[SQUEAKING]

[RUSTLING]

[CLICKING]

BERTHOLD HORN: We talked some about noise gain. And in the 1D case where we have one unknown and one measurement, it was pretty simple because we just looked at the inverse of the slope. So we had some small increment here and the relationship with the small increment over here. So x, in this case, is the unknown quantity, and y is the measurement we're making, and we're trying to estimate x. And obviously, if this curve is very low slope, then a small error here can be amplified into a large error there.

So that's very simple in the 1D case. But of course, we'll be looking at more complicated cases where we're trying to estimate three-dimensional quantities. And so I want to take a little detour now, and talk about at least the 2D case, which comes up in the optical mouse and some other things that we'll be talking about pretty soon. So we're going to talk a little bit about eigenvectors an eigenvalues.

And if you are familiar with them, then just go to sleep for the next 10 minutes, but I think it's important to get everyone up to speed on this. So what are these? Well, if we take matrix, m, and we multiply some vector, we're going to get a new vector, which is a different size, typically, and pointing in a different direction. And so an interesting question is, do we ever have a situation where the vector we get is pointing in the same direction as the vector that we multiply the matrix with? And so those are considered special.

And in some sense, they are characteristic of that matrix, and so sometimes they're called characteristic vectors with characteristic values. And I guess "eigen" is the German word for "own". It's the vectors and values that the matrix owns. And so we're looking at a situation where the vector we get is parallel to the vector that we pump into the matrix, and it may be a different size. So that's the definition. And we're mostly going to focus on real symmetric matrices.

We can generalize this. But for the moment, that's all we need. And so obviously this isn't going to be true for typical vectors. Also notice that the scale doesn't matter because if e is an eigenvector and I multiply it by k, the equation still holds. It's still an eigenvector. So for many purposes, we'll either just ignore the magnitude or make it a unit vector for simplicity. So the size doesn't matter. And by the way, that includes multiplication by minus 1. So if V is an eigenvector, then minus V is also an eigenvector.

OK. So then the question is, how do we find these things? And how many are there? And things of that nature. Well, one thing we can do is we can try and solve this equation by bringing this over to the other side. So I is just the identity matrix of the appropriate size, m is m by n, then I will be n by n. And so this is exactly the same equation, M times e minus lambda times e is 0. OK. So there's a set of linear equations, and we've been bragging about how linear equations are so easy.

And obviously, what we can say is that. We just invert that matrix and we multiply it by-- and what do we get? 0, right? So that's called the trivial solution. And if we can do this, then that's not going to be helpful. So what is it we're looking? We're looking for the case where actually this inverse doesn't exist. So usually, we're very keen to make sure that inverses exist. In this case, we're not. So first of all, terminology.

Here we have a set of equations that are homogeneous. That means that the right-hand side is 0. And amongst other things, that means that if you have a solution, then any multiple of that solution is also a solution, which is not the case normally when we're solving linear equations. There's a unique solution, and any multiple of it is not a solution, so any non-one multiple. OK. So what we're looking for, then, is that we have a singular matrix.

So that means that the determinant of that matrix is 0. And so what does that mean? Unfortunately, the determinant is this complicated, messy thing. But if you think about it, we can say something about the determinant. But if we write this thing out, 3 minus lambda, what we've done is we've taken the matrix, m, which is some n by n real symmetric matrix, and we subtracted lambda off the diagonal.

And now we're talking about the properties of this matrix, and we actually want this one to be singular because otherwise that equation has nothing but the obvious solution. So what's the determinant of this? Well, I don't know. You might remember that the determinant involves taking this and multiplying it by the determinant of that matrix plus this times the determinant of some submatrix. Or you may remember it as-- take one from column A, one from column B, and one from column C.

And depending on whether the number of switches in direction is even or odd, you add a minus sign. Don't need to remember the exact formula, the key thing is that we're taking something from here, and something from there, and so on. Never repeating columns or rows, and there are large numbers of ways of doing that, which is why it's actually computationally expensive. But the key thing is that we can get a term as large as this one along the diagonal in terms of lambda. So if we look at the product of all of the terms on the diagonal, that'll be part of the determinant, one of many parts.

And it's got lambda to the n in it, right? We're taking m11 minus lambda, m22 minus lambda, and so on. So what's the determinant? Well, it's a polynomial in lambda, and it's an n-th order polynomial in lambda. And polynomials of n-th order have how many roots? N, good. So we're going to n roots. And so that means that not only is there a solution to this, but there are going to be n solutions, and those are the things that we'll be looking for.

So just to make this concrete, let's look at a very simple example, also because that's the one that we're dealing with right now with, say, the optical mouse. So we already mentioned that-- just saying that the solution is unstable that the noise gain is high isn't enough when we're dealing with multidimensional problems. It's fine for-- this is it. The noise gain is a scalar. We're done. But in the case of the optical mouse, we're recovering u and v, so it's a three-dimensional problem, and it may very well be that the error in certain directions is very different from the error in other directions.

So we want to have a more nuanced picture. We can have a kind of a gross statement that says, OK, it's bad if the determinant is small. That's a good start, but that's a scalar constraint, and it doesn't tell you that actually if you move in this direction, you have very good knowledge of the motion. But if you move in some other direction, you don't. OK. So let's look at this. Well, following what we did over there, that's the condition for that set of homogeneous equations to have a non-trivial solution.

