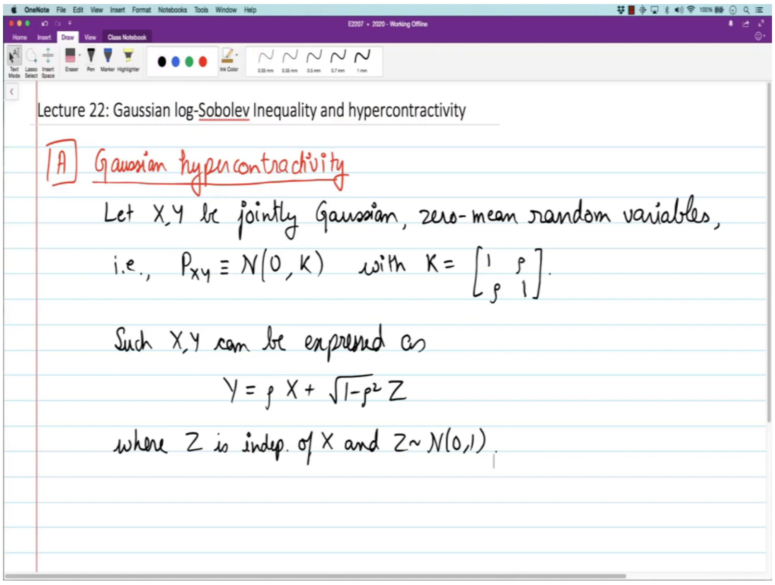
Concentration Inequalities

Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 22
Gaussian log- Sobolev Inequality and hypercontractivity

(Refer Slide Time: 00:20)

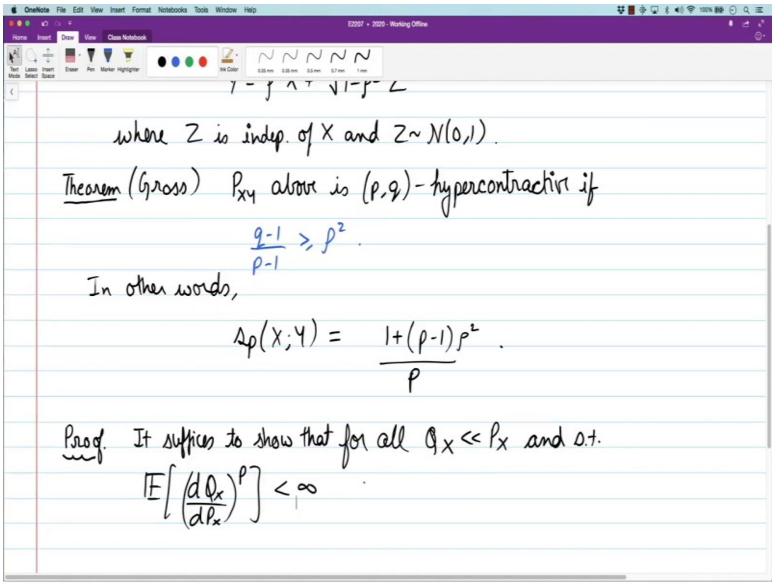


In the last lecture we saw the notion of hypercontractivity and so, how it can have amazing consequences in terms of strong data processing inequalities. And in fact, there are many more which we did not discuss. We saw different equivalent forms of hypercontractivity. And the question is this really a new phenomenon, which have which has not been seen before and what we will now see is that in fact, for some distributions like Gaussian distributions hypercontractivity is exactly equivalent to log-Sobolev inequality.

So, before we proceed, let us just first describe Gaussian hypercontractivity ok. So, we consider generic Gaussian random variable P X, Y. Let X, Y be jointly Gaussian, zero mean random variables ok. What does it mean? That is P X, Y is Gaussian with mean 0 and covariance matrix K, with K given by yeah. Let us also assume that there are they are standard normal. So, we normalize them to unit variance and rho is their correlation coefficient ok.

Such X, Y can be expressed as this is something that you may know already Y equals to rho X plus root 1 minus rho square Z, where Z is independent of X and its again 0 mean and unit variance. Note that X and Y are both 0 mean and unit variance Gaussian random variable ok.

(Refer Slide Time: 02:57)

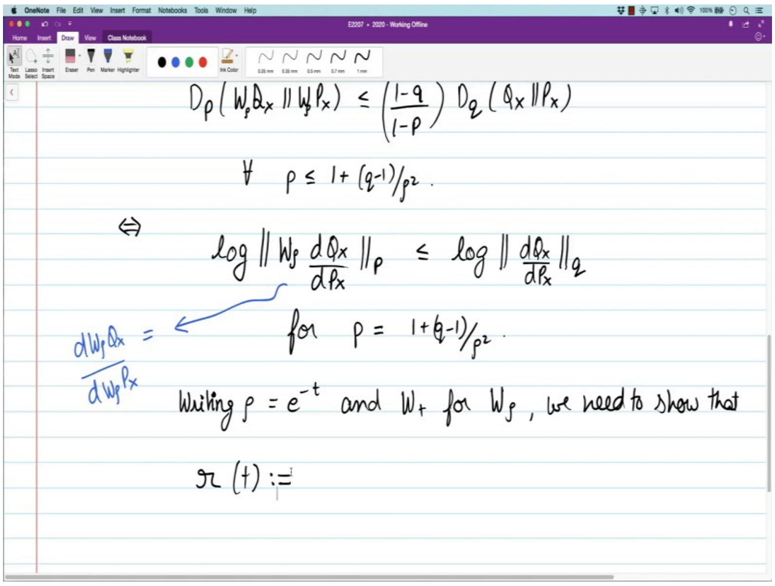


So, here is this Gaussian hypercontractivity theorem from (Refer Time: 03:04) gross. It says that P X, Y is p, q hypercontractive if and only if q minus 1 by p minus 1 exceeds rho square ok this. We do not even claim by the only if part if this exceeds. Or so, in other words quantity s P X, Y that we had seen is exactly equal to 1 plus p minus 1 rho square by p ok. Not an easy form to remember, I think this is an easier form to q minus 1 by p minus 1 exceeds rho square ok.

This is the Gaussian hypercontractivity and now we will like to prove it. In fact, to prove it we will develop little bit of a we will use a little bit of heavy machinery eventually we will do that, but to begin with we will start from some elementary proof. So, the proof proceed this follows.

We first recall that by using the equivalence between pq hypercontractivity and the strong data processing inequality that we saw for any divergence. It suffices to show that for all Q x that have density with respect to just Gaussian density P x and such that this density has d Q x by d P x to the power be the P th moment finite.

