, we want to discuss ordinary differential equations, which will be a separate topic. Stochastic approximation, in some sense, is a combination of understanding ordinary differential equations and martingale theory put together. So, with that, let me stop. Thank you. Hope to see you next week.