So homogeneous equations, we don't run across a whole lot. We almost always have inhomogeneous equations, then they have solution unless the determinant is 0. And homogeneous equations had some very interesting properties, and we'll just look at them a little bit. OK. So what is the determinant of that? Well, it's just that times that minus that times that. And if I multiply that out, I get a second order polynomial in lambda that I can solve using the usual formula for quadratic.

Actually, I simplified something here. So before I simplified it, this was a plus c squared minus 4ac minus b squared. So it's B squared, which is now this term, minus 4 times that term. And then when I multiply that out and rearrange the terms, I get this. So those are the eigenvalues. So there will be two vectors that have this property that when you multiply that matrix by the vector, you get a vector in the same direction. And the length will be changed by this quantity, that's how much they'll be magnified or demagnified.

And we're very interested in that because typically what's on the right-hand side is the measurement, including the error, and so these eigenvalues will determine how much the error will be magnified. So just for cutting down on the writing, let me give this a name so that I can abbreviate things. OK. So those are the eigenvalues, and I'll be interested in how big they are. Now in our case with optical mouse, these were integrals of ex squared and integrals of ex, ey, and stuff like that, and then we can plug that into the formula to find out what those eigenvalues actually are.

Now in practice, if someone's finding eigenvalues and eigenvectors, they just go to MATLAB or whatever. But I want you to get a feel for this at this trivial level of 2 by 2. And then if you're going to do a 13 by 13 matrix, of course you're not going to do it by hand, but it's useful to do this and get a clear idea of what it's all about. OK. So those are the eigenvalues, what about the eigenvectors? So they are going to be special directions in which this property holds.

And so how do we solve for those? Well, now we have to solve the homogeneous equations. So we got this matrix, a minus lambda b. So this is going to be our eigenvector, the one that corresponds to the lambda. We have two choices that we pick. And so we're now assuming that we'll be plugging in a specific value of lambda, namely one of these two. So what does this mean? Well, it means that there's a relationship between this vector and that vector, and between this vector and that vector, namely they're orthogonal.

So why is that? Well, this 0 up here comes from multiplying this row vector by this column vector. So it's a minus lambda times x plus by is 0. That's the first equation. And then the other 0 comes from multiplying this row vector by that vector. So I get bx. And so I can think of this as a dot product, and I can think of this then as vectors being orthogonal to one another. So this means that a minus lambda b is perpendicular to xy.

And by the way, b minus c lambda, c minus lambda, is also perpendicular to x and y. So in solving these homogeneous equations, I can note that basically I'm saying that the solution is perpendicular to the rows. And so I can write down the answer very easily, particularly in the 2 by 2 case. All I need to do is find something that's perpendicular to a minus lambda b, and so how about this?

If I multiply that by a minus lambda b, I just get two terms with opposite sides, and they cancel out. So there's an eigenvector. How big is it? As we said, in a sense it doesn't matter because any multiple of an eigenvector is also an eigenvector. OK. So that's one of them. Actually, that's two of them because we can plug-in the two different values of lambda and we get two different vectors. But why focus on the first row? This should be true of the other row as well.

So let's try it on the other row. Well, on the other row-- here's a perpendicular to the other row. If I take the dot product of that with the second row, I get two terms that are equal in magnitude and opposite in sign. So they cancel it out. So that's also an eigenvector. And now that's getting a little confusing because now I can plug-in the two different values of lambda, so I get four eigenvectors.

Well, it turns out those actually point in the same direction, and they are the same if lambda is given by that expression up there. And I'm not going to bore you with algebra, it's pretty straightforward to show that. So altogether, we can actually write the result-- if we want the unit, eigenvector, we can normalize this. So we just divide by the square root of the sum of squares of these two terms, and we get something like this.

And the plus minus applies to the two different solutions, the top sign always corresponding to one case-- I should mention that this is all written up in this little four-page pamphlet that's on the materials. So you can check on it there. I'm leaving out some details. OK. You might think, well, gee, this is what it is for 2 by 2. This must be pretty complicated if I get to 10 by 10, and it is, and that's why typically you use some prepackaged tools.

But it's good to see how this works. So for an m by n matrix, real symmetric matrix, we're going to have n of these, typically, and there'll be corresponding eigenvalues and eigenvectors. And they allow us to talk about error amplification, and let's see how that works, what's the relationship to error amplification. So I think I mentioned this before, but we were thinking of our vectors as column vectors, and we can alternatively, or equivalently, think of them as skinny matrices.

And so I can, on the one hand, write a dot product this way, or I can write it like this. And so what is that? Well, a1, a2. And obviously, if I multiply these two skinny matrices, I get a scalar, which is just the product of a1 and b1 plus the product a2 and b2, and so on. So that's just the dot product. And that's a convenient notation, and we extend that to matrices as well. So in our context over here, what we can show is-- there's a line of algebra which I'm leaving out, which is in this pamphlet. So you can verify it.

So that's surprising. What this is saying is, OK, I can the 1, multiply the matrix by that, and then I get something new, and I take the dot product with the other eigenvector, and it's the same as flipping it. Like, I don't know, m could be a rotation. So it's like I'm rotating e1 and then taking the dot product with e2, and that's the same as rotating e2 and taking the dot product with e1. Well, if that's the case, then we can show that these have to actually be orthogonal. So that's where we're going. We want to show that these eigenvectors are actually orthogonal.

So let's first look at this. How do I know that? Well, because the whole point of these eigenvectors, eigenvalues was that Me1 is going to be in the direction e1, but a different length multiplied by the eigenvalue. And this one here is going to be-- Me2 is the vector e2 just multiplied by the eigenvector. So the part that takes a few steps is proving that these are actually equal, but it's in the paper. OK. Now what does this say?