(Refer Slide Time: 06:09)



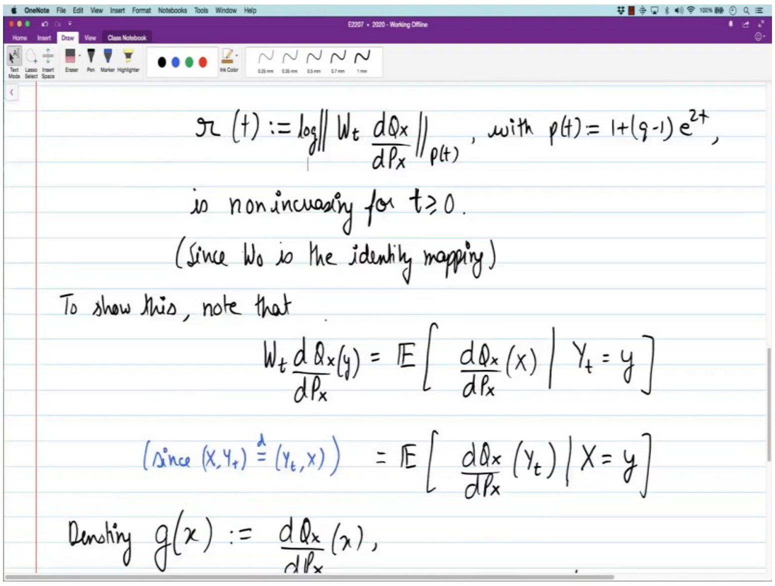
What we need to show is that for all such; for all such Q x the P th divergence between Q x and P x when pass through this channel ok. By the way this channel will be parameterized by rho because that is what brings in this rho. Remember this channel can just be described as this additive Gaussian noise channel. It multiplies with the gain rho and this 1 minus rho square. So, this is a channel parameterized by rho.

And this guy is less than equal to 1 minus q by 1 minus p D q Q x p x if or this is true for all p less than equal to 1 plus q minus 1 by rho square. In fact, we use a slightly different parameterization, but before that this last inequality by the way is equivalent to log. This is something we saw earlier also. W rho of p is less than equal to log D q x by d P x and this is q. And this must hold for all p less than equal to 1 plus q minus 1 by rho square. In fact, it suffices to have it for p equal to this by rho square.

Now, just a side remark; we saw earlier that this guy here the W that the Markov kernel of the channel applied to the log likelihood to the likelihood ratio to the density is exactly the same as this guy remember that ok. And this is by the way the Gaussian density, when you apply this channel to Gaussian input you get Gaussian output ok.

Now, we will see this thing and we will use a slightly different parameterization. Writing rho equals to e to the power this only depends on rho square. So, we will write rho square as e to the power actually we assume rho is positive because otherwise the same proof can be repeated with x this is minus x does not change anything. So, we set rho to e to the power minus t and W t for W rho what we need to show is that if you look at this function g t or maybe something else maybe call it r t defined.

(Refer Slide Time: 09:57)



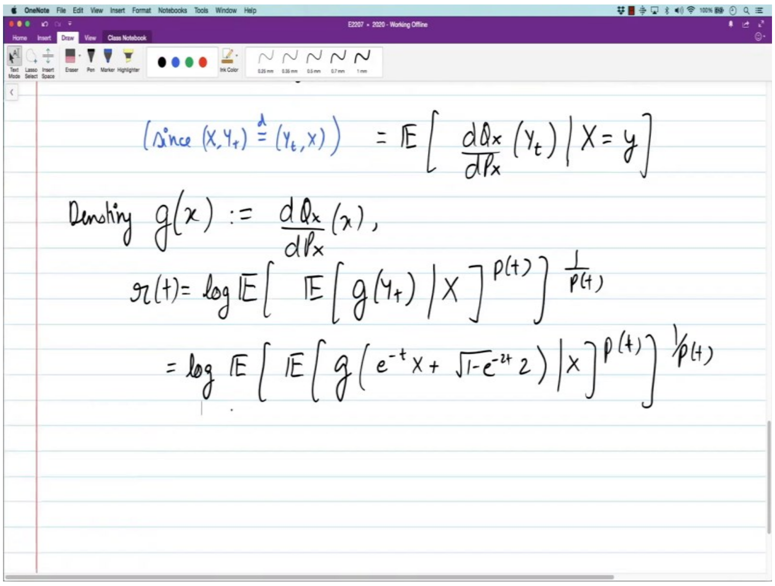
So, this logs can also by the way. We do not really need to have this logs here. So, it sufficed to show that this function r t defined as W t d Q x d P x p t. This is the quantity we are calling p yeah p t, with p t defined as 1 plus q minus 1 e to the power 2 t. This quantity here is non increasing for t greater than 0 ok.

And this is true since W 0. What is W 0? W 0 basically corresponds to having rho equal to 1 and in which case y equals to x ok. So, this channel is just the identity channel. So, its W 0 is the identity mapping ok that is what you have to show.

So, let us take the derivative of this function and show that it is less than 0 that is all we need to show ok. But how do we differentiate this function, alright? So, to do that to show this note that W t of this Q x d P x this is exactly the same as the conditional expectation of d Q x d P x.

Now, this is a function of x, x is the standard Gaussian random variable here given y ok that is something we have seen earlier and you should say it is given Y t denoting by Y t the random variable y corresponding to e to the power minus t ok. But in fact, there is a symmetry in the channel from y to x and x to y, this distribution is symmetric this distribution here is symmetry.

(Refer Slide Time: 13:12)



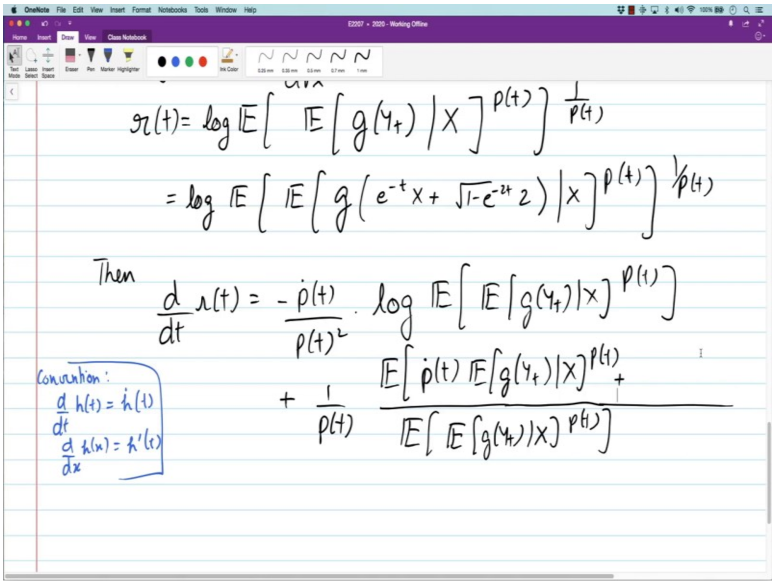
And therefore, this can also be written as expected value this is by symmetry, symmetry of the distribution. So, this is since X, Y t has the same distribution as T t, X. So, this can be written as d Q x by d P x. Now, this symmetry is actually pretty crucial to this proof, but it holds here. It holds for the symmetric channel. This is this Y t given X equal to y ok. So, this guy has now this kind of trajectory alright.

