STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

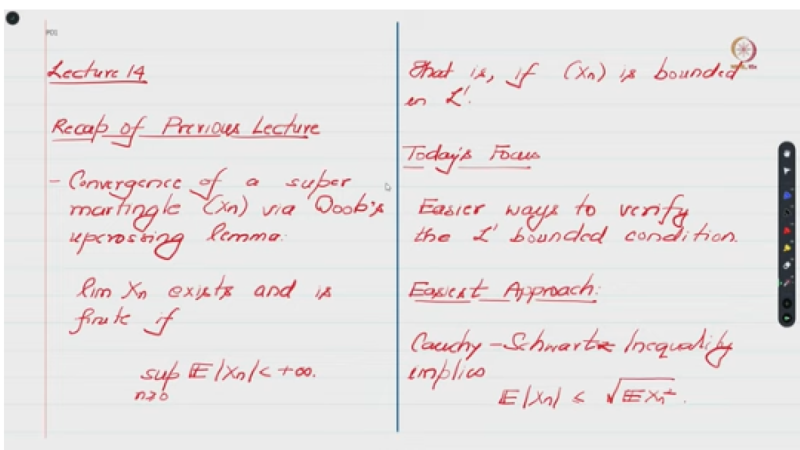
Indian Institute of Science

Week 3

Lecture 14

L^2 Martingales

Hello and Namaste, everyone. Welcome to Lecture 14 of this NPTEL course on Stochastic Approximation. So, let us first do a quick recap of what we did in the previous class, right? So, as I mentioned to you, martingales will play a very important role in our discussion on the convergence of stochastic approximation algorithms, right? And towards that, in the previous class, we discussed the convergence of



supermartingale Xn by making use of Doob's upcrossing lemma. This lemma stated that the limit of a supermartingale exists and is finite if it is bounded in L1. That is, the supremum of the expected value of the absolute value of Xn is less than infinity. Right.

$$E\left|X_n\right| <+ \infty$$

So, in today's class, what we will do is try to figure out how to ensure a condition like this or perhaps find some alternative condition, okay?

So, one of the natural things to consider is, instead of working with boundedness in L1, we will look at boundedness in L2. Let me elaborate. So, by a simple application of the Cauchy-Schwarz inequality, it is easy to see that the expected value of the absolute value of Xn is less than the square root of the expected value of Xn squared. Now, because of this inequality, if we somehow have that the sup of the expected value of Xn squared is less than infinity, then, by making use of this previous inequality, we can immediately conclude that Xn is also bounded in L1, right?

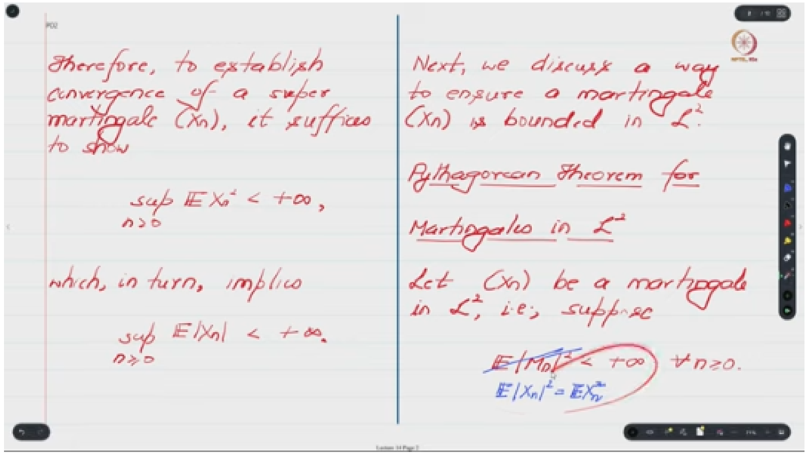
So, one way to show that your super martingale is bounded in L1, right? One can instead show that that super martingale is actually bounded in L2.

$$E\ X_n^2 < +\ \infty$$

$$E\ \left|X_n\right| < +\ \infty$$

Now, what we are going to do now is that we will first restrict our attention to martingales not super martingales. So, we will restrict our attention to martingales and discuss ways how to show that Xn is bounded in L2 and later on we will also see if this can be further relaxed. So, towards discussing a condition to show that or to check when Xn is bounded in L2, we will discuss a very nice property satisfied by martingales, right, which is referred to as the Pythagorean theorem for martingales.

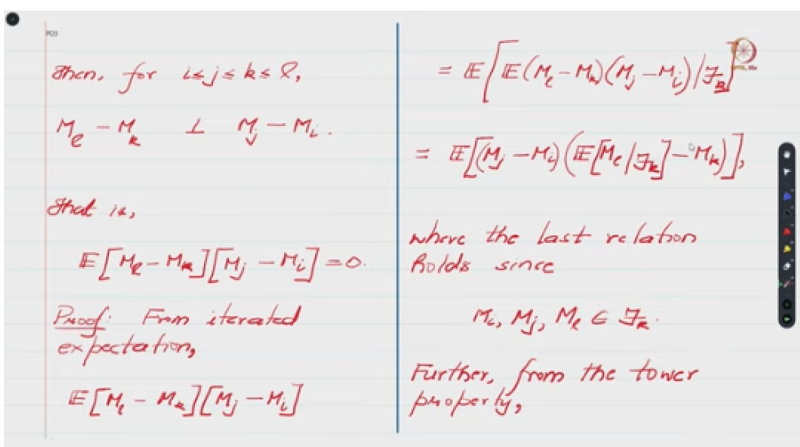
And we will consider martingales which are in L2 by that I mean you know if you have this Xn process we will you know require that each of this you know so I should actually there is a typo here. So, this should be expected value of absolute value of Xn square which is expected value of Xn square ok. So, we require that this be less than infinity. So, we require that this quantity be less than infinity. So, notice that here I am not you know a priori insisting that the supremum should be less than infinity rather I only require that for every n expected value of xn square is less than infinity and when such a thing holds we will say that xn is a martingale in L2.



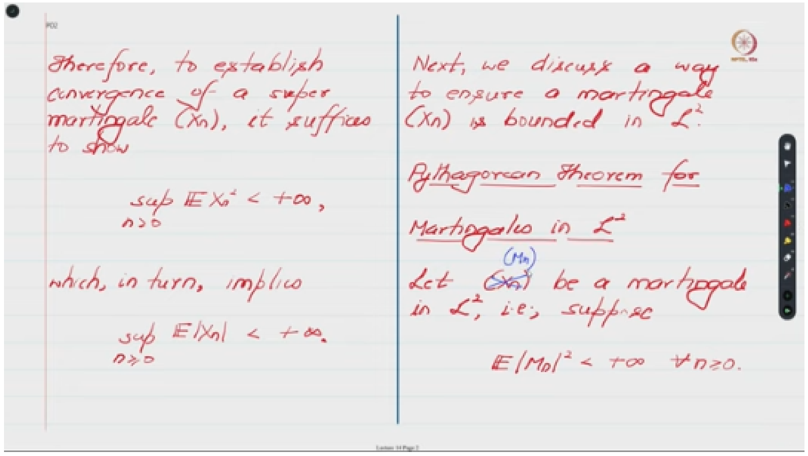
If it was uniformly bounded in L2, I would say bounded in L2. If not, I will just say it is a martingale in M2.

$$E\left|M_n\right|^2 <+ \infty \ \forall n \geq 0$$

And our goal is to come out with some condition to check when your given martingale is bounded in L2. So, the Pythagorean theorem relies on a very, very interesting and crucial fact. It says that



Okay, I think I have changed the notation over here. So, you know, in this whole discussion, I will not keep changing your m's. But, you know, maybe I can go back here and let me just change all of this to m itself. Okay, so that there is no confusion. Okay, so this I think should be m n.



