So far we have learned about partial derivatives and how to use them to find minima and maxima of functions of two variables or several variables. And now we are going to try to study, in more detail, how functions of several variables behave, how to compete their variations. How to estimate the variation in arbitrary directions. And so for that we are going to need some more tools actually to study this things. More tools to study functions. Today's topic is going to be differentials. And, just to motivate that, let me remind you about one trick that you probably know from single variable calculus, namely implicit differentiation. Let's say that you have a function y equals f of x then you would sometimes write dy equals f prime of x times dx. And then maybe you would -- We use implicit differentiation to actually relate infinitesimal changes in y with infinitesimal changes in x. And one thing we can do with that, for example, is actually figure out the rate of change dy by dx, but also the reciprocal dx by dy. And so, for example, let's say that we have y equals inverse sin(x). Then we can write x equals sin(y). And, from there, we can actually find out what is the derivative of this function if we didn't know the answer already by writing dx equals cosine y dy. That tells us that dy over dx is going to be one over cosine y. And now cosine for relation to sine is basically one over square root of one minus x^2. And that is how you find the formula for the derivative of the inverse sine function. A formula that you probably already knew, but that is one way to derive it. Now we are going to use also these kinds of notations, dx, dy and so on, but use them for functions of several variables. And, of course, we will have to learn what the rules of manipulation are and what we can do with them. The actual name of that is the total differential, as opposed to the partial derivatives. The total differential includes all of the various causes that can change -- Sorry. All the contributions that can cause the value of your function f to change. Namely, let's say that you have a function maybe of three variables, x, y, z, then you would write df equals f sub x dx plus f sub y dy plus f sub z dz. Maybe, just to remind you of the other notation, partial f over partial x dx plus partial f over partial y dy plus partial f over partial z dz. Now, what is this object? What are the things on either side of this equality? Well, they are called differentials. And they are not numbers, they are not vectors, they are not matrices, they are a different kind of object. These things have their own rules of manipulations, and we have to learn what we can do with them. So how do we think about them? First of all, how do we not think about them? Here is an important thing to know. Important. df is not the same thing as delta f. That is meant to be a number. It is going to be a number once you have a small variation of x, a small variation of y, a small variation of z. These are numbers. Delta x, delta y and delta z are actual numbers, and this becomes a number. This guy actually is not a number. You cannot give it a particular value. All you can do with a differential is express it in terms of other differentials. In fact, this dx, dy and dz, well, they are mostly symbols out there. But if you want to think about them, they are the differentials of x, y and z. In fact, you can think of these differentials as placeholders where you will put other things. Of course, they represent, you know, there is this idea of changes in x, y, z and f. One way that one could explain it, and I don't really like it, is to say they represent infinitesimal changes. Another way to say it, and I think that is probably closer to the truth, is that these things are somehow placeholders to put values and get a tangent approximation. For example, if I do replace these symbols by delta x, delta y and delta z numbers then I will actually get a numerical quantity. And that will be an approximation formula for delta. It will be the linear approximation, a tangent plane approximation. What we can do -- Well, let me start first with maybe something even before that. The first thing that it does is it can encode how changes in x, y, z affect the value of f. I would say that is the most general answer to what is this formula, what are these differentials. It is a relation between x, y, z and f. And this is a placeholder for small variations, delta x, delta y and delta z to get an approximation formula. Which is delta f is approximately equal to fx delta x fy delta y fz delta z. It is getting cramped, but I am sure you know what is going on here. And observe how this one is actually equal while that one is approximately equal. So they are really not the same. Another thing that the notation suggests we can do, and they claim we can do, is divide everything by some variable that everybody depends on. Say, for example, that x, y and z actually depend on some parameter t then they will vary, at a certain rate, dx over dt, dy over dt, dz over dt. And what the differential will tell us then is the rate of change of f as a function of t, when you plug in these values of x, y, z, you will get df over dt by dividing everything by dt in here. The first thing we can do is divide by something like dt to get infinitesimal rate of change. Well, let me just say rate of change. df over dt equals f sub x dx over dt plus f sub y dy over dt plus f sub z dz over dt. And that corresponds to the situation where x is a function of t, y is a function of t and z is a function of t. That means you can plug in these values into f to get, well, the value of f will depend on t, and then you can find