But so, let us abbreviate this did on Nickelodeon derivative ok. Denoting g of y as or maybe g of does not matter x as d Q x d P x of x r t is given by expected value expected value of g of Y t given X yeah. The entire distribution is symmetric. So, that we can take the outer expectation with respect to we can take the outer expectation also with respect to with respect to x ok. Remember that the original thing had this outer expectation with respect to y and the inner one with respect to x. So, this is r t therefore.

So, we can write a little bit more about it. This is equal to expected value of expected value of now we know what g looks like given X looks like e to the power minus t X plus square root 1 minus e to the power minus 2 t is it ok and given X this expectation is over Z then by the way given X to the power p t p t and then to the power 1 by p t right.

Actually it will help keeping the log sorry above that. Let us keep this log here ok. So, r t was this, r t is defined as log of this same thing and then this guy has log here and this also has a log here right ok.

(Refer Slide Time: 17:06)

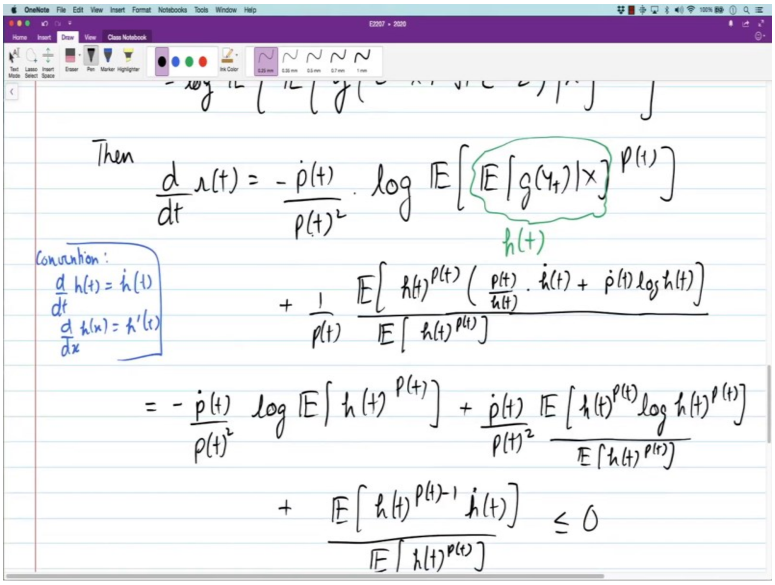


So, then if you take derivative of r with respect to t, so, this is this by this. So, it this expression here this is equal to first we take derivative of this 1 by p t. So, that is minus. So, one convention when we take derivative with respect to t we use a dot ok. So, some convention here d by dt of g t of g t or any function actually d by dt of h t is h dot and d by d x of h x is h prime ok. We will use this dot and prime notation

So, here we have p dot t by p t square * of course, the log natural log of expected value of this whole thing; expected value of g Y t given x to the power p t ok. And plus 1 by p t and now this second derivative part of the derivative. So, log of this we can get this in the denominator and derivative of this guy; now we can take derivative inside and then there are two terms given X and we have to now focus on this one.

So, first term is taking the derivative again of p t, but this guy will remain right. So, that is expected. So, that is p dot t. So, just taking it this is just taking derivatives plus now we take the term of this guy p dot t p dot t yeah maybe its convenient to use a little bit more notation right.

(Refer Slide Time: 20:31)

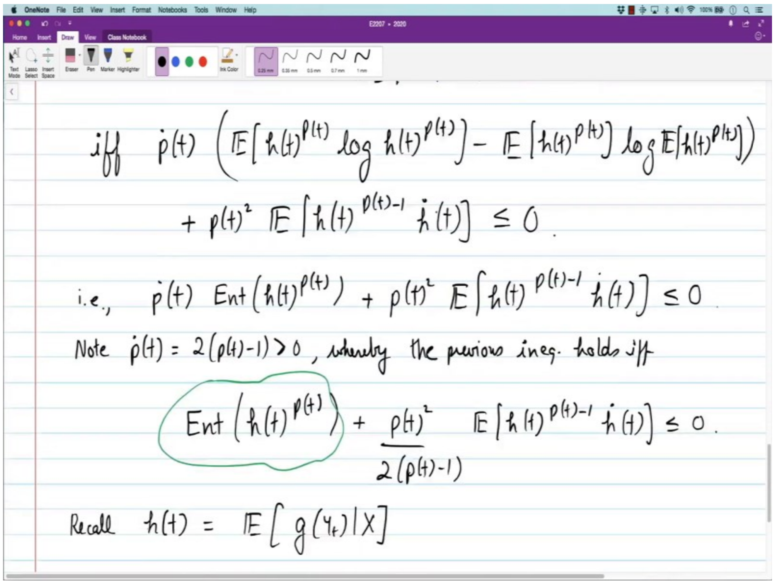


So, let us call this name this function something. We will call this h t ok. So, what we see is for the second term? What we see is plus 1 by p t expected value of h t is a random variable h t to the power p t no problem and then in the numerator we can take derivative inside h t to the power p t. What is the derivative? That is e to the power h t log h t p t.

So, first it is e to the power h t p t. So, that is h t p t * two terms will come. First one is p t by h t into h dot t plus p dot t into log h t ok. These two terms come here. And therefore, this is minus p dot t p t square log expected value of this h t to the power p t plus p dot t by p t. Let us say p t square again, expected value of log. So, this is the second term here this one.

So, h t to the power p t log, I multiplied and divided by pt. So, you have this and plus this p t cancels here it is the expected value of h t p t minus 1 h dot t by expected value of this expected value remains here as well expected value of h t to the power p t h t to the power p t.

(Refer Slide Time: 23:30)



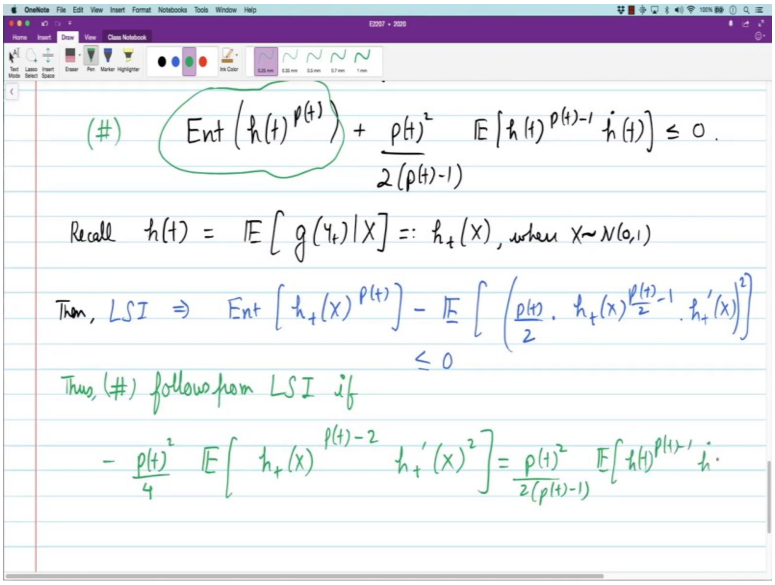
And we would like to claim that this is less than equal to 0, if p dot t; so, there is so, we multiply and divide this guy by p t square p dot t * expected value of h t p t log h t p t minus we have multiplied divided by this denominator here, h t p t log expected value of h t p t ok. This is the first term plus p t square expected value of h t p t minus 1 h dot t is less than equal to 0.