So, you know, one of the crucial properties satisfied by a martingale is that if you have these four indices, which are non-decreasing—that is, you have i, j, k, and l such that i is less than j, j is less than k, and k is less than l—then the first crucial fact is that if you look at these differences. M L minus M K and M J minus M I, then the claim is that this difference is orthogonal to M J minus M I. So, this notation over here means that the two terms are orthogonal to each other. Right. And formally, this means that if you look at the expectation of the product.

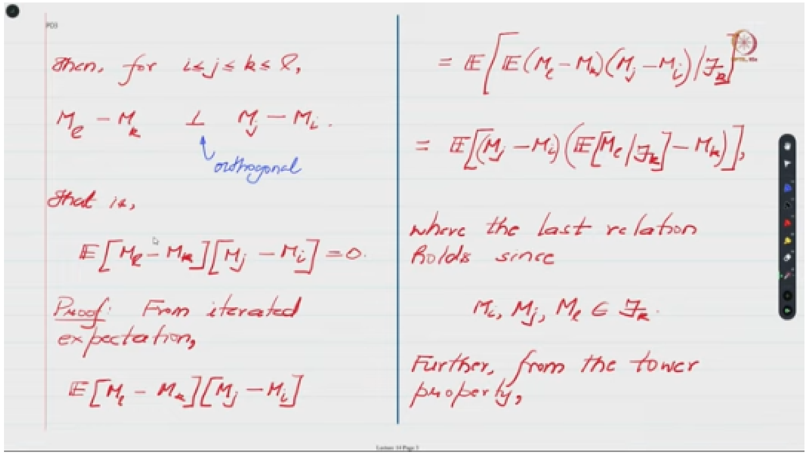
OK. So, if you look at the expectation of the product, then that expectation will be zero.

$$i \leq j \leq k \leq l$$

$$M_l - M_k \perp M_j - M_i$$

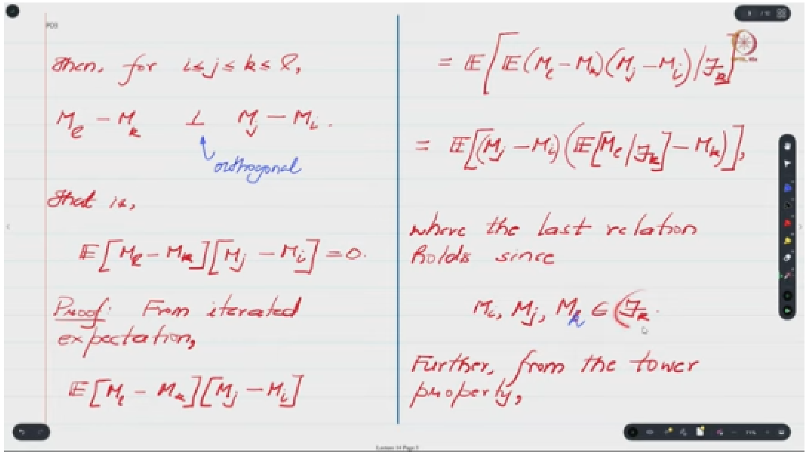
$$E\left[M_l - M_k\right]\left[M_j - M_i\right] = 0$$

So, let us quickly prove this. It is very straightforward, and it will also give us, you know, some practice with working with martingales. Right.



So, from the iterated expectation property, right, we can, you know, take an expectation and, you know, write it as the expected value of the expected value with respect to some sigma field. Is that OK? So, in this case, we take the sigma field Fk, which is the third element in this increasing sequence, right. So, we take Fk here. Now, because k is larger than j and j is larger than i, it is easy to see that mi, mj, and mk. So, I should have mk here.

So, mi, mj, and mk, all these three elements would be measurable with respect to Fk. In particular, this mj minus mi would be measurable with respect to Fk, and hence I can take it outside the expectation and notice that I have written it over here. So, then I will be left with the conditional expectation of this expression with respect to Fk. So, by invoking linearity of expectation, you know, this results in the expected value of ML given FK, and since MK is measurable with respect to FK, you know, the conditional expectation of MK with respect to FK results in only MK, right? So, this is the first observation we make.

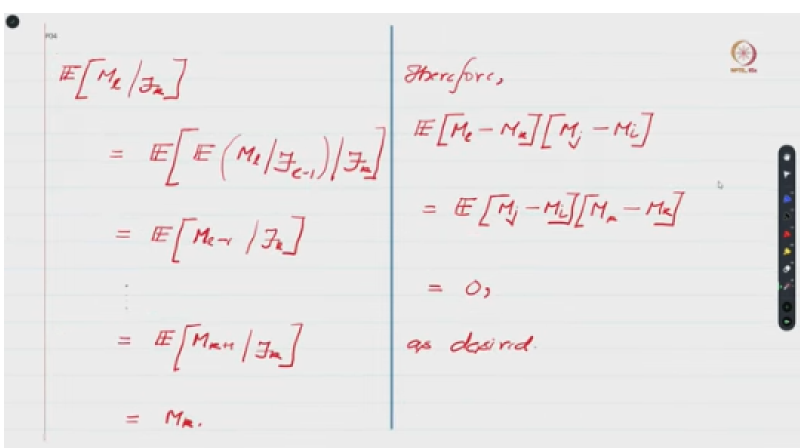


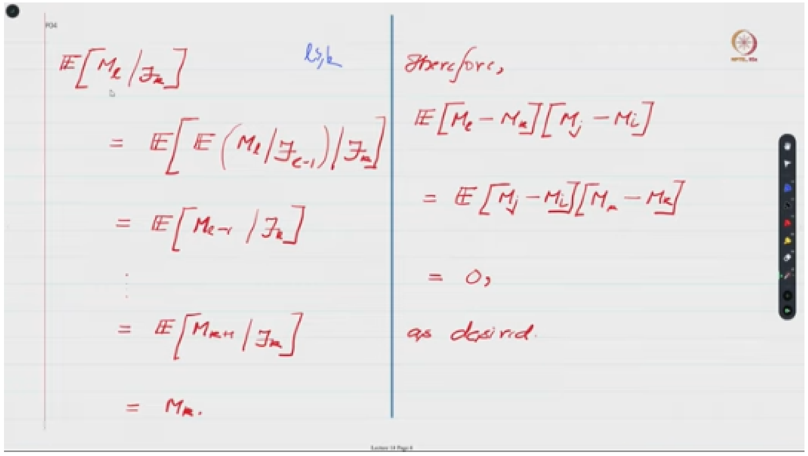
$$E\left[M_l - M_k\right]\left[M_j - M_i\right] = E\left[E\left(M_l - M_k\right)\left(M_j - M_i\right)|_{F_k}\right]$$

$$= E\left[\left(M_j - M_i\right)\left(E\left[M_l|_{F_k}\right] - M_k\right)\right]$$

