PROFESSOR: So far we had a function, and then we differentiated to get an equation of this type. But now, we're given this equation. And we have to go backwards, want to find the stochastic process that satisfies this equation. So goal is to find a stochastic process, xt, satisfying this equation.

In other words, we want x of t to be the integral of mu ds plus sigma [? dst. ?] The goal is clear. We want that.

And so these type of equations are called differential equations, I hope you already know that, also for PDE, partial differential equations. So even when it's not stochastic, these are not easy problems. At least not easy in the sense that, if you're given an equation, typically you don't expect to have a closed form solution. So even if you find this x, most of the time, it's not in a very good form.

Still, a very important result that you should first note before trying to solve any of the differential equation is that, as long as mu and sigma are reasonable functions, there does exist a solution. And it's unique.

So we have the same correspondence with this PDE. You're given a PDE, or given a differential equation, not a stochastic differential equation, you know that, if you're given a reasonable differential equation, then a solution exists. And it's unique.

So the same principle holds in stochastic world. Now, let me state it formally.

This stochastic equation, star, has a solution that is unique, of course with [? boundary ?] condition, had a solution. And given the initial points-- so if you're given the initial point of a stochastic process, then a solution is unique. Just check, yes-- as long as mu and sigma are reasonable.

One way it can be reasonable, if it satisfies this conditions. These are very technical conditions. But at least let me parse to you what they are.

They say, if you fix a time coordinate and you change x, you look at the difference between the values of mu when you change the second variable and the sigma. Then the change in your function is bounded by the distance between the two points by some constant k. So mu and sigma cannot change too much when you change your space variable a little bit. It can only change up to the distance of how much you change the coordinate. So that's the first condition.

Second condition says, when you go to your x, very similar condition. Essentially it says it cannot blow up too fast, the whole thing.

This one is something about the difference between two values. This one is about how it expands as your space variable grows.

These are technical conditions. And many cases, they will hold. So don't worry too much about the technical conditions.

Important thing here is that, given a differential equation, you don't expect to have a good closed form. But you do expect to have a solution of some form.

OK, let's work out some examples. Here is one of the few stochastic differential equations that can be solved.

This one can be solved. And you already know what x is. But let's pretend you don't know what x is. And let's try to solve it.

I will show you an approach, which can solve some differential equations, some SDEs. But this really won't happen that much. Still, it's like a starting point.

There's x. Going to assume x0 [INAUDIBLE]. And mu [? and ?] sigma are constants.

Just like when solving differential equations, first thing you'll do is just guess, suppose, guess. If you want this to happen, then d of x of t is-- OK. x is just a second variable.

And then these two have to match. That has to be equal to that. That has to be equal to that. So we know that del f over del t, mu of x of t is equal to mu of x.

So we assumed that this is a solution then differentiated that. And then, if that's a solution, these all have to match.

You get this equation. If you look at that, that tells you that f is an exponential function in the x variable. So it's e to the sigma times x. And then we have a constant term, plus some a times a function of t.

The only way it can happen is if it's in this form. So it's exponential function in x. And in the time variable, it's just in constant.

When you fix a t, it's a constant. And when you fix a t and change x, it has to look like an exponential function. It has to be in this form, just by the second equation.

Now, go back to this equation. What you get is partial f over partial t is now a times g prime of t [? prime of ?] f.

AUDIENCE: Excuse me.

PROFESSOR: Mm-hm.

AUDIENCE: That one in second to last line, yeah. So why is it minus mu there at the end?

PROFESSOR: It's equal.

AUDIENCE: Oh, all right.

PROFESSOR: Yeah.

OK, and then let's plug it in. So we have a of g prime t f plus 1/2 of sigma square f equals mu of f. In other words, a of t prime of t is mu minus square. gt is mu times some constant, c1 plus c2.

OK, and then what we got is original function, f of t of x is e to the sigma x plus mu minus 1 over 2 sigma square t plus some constant. And that constant can be

chosen because we have the initial condition, x of 0 equals 0.