the rate of change with t of a value of f. These are the basic rules. And this is known as the chain rule. It is one instance of a chain rule, which tells you when you have a function that depends on something, and that something in turn depends on something else, how to find the rate of change of a function on the new variable in terms of the derivatives of a function and also the dependence between the various variables. Any questions so far? No. OK. A word of warming, in particular, about what I said up here. It is kind of unfortunate, but the textbook actually has a serious mistake on that. I mean they do have a couple of formulas where they mix a d with a delta, and I warn you not to do that, please. I mean there are d's and there are delta's, and basically they don't live in the same world. They don't see each other. The textbook is lying to you. Let's see. The first and the second claims, I don't really need to justify because the first one is just stating some general principle, but I am not making a precise mathematical claim. The second one, well, we know the approximation formula already, so I don't need to justify it for you. But, on the other hand, this formula here, I mean, you probably have a right to expect some reason for why this works. Why is this valid? After all, I first told you we have these new mysterious objects. And then I am telling you we can do that, but I kind of pulled it out of my hat. I mean I don't have a hat. Why is this valid? How can I get to this? Here is a first attempt of justifying how to get there. Let's see. Well, we

Why is this valid? How can I get to this? Here is a first attempt of justifying how to get there. Let's see. Well, we said df is f sub x dx plus f sub y dy plus f sub z dz. But we know if x is a function of t then dx is x prime of t dt, dy is y prime of t dt, dz is z prime of t dt. If we plug these into that formula, we will get that df is f sub x times x prime t dt plus f sub y y prime of t dt plus f sub z z prime of t dt. And now I have a relation between df and dt. See, I got df equals sometimes times dt. That means the rate of change of f with respect to t should be that coefficient. If I divide by dt then I get the chain rule. That kind of works, but that shouldn't be completely satisfactory. Let's say that you are a true skeptic and you don't believe in differentials yet then it is maybe not very good that I actually used more of these differential notations in deriving the answer. That is actually not how it is proved. The way in which you prove the chain rule is not this way because we shouldn't have too much trust in differentials just yet. I mean at the end of today's lecture, yes, probably we should believe in them, but so far we should be a little bit reluctant to believe these kind of strange objects telling us weird things. Here is a better way to think about it. One thing that we have trust in so far are approximation formulas. We should have trust in them. We should believe that if we change x a little bit, if we change y a little bit then we are actually going to get a change in f that is approximately given by these guys. And this is true for any changes in x, y, z, but in particular let's look at the changes that we get if we just take these formulas as function of time and change time a little bit by delta t. We will actually use the changes in x, y, z in a small time delta t. Let's divide everybody by delta t. Here I am just dividing numbers so I am not actually playing any tricks on you. I mean we don't really know what it means to divide differentials, but dividing numbers is something we know. And now, if I take delta t very small, this guy tends to the derivative, df over dt. Remember, the definition of df over dt is the limit of this ratio when the time interval delta t tends to zero. That means if I choose smaller and smaller values of delta t then these ratios of numbers will actually tend to some value, and that value is the derivative. Similarly, here delta x over delta t, when delta t is really small, will tend to the derivative dx/dt. And similarly for the others. That means, in particular, we take the limit as delta t tends to zero and we get df over dt on one side and on the other side we get f sub x dx over dt plus f sub y dy over dt plus f sub z dz over dt. And the approximation becomes better and better. Remember when we write approximately equal that means it is not quite the same, but if we take smaller variations then actually we will end up with values that are closer and closer. When we take the limit, as delta t tends to zero, eventually we get an equality. I mean mathematicians have more complicated words to justify this statement. I will spare them for now, and you will see them when you take analysis if you go in that direction. Any questions so far? No. OK. Let's check this with an example. Let's say that we really don't have any faith in these things so let's try to do it. Let's say I give you a function that is x ^2 y z. And let's say that maybe x will be t, y will be e^t and z will be sin(t). What does the chain rule say? Well, the chain rule tells us that dw/dt is, we start with partial w over partial x, well, what is that? That is 2xy, and maybe I should point out that this is w sub x, times dx over dt plus -- Well, w sub y is x squared times dy over dt plus w sub z, which is going to be just one, dz over dt. And so now let's plug in the actual values of these things. x is t and y is e^t, so that will be 2t e to the t, dx over dt is one plus x squared is t squared, dy over dt is e over t, plus dz over dt is cosine t. At the end of calculation we get 2t e to the t plus t squared e to the t plus cosine t. That