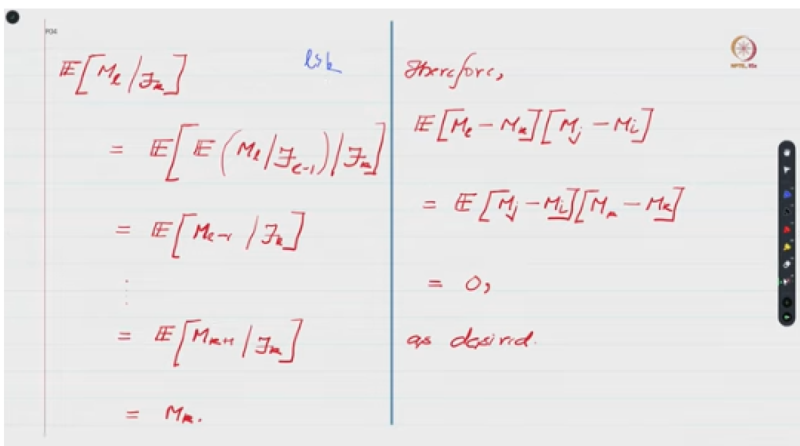
$$M_i,\ M_j,\ M_k \in F_k$$

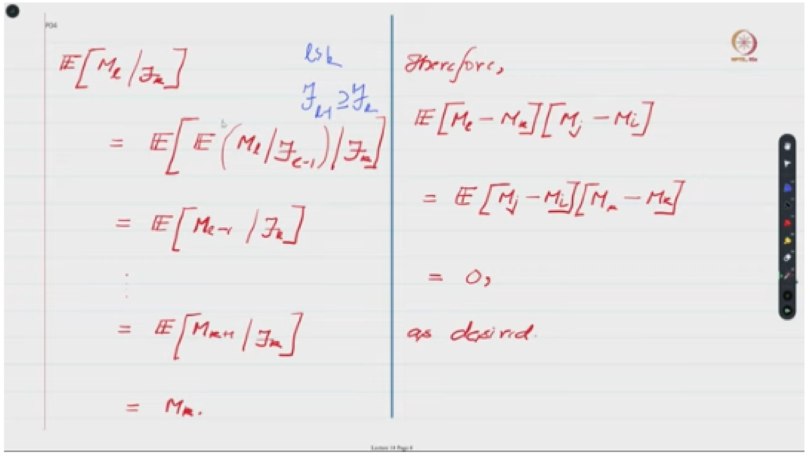
Now, we will focus on this term in the product. So, from the tower property, it is easy to see that if I have the expected value of ML given FK. So, recall that L is bigger than or equal to K. So, if I look at, you know, this expression for L bigger than or equal to K. So, if L is equal to K. Then, you know, your ML itself will become measurable with respect to FK, and this conditional expectation will equal MK, right? On the other hand, suppose L is strictly bigger than





So, let us suppose that L is strictly bigger than K. In that case, by making use of the tower property, one can see that this conditional expectation can be written as the conditional expectation of the expected value of ML given FL minus 1. And because of this condition, it is easy to see that your FL minus 1. And FK will satisfy some relation like this. And hence, we can invoke the tower property, right? And over here, right?





We know that your ML is a term in the martingale sequence, right? And hence, this conditional expectation will equal ML minus 1, right? And then we will end up with the conditional expectation of ML minus 1 given FK. Right. And now we can repeat the same trick.

So notice that you started with ML and ended up with ML minus 1. And if L minus 1 is again strictly bigger than K, we can, you know, repeat this whole process until we end up with MK plus 1. So this K over, I mean, this index over here is 1 more than K. Right. And hence, we can finally invoke the martingale property to conclude that this conditional expectation equals MK. So, what is the summary of this short discussion? The conditional expectation over here actually equals MK, and if you go back over here, we have this conditional expectation subtracted from MK.

And if this conditional expectation equals mk, then this term over here will be 0, from which we can conclude that indeed the expectation of the product will be 0, as desired, which helps us conclude that if you have four indices in the following way, then indeed ML minus MK is orthogonal to MJ minus MI.

$$E\left[M_l|_{F_k}\right] = E\left[E\left(M_l|_{F_{l-1}}\right)|_{F_k}\right]$$

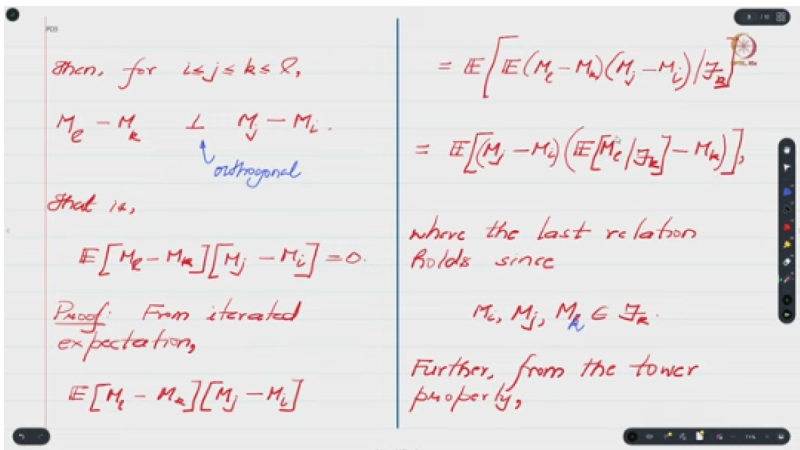
$$= E\left[M_{l-1}|_{F_k}\right]$$

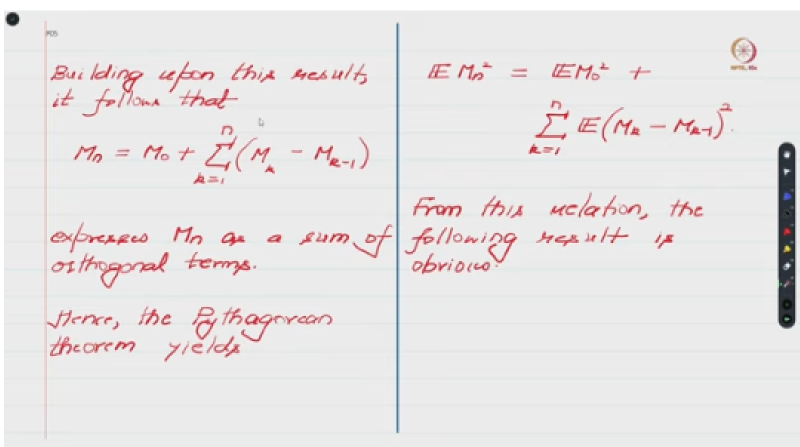
$$= E\left[M_{k-1}|_{F_k}\right]$$

$$= M_k$$

$$E\left[M_l - M_k\right]\left[M_j - M_i\right] = E\left[M_j - M_i\right]\left[M_k - M_k\right] = 0$$

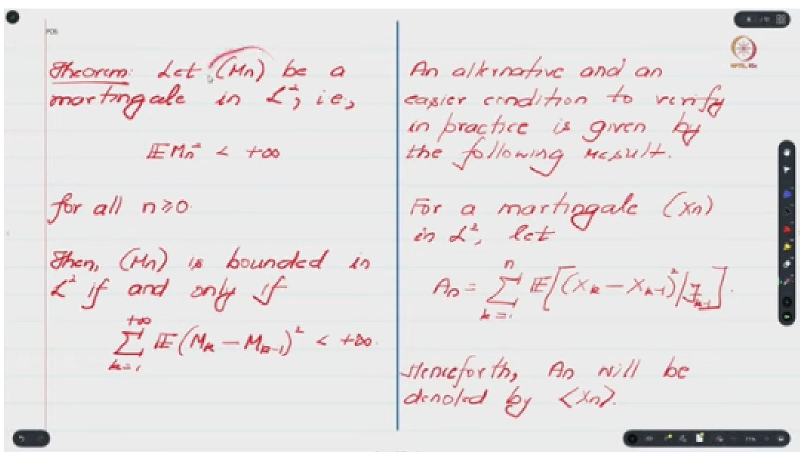
So, by building upon this proof technique and result, it is easy to see that if we express Mn in this fashion. So, of course, notice that this is just a telescopic sum. So, Mn indeed satisfies a relation like this.