That means if t is equal to 0, f 0, 0 is equal to e to the c. That has to be x0.

In different words, this is just x0 times e to the sigma x plus mu minus 1 over 2 sigma square of t. Just as we expected, we got this.

The sum of two stochastic differential equations can be solved by analyzing it. But I'm not necessarily saying that this is a better way to do it than just guessing. Just looking at it, and, you know, OK, it has to be exponential function. You know what? I'll just figure out what these are. And I'll come up with this formula without going through all those analysis.

I'm not saying that's a worse way than actually going through the analysis. Because, in fact, what we did is we kind of already knew the answer and are fitting into that answer. Still, it can be one approach where you don't have a reasonable guess of what the xt has to be, maybe try to break it down into pieces like this and backtrack to figure out the function.

Let me give you one more example where we do have an explicit solution. And then I'll move on and show you how to do when there is no explicit solution or when you don't know how to find an explicit solution. Maybe let's keep that there.

Second equation is called this. What's the difference? The only difference is that you don't have an x here.

So previously, our main drift term also was proportional to the current value. And the error was also dependent on the current value or is proportional to the current value.

But here now the drift term is something like an exponential minus the exponential. But still, it's proportional to the current value. But the error term is just some noise. Irrelevant of what the value is, this has the same variance as the error.

So it's a slightly different-- oh, what is it? It's a slightly different process. And it's

known as Ornstein-Uhlenbeck process.

And this is used to model a mean reverting stochastic process. For alpha greater than 0, notice that if x deviates from 0, this gives a force that drives the stochastic process back to 0. This negatively proportional to your current value.

So yeah, this is used to model some mean reverting stochastic processes. And they first used it to study the behavior of gases. I don't exactly see why. But that's what they say.

Anyway, so this is another thing that can be solved by doing similar analysis. But if you try the same method, it will fail. So as a test function, your guess, initial guess, will be-- or a0 is equal to 1.

Now, honestly, I don't know how to come up with this guess. Probably, if you're really experienced with stochastic differential equations, you'll see some form, like you'll have some feeling on how this process will look like. And then try this, try that, and eventually something might succeed. That's the best explanation I can give. I can't really give you intuition why that's the right guess. Given some stochastic differential equation, I don't know how to say that you should start with this kind of function, this kind of function.

And it was the same when, if you remember how we solved ordinary differential equations or partial differential equations, most of the time there is no good guess. It's only when your given formula has some specific form such a thing happens.

So let's see what happens here. That was given. Now, let's do exactly the same as before. Differentiate it, and let me go slow.

So we have a prime of t. By chain rule, a prime of t and that value, that part will be equal to xt over at. This is chain rule.

So I differentiate that to get a prime t. That stays just as it was. But that can be rewritten as xt divided by at. And then plus at times the differential of that one. And that is just b(t)dB(t).

You don't have to differentiate that once more, even though it's stochastic calculus, because that's a very subtle point. And there's also one exercise about it in your homework. But when you have a given stochastic process written already in this integral form, if we remember the definition of an integral, at least how I defined it, is that it was an inverse operation of a differential. So when you differentiate this, you just get that term.

What I'm trying to say is, there is no term, no term where you have to differentiate this one more. Prime dt, something like that, we don't have this term. This can be confusing. But think about it.

Now, we laid it out and just compare. So minus alpha of x of t is equal to a prime t over at xt. And your second term, sigma dBT is equal to at times bt.

But these two cancel. And with these, at has to be e to the minus alpha t. This is [? a decent ?] explanation.

Now, plug it in here. You get bt. And that's it.

So plug it back in. x of t is e to the minus alpha t of x of 0 plus 0 to t sigma e to the alpha s.

So if this is a variance, expectation is 0. Because that's a Brownian emotion, this term, as we expected, as time passes, goes to 0, exponential decay. And that is kind of hinted by this fact, the mean reversion. So if you start from some value, at least the drift term will go to 0 quite quickly. And then the important term will be the noise term or the variance term.