is what the chain rule tells us. How else could we find that? Well, we could just plug in values of x, y and z, x plus w is a function of t, and take its derivative. Let's do that just for verification. It should be exactly the same answer. And, in fact, in this case, the two calculations are roughly equal in complication. But say that your function of x, y, z was much more complicated than that, or maybe you actually didn't know a formula for it, you only knew its partial derivatives, then you would need to use the chain rule. So, sometimes plugging in values is easier but not always. Let's just check quickly. The other method would be to substitute. W as a function of t. Remember w was x^2y z. x was t, so you get t squared, y is e to the t, plus z was sine t. dw over dt, we know how to take the derivative using single variable calculus. Well, we should know. If we don't know then we should take a look at 18.01 again. The product rule that will be derivative of t squared is 2t times e to the t plus t squared time the derivative of e to the t is e to the t plus cosine t. And that is the same answer as over there. I ended up writing, you know, maybe I wrote slightly more here, but actually the amount of calculations really was pretty much the same. Any questions about that? Yes? What kind of object is w? Well, you can think of w as just another variable that is given as a function of x, y and z, for example. You would have a function of x, y, z defined by this formula, and I call it w. I call its value w so that I can substitute t instead of x, y, z. Well, let's think of w as a function of three variables. And then, when I plug in the dependents of these three variables on t, then it becomes just a function of t. I mean, really, my w here is pretty much what I called f before. There is no major difference between the two. Any other questions? No. OK. Let's see. Here is an application of what we have seen. Let's say that you want to understand actually all these rules about taking derivatives in single variable calculus. What I showed you at the beginning, and then erased, basically justifies how to take the derivative of a reciprocal function. And for that you didn't need multivariable calculus. But let's try to justify the product rule, for example, for the derivative. An application of this actually is to justify the product and quotient rules. Let's think, for example, of a function of two variables, u and v, that is just the product uv. And let's say that u and v are actually functions of one variable t. Then, well, d of uv over dt is given by the chain rule applied to f. This is df over dt. So df over dt should be f sub q du over dt plus f sub v plus dv over dt. But now what is the partial of f with respect to u? It is v. That is v du over dt. And partial of f with respect to v is going to be just u, dv over dt. So you get back the usual product rule. That is a slightly complicated way of deriving it, but that is a valid way of understanding how to take the derivative of a product by thinking of the product first as a function of variables, which are u and v. And then say, oh, but u and v were actually functions of a variable t. And then you do the differentiation in two stages using the chain rule. Similarly, you can do the quotient rule just for practice. If I give you the function g equals u of v. Right now I am thinking of it as a function of two variables, u and v. U and v themselves are actually going to be functions of t. Then, well, dg over dt is going to be partial g, partial u. How much is that? How much is partial g, partial u? One over v times du over dt plus -- Well, next we need to have partial g over partial v. Well, what is the derivative of this with respect to v? Here we need to know how to differentiate the inverse. It is minus u over v squared times dv over dt. And that is actually the usual quotient rule just written in a slightly different way. I mean,

squared times u over a. And that is actually the usual question rule just written in a slightly different way. I mean, just in case you really want to see it, if you clear denominators for v squared then you will see basically u prime times v minus v prime times u. Now let's go to something even more crazy. I claim we can do chain rules with more variables. Let's say that I have a quantity. Let's call it w for now. Let's say I have quantity w as a function of say variables x and y. And so in the previous setup x and y depended on some parameters t. But, actually, let's now look at the case where x and y themselves are functions of several variables. Let's say of two more variables. Let's call them u and v. I am going to stay with these abstract letters, but if it bothers you, if it sounds completely unmotivated think about it maybe in terms of something you might now. Say, polar coordinates. Let's say that I have a function but is defined in terms of the polar coordinate variables on theta. And then I know I want to switch to usual coordinates x and y. Or, the other way around, I have a function of x and y and I want to express it in terms of the polar coordinates r and theta. Then I would want to know maybe how the derivatives, with respect to the various sets of variables, related to each other. One way I could do it is, of course, to say now if I plug the formula for x and the formula for y into the formula for f then w becomes a function of u and v, and it can try to take partial derivatives. If I have explicit formulas, well, that could work. But maybe the formulas are complicated. Typically, if I switch between rectangular and polar coordinates, there might be inverse trig, there might be maybe arctangent to express the polar angle