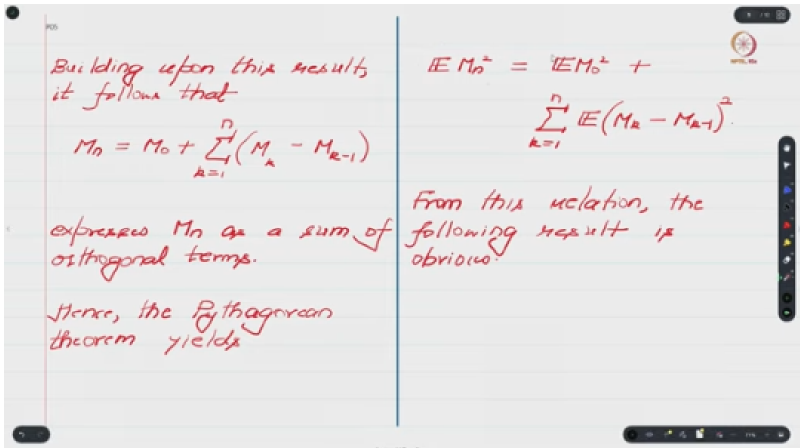
But you know, because of this property that we have proved previously, this expression can be viewed as a sum of orthogonal terms, which means that if you look at any of the cross terms over here and take their expectation, the result would be 0. And because this result is 0, if you take the square of the left-hand side and take its expectation, you will only end up with the expectation of the squares of the individual terms. So, notice that we do not have any cross terms over here, and that holds because of that orthogonality property that we proved previously. And this relation that we have is known as the Pythagoras relation or Pythagorean result for martingales. So, now what we have shown is that the expected value of Mn squared equals the expected value of M0 squared plus the sum of expectations of successive differences—or, let me repeat that again, the sum of expectations of squares of successive differences, right.

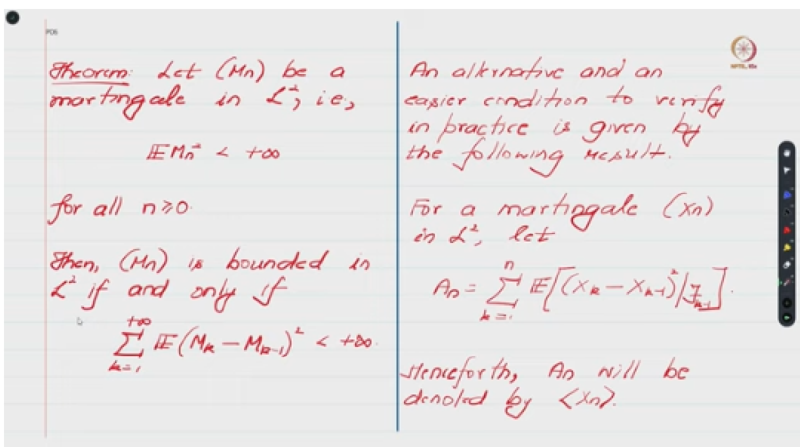
So, you know, this relation is very easy to see. Because of this relation, we can now state the following result. So, the result says that suppose Mn is a martingale in L2, which means that for every n, the expected value of Mn squared is less than infinity. This is what a martingale in L2 means. The sequence Mn is bounded in L2, which means that, you know, the supremum of these expectations is less than infinity.



So, it is bounded in L2 if and only if you know this infinite sum of expectations of squares of these successive differences. So, notice that there is an infinite sum over here. So, this infinite sum should be less than infinity. So, this is very easy to see because of this relation. As you increase n over here, this expression keeps growing.

So, if you take the supremum of this, then it will be upper bounded by this plus this expression where n is replaced by infinity. Hence, if this quantity is less than infinity, then indeed your Mn will be bounded in L2. Is that okay? So, you know this result provides us with one way to check whether your given martingale sequence is bounded in L2, and once we check that, we can conclude that Mn is also bounded in L1. Since it is bounded in L1, we can invoke Doob's upcrossing lemma, right, and its consequences to conclude that the limit of Mn exists and is finite. So, unfortunately, you know this result by itself will not be enough to work with the kind of things needed for proving convergence of the stochastic approximation algorithm.

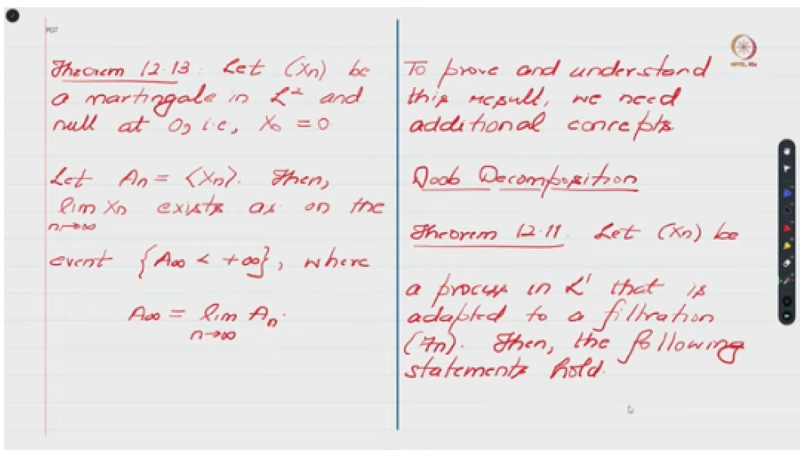




We will need a relaxed version of this result, which I am going to state now. So, to state this result, we first need some notion of what is called the angle bracket process associated with a martingale. So, let me first define that. So, given a martingale Xn—I actually change back to the Xn notation. So, given a martingale Xn in L2.

Let us define a_n to be this expression over here, right? So, this expression over here. So, what is this expression? It is the expected value or the conditional expectation of the square of successive difference conditioned on f_{k-1}, right? So, here we did not have any conditioning, but here on the other hand we have some conditioning.

And then we sum it from k equals 1 to n, and whatever is the result, we refer to it as a_n. So, whenever you have a process X_n which is a martingale, this process for different values of n, right, you know, defined in this way will be referred to as the angle bracket process. Angle bracket meaning, you know, you have these angle brackets, right. So, this a_n process will be referred to as the angle bracket process associated with your martingale X_n and will be denoted in the following way. So, the relaxed condition to discuss or show convergence of X_n is stated now, right.



So, you know, this theorem again is taken from this textbook called Probability with Martingales by David Williams, and this number refers to the corresponding result in that textbook. So, in case you want to see more details, I encourage you to, you know, look up look up this result in that textbook, right? So, what does this result say? This result says that suppose your X_n is a martingale in L^2, and we will just make some simplifying assumption that it is null at 0, okay?

I will later on show that this is not such a, you know, restrictive assumption. In fact, you know, every sequence Xn is you know, by a small modification, can be ensured to be null at 0. So, what do I mean by a martingale sequence being null at 0? We require that X0 be 0, and hence the expected value of X0 is 0, right?

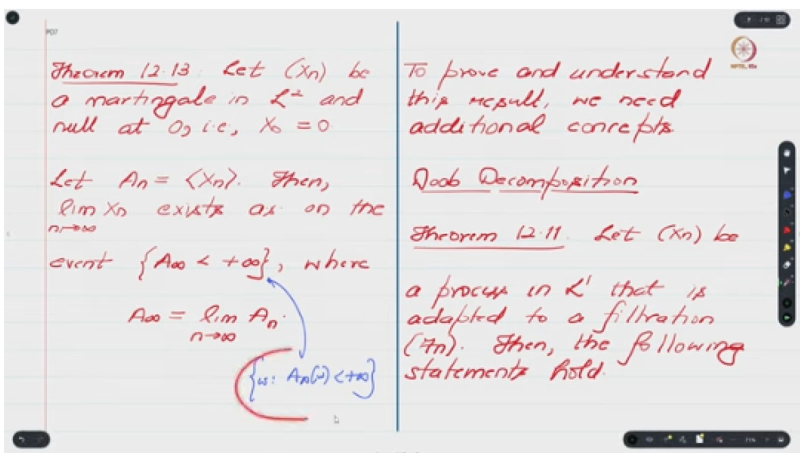
So, this property we will be repeatedly using, right? So, let Xn be a martingale. which is null at 0, and let An be the angle bracket process. So, then your limit Xn exists almost surely, so this is a.s., almost surely on the event A∞ less than infinity, right?