So, this term if you see this term this exactly look like p dot t into entropy of h t p t plus p t square expected value of h t p t minus 1 h dot t is less than equal to 0. Note that for our choice of p t this p dot t is nothing but you can show this is 2 p t minus 1 you can verify this. Remember p t was 1 plus q minus 1 e to the power 2 t.

So, you get from this formula and this must be greater than equal to. In fact, must be strictly greater than 0, this is strictly greater than 0. And this implies this previous inequality holds if and only if the entropy of h t p t plus p t square by 2 into p t minus 1 expected value of h t p t minus 1 h dot t is less than equal to 0 ok.

So, that is what we have to show and now that we see this nice entropy term here, this one here ok. It is time to bring in the logs of all of you. Let us introduce a slightly more a heavy notation. So, let so, remember that h t h and recall. So, recall that this h t that we using here. This h t was given by the conditional expectation of g of Y t given X yeah that is what h t was and right. So, this is a function of X basically.

(Refer Slide Time: 27:41)

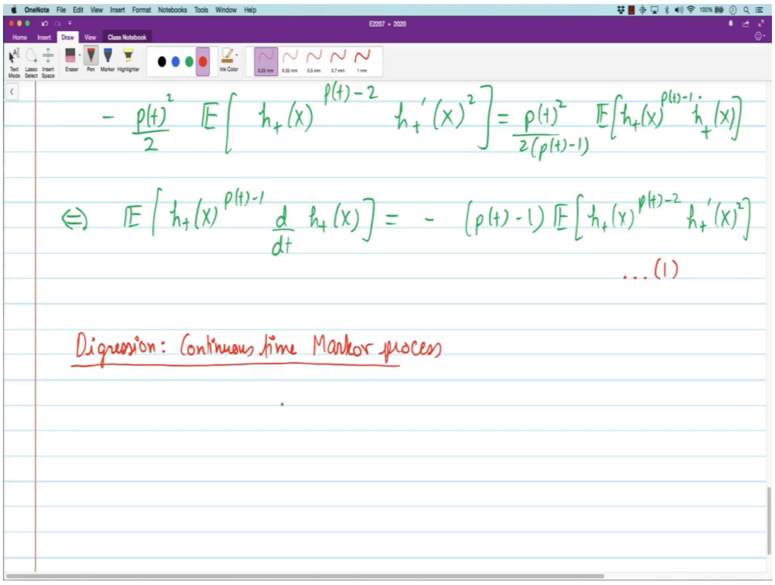


So, we will now use the slightly heavier notation. Let me call this h t of x ok. So, that I see this dependence on X explicitly it depends on t and it is a function which depends on t and then it depends on X. Then with where X is the Gaussian random variable here, X is normal 0, 1. Then log-Sobolev inequality implies; what would be log-Sobolev inequality?

It says that expected value of h t X sorry the entropy of h t X to the power p t f square minus the gradient of the gradient of this guy. So, what is this function? This is d by so, so we are looking at h t X to the power p t by 2. What is its derivative with respect to X? It is p t by 2 into h t X to the power p t by 2 minus 1 into h t X prime. So, this is the derivative with respect to the argument whole square this is log-Sobolev inequality and this is the inequality we want to show.

So, let us number this guy. Let us call this hash. Thus hash is equivalent not equivalent then hash follows from log-Sobolev inequality, if so, we will add and subtract this term. So, this term must equal this term ok. So, this plus this should be 0, if minus expected value of. So, we will take this out. p t square by 4 h t X p t minus 2 h t prime X whole square equals to this term now p t square by 2 p t minus 1 expected value of h t p t minus 1. I am sorry, h t X t minus 1 h dot t X.

(Refer Slide Time: 31:07)



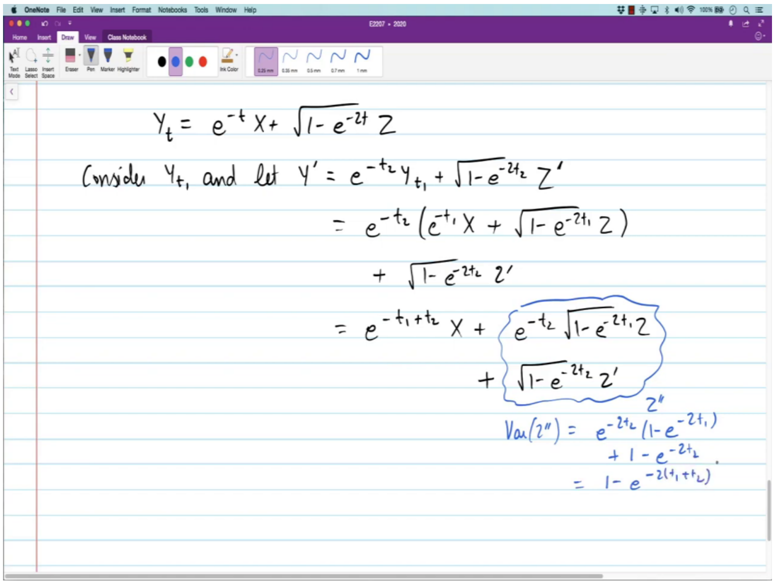
So, if you show these two are equal then this will follow from log-Sobolev inequality. And in fact, the opposite implication the fact that log-Sobolev inequality follows from hypercontractivity also can be shown similarly by somehow using a t for which this function becomes your required function ok. So, we will not discuss that part. We will just talk about this part of the implication ok.

So, how do we show this now? This is what it has come down to. So, this holds if and only if sorry log-Sobolev inequality at a factor of I think there is a factor of 2 that I have to bring in. It is my factor rho right. This is the log-Sobolev inequality. So, this guy here will be just by 2 ok. So, this holds if and only if this guy here expected value of h t X to the power p t minus 1 h d by dt of h t X is exactly equal to minus 2 p t minus 1.

Expected value of ok; that is what this formula is what you have to show ok yeah. So, that is what you have to show. Now, in showing this formula we will use a little bit of heavy machinery ok. So, till now I think this was an elementary calculation and the big observation here is that these two quantities are the same and once we show that we are done our proof is complete.

So, to show this we develop we recall some basic concepts from the from continuous time Markov chains ok. So, we will call this we will number this inequality. This is this identity, this is 1, for us we will show this ok.