Well, I can gather this up, and that tells me that e1 dot e2 is 0. So that means they're perpendicular. Well, there's another thing that can happen. If lambda 2 is the same as lambda 1, then that doesn't follow. So when I take the roots of this polynomial, and the roots are all different, that means that all of the eigenvectors are orthogonal. So that's what that says. And if they're not all different, if there's a multiplicity, it turns out I can pick the eigenvectors to be orthogonal.

So the example would be-- the eigenvectors are in a plane, and I can pick any two vectors in that plane. They'll all work, all of the vectors in that plane, eigenvectors, but I can pick two of them that are orthogonal. So yes, if two of the roots happen to be the same, then this doesn't force the eigenvectors to be orthogonal, but I can just pick out of all the possible ones too that are. And the idea is to construct a whole coordinate system. So I've got all of these orthogonal vectors, and they, of course, define a basis for the vector space.

So in the 2 by 2 case, what happens? Well, I get a eigenvector here, and the other one better be perpendicular to it. And of course, I can use those to talk about points in the plane just as I can use my original x and y-axes. They form a basis. And so that means that I can write any vector as a weighted sum of these eigenvectors because they form a basis. So what are these alpha I's? How do I find them?

Well, that's pretty easy because I just do v dot e-- I don't know, let's call it j. We don't want to get in conflict with I, which is a dummy variable in the sum. And we said that the eigenvectors are orthogonal to each other. So that means that all of these dot products are 0 except one, which is ei dot ei. And so this is going to give me alpha I, right? If I pick the unit vector version.

OK. So this is the sum, but I don't care about most of the terms because they're multiplied by 0, this dot perfect. The only one I care about is where these two are the same vectors. And in that case, if I pick unit vectors, such as 1, so I just get alpha I. So it's very easy to re-express any vector in terms of the eigenvectors. OK. So then let's look at Mv is sigma alpha i, M--

So now we get to the juicy part, the conclusion here, which is that if we take an arbitrary vector measurement, and we multiply the matrix by that measurement to obtain our unknown variables, what happens is that different components are magnified by different amounts. So those directions are special in that along those directions, we know how much the error is magnified. And so in this case, obviously if we have a large eigenvalue, the component in that direction will be magnified a lot.

If we have a small eigenvalue, it will be minified, it'll be diminished, and we'll be happy. OK. So that's the connection to the error gained. But we mostly are dealing with inverses. So this is what happens if we multiply by a matrix, m, but we're typically solving inverse problems where we're dealing with the inverse matrix. And so what are the eigenvalues and eigenvectors of the inverse matrix? So to see that I want to first introduce something else we're going to need.

So that looks like what we did for dot product, but not quite. Now it's the other one that has a transpose on it, and we'll use this notation quite a bit. And so let's write this out, a1, a2, an. And that's not a scalar, it's not a dot product because what you're going to do is multiply the first term here by the first term here, a1, b1, and then the first term here by the first term there, a1, b2 dot dot dot a1, bn.

And then we go down to the next row here, we get a2-- so again, if we treat these vectors as skinny matrices, this is what we get. So altogether, the dyadic product of n vectors is an n by n matrix, and that'll be handy for us. So let's apply that idea. So we've got V, we expanded it. Right over there, we said that because these eigenvectors form a basis, we can express an arbitrary vector in terms of it, and then we found the actual weights. And when we plug that in, we get this formula.

And so we can now rewrite that in various ways. So one way to rewrite this is v transpose times eI. Or another way of writing it is eI, eI transpose, v. So this is just taking the dot product and rewriting it. And how do we get here? In the dot product, it's commutative. We can flip the v and the eI. And therefore, we can get this expression for the dot product just as easily as that one.

And then taking a scalar times the vector is the same as multiplying the vector by the scalar, so we get this. And why is this interesting? Well, because when we do these matrix products, they are associative so we can rewrite this as eI, eI transpose, v. And up there, we have a sum over I, and v is not dependent on I. So we can actually factor this out like this. So these terms all depend on I.

So in the sum, every term has a different one. But the v is the same, so I can separate that out. And wait a minute, I'm saying v equals something times v? So what is this? That's the identity matrix. So this is a very long way around to write the identity matrix. And the reason we're doing it is because with a slight change, we can now write the matrix, m, using the same idea. And so we can actually get to this version of m. And why is this interesting?

Because now we can see the properties of the eigenvalues that we're interested in. And then how do we check this? Well, we can perform operations of m on some vector and see whether it produces the same result. So we can check this by doing, for example, m times eI, checking that this is true. Let's make it-- I guess we got eI here, so we'll make this j. That's very easy to check. And if that's true, then this is true.

More interestingly is this one. So this is the one we finally get to that's the most interesting. And how do we check this? Well, an easy way is to take this expression for m, multiply it by this expression for m inverse, and show that you get this expression, the identity matrix. OK. So this may be going a little bit fast. But remember, it's all there, and it all fits on four pages like those Facebook comments that say, oh, and it's only three pages long.

OK. So why are we interested in this? Here's the key. This matrix we're using to solve our vision problem, which takes a measurement and turns it into some quantity of interest to us, displacement, velocity, whatever, it multiplies components of the signal by 1 over lambda I. So it's bad if lambda I is small. So that's where we're going with this. So often, the way we can understand the performance of one of these methods is to look at what the noise gain is.

And when we get up from one dimension, this is the way to do it, we take that matrix, we find its eigenvalues. And if some of them are small, we know that it's an ill-posed problem. It's not going to have a stable solution. If you make a small change in the measurement, you'll get a big change in the result. So I know I'm saying this again, and again, and again, and it's because key. It's important.