Where A∞ is the limit of An.

$$A_{\infty} = A_{n}$$

So, let us take some time to digest this result, right? So, first let us ask why should this, you know, quantity over here exist, right? We will soon see that this An term that we have right is actually, you know, an increasing sequence. So, because this is an increasing sequence, this limit actually exists always, right?

So, this limit always exists, and hence A infinity is well-defined. And, because A infinity is a well-defined random variable, we can consider this event: A infinity less than infinity. So, just to recap what this means, it is basically the collection of all omegas such that A infinity of omega is less than infinity. So, you collect all those omegas, and the claim is that except for a zero-measure subset of this set, on all other omega values, your limit Xn will exist. So, let us go over this result one more time just to get a sense of what exactly is happening.

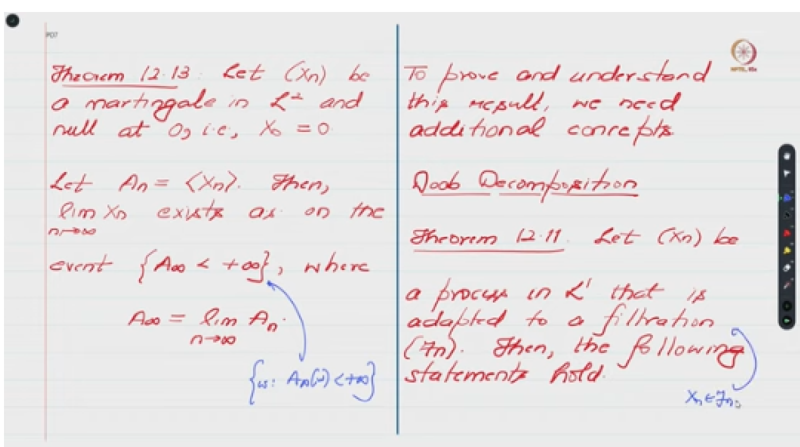


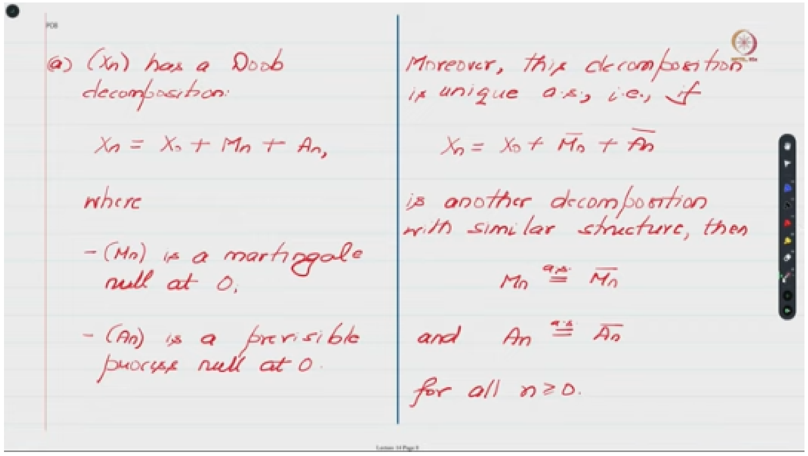
So, this result, as I said, gives a relaxed condition under which we can discuss the existence of a limit of a martingale. So, this result says that suppose you have a martingale in L2 which is null at 0, and suppose you have A_n denoting the angle bracket process at X_n, then the limit X_n exists almost surely on the event A infinity less than infinity, right? So, just to give you a heads-up, we will be using this result concretely while talking about the convergence of a suitable noise term in the analysis of stochastic approximation algorithms. So, keep this in mind when we formally move to discussing the convergence; you will see me invoking this result.

So, we will now prove this result formally just to get a feel for how these results are proved, and in order to prove this result, we need to understand a few additional concepts. The first among these additional concepts that we need to understand is that of the Doob decomposition. Again, this Doob decomposition plays a very crucial role in the analysis of martingales and hence in the analysis of stochastic approximation algorithms as well, right? So, this Doob decomposition is numbered Theorem 12.1 in the 'Probability with Martingales' textbook. So, what does this result say?

This result says that suppose Xn is a process in L1. So, this is a process in L1, which means that the expected value of the absolute value of Xn is less than infinity. I am not saying it is bounded in L1. I am only saying it is a process in L1. And also notice that I am not presuming that this Xn is either a martingale, supermartingale, or anything like that.

It is just some process, some stochastic process that is in L1. And that is adapted to this filtration Fn, which, recall, means that your Xn belongs to Fn, right? This is what it means. And then this Doob decomposition says that the following statements hold, right? So, what are the following statements? So, there are two statements, and on this page, I have stated the first of these two statements. So, the first statement says that

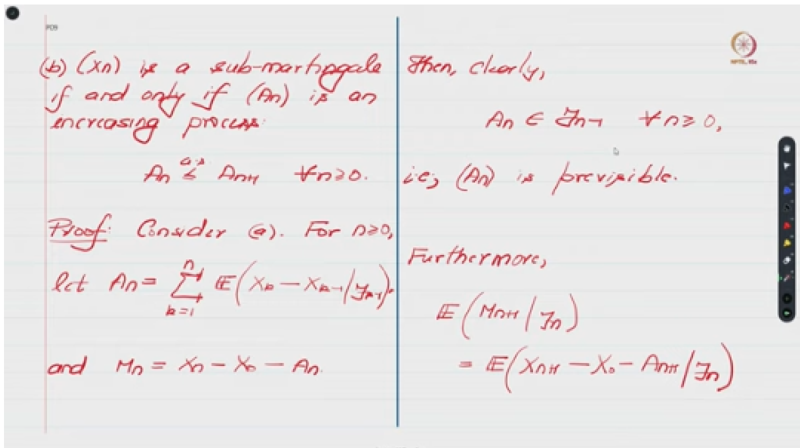


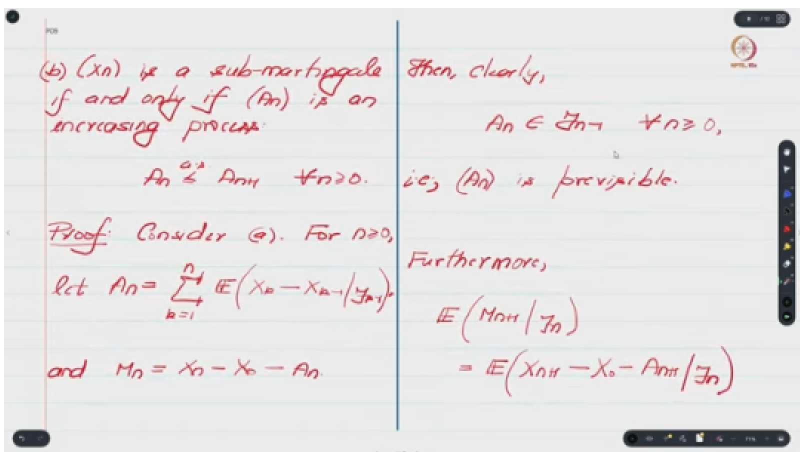


Whenever your Xn is a process in L1, whenever it is a process in L1 and you have some filtration with respect to which it is adapted, then this Xn process has a Doob decomposition. So, what does the Doob decomposition state? It says that you will be able to express Xn as a sum of X0, which is the initial value of your Xn process, plus some value Mn and An such that the sequence of Mn values will form a martingale which is null at 0, which means that M0 is 0, right? And this An process will be your previsible process null at 0, right?