(Refer Slide Time: 35:06)



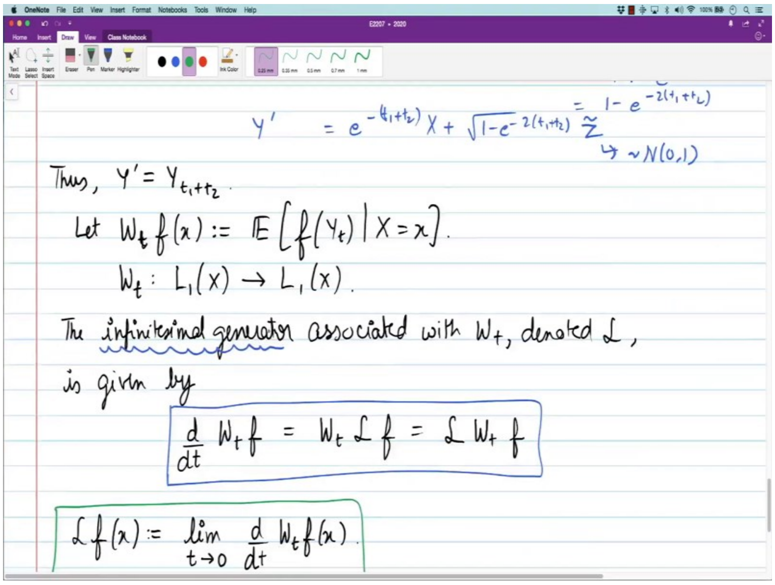
So, a Markov process can be described by its maybe we describe only in the context of the process that we work with. So, for us this Y t is given by e to the power minus t X plus 1 minus e to the power minus 2 t Z, where Z is the Gaussian noise and X can be any random variable we start with a Gaussian X, but in general we can start with any other X tilde ok. This is the process we are looking for looking at.

So in fact, this process can be viewed, it can be viewed as follows. So, consider this random variable Y t and apply look at Y t 1. Consider Y t 1 and let Y prime be given by e to the power minus t 2 Y t 1 plus minus t 2. Let us call it Z prime which is an independent Gaussian noise independent of everything else in the past. So, this becomes e to the power minus t 2 e to the power minus t 1 X plus 1 minus e to the power minus 2 t 1 Z plus square root 1 minus e to the power minus 2 t 2 Z prime.

And this is equal to e to the power minus t 1 plus t 2 X plus e to the power minus t 2 into root 1 minus 2 t 1 Z plus root 1 minus e to the power minus 2 t 2 Z prime. So, if you look at this guy here, this is again a 0 mean Gaussian random variable here with variance. So, its variance is given by let us call this guy Z prime.

So, variance of Z prime prime equals to expected value of Z prime prime square which is just the sum of these two variants because these two are independent. So, its e to the power minus 2 t 2 into 1 minus e to the power minus 2 t 1 plus 1 minus e to the power minus 2 t 2; so, that is 1 minus e to the power minus 2 t 1 plus t 2 ok that is the variance of Z prime prime. This whole thing is what we calling.

(Refer Slide Time: 38:51)



So, essentially I can view this random variable here as this is equal to e to the so, Y which is Y prime is equal to e to the power minus t 1 plus t 2 X plus square root 1 minus 2 t 1 plus t 2. And then there is a random variable which is independent of X and is normal 0, its Gaussian 0, 1 ok. Thus Y prime exactly is equal to has the same distribution as Y t 1 t 2 t 1 plus t 2.

So, this process actually if you can from this you can sort of convince yourself this is just some quick heuristic calculation that. This process is a Markov process and therefore, it has it has a semi group linear operator associated with it which is given by the following. Let W t, this name is not important that I will just define this operator.

W t of f is of x is defined as the conditional expectation when you run through this Markov chain of f given X equal to x and this operator captures this operator W 2 ok, so, this operator W 2. W 2 is a mapping from all I think functions with finite expectation to gain functions with finite expectations alright.

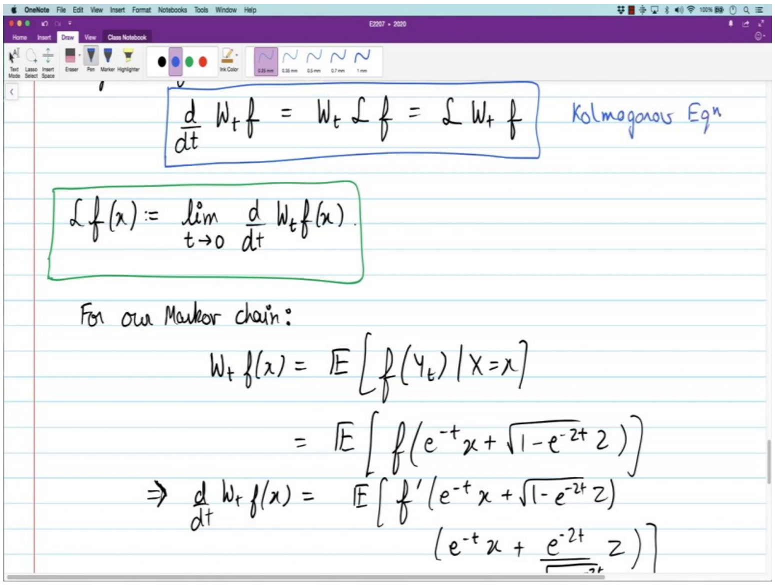
So, this W 2 actually captures the evolution of the Markov chain and in fact, there is another operator that one can associate with the Markov chain. It is the generator or the let us say the some* called the infinite decimal generator associated with W t. Let us denote it by L is given by.

So, this guy L again acts on the same functions is given by if you take W t of a function and you take its derivative then that is equal to almost surely equal to W t of L of f ok which is also the same as L of W t f. There is an operator which will satisfy this equation and that is this infinite decimal operator ok.

Actually a better definition of this L of f is the following. L of f x equals to is defined as so, I will define it for every x, limit t going to 0 d by dt of W t of f x ok. So, this guy this is maybe a better definition. This can also be treated as a definition. The operator for L should be L and W t are related to this equation this some* called Kolmogorovs equation, but perhaps this is a better definition ok.

So, you can associate this generator matrix with it right. Its telling you how the distribution is changing per time roughly that is what it tells you, ok. Now, this operator has some very nice properties ok so. Firstly, we can firstly, we can find this operator associated with our own continuous time Markov chain.