So for example, in our optical mouse situation where we just have the 2 by 2 case, we end up with a diagram like this where we have two directions that are eigenvectors. And if one of them has a small eigenvalue, then it's going to be hard for us to accurately compute the motion of the mouse. And it turns out that in that case, one of them is the direction of the isophote, and the other one is the direction of the gradient. So remember that isophote are just lines of brightness is constant, and they're very handy for drawing things on the blackboard because I can't draw gray levels.

And they have the property that they're perpendicular to the gradient where the gradient is just the two vector of derivatives with respect to x and y. And it turns out that the eigenvalue that corresponds to the isophote direction is very small, in the ideal case, 0, and the one in the gradient direction is not. So invert that, that means that in the isophote direction, any small error will be magnified hugely whereas in the other direction, the gradient direction, it's OK.

And so that corresponds with our understanding that if you move this image in the gradient direction, things change. This isophote is now down here, and the brightness here has changed. Whereas if you move it in the direction of the isophote-- by definition, isophote meaning constant brightness, the brightness tends not to change, or not change by much, and so that's then a motion that is not easy to detect accurately. So this is kind of a little story about eigenvectors and eigenvalues.

And we'll see that they'll play a role, and it's not just in error analysis, but this is one of their main uses. OK. Back to slightly less abstract stuff, photometric stereo, and we discussed that somewhat last time. So the idea is that a single brightness measurement doesn't give us enough information to recover surface orientation because surface orientation has two degrees of freedom, and so we need more constraints.

And one way to get more constraint is to take more pictures. But of course, if they're under the same conditions, they'll be just the same picture with a little bit of different noise. But the difference in noise isn't going to buy you much other than maybe you can average them to reduce the noise. So we ended up with a system where we took three pictures to make things simple under three different lighting conditions. OK.

And we started off with a real simple case where we said that the brightness was proportional to cosine of the incident angle, and that, of course, is the dot product of the surface normal and the direction to the light source. And in a minute, we'll talk about, what if the surface doesn't satisfy that constraint? Because obviously, real surfaces don't. They may approximate this. Some surfaces approximate this rather well, but we want to deal with arbitrary surfaces. OK.

So we might have, for example, one brightness measurement. So we're now talking about a particular pixel, right? So we're going to focus on doing this at every pixel, and so looking at a particular pixel, and this is the brightness, the gray level we measure there. And I've put in this row, which I'll call the albedo, as a way of describing how much light the surface reflects. So a white surface, the row would be 1, and a black surface row would be 0, and a real surface would be somewhere in between.

And one reason I do that is because I can, and because it makes the problem actually easier-- because now I've got three unknowns. And if I have three constraints that are linear, I can use my great linear equation solving methods to solve them. In a way, two images would be enough because if we have two equations and two unknowns, there's usually a finite number of solutions. The trouble is the finite number may not be 1. And in this case, the finite number is 2.

And so I've disambiguated hugely. Before I didn't know what the orientation was at all, now I know it's one of these two possibilities. But it's easier to deal first with a case where we can completely disambiguate it by introducing a problem with three unknowns and three measurements. OK. So then I use a different light source, and I get a second measurement. And then I use a third light source, and I get a third measurement, and we did this last time.

And this is where we use that notation for a dot product, and we use it to talk about-- let's do that [INAUDIBLE]. So what's this? This is a 3 by 3 matrix. S1 transpose is the vector to the light source one just flipped from being a column vector into a row vector. So the first row of this matrix is S1 just turned on its side, and this is S2, and so on. And so when you multiply this matrix by that vector, the first thing you get is the dot product of S1 and n, and that of course is E1.

OK. And here, I'm using the shorthand notation. So I'm absorbing that lambda into the n. And that's convenient because now, instead of having to deal with a unit vector, I have an arbitrary 3-vector. And that means I don't need to deal with a nasty nonlinear constraint. The thing that leads to two solutions is a quadratic, and why do we get a quadratic? Well, because we have this second order constraint on n.

But by doing this, we avoid having to do that. OK. So that, I can just write as a matrix s times vector n is vector E. So the vector E is just I'm stacking up my three measurements. So again, to be clear, so I take one picture, I look at this pixel, I get E1. Then I turn on a different light source, I take a picture. At that same pixel, I get E2. And then I turn on the third light source, I look at that particular pixel, and I get E3, and I stack them together to make this vector, E1.

And so the solution is very simple and extremely cheap to compute. In particular, I can precompute this matrix assuming that the light sources are in fixed positions. So if I know where the light sources are, I know S1, S2, and S3, I can just construct this matrix and invert it ahead of time, and then this is just a multiplication of the 3 by 3 matrix by a 3 vector. And we talked about this last time, and we also said that's assuming that this matrix doesn't give us problems.

So the problem would be where the matrix is singular, then we can't invert it. Or if it's nearly singular, we can't invert it. So when does that happen? Well, it happens when the rows are not linearly independent. So that's when this blows up. And for example, let's try this, S3 is some combination of S1, some linear combination of S2. So that's how we create problems, right? If the third row is just a combination of the first two rows. Or in general, if there's a linear combination of the three rows, that gives us 0. OK.

So why is this bad? Well, because this means that E3 is alpha E1 plus beta E2. I just need to take the dot product of this equation with n, and I get this result, and that tells me that the third measurement is redundant. It's not telling me anything new. So no wonder it blows up, right? It's like you're cheating. For example, if you made the third row the same as the first row, it's pretty clear that the matrix is singular, and it's also pretty clear that you're not getting any new information. So it all makes intuitive sense.