So, at this point, you know, please do not imagine that this A_n is actually this angle bracket process or anything like that. Right now, I am only saying that you give me some process, then I will be able to express it in this fashion where M_n is a martingale and A_n is some previsible process, right? And the significance of this Doob decomposition is that this decomposition is unique almost surely, which means that if you manage to come up with another decomposition for X_n, right, which has a similar structure. So, let us say X_n equals X_0 plus M̄_n plus Ā_n, where this sequence M̄_n is a martingale and Ā_n is a previsible process. Then one can show that M_n equals M̄_n almost surely and A_n equals Ā_n almost surely for all n greater than or equal to 0.

Which means that except for a zero-measure set, you would not be able to have any differences between M̄_n and M_n and similarly between A_n and Ā_n, right? And the second part of this Doob decomposition result is that if X_n is a submartingale, then your process A_n is an increasing process. So, a submartingale or martingale is a special case of a submartingale. So, this conclusion also holds for martingales. So, as I told you before, you know, there is this limit, and I told you that this has to exist and so on.





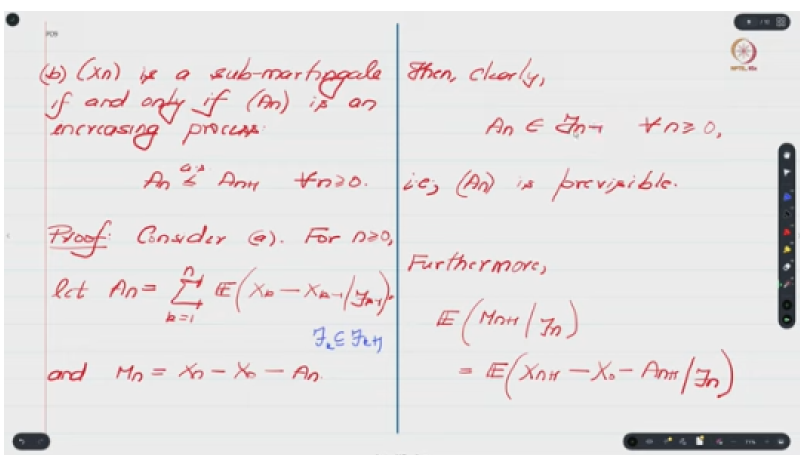
This actually will be a consequence of the fact that this A_n will be an increasing process; you will soon see that. So, the second part of this Doob decomposition says that X_n is a submartingale if and only if this A_n process is an increasing process, which basically means that A_n is less than or equal to A_{n+1} almost surely. Is that okay? So, now we will prove this result in the rest of this class. So, let us consider the first statement and let us define.

An in this way, right. So, we have been given this sequence xn, right? And from this sequence xn, we will define an in the following way. So, notice that I have not put any squares or anything over here, and hence this an is not the angle bracket process, right. So, that is why I highlighted that part before. So, an is defined in this way.

Right? And Mn is defined in this way, right? So, you take xn, subtract from it x0, and subtract from it an, where an is defined in this way, right? So, as a reader, one may ask, okay, you know, how can one think of such a decomposition or such a definition for an and Mn, right? So, you know, once you look at the result, you would

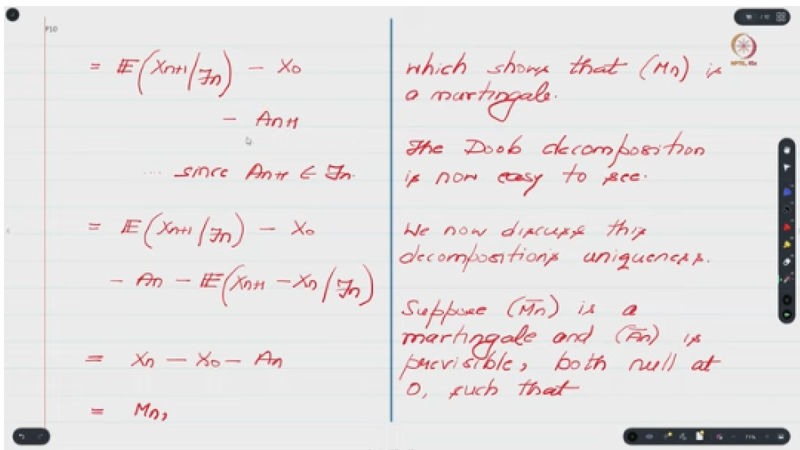
very easily see that, you know, we want to come up with an Mn which has a martingale nature, right? And you will soon see that, you know, if you define something like this, where an is defined in this way, then Mn indeed has a martingale property, right? And that is what we will now see. So, from this definition and because of this conditioning over here, you can see that if you substitute k equals n, then this term will be the expected value of xn minus xn minus 1 given fn minus 1. So, the last conditioning is fn minus 1. And, you know, all previous summands will be like conditioned on fk minus 1, where k is strictly less than n. And, you know, because you have a filtration, you have that your fk is a, you know, sub-sigma-algebra of fk plus 1. So, from this property, one can indeed conclude that an is measurable with respect to fn minus 1.

And because there is n minus 1 over here, one can conclude that this a_n is actually a pre-visible process. Is that okay? It is a pre-visible process. And now, you know, this is as desired. So, if you remember the decomposition, we require that V.



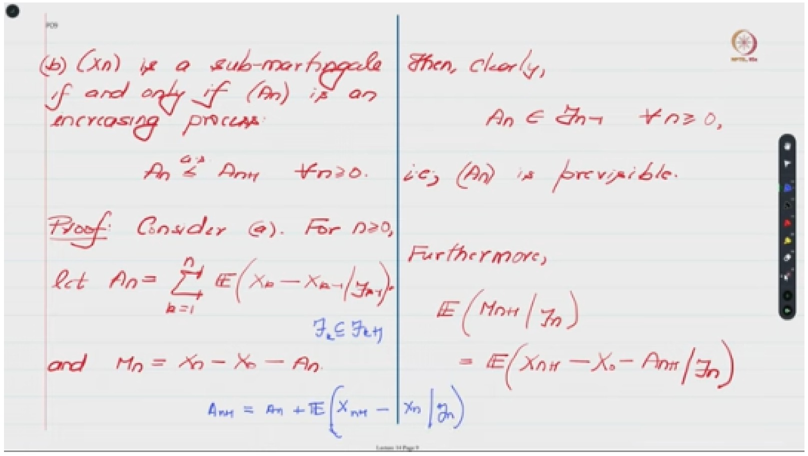
M_n be a martingale and A_n be a pre-visible process. So, right we have verified that it is a pre-visible process, and one can, you know, easily see that when you substitute N equals 0, the upper index will be strictly less than the lower index, and in that case, we will define this sum to be 0. Hence, this A_n is actually null at 0, right. So, this is as desired. Now, it remains to check whether your M_n process is actually a martingale or not. So, let us do that check.