in terms of x and y. And when I don't really want to actually substitute arctangents everywhere, maybe I would rather deal with the derivatives. How do I do that? The question is what are partial w over partial u and partial w over partial v in terms of, let's see, what do we need to know to understand that? Well, probably we should know how w depends on x and y. If we don't know that then we are probably toast. Partial w over partial x, partial w over partial y should be required. What else should we know? Well, it would probably help to know how x and y depend on u and v. If we don't know that then we don't really know how to do it. We need also x sub u, x sub v, y sub u, y sub v. We have a lot of partials in there. Well, let's see how we can do that. Let's start by writing dw. We know that dw is partial f, well, I don't know why I have two names, w and f. I mean w and f are really the same thing here, but let's say f sub x dx plus f sub y dy. So far that is our new friend, the differential. Now what do we want to do with it? Well, we would like to get rid of dx and dy because we like to express things in terms of, you know, the question we are asking ourselves is let's say that I change u a little bit, how does w change? Of course, what happens, if I change u a little bit, is y and y will change. How do they change? Well, that is given to me by the differential. dx is going to be, well, I can use the differential again. Well, x is a function of u and v. That will be x sub u times du plus x sub v times dv. That is, again, taking the differential of a function of two variables. Does that make sense? And then we have the other guy, f sub y times, what is dy? Well, similarly dy is y sub u du plus y sub v dv. And now we have a relation between dw and du and dv. We are expressing how w reacts to changes in u and v, which was our goal. Now, let's actually collect terms so that we see it a bit better. It is going to be f sub x times x sub u times f sub y times y sub u du plus f sub x, x sub v plus f sub y y sub v dv. Now we have dw equals something du plus something dv. Well, the coefficient here has to be partial f over partial u. What else could it be? That's the rate of change of w with respect to u if I forget what happens when I change v. That is the definition of a partial. Similarly, this one has to be partial f over partial v. That is because it is the rate of change with respect to v, if I keep u constant, so that these guys are completely ignored. Now you see how the total differential accounts for, somehow, all the partial derivatives that come as coefficients of the individual variables in these expressions. Let me maybe rewrite these formulas in a more visible way and then re-explain them to you. Here is the chain rule for this situation, with two intermediate variables and two variables that you express these in terms of. In our setting, we get partial f over partial u equals partial f over partial x time partial x over partial u plus partial f over partial y times partial y over partial u. And the other one, the same thing with v instead of u, partial f over partial x times partial x over partial v plus partial f over partial u partial y over partial v. I have to explain various things about these formulas because they look complicated. And, actually, they are not that complicated. A couple of things to know. The first thing, how do we remember a formula like that? Well, that is easy. We want to know how f depends on u. Well, what does f depend on? It depends on x and y. So we will put partial f over partial x and partial f over partial y. Now, x and y, why are they here? Well, they are here because they actually depend on u as well. How does x depend on u? Well, the answer is partial x over partial u. How does y depend on u? The answer is partial y over partial u. See, the structure of this formula is simple. To find the partial of f with respect to some new variable you use the partials with respect to the variables that f was initially defined in terms of x and y. And you multiply them by the partials of x and y in terms of the new variable that you want to look at, v here, and you sum these things together. That is the structure of the formula. Why does it work? Well, let me explain it to you in a slightly different language. This asks us how does f change if I change u a little bit? Well, why would f change if u changes a little bit? Well, it would change because f actually depends on x and y and x and y depend on u. If I change u, how quickly does x change? Well, the answer is partial x over partial u. And now, if I change x at this rate, how does that have to change? Well, the answer is partial f over partial x times this guy. Well, at the same time, y is also changing. How fast is y changing if I change u? Well, at the rate of partial y over partial u. But now if I change this how does f change? Well, the rate of change is partial f over partial y. The product is the effect of how you change it, changing u, and therefore changing f. Now, what happens in real life, if I change u a little bit? Well, both x and y change at the same time. So how does f change? Well, it is the sum of the two effects. Does that make sense? Good. Of course, if f depends on more variables then you just have more terms in here. OK. Here is another thing that may be a little bit confusing. What is tempting? Well, what is tempting here would be to simplify these formulas by removing these partial x's. Let's simplify by partial x. Let's simplify by partial y. We get partial f over partial u equals partial f over partial u plus partial f over partial u. Something is not working properly. Why doesn't it work? The answer is precisely because these are partial derivatives. These are not total derivatives. And so you cannot simplify them in that way. And that is actually the reason why we use this curly d rather than a straight d. It is to remind us, beware, there are these simplifications that we can do with straight d's that are not legal here. Somehow, when you have a partial derivative, you must resist the urge of simplifying things. No simplifications in here. That is the simplest formula