So I can think about this as a picture. This is a case where S1 plus S2 plus S3 is 0, and that's clearly a bad case. And I can be fancier, and I can put multipliers on these, I can put alpha on this and beta on that, and if you'd like, gamma on that. So if some multiple of those three vectors adds up to 0, this won't work. And what is that condition? How can I geometrically say what the condition is for this to go wrong?

So the vectors S1, S2, and S3 are-- for this loop to close. Remember, this is now in 3D. So we've got a vector going a certain way, of a certain length, then another vector. You start at the tip of that first vector and you put the tail of the second vector there. And then when you get to the third vector, it closes. So what is the condition on those three vectors, geometrically, that makes that possible?

AUDIENCE: [INAUDIBLE].

BERTHOLD HORN: They're in the same plane, right? If I had x, y, and z axes, obviously you can't do that. One goes off in the x direction, one goes off in the y direction, one goes off in the z direction, they're going to close. So the problem is if they're coplanar, so that's bad. That's when this method fails. And so obviously when you do this, you should place them so that they're kind of as far as possible from being coplanar.

For example, you could put them at x, y, z axes. So you could have an arrangement where the object is down here. And then you can erect some rectangular coordinate system, and S1 goes there, and S2 goes there, and S3 goes there. And now they're not coplanar so this won't happen, they won't fall apart. And then there are questions of, well, which is best? Obviously if I make them almost coplanar, I'm going to get unstable results, some eigenvalue will be small.

And if I take the inverse, the inverse of the eigenvalue will be large, and so the noise amplification will be large. OK. So if you're doing this in an industrial setting, you are in control of where the light sources go. So that's all you need to know, don't make them coplanar. And probably, the closer you can get to this arrangement, the better where they're actually at right angles to each other. So this is where I want to talk briefly about the Earth, moon, and the sun.

So let's suppose that the moon is made of green cheese, and green cheese has a Lambertian-reflecting property. And so we are on Earth, and we're trying to get a topographic map of the moon, which would be a useful thing to do before people land there. Of course now we did a long time ago. But before people landed there, there was a lot of uncertainty. People didn't know, for example, whether the solar wind had pummeled the surface to such an extent that there was 10 meters of dust. And if you were landing on it, you'll just disappear into that dust.

Anyway, so there was a lot of interest in trying to figure out how tall are these craters. We can see these nice craters, but what's the slope? We don't want to land on something that has a very large slope. So it would be really great if we could use photometric stereo because the sun does illuminate the moon in different ways, different parts of the cycle. So here, as you know, the moon is always showing the same face to the Earth, and the other side, which is stupidly called the dark side of the moon, is not visible from the Earth.

And so if you think about-- let's take a particular point here-- while it's in the cycle here, the sun is over there. But here, the sun is in that direction. And here, the sun is in that direction. So we could easily arrange three, 10, however many measurements at different positions in the orbit, and we can use this method. Well, there's some assumptions. One of them is that it has Lambertion reflectance, which it doesn't, but we'll try and fix that later.

And annoyingly, it doesn't work, and the reason is that this plane here that contains the moon's orbit is pretty much the same plane as the plane that contains the orbit of the Earth around the sun. So you can see what's going to happen that these three vectors, or more, are coplanar, or almost coplanar. The moon's orbit is a couple of degrees off the Earth orbit around the sun. So you get a tiny change, but it's not enough to make a useful measurement.

And so it's amazing. We've just done something very simple here, and we've already reached a very profound conclusion, which is that you can't get the moon's topography from Earth measurements as it goes through its orbit, and different parts are illuminated differently as it does so, which is a pretty amazing thing. OK. So let's talk a little bit about the Lambertion assumption. So Lambert was this monk, and he did these experiments.

Now why was he a monk? Well, because there was a time in our history in the Western world where the only people who could learn anything were in the religious domain. Ordinary people couldn't read or write, and they weren't allowed to read or write, and monks did all of these interesting things like figure out how to brew rum and whatnot. And this particular guy, he had wrapped up his lunch in paper, and it contained some oily fish. And so a piece of the paper was infused with oil. And I think you've seen this, white piece of paper with some oil on it, the oily area looks kind of darker.

And then if you hold it up and move it around, you see that it has different reflecting and transmitting properties. And so what he discovered was that he could make measurements of light using this tool. Today, of course, we'd use an $100,000 computer and PIN photodiodes, but he used a piece of paper. And so the idea is this. Here's our piece of paper and here's the fatty spot, and what happens is that the paper is not absorbing any light.

It's a white paper. Ideally-- it could absorb a little bit, but let's assume it absorbed none. So it only has two choices. The one is sunlight arrives and is reflected back, and a little bit of it goes through to the other side. Now in the fatty part, sunlight arrives and a little bit of it is reflected back, but a lot goes through. OK? So that's the difference about the two parts. The fatty substance basically fills in the air voids, and so the surface is no longer as reflective as before, but is not an absorbing material so what happens is the light just go through.

OK. So why would this be of interest? It allows you to compare two illumination intensities, and you do it basically by illuminating this side with one source, and this side with the other source, and then you balance it so that you can't see the fatty spot. When can that happen? Well, we could write out equations, but I think you can see what's going to happen is if the same amount of light is coming in from this side as from that side, then these will balance.