Any questions?

And I'm really emphasizing a lot of times today, but really you can forget about what I did in the past two boards, this board and the previous board. Because most of the times, it will be useless. And so now I will describe what you'll do if you're given a stochastic differential equation, and you have a computer in front of you.

What if such method fails? And it will fail most of the time. That's when we use these techniques called finite difference method, Monte Carlo simulation, or tree method.

The finite difference method, you probably already saw it, if you took a differential equation course. But let me review it.

This is for PDEs or ODEs, for ODE, PDE, not stochastic differential equations. But it can be adapted to work for stochastic differential equations. So I'll work with an example.

Let u of t be u prime of t plus-- u prime of t be ut plus 2 where u0 is equal to 0.

Now, this has an exact solution. But let's pretend that there is no exact solution. And if you want to do it numerically, you want to find the value of u equals u1 numerically. And here's what you're going to do.

You're going to chop up the interval from 0 to 1 into very fine pieces. So from 0 to 1, chop it down into tiny pieces. And since I'm in front of a blackboard, my tiniest piece will be 1 over 2 and 1. I'll just take two steps. But you should think of it as really repeating this a lot of times.

I'll call my step to be h. So in my case, I'm increasing my steps by 1 over 2 at each time.

So what is u of 1 over 2? Approximately, by Taylor's formula, it's u0 plus 1/2 times u prime of 0. That's Taylor approximation.

OK, u0 we already know. It's given to be equal to 0.

u prime of 0, on the other hand, is given by this differential equation. So it's 1 over 2 times 5 times u0 plus 2. u0 is 0. So we get equal to 1.

Like we have this value equal to 1, approximately. I don't know what happens. But it will be close to 1.

And then for a next thing, u1. This one is, again by Taylor approximation, is u of 1

over 2 plus 1 over 2 u prime of 1 over 2.

And now you know the value of u of 1 over 2, approximate value, by this. So plug it in. You have 1 plus 1 over 2 and, again, 5 times u 1 over 2 plus 2. If you want to do the computation, it should give 9 over 2.

It's really simple. The key idea here is just u prime is given by an equation, this equation. So you can compute it once you know the value of u at that point.

And basically, the method is saying take h to be very small, like 1 over 100. Then you just repeat it 1 over 100 times.

So the equation is the i plus 1 step value can be approximated from the i-th value plus h times u prime of h. Now repeat it and repeat it. And you reach u of 1. And there is a theorem saying, again, if the differential equation is reasonable, then that will approach the true value as you take h to be smaller and smaller.

That's called the finite difference method for differential equations. And you can do the exact same thing for two variables, let's say.

And what we showed was for one variable, finite difference method, we want to find the value of u function u of t. We took values at 0, h, 2h, 3h. Using that, we did some approximation, like that, and found the value.

Now, suppose we want to find, similarly, a two variable function, let's say v of t and x. And we want to find the value of v of 1, 1.

Now the boundary conditions are these. We already know these boundaries.

I won't really show you by example. But what we're going to do now is compute this value based on these two variables.

So it's just the same. Taylor expansion for two variables will allow you to compute this value from these two values. Then compute this from these two, this from these two, and just fill out the whole grid like that, just fill out layer by layer. At some point, you're going to reach this. And then you'll have an approximate value of that.

So you chop up your domain into fine pieces and then take the limit. And most cases, it will work.

Why does it not work for stochastic differential equations? Kind of works, but the only problem is we don't know which value we're looking at, we're interested in. So let me phrase it a little bit differently.

You're given a differential equation of the form dx equals mu dt plus t dB of t and time variable and space variable. Now, if you want to compute your value at time 2h based on value h, in this picture, I told you that this point came from these two points. But when it's stochastic, it could [? depend ?] on everything. You don't know where it came from. This point could have come from here. It could have came from here. It could have came from here, came from here. You don't really know. But what you know is you have a probability distribution.