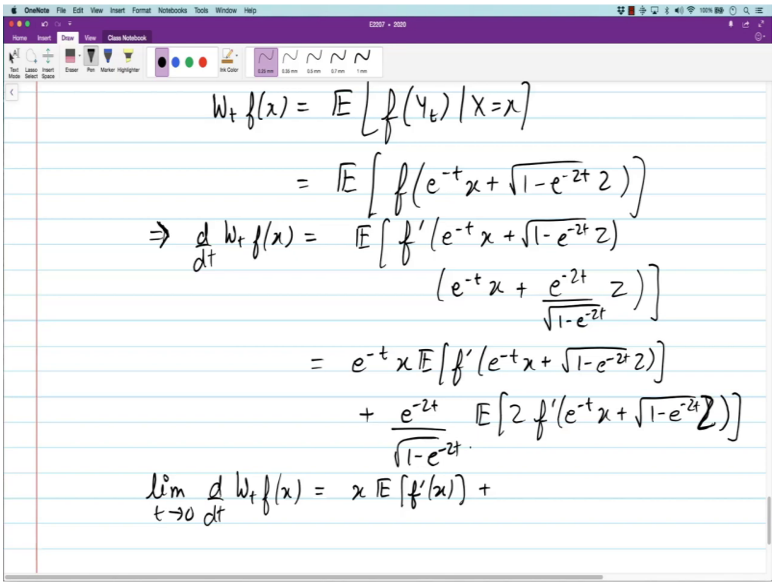
(Refer Slide Time: 43:51)



So, for us for our Markov chain, let us try to compute this derivative. W t of f(x) equal to conditional expectation of f of Y t given X equals to x, which is equal to expected value of f of e to the power minus t x plus square root 1 minus 2 t Z and that is. So, that is the just W of x and therefore, d by dt of W t f x equals to. So, you can take this derivative inside, so, first you will get this x part.

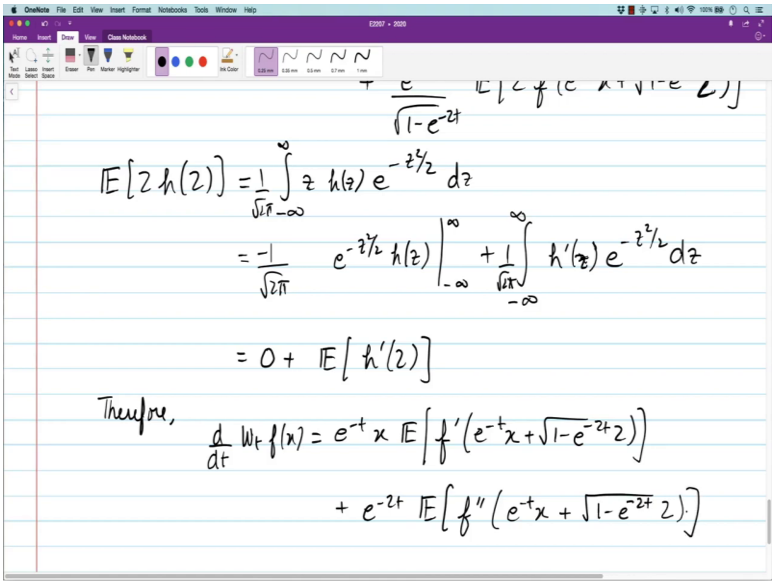
So, it is expected value of f prime to the power minus t x plus square root 1 minus 2 t Z and then the derivative of this part with respect to t. So, that is given by e to the power minus t x plus; what is the derivative of this part that is e to the power minus 2 t x by square root 1 minus e to the power minus 2 t x yeah so, minus minus cancel. So, it becomes plus * Z ok that is what we get. So, this is exactly equal to e to the power minus t x expected value of f prime e to the power minus t x plus square root 1 minus Z.

(Refer Slide Time: 46:09)



Plus e to the power minus 2 t by square root 1 minus 2 t expected value of Z f prime e to the power minus t x plus ok. So, taking the limit t going to 0, actually t goes from 0 from the positive side of this limit. So, this one remains as so, this guy as t goes to 0, this becomes x expected value of f prime x plus; this again goes to 1 here yeah. Actually before taking the limit we will simplify the second term now. So, the second term is actually quite interesting.

(Refer Slide Time: 47:31)



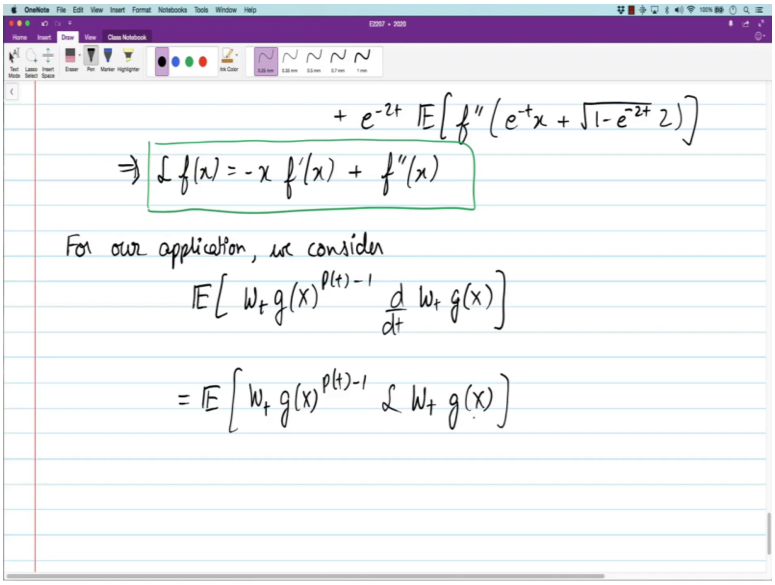
So, if you are given any function let us say h and you look at this guy here this is exactly equal to. So, we will use an integration by part formula here. So, this is Gaussian density minus 1 by root 2 pi z h z e to the power minus z square by 2 dz. Let us use integration by part. So, this is equal to 1 by root 2 pi minus infinity to infinity.

If you take the if you take the integral of this function z e to the power minus z square by 2 then that is just again e to the power minus z square by 2. So, you get this and minus right you get minus of this and then plus h prime z e to the power minus x square by 2 by root 2 pi dz.

Now, turns out that we can restrict our attention to those functions for which so, we have this function f here, an arbitrary function. Turns out by some approximation arguments it is enough to restrict our attention to these functions f which are compactly supported and continuously differentiable. So, they are in fact, we can assume that this guy goes to 0 as z goes to infinity ok.

So, this is 0 and what you are left here is expected value of h prime of Z ok. So, we will apply this formula here. When you take derivative of this with respect to Z now this factor comes out. Therefore, this guy here t by dt W t of f x we already saw the first term that was e to the power minus t x expected value of f prime e to the power minus t x plus this. And now the second term is e to the power minus 2 t expected value of f prime prime.

(Refer Slide Time: 50:23)



So, when the limit as t goes to 0, this thing goes to x expected value of f prime x. So, which is just f prime x and similarly the second term goes to I missed a minus sign somewhere ok, there is a minus sign here right. So, this minus sign is here plus f prime prime x ok. So, this gives us the infinitesimal generator matrix for our Markov chain of interest, this particular operator.