This will appear equally bright when looked at from here as that. So that's a very powerful idea. And he didn't really do this with these lunch paper, but that was where he went. He had white paper, which was pretty precious at the time. You may remember that people would sometimes write something-- and then because paper was expensive, they would write something at 45 degrees on top of it and maybe more.

So anyway, he had this nice piece of paper. And this way, he could compare brightness. And one of the things you can do, then, is get the inverse square law. So he might put four candles on this side and one at twice the distance as one candle on this side, right? And so it should match. So he was able to do all of these amazing experiments with this very simple apparatus and get the inverse square law. And of course people like Newton would say, well, it's obvious, you don't need to do an experiment.

The energy is going over the surface of a sphere, and the surface of the sphere is 4 pi r squared rr squared. So anyway, then the next thing he did was he wondered how surfaces reflect light when they illuminated from different directions. And using methods like this, he came up with what's now called Lambert law, which is that it's cosine of incident angle-- the brightness is proportional to cosine of incident angle. Now of course, part of that is, again, not something that you need to do an experiment about.

We already talked about foreshortening, and we know that the amount of light falling on the surface varies as to cosine of the incident angle. But what he talked about was, how bright does it look? In other words, how much light does it reflect, and in what direction? OK. Now to talk about this in more detail, we need to have a way of talking about surface orientation because the brightness is going to depend on surface orientation, and it may do so in a more complicated way than Lambert.

I put law in quotation marks because it's not a law, it's a phenomenological model. What does that mean? That means that you postulate a particular way something behaves and then give it a name. And it's not like there's a real surface that does exactly this. Many real surfaces like paper are good approximations, but they don't do exactly that. OK. We already talked a little bit about surface orientation. We said that we can erect a unit normal.

So here's a little patch of the surface, and we've erected something that's perpendicular to the surface. And that's a way of talking about the orientation, and we mentioned that it has two degrees of freedom because it's a 3-vector, but there's a constraint that it's a unit vector. So 3 minus 1 is 2. So that's one way of talking about it. And for an extended surface, you can imagine that we do this for little facets all over the surface, and they'll all be pointing different ways.

But if you pick a small enough area, as long as the surface is reasonably smooth, then we can reduce it to this case. We have to exclude things that mathematicians will construct like a surface that 1 where x is a rational number and 0 where x is an irrational number, we can't do that there. But for real surfaces, we can facet it. And for different facets, we can do that. Then we also mentioned since it's a unit vector, we can talk about orientation in terms of point on the unit sphere, and we'll find that pretty handy later on.

For example, if you want to talk about all possible orientations, that's the whole surface of the sphere, or there's certain operations that we'll be doing on the surface of the sphere. So those are representations, but we need something else. So let's look at this. So here's a Taylor series expansion where the dots indicate higher order terms, which we can ignore as long as we make the infinitesimally small enough. OK.

So the difference of these two, we can write as delta z. So that's the same equation. So I'm introducing p and q as shorthands for these derivatives. And this is just analogous to what we've done before where we introduced u and v as shorthand for dx/dt and dy/dt, partly because it's too much bother writing all that stuff, and partly because it makes it look too intimidating when it's actually very simple.

So p and q are slopes. And in fact, p and q is the gradient on the surface. So remember how I said that one way to think about an image is that it's height above some ground level. So we have x and y in the image plane, and then we can plot the brightness as height above that. And then when we talked about the brightness gradient, ex, ey, I was saying that it's gradient of that surface.

Well, here's a case where we actually are talking about a surface in 3D, and this is its gradient. It's dz/dx, dz/dy. OK. One way to make a picture of that is this. OK. So this is our surface with normal vector, and this edge is delta x, and this edge is delta y. And this part here is q delta y, and this part is p delta x. So this is a diagram that basically illustrates this idea.

I take a small step in the x direction, delta x, and the surface goes up by dz/dx times the delta x. Then I take a small step in the y direction, and the surface goes up by dz/dy times delta y because it might go down. Just in this particular picture, it goes up. So this is another way of understanding what this equation is saying. Now we'll find that the places where the unit normal notation works for us, and there are places where the gradient notation works for us, and so we need to have ways of switching back and forth.

And this is very common in computational problems where some problems are easily done in one domain, and some are easily done in another domain. So it ends up being a problem of conversion. Like Cartesian coordinates and polar coordinates, some things are easy to do in polar coordinates, some are easy to do in Cartesian coordinates so you just want to be able to convert back and forth, and so same here. So how are these related? How is n related to p and q?

Well, one thing we can do is look at this surface, which has a normal n. Any line in the surface has to be perpendicular to n, right? That's the whole idea of n is perpendicular to the surface, meaning it's perpendicular to any line in that surface. And if I have any two lines in the surface that are not the same, then I'm done because if n is perpendicular to two lines, I can just take the cross product, right? Because the cross product of two vectors is perpendicular to both of those vectors.

So I need to find some tangents in the surface. Well, here's one. This is an edge which lies in that surface, and what is its direction? Well, the x component of that vector is delta x, it has no y component so that's 0, and the z component is p delta x. So that's a vector in the red direction. And I can take the delta x outside because it's arbitrary length, and I get that vector. So that's a tangent.

I can get another one. I can get this one here. And that one, as I move along that edge, there's no change in x. So this is 0. There's a change in y, right? I'm moving along here, so delta y. And there's a change in height, which is q delta y, and so I can take the delta y of-- OK. So what have I got? I have two directions that lie in the surface. I could have picked some other direction, those just happen to be very convenient to calculate.