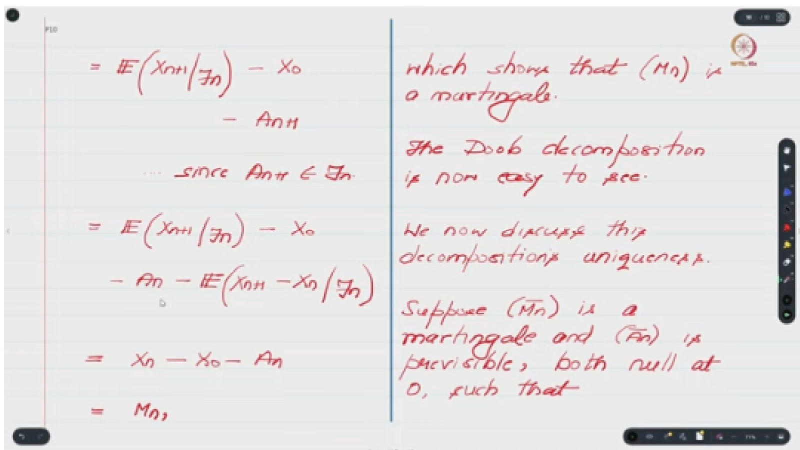
So, let us look at the conditional expectation of M_{n+1} given F_n, right. So, M_{n+1} by definition is X_{n+1} minus X_0 minus A_{n+1}, right? And A_{n+1}, because of its pre-visible nature, I can pull it out as it is. Similarly, X_0 is measurable with respect to F_0, which is a subset of F_n. Hence, X_0 is also measurable with respect to F_n, and I can pull that out as well, and we will be left with something like this.



So, this term that we have will leave us with something like this. Right. And now what we will do is, you know, make use of the definition of a_{n+1}, right? So, a_{n+1} has this sum over here, right? So, one can easily see that a_{n+1} will equal a_n plus the last term, which is the conditional expectation of X_{n+1} minus X_n given F_n.

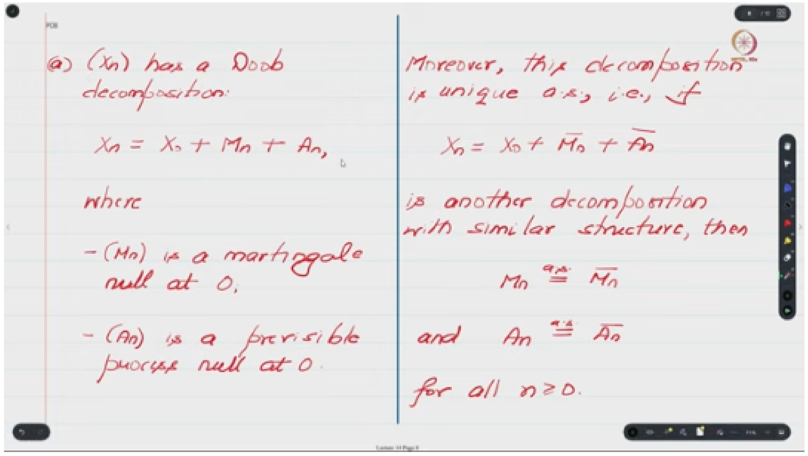
So, I can substitute that thing over here. So, wherever you have a_{n+1}, I can write it as minus a_n minus this conditional expectation. And one can see that this whole sum now equals this. So, let us see why that is the case. You see that you have this conditional expectation of X_{n+1} given F_n.



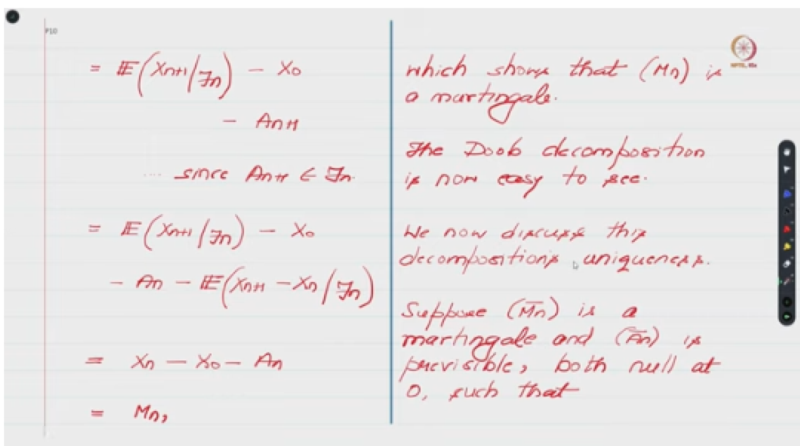


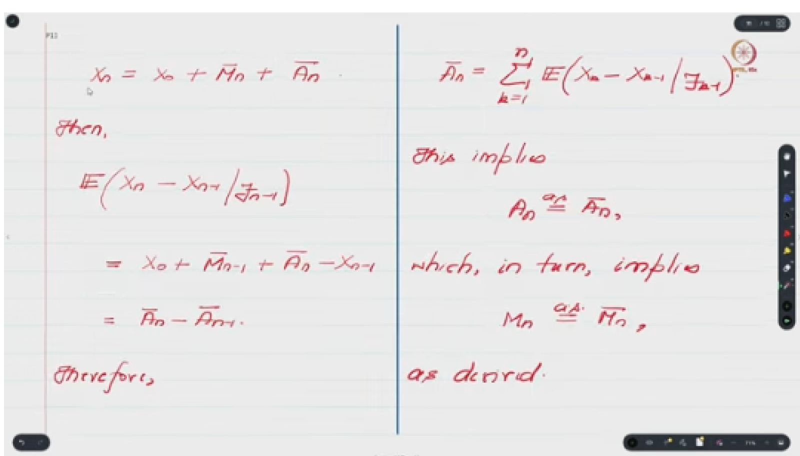
So, if you invoke linearity of conditional expectations over here, you will get this same quantity like this. So, this and this will cancel off. And then the only term that remains is the conditional expectation of X_n given F_n. And since X_n is measurable with respect to F_n, we will end up with X_n over here. Hence, this whole expression equals this, which is M_n.

And by this verification, we can now conclude that indeed this sequence Mn is a martingale. And hence, we have Fn. The existence of the Doob decomposition, right? So, what we have managed to show is that if you give me an arbitrary sequence Xn which is measurable with respect to a filtration Fn, then you know one can express Xn in the following way where Mn is some martingale and An is some previsible process.



So, now what remains to discuss is why this decomposition should be unique. So, that is what remains to be discussed. So, towards showing that it is unique. Let us suppose that Mn bar is a martingale and An bar is previsible and both are null at 0, and let us say you know Xn has a relation like this.



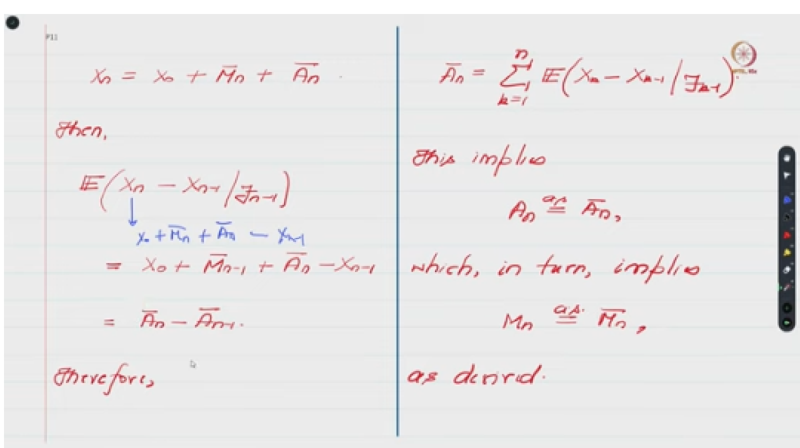


Is this okay? So, we are supposing—I mean, this is like a proof by contradiction technique. So, we are supposing that let us say there is another decomposition which has the same structure. Right. And, you know, then one can conclude that if you look at the difference of Xn and Xn minus 1 and look at the conditional expectation of that with respect to Fn minus 1, right.