you can get. Any questions at this point? No. Yes? When would you use this and what does it describe? Well, it is basically when you have a function given in terms of a certain set of variables because maybe there is a simply expression in terms of those variables. But ultimately what you care about is not those variables, z and y, but another set of variables, here u and v. So x and y are giving you a nice formula for f, but actually the relevant variables for your problem are u and v. And you know x and y are related to u and v. So, of course, what you could do is plug the formulas the way that we did substituting. But maybe that will give you very complicated expressions. And maybe it is actually easier to just work with the derivates. The important claim here is basically we don't need to know the actual formulas. All we need to know are the rate of changes. If we know all these rates of change then we know how to take these derivatives without actually having to plug in values. Yes? Yes, you could certain do the same things in terms of t. If x and y were functions of t instead of being functions of u and v then it would be the same thing. And you would have the same formulas that I had, well, over there I still have it. Why does that one have straight d's? Well, the answer is I could put curly d's if I wanted, but I end up with a function of a single variable. If you have a single variable then the partial, with respect to that variable, is the same thing as the usual derivative. We don't actually need to worry about curly in that case. But that one is indeed special case of this one where instead of x and y depending on two variables, u and v, they depend on a single variable t. Now, of course, you can call variables any name you want. It doesn't matter. This is just a slight generalization of that. Well, not quite because here I also had a z. See, I am trying to just confuse you by giving you functions that depend on various numbers of variables. If you have a function of 30 variables, things work the same way, just longer, and you are going to run out of letters in the alphabet before the end. Any other questions? No. What? Yes? If u and v themselves depended on another variable then you would continue with your chain rules. Maybe you would know to express partial x over partial u in terms using that chain rule. Sorry. If u and v are dependent on yet another variable then you could get the derivative with respect to that using first the chain rule to pass from u v to that new variable, and then you would plug in these formulas for partials of f with respect to u and v. In fact, if you have several substitutions to do, you can always arrange to use one chain rule at a time. You just have to do them in sequence. That's why we don't actually learn that, but you can just do it be repeating the process. I mean, probably at that stage, the easiest to not get confused actually is to manipulate differentials because that is probably easier. Yes? Curly f does not exist. That's easy. Curly f makes no sense by itself. It doesn't exist alone. What exists is only curly df over curly d some variable. And then that accounts only for the rate of change with respect to that variable leaving the others fixed, while straight df is somehow a total variation of f. It accounts for all of the partial derivatives and their combined effects. OK. Any more questions? No. Let me just finish up very quickly by telling you again one example where completely you might want to do this. You have a function that you want to switch between rectangular and polar coordinates. To make things a little bit concrete. If you have polar coordinates that means in the plane, instead of using x and y, you will use coordinates r, distance to the origin, and theta, the angles from the x-axis. The change of variables for that is x equals r cosine theta and y equals r sine theta. And so that means if you have a function f that depends on x and y, in fact, you can plug these in as a function of r and theta. Then you can ask yourself, well, what is partial f over partial r? And that is going to be, well, you want to take partial f over partial x times partial x partial r plus partial f over partial y times partial y over partial r. That will end up being actually f sub x times cosine theta plus f sub y times sine theta. And you can do the same thing to find partial f, partial theta. And so you can express derivatives either in terms of x, y or in terms of r and theta with simple relations between them. And the one last thing I should say. On Thursday we will learn about more tricks we can play with variations of functions. And one that is important, because you need to know it actually to do the p-set, is the gradient vector. The gradient vector is simply a vector. You use this downward pointing triangle as the notation for the gradient. It is simply is a vector whose components are the partial derivatives of a function. I mean, in a way, you can think of a differential as a way to package partial derivatives together into some weird object. Well, the gradient is also a way to package partials together. We will see on Thursday what it is good for, but some of the problems on the p-set use it.