So what I'm trying to say is now, if you want to adapt this method, what you're going to do is take a sample Brownian motion path. That means just, according to the distribution of the Brownian motion, take one path and use that path. Once we fix a path, once a path is fixed, we can exactly know where each value comes from. We know how to backtrack.

That means, instead of all these possibilities, we have one fixed possibility, like that. So just use that finite difference method with that fixed path. That will be the idea.

Let me do it a little bit more formally. And here is how it works.

If we have a fixed sample path for Brownian motion of Bt, then x at time i plus 1 of h is approximately equal to x at time a of h plus h times dx at that time i of h, just by the exact same Taylor expansion. And then d of x we know to be that is equal to mu of dt plus-- oh, mu of dt is h. So let me write it like that, sigma times d of x dB. And these mu depend on [? their paths, ?] x at i of h dt sigma.

With that, here to here is Taylor expansion. Here to here I'm going to use the differential equation d of x is equal to mu dt plus sigma dBT.

Yes?

AUDIENCE: Do we need that h for [INAUDIBLE]?

PROFESSOR: No, we don't actually. Oh, yeah, I was-- thank you very much. That was what confused me. Yes, thank you very much.

And now we can compute everything. This one, we're assuming that we know the value.

That one can be computed from these two coordinates. Because we now have a fixed path x, we know what x of ih is.

dt, we took it to be h, approximated as h, or time difference. Again, sigma can be computed.

dB now can be computed from Bt. Because we have a fixed path, again, we know that is equal to B of i plus 1 of h minus B of i of h, with it's fixed path.

They're basically exactly the same, if you have a fixed path B. The problem is we don't have a fixed path B. That's where Monte Carlo simulation comes in.

So Monte Carlo simulation is just a way to draw, from some probability distribution, a lot of samples. So now, if you know how to draw samples from the Brownian motions, then what you're going to do is draw a lot of samples. For each sample, do this to compute the value of x0, can compute x of 1.

So, according to a different B, you will get a different value. And in the end, you'll obtain a probability distribution.

So by repeating the experiment, that means just redraw the path again and again, you'll get different values of x of 1. That means you get a distribution of x of 1, obtain distribution of x of 1. And that's it. And that will approach the real distribution of x of 1. So that's how you numerically solve a stochastic differential equation.

Again, there's this finite difference method that can be used to solve differential

equations. But the reason it doesn't apply to stochastic differential equations is because there's underlying uncertainty coming from Brownian motion. However, once you fix a Brownian motion, then you can use that finite difference method to compute x of 1.

So based on that idea, you just draw a lot of samples of the Brownian path, compute a lot of values of x of 1, and obtain a probability distribution of x of 1. That's the underlying principle. And, of course, you can't do it by hand. You need a computer.

Then, what is tree method? That's cool.

Tree method is based on this idea. Remember, Brownian motion is a limit of simple random walk. This gives you a kind of approximate way to draw a sample from Brownian motions. How would you do that?

At time 0, you have 0. At time really tiny h, you'll have plus 1 or minus 1 with same probability. And it goes up or down again, up or down again, and so on. And you know exactly the probability distribution.

So the problem is that it ends up here as 1/2, ends up here as 1/2, 1/4, 1/2, 1/4, and so on. So instead of drawing from this sample path, what you're going to do is just compute the value of our function at these points.

But then the probability distribution, because we know the probability distribution that the path will end up at these points, suppose that you computed all these values here. I draw too many, 1,2, 3, 4, 5. This 1 or 32 probability here. 5 [INAUDIBLE] so 5 over 32, [? times ?] just [? 2 ?] is 10 over 17.

Suppose that some stochastic process, after following this, has value 1 here, 2 here, 3 here, 4, 5, and 6 here. Then, approximately, if you take a Brownian motion, it will have 1 with probability 1 over 32, 2 with probability 5 over 32, and so on.