And so what I need to do is take the cross product, and that should be parallel to n. It'll have some size, which we don't care about. We're only worried about the direction of that unit vector. So what's that cross product? I think it's minus p minus q1. So that's the connection between the two representations, and now I can go a little bit further. I can say the unit vector is-- just normalize it.

So if you give me p and q, I can compute the unit normal n, and I guess we also want to be able to go the other direction. So p is minus n dot x. So if I have to ever go the other direction, I can do this. And it looks intimidating, but it's just saying take the first component of n, namely this minus p, and divide it by the last component of n, which is 1. So why did I do that? Because what if n isn't the unit vector? Then this will take care of it.

If it is a unit vector, I don't need to be this fancy. OK. So I can go back and forth between the two notations representing surface, and what do I do with this? Well, the great thing now is I have a way of mapping, in the plane, all possible surface orientations. I already kind of had that because I had the sphere. I said that all possible surface orientations correspond to points on the sphere. Trouble is, I don't have spherical paper and I don't have a spherical blackboard.

So this is a projection of the surface of that sphere into the plane that's particularly easy to understand because this is just the dz/dx and this is just dz/dy. OK. So as with velocity space, this is a very useful construct, but it takes some getting used to. So points in this plane are not points in the unit. Points in this plane correspond to different p and q, i.e. different surface orientations. So for example, let's consider that point.

So that's the point where p is 0 and q is 0, and that's a plane. Now suppose that the flaw here has an xy coordinate system and z comes up out of the floor, what type of a surface would have this point in this representation? Yes. The floor, for example. Anything that's level. Why? Well, because the zdx is 0 and the zdy is 0. So if it had any tilt at all, then the zdx or the zdy would be non-zero. So the floor has this property.

But actually, any surface above the floor that has the same orientation-- so there's an ambiguity here. It's not telling us where this thing is. It's only telling us how it's oriented in space, and so that's going to be a secondary problem. Suppose that we come up with a machine vision method, which for every pixel, allows us to recover surface orientation either this way or in p and q. We still have to patch it together to make a complete surface, but that turns out to be easy because it's overdetermined unlike most of the problems we deal with which are underdetermined.

OK. So that's that surface. Now if I go over here to p equals 1, let's suppose that x goes to the right and y goes forward, that corresponds to the surface where the slope going to the right is 1, 45 degree up. That would be pretty steep, I would slide off it with these shoes. So then if I go over here, what's that? Well, that's a surface where the slope in the y direction is 1. So it's the kind of thing you find at EMS to check out your rock-climbing boots, make sure that they are fitting well, and you stand with your toes up like this.

And then this one, of course, is a combination where we have a slope to the right and a slope to the forward, and so on. And the further out I go, the steeper it gets. So that's what this plane is. Every point in this plane corresponds to a particular surface orientation. Now in application to machine vision, we find that the brightness depends on surface orientation. So this is a wonderful tool to plot brightness. All right? So just experimentally, I could take a patch of this material, I could orient it flat, parallel to the ground. I measure how bright it appears, and I put that number here, E1.

Then I tilt it up 45 degrees, and I put that number over here. And I tilt it up-- what's the slope of 2? The inverse tangent of 2, and I plot it here. So I can plot my brightness values as a function of surface orientation. So this becomes kind of an image because at every point there's a brightness, but it's not it's not a transformation of an image that you take with an optical system at all. That's confusing, but it's not. So everything in here corresponds to orientation, and then we can plot whatever we want, such as brightness. OK.

So where is this going? Well, one idea is we can then invert this. Suppose that we've made this map, then you measure the certain brightness. And then you go back to it and say, oh, that means the surface orientation is such and such. So that's the idea. And you're probably saying, well, really? Because maybe brightness down here is the same as there, and maybe the brightness down here is the same as there.

And in fact, maybe there's a whole line of points that have the same brightness and a whole line of points that have the same brightness, and just counting equations and constraints tells you what the problem is. If we make one brightness measurement, we can't recover two unknowns. So just as with velocity determination, a single measurement won't be enough. It will be a dramatic improvement of a no measurement. So if we don't take a measurement, we don't know where we are in this plane at all.

We could have any orientation for a surface element. If we take one measurement, we will be constrained to some curve, and then we need more constraint to actually pin it down to a particular orientation. So let's relate that to what we did in our example of photometric stereo, and also our discussion of Lambert. Let's suppose that we have a Lambertion surface. So there's a common confusion, which is that this stuff only applies to Lambertion surfaces. Why are we doing Lambertion surfaces?

Because for Lambertion surfaces, I can show you nice diagrams, I can solve equations. In the real world, nothing is exactly Lambertion. So if you want accurate result, you will have to measure, calibrate, and we'll see how to do that. But for the moment, let's just assume that, magically, we're dealing with a Lambertion surface. So then we have the surface normal proportional to brightness. So brightness is proportional to the cosine of the incident angle, and that's the dot product of the surface normal and the incident direction.

And now we want to translate that into this notation, p and q. But one thing that's important is we're looking for isophotes in here. We're looking for these curves. And so we'll be looking at places where this is a constant, but let's just replace-- so the unit vector, n, is minus p minus q1. And we can take the dot product of that with some light source direction, and then we can plot that.

But there, it'll be handy to introduce another little shortcut, which is a different way of writing the direction to the light source. So we thought of that equation there as kind of a mixture. We've halfway gone from unit vectors to pq space. Let's go fully. And so the full way is to say, well, to perform the same transformation on that unit vector that we did on n, it's just the same equation. It's just this time we're talking about the vector to the light source rather than the unit normal.

