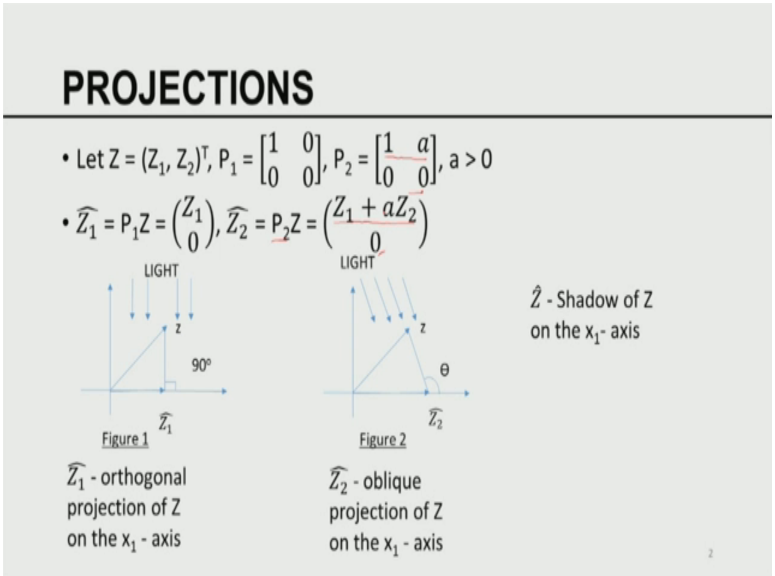
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture – 11
A Geometric View – Projections

So, far we have applied the least square method to solving static deterministic least square problems; both well posed and ill posed. Given the importance of least squares starting from the days of Gaussi, I think it is worthwhile to get a geometric view of the nature of the least square solution. Thus for we use the analytical methods to derive the least square solution, by formulating a problem as a minimization problem; both unconstrained and constrained.

(Refer Slide Time: 01:01)



This geometric view enables us to be able to look at least squares from a very simple perspective of the notion of projections.

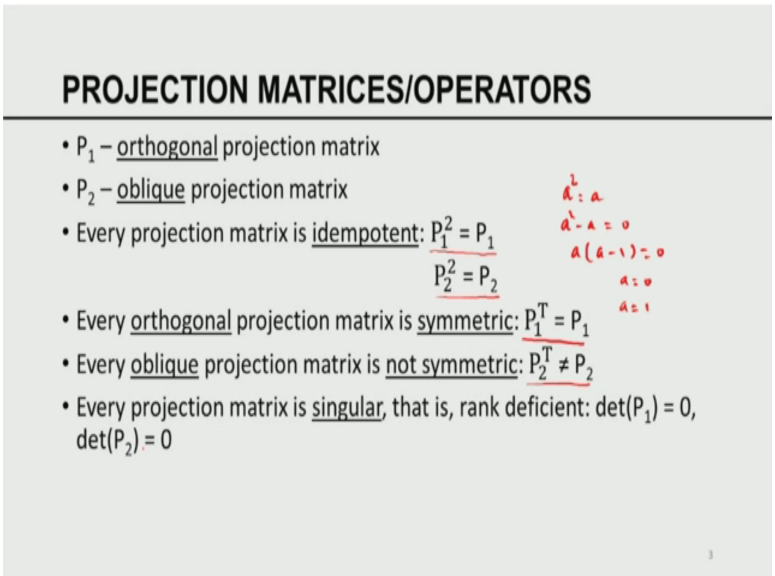
So, let us consider a vector Z 1, Z 2 in the 2 dimensional plane; this is the vector Z if you shine lights parallel to the y axis; a shadow will be cast on the x axis. This arrow segment gives you the shadow of Z; the shadow is called Z 1 hat. The property of this shadow in figure 1 is that; it is an orthogonal projection, in the sense that if I join the tip of the vector Z and Z 1 hat; the angle between the 2 vectors is 90 degrees. On the other hand, if you shine light in a direction not parallel to the y axis; the shadow cast by Z on the x axis is Z 2 hat; if I join the tip of these 2 vectors, the angle is theta; in this case theta is not equal to 90; so, this is called oblique projection.

So, Z 1 hat is called an orthogonal projection; projection of Z onto the x 1 axis, Z 2 bar is the oblique projection of Z on to the x 1 axis. Oblique and orthogonal decided simply by the direction in which light is shown on the vector Z. Mathematically, this operation of shining light and projection can be thought of as a matrix P 1. If P 1 operates on Z, you get Z 1 hat; the P 1 has a form which is 1, 0; the first row 0, 0 in the second row.

So, in this case P 1 Z is essentially Z 1, 0; the second component is annulled the first component is nonzero, so Z 1 hat is equal to Z 1; so, that is an orthogonal projection. On the other hand, if you consider P 2 to be P 1, a and 0, 0 and apply that operator P 2 and Z, you get Z 1, a Z 2 and the second component is 0.

So, now you can see when a is 0, P 2 becomes equal to P 1; when a is not equal to 0, for example, a is greater than 0, the first component Z 2 hat is Z 1 plus a times; Z 1 is the first component of Z, say Z 2 is the second component of Z; a is positive; the shadow is longer. So, you can think of projection as a geometric operation, algebraically the operators; matrices P 1 and P 2 essentially generate this orthogonal on oblique projection. This is very fundamental geometric point of view and it is a very close and intimate relation to the properties of least square solutions.

(Refer Slide Time: 04:51)