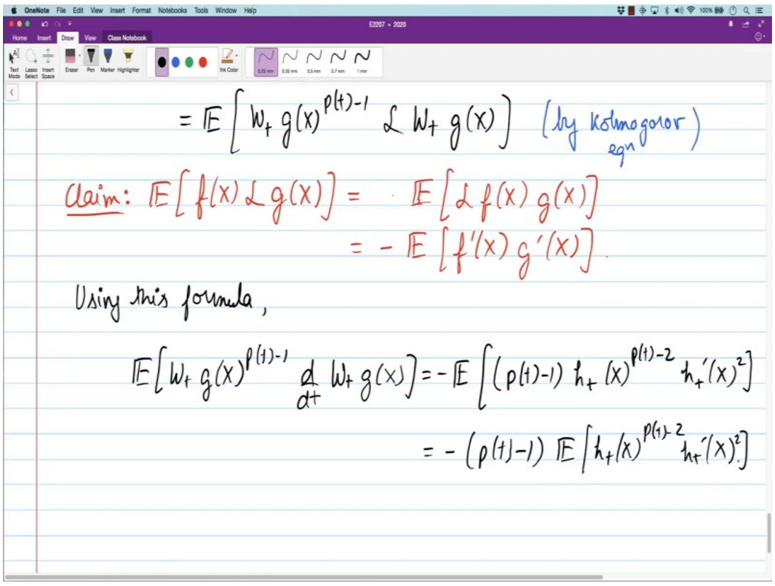
And in fact, this Kolmogorov equation that I wrote here it is quite useful. It tells you the derivative of this function is actually the L of that derivative L of that function and the reason and I have not derived this formula. You can explicitly verify this formula by the way for our case, but this holds in generality the this matrix must satisfy this.

It is easy to show this. How? It is easy to show this formula by just using the fact that by using the fact that this W t of f is this W t is semi group. So, W t plus s equal to W t f and then W s applied to it using that you can establish this formula, but I am not deriving that formula. And if you are; if you really are interested for the specific distribution for our specific Markov chain you can verify this formula the derivative of W t f is L W t of f.

Now, the reason this formula is important for us is because in our calculation we see this guy here d by dt of h t of x where h t of x if you remember was a similar conditional expectation. This is basically W t of g of x. So, this is the derivative of W t of g of x and that must be L of t of x ok.

So, I will name this guy. This is Kolmogorov equation. So, for our application, we consider this guy here. So, that is expected value of W t of g of X to the power p t minus 1 * d by d t of W t of g of X that is what we look at, which is equal to now we apply this Kolmogorov equation. So, this guy here is p t minus 1 L of W t of g X because this derivative is L of W t of g X ok.

(Refer Slide Time: 54:14)



And this is by once again you can just verify it by differentiating ok. So, now what we have shown this. Now, it turns out that for this particular Markov chain now this is the next thing we will need about our particular Markov chain frame expected value of f X L g X for our Markov chain is equal to minus expected value of actually first thing. This is true for any Markov chain when X is the stationary measure.

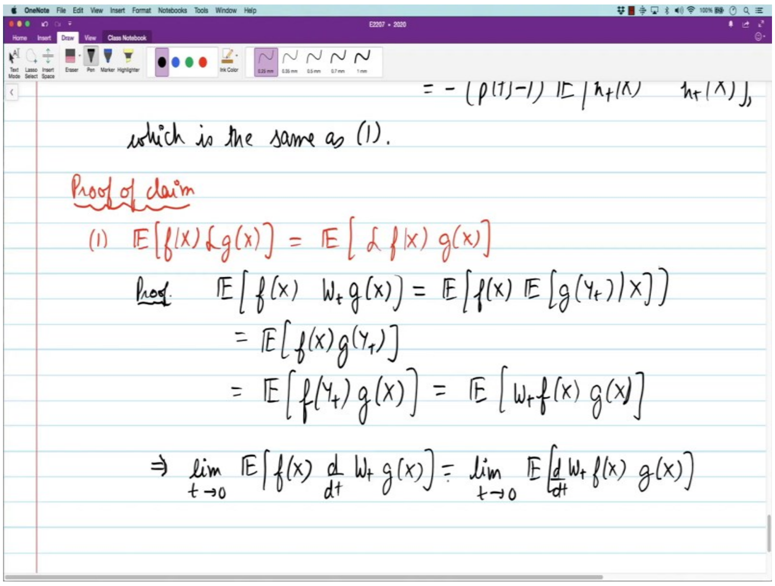
Note that X is the stationary measure for Markov chain because if you start with Gaussian you the distribution always remains to be the same Gaussian ok and for that we must have this, it is true. So, this operator L is so called self adjoint. And in fact, this is equal to expected value of f prime X g prime X ok that is the claim.

We will show this, but using this formula above in this formula with using this formula. This expected value here W t of g X to the power p t minus 1 d by d t W t g X. This is the quantity we were looking at here and this is this we now have just take derivative of two things; derivative of W t g X and derivative here. So, that derivative is that is derivative with respect to X.

So, first we take this derivative with respect to X that becomes p t minus 1 * the derivative of this guy with respect to X. So, W t of g X to the power p t minus 2 * derivative of this with respect to X will which will bring back yeah. Remember we are not using this notation h t X for this. So, this notation is more convenient sorry, we have to switch between these notations so many * and so, its derivative h t prime X.

This is good yeah and then also the derivative of this guy this is just h t. So, derivative of this is again ht. So, you get this. But this is exactly what we wanted to show ok. So, we have shown this and you can go back and check. Here this is same as our 1 ok.

(Refer Slide Time: 57:52)



So, which is the same as 1. So, it only remains to establish the claim. Proof of claim, this is true for any f g for our particular generator matrix for our particular infinite decimal generator here. So, first thing we will show is that this is equal to this ok that is the first thing we will show.

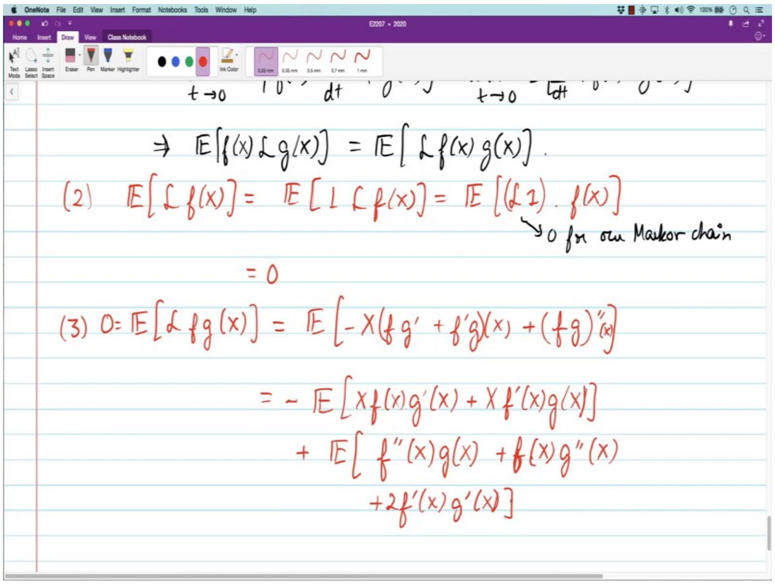
First expected value of f X L g X equals to expected value of L f X g X. So, how do we prove this? Proof: Expected value of f X, this is for any f g not the particular g function that we have about W t of some function g X is equal to expected value of f X expected value of g Y t given X, which is also equal to expected value of f X g Y t.