Maybe I didn't explain it that well. But basically, tree method just says, you can discretize the outcome of the Brownian motion, based on the fact that it's a limit of

simple random walk. So just do the exact same method for simple random walk instead of Brownian motion. And then take it to the limit. That's the principal.

Yeah. Yeah, I don't know what's being used in practice. But it seems like these two are the more important ones. This is more like if you want to do it by hand. Because you can't really do every single possibility. That makes you only a finite possibility.

Any questions? Yeah.

AUDIENCE: So here you said, by [INAUDIBLE] experiment we get variable distribution for x1. I was wondering if we could also get the distribution for not just x1 but also for [? xih. ?]

PROFESSOR: All the intermediate values?

AUDIENCE: Yeah.

PROFESSOR: Yeah, but the problem is we're taking different values of h. So h will be smaller and smaller. But for those values that we took, yeah, we will get [INAUDIBLE] distribution.

AUDIENCE: Right, so we might have distributions for x of d for many different points, right?

PROFESSOR: Yeah.

AUDIENCE: Yeah. So maybe we could uh-- right, OK.

PROFESSOR: But one thing you have to be careful is let's suppose you take h of 1 over 100. Then, this will give you a pretty fairly good approximation for x of 1. But it won't give you a good approximation for x of 1 over 50.

So probably you can also get distribution rates of 1 over 3, 1 over 4. But at some point, the approximation will be very bad. So the key is to choose a right h. Because if you pick h to be too small. You will have a very good approximation to your distribution. But at the same time, it will take too much time to compute it.

Any remarks from a more practical side?

OK, so that's actually all I wanted to say about stochastic differential equations. Really the basic principle is there is such thing called stochastic differential equation. It can be solved. But most of the time, it won't have a closed form formula. And if you want to do it numerically, here are some possibilities. But I won't go any deeper inside.

So the last math lecture I will conclude with heat equation.

Yeah.

AUDIENCE: The mean computations of [INAUDIBLE], some of the derivatives are sort of path-independent or have path-independent solutions so that you basically are looking at say the distribution at the terminal value and that determines the price of the derivative. There are other derivatives where things really are path-dependent, like with options where you have early exercise possibilities. When do you exercise, early or not? Then the tree methods are really good because at each element of the tree you can condition on whatever the path was. So keep that in mind, that when there's path dependence in the problem, you'll probably want to use one of these methods.

PROFESSOR: Thanks.

AUDIENCE: I know that if you're trying to break it down into a sample random walks you can only use [INAUDIBLE]. But I've heard of people trying to use, instead of a binomial, a trinomial tree.

PROFESSOR: Yes, so this statement actually is quite a universal statement. Brownian motion is a limit of many things, not just simple random walk. For example, if you take plus 1, 0, or minus 1 and take it to the limit, that will also converge to the Brownian motion. That will be the trinomial and so on.

And as Peter said, if you're going to use tree method to compute something, that will increase accuracy, if you take more possibilities at each step. Now, there is two ways to increase the accuracy is take more possibilities at each step or take smaller

time scales.

OK, so let's move on to the final topic, heat equation. Heat equation is not a stochastic differential equation, first of all. It's a PDE.

That equation is known as a heat equation where t is like the time variable, x is like the space variable. And the reason we're interested in this heat equation in this course is, if you came to the previous lecture, maybe from [? actually ?] last week, Black-Scholes equation, after change of variables, can be reduced to heat equation. That's one reason we're interested in it.

And this is a really, really famous equation also in physics. So it was known before Black-Scholes equation.

Particularly, this can be a model for equations that model this situation. So you have an infinite bar, very long and thin. It's perfectly insulated. So heat can only travel along the x-axis. And then at time 0, you have some heat distribution. At time 0, you know the heat distribution.

Then this equation tells you the behavior of how the heat will be distributed at time t. So u of t of x, for fixed t, will be the distribution of the heat over the x-axis. That's why it's called the heat equation. That's where the name comes from.