So, from this expression, one can see that Xn minus Xn minus 1, right, will basically be X0 plus, so let me just write this. So, this will be like X0 plus Mn bar plus An bar, right, minus Xn minus 1, right. Now, X0 is measurable with respect to Fn minus 1. So, that will come out as it is.

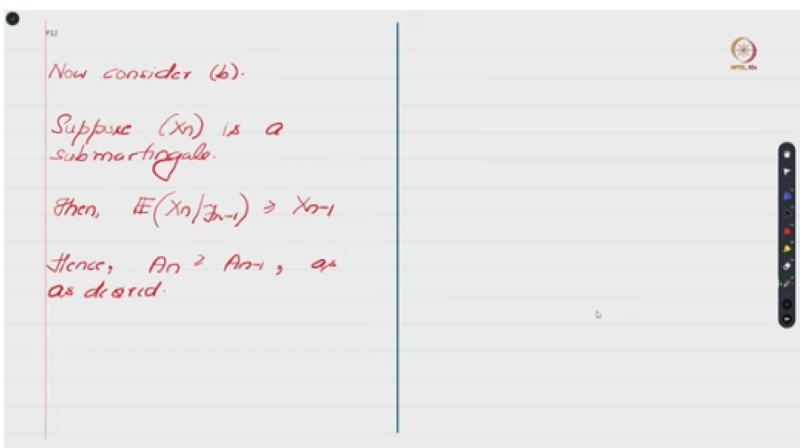
Now, Mn bar is a martingale. Hence, if you take its conditional expectation with respect to Fn minus 1, you will end up with Mn minus 1 bar. Now, this is a previsible process. Hence, you will be left with this An bar itself. And whatever you have over here, I leave it as it is.

Now, this quantity and if you add An minus 1 bar will give me Xn minus 1. Hence, from the definition of Xn minus 1, one can conclude that this expression equals An bar minus An minus 1 bar. So, let us summarize what we have done so far. We have said that suppose Xn has this decomposition, then your An bar minus An minus 1 bar should satisfy this conditional expectation. And hence, if you add such terms for different values of n, one can see that An bar would actually equal the sum of these conditional expectations.

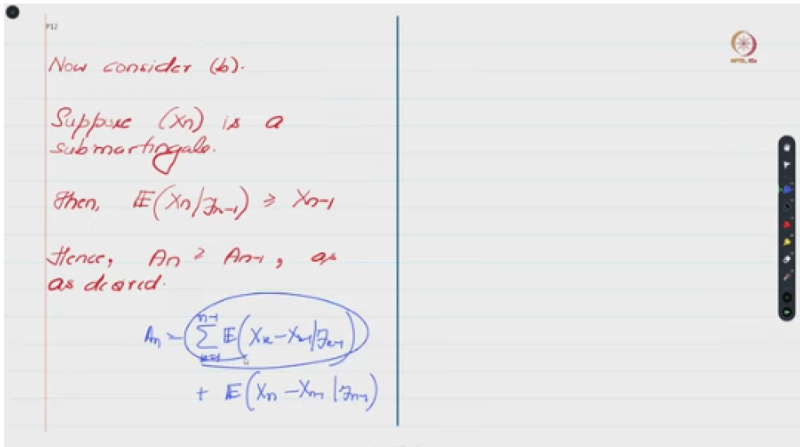


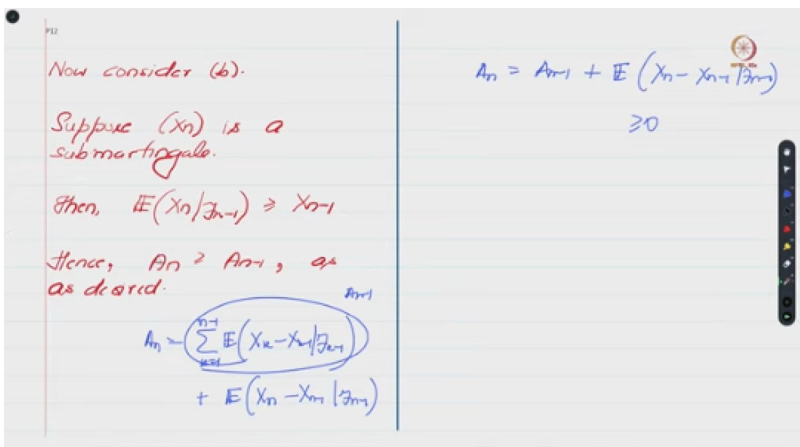
But if you see this definition and the definition that we had for AN, they are one and the same, right? So, they are one and the same, and because conditional expectations can only differ on a zero-measure set, we can now conclude that AN and AN bar must be almost surely the same, right? And once... you have a decomposition where this, this are the same, sorry, this, this, and this are the same almost surely, okay? It should follow that your Mn and Mn bar should also be the same almost surely.

Is that okay? So, from this, we have managed to show that your decomposition is almost surely unique, right? And now comes the last part wherein we had to show that, you know, whenever your Xn is a supermartingale, right? Whenever Xn is a supermartingale, right? Your previsible process An will have this monotonicity property, okay? So, that is very easy to see. Right, you know, because your Xn, when Xn is a submartingale, right? Because it is a submartingale, this conditional expectation is greater than or equal to Xn minus 1, right. And if you, you know, recall the definition of An, the definition of An is that it is the sum of k equals 1 till n minus 1 expectation of Xk minus Xk minus 1 given Fk minus 1.



Plus the conditional expectation of, you know, Xn minus Xn minus 1 given Fn minus 1. So, this term over here is actually An minus 1; hence, we can write An equals An minus 1 plus the conditional expectation Xn minus Xn minus 1 given Fn minus 1. Now, because your Xn is actually a submartingale, this expression over here will be non-negative, and hence one can conclude that your An will be greater than or equal to An minus 1 almost surely. And I have shown the proof in one direction, but statement B says that if this holds, then this holds and vice versa as well. That is, if this holds, then this holds, and the argument actually proceeds similarly.





If a_n is greater than or equal to a_n minus 1, then this conditional expectation should be greater than or equal to 0. On the other hand, if this is greater than or equal to 0 on account of X_n being a submartingale, then a_n should be greater than or equal to a_n minus 1.

$$E\left(X_n|_{F_{n-1}}\right) \geq X_{n-1}$$

So, this brings me to the end of today's class. So, let me quickly summarize what we have discussed in today's class. So, in the previous class, we looked at a condition called boundedness in L1 to talk about the convergence of supermartingales and also martingales.

And then, in this class, we asked if there is some alternative way to look at this boundedness in L1, and we said that we can talk about this boundedness in L2. And so, we sort of came up with this result that a martingale sequence is bounded in L2 if the infinite sum of the expectation of the squares of successive differences is bounded. So, that was one result. However, as we will see later, that result by itself is not enough to show or to be of use in the discussion on the convergence of stochastic approximation algorithms. So, in the second part, we began our discussion on relaxing a condition that ensures the convergence of martingales.

In particular, we looked at this angle bracket process and then we said as long as your A_infinity is less than infinity on that event, your process actually will be finite. Is this okay? So, let me just make sure if this is the angle bracket process. So, you know you had a martingale, and I said that if your X_n is the angle A_n is your angle bracket process, then your martingale actually converges almost surely on the event where A_infinity is less than infinity. So, the proof of that result is not complete; we will continue this discussion in the next class.

In today's class, I introduced a very important concept called the Doob decomposition, which we will be utilizing crucially in the next class. Having said that, let me say thank you, and I hope you will join me in the next class. Thank you.