And so to make that clearer, that point, ps, qs, it's in that plane. And what is it? Well, it's the orientation where the incident light rays are parallel to the surface normal, right? So this is the point where-- and so for the Lambertian surface, that's going to be the brightest spot, right? Because the angle between those two vectors is 0. So the cosine of the incident angle is 1, and that's as big as cosine can get.

So we picked this particular point, ps, qs because it's in that plane-- because it's the one that gives us the brightest surface. So it has some real meaning. Other than geometrically, it just means that we're illuminating the surface right down the normal. Another way to think about it is the foreshortening, there's no foreshortening. We haven't tilted the surface relative to the light source. So we don't have the same power spread over a larger area.

Here it's concentrated in the smallest possible area. It's the most efficient. It's what you do with your solar collectors. OK. So that means we can rewrite n dot s in this form. So this is starting to look a little bit messy. And what are we doing? We want isophote. So we want to know, where is this quantity constant? What are the curves in that pq space, in gradient space where that quantity is constant? Well, I can square this and move things around a bit.

Now this quantity here, this is just some constant, right? Because I'm keeping the light source in a fixed position. And then the question is, what kind of curve does this define in pq space? If you multiply it out, you're going to get some constant terms, some terms proportional to p, some proportional to q. The highest order terms you're going to get are second order. So when you multiply it all out, which is messy, you'll have something that's second order in p and q.

So the question is, what kind of a curve does that define? And it may be hard to think about it in terms of p and q. We're not just talking about geometry in the plane. So imagine you have an x and y-- you have an equation that's got x squared, y squared, x times y, x, y, and a constant in it all added up, what curve would that correspond? Yes.

AUDIENCE: [INAUDIBLE] parabola.

BERTHOLD HORN: It could be a parabola, yeah. Anything else?

AUDIENCE: Ellipse.

BERTHOLD HORN: Ellipse, Yes. OK, great. Generalize it a little bit more.

AUDIENCE: Conic section.

BERTHOLD HORN: Sorry?

AUDIENCE: Conic section.

BERTHOLD HORN: It's a Conic section. OK. Yes, those are great examples. Overall, they're conic sections. And so yes, we can have a parabola. We can have an ellipse. We can even have a circle. We can have a line, a special degenerate case. We can have a point, even more special degenerate case, and we can even have a hyperbola. OK. So I want to plot this thing, and this is like a preview of what it's going to look like.

So if c is 0-- let's look at that special case. If c is 0, then this is 0. And that means that 1 plus psp plus qsq is 0. And what kind of an equation in p and q is that? That's a line. It's a linear equation so it's just a straight line. So that's one special case. So there's going to be some sort of line in here, and it's where the brightness is 0. OK. And then another special case is where p equals ps and q equals qs.

And that's a special case where-- we talked about here where the normal is pointing straight at the light source, and we're getting the maximum brightness. So there's some point over here where E is 1 if we normalize it appropriately. OK. And then the rest, you can plot from using some sort of program, if you'd like. OK. So that's a very handy diagram for graphics because if I have a surface that I'm plotting, I can easily determine the unit normal from it, I can get p and q, or I can get p and q directly.

And then I'd just go to this diagram and I read off whatever the brightness is here, and I use that as a gray level, or color, in the image that I'm plotting. What we're doing is kind of the other way around. What we want to do is say, OK, I've measured E as point two, what's the orientation? In this case, that's that curve. So I don't get a unique answer, but it's heavily constrained. It has to be on that curve. OK.

Now if I had more constraint, I can improve on this. For example, suppose now I move the light source, then this whole diagram changes, right? Because remember this point here is one that basically depends on the position of the light source. It's ps, qs from this equation, it's related to the direction to the light source. OK. I don't want to mess up this diagram, but imagine that I mirror image this by moving the light source over here.

So I'll have a second set of isophote that now intersect with these. And if I make a measurement under those other light conditions, then the answer has to be on both curves, and then I get the solution from that. So just for fun, let's suppose that the other curve was like this. They're curves, they're not lines. So it is quite possible for them to intersect in two places. So I have a finite number of solutions in general, not one.

And that's why we focus more on the case where we use three light sources instead of two. So just a note on, why are they conic sections? Well, that actually also has an easy answer, which is suppose I take a brightness measurement of a Lambertian surface-- so here's my light source, here's a surface. And from the brightness, I can calculate this angle. But of course, I can spin this vector around this line to the light source. I can spin that around, and what do I get? I get a cone.

If I measure a different brightness, it'll be a different angle, I'll get a different cone, and so on. So again, imagine some third surface element, I measure yet another angle, and I get, say, this cone. So there are these nested cones, and now imagine that you cut this with a plane. So this is our pq plane, and ta-da, conic sections. And yes, you won't just get ellipses, you may get a hyperbola.

As long as this bottom edge of the cone is actually below this plane, you will not get a closed curve, so parabolas are possible. Yeah. OK. Let me first address that one, other side of the line. So we said that it's cosine theta except when it's negative. This is where cosine theta I goes negative. And I haven't purposefully drawn this part of the diagram because in practice, brightness doesn't go negative. It's a measure of power, so it can't be negative.

So if I were to just plot cosine theta, it would continue. But we're having max of 0 and cosine theta, so we have this part. The other part of the question is, where does it turn from being closed curve to being open? I'll leave that as a puzzle for a future homework problem. Why? Because I don't know the answer. So I'll let you figure it out. OK. That's it for today. So I guess you all know there's a homework problem out that will be due.

And please make sure you are signed up on Piazza because a lot of announcements are on there about office hours, homework problems, and stuff like that.