But note that since these random variables are exchangeable the distribution of X, Y t in this case is the same as distribution of Y t X. In general this will be true for a reversible

Markov chain. So, there we can write it as f of Y t g X ok. You can this symmetry we have mentioned earlier also, but then this becomes expected value of W t f X g X and so, now, what we do is we take derivative with respect to t and take limit t going to 0. This derivative I am taking it inside freely and that assumes some regularity conditions on this f and g something like they are compact and continuously differentiable and all that is true here.

So, we can take it inside no problem. So, this guy then becomes equal to limit t going to 0. And In fact, again this limit can also be taken inside and this implies expected value of f X L g X is equal to expected value of L f X g X, ok.

(Refer Slide Time: 61:12)



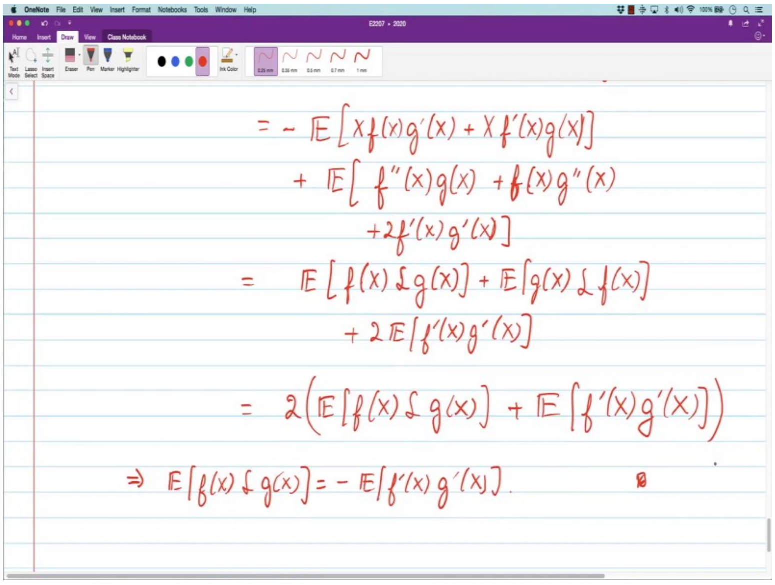
So, this function is symmetric this way ok. So, then 2nd property: Expected value of L of f X is equal to expected value of 1 * L of f X. This first property did not really use, the only thing we use was the reversibility here does not did not use anything else. Induce the Gaussian assumption. This is exactly equal to from the previous property L of 1 * f X and now we use the Gaussian property.

This L of 1 for is basically the derivative of inverse derivative of constant and. So, this is 0, 0 for our Markov chain ok because L X is given by this operator here. We have both f prime and f prime prime will be 0. So, this is 0 and therefore, this is 0.

So, for our Markov chain L of any f X is now the expected value is 0. So, L of X is a 0 mean random variable ok. And now the 3rd step is to apply this property to. So, we consider expect 0 which is expected value of L of f X g X maybe better way to write this. Product function f g x, but this let us write what is L of f g X that is minus X f g prime plus f prime g X plus f g prime prime.

This is equal to minus expected value of X f X g prime X plus X f prime X gx ok plus expected value of f prime prime X g prime X plus f prime X g prime prime X sorry f prime X g X f X g prime prime X plus 2 f prime X g prime X ok.

(Refer Slide Time: 65:34)



So, this is exactly equal to you can simplify this collect the terms with g and collect the term with f. So, this is minus expected value of its not to minus expected value of if you collect these terms; f X L of g X plus expected value of g X L of f X plus 2 * expected value of f prime X g prime x, but these two are equal we have already seen. This in step one above.

So, this is 2 * expected value of f X L g X plus expected value of f prime X g prime X ok and this is 0. Let us do a step here. So, this guy here is equal to minus this ok. So, that that completes the proof. So, this implies though the claim holds. So, expected value of f X L g X is equal to minus expected value of f prime X g prime X ok. So, we worked hard for this last differential inequality, but we are there.

So, many of these steps are very general and this proof is very sophisticated, it just looks like some manipulations, but the way this the things that we have been differentiating in this differential equation that we wrote we exploit a lot this structure in the distribution, the Gaussian distribution.

We realize that the we saw our distribution as a process ok although we wanted to work with a single t, but we saw it as a process which evolves like this. In fact, this process is a name this is the OU process and this OU process has is it is a Markov process and it has a associated with this process is this semi group operator W t.

This family this semi group operator W t and with this Markov process we have this nice infinite decimal generator L f X given by f minus X f prime X plus f prime prime X and an any such infinite decimal generator and the semi group must satisfy this Kolmogorov equation.

So, once we have this relation and once we have this formula we have this nice formula. This is only true for Gaussian case for this particular OU process. We have this nice formula for the OU process ok for f X L g X and basically we use this formula this is I think called the integration by parts formula. And we use this one here to get this equality. And essentially it is this integration by part formula which establishes the equivalence of log-Sobolev inequality and hypercontractivity.

Not note that we have only shown that log-Sobolev inequality implies hypercontractivity. Now, for the other way round we have to have any arbitrary function any arbitrary function here and we can start with our favourite function f X by if we can start with our favourite function f and if will let us, but we should be able to express that function f in this form in this particular form. And that can be done that can always be done.

I think if you set this. I think we can do that by even taking this limit t to be t 2 equal to 0. I think maybe we will be able to do that right, but that is a small technical part no big deal about that part. The fact that, so, the fact that log-Sobolev inequality implies this is the critical part of the proof and yeah.

So, to summarize we have seen that the Gaussian log-Sobolev inequality is also equivalent to Gaussian hypercontractivity. By the way which also gives you the Gaussian strong data processing inequality and all the other implications that we saw for hypercontractivity. So, log-Sobolev inequality for Gaussian case is indeed very powerful.

In fact, many of the results that we saw for the Gaussian case not the stamps inequality, but the equivalence between hypercontractivity and log-Sobolev inequality also hold for the binary symmetric case. So, instead of this Gaussian if you had so, here we were adding this Gaussian noise that is how that is how our correlation was, but you can have a binary symmetric channel where you have some X and you flip it with some probability.

Let us say you have a uniform by 0, 1 valued random variable X and you flip that bit with some probability and symmetrically for both 0 and 1 that is a binary symmetric channel. For that also we can establish an equivalence between Gaussian an equivalence between log-Sobolev inequality and hypercontractivity ok.

This concludes this lecture. And from the next lecture onwards I think Auditya, actually no. So, I will see you in the next lecture where I will talk to you about some more inequalities and I will describe the I will further discuss the connection of log-Sobolev inequalities to concentration bounds. And I will also discuss the I will revisit the transportation information inequalities. See you in the next lecture.