And this equation is very well understood. It does have a closed form solution. And that's what I want to talk about.

OK, so few observations before actually solving it. Remark one, if u1 and u2 satisfies heat equation, then u1 plus u2 also satisfies, also does. That's called linearity.

Just plug it in. And you can figure it out.

More generally that means, if you integrate a family of functions, ds, where us all satisfy the star, then this also satisfies star, as long as you use reasonable function. I'll just assume that we can switch the order of integration and differentiation. So it's the same thing. Instead of summation, I'm taking integration of lot of solutions.

And why is that helpful? This is helpful because now it suffices to solve for-- what is it? Initial condition, u of t of x equals delta, delta function, of 0. That one is a little bit subtle.

The direct delta function is just like an infinite method. x equals 0. It's 0 everywhere else.

And basically, in this example, what you're saying is, at time 0, you're putting like a massive amount of heat at a single point. And you're observing what's going to happen afterwards, how this heat will spread out. If you understand that, you can understand all initial conditions.

Why is that? Because if u sub s t, x-- u0-- is such solution, then integration of-- let me get it right-- u of t of s minus x ds is a solution with initial condition x 0, x. What is it? Sorry about that.

So this is really the key. If you have a solution to the direct delta initial condition, then you can superimpose a lot of those solutions to obtain a solution for arbitrary initial condition.

So this is based on that principle because each of them is now a solution. If you superpose this, then that is a solution. And then if you plug it in, you figure out that actually it has satisfied this initial condition. That was my first observation.

Second observation, second remark, is for the initial value u 0,x equals a direct delta function, u of-- is a solution. So we know the solution for the direct delta part.

First part, we figured out that if we know the solution for the direct delta function, then we can solve it for every single initial value. And for the initial value direct delta, that is the solution that solves the differential equation.

So let me say a few words about this equation, actually one word. Have you seen this equation before? It's the PDF. Of normal distribution. So what does it mean?

It means, in this example, if you have a heat traveling along the x-axis, perfectly

insulated, if you put a massive heat at this 0, at one point, at time 0, then at time t your heat will be distributed according to the normal distribution. In other words, assume that you have a bunch of particle. Heat is just like a bunch of particles, say millions of particles at a single point. And then you grab it. And then time t equals 0 you release it. Now the particle at time t will be distributed according to a normal distribution.

In other words, each particle is like a Brownian motion. So for particle by particle, the location of its particle at time t will be kind of distributed like a Brownian motion. So if you have a massive amount of particles, altogether their distribution will look like a normal distribution. That's [? selective ?] content.

So that's also one way you see the appearance of a Brownian motion inside of this equation. It's like a bunch of Brownian motions happening together at the exact same time.

And now we can just write down the solution. Let me be a little bit more precise.

OK, for the heat equation delta u over delta t, with initial value u of 0, x equals some initial value, let's say v of x, and t greater than equal to 0. The solution is given by integration. u at t of x is equal to e to the minus-- let me get it right.

Basically, I'm just combining this solution to there. Plugging in that here, you get that. So you have a explicit solution, no matter what the initial conditions are, initial conditions are given as, you can find an explicit solution at time t for all x.

That means, once you change the Black-Scholes equation into the heat equation, you now have a closed form solution for it. In that case, it's like a backward heat equation. And what will happen is the initial condition you should think of as a final payout function. The final payout function you integrate according to this distribution. And then you get the value at time t equals 0.

The detail, one of the final project is to actually carry out all the details. So I will stop here.

Anyway, we didn't see how the Black-Scholes equation actually changed into heat equation. If you want to do that project, it will be good to have this in mind. It will help.

Any questions?

OK, so I think that's all I have. I think I'll end a little bit early today.

So that will be the last math lecture for the semester. From now on you'll only have application lectures. There are great lectures coming up, I hope, and I know.

So you should come and really enjoy now. You went through all this hard work. Now it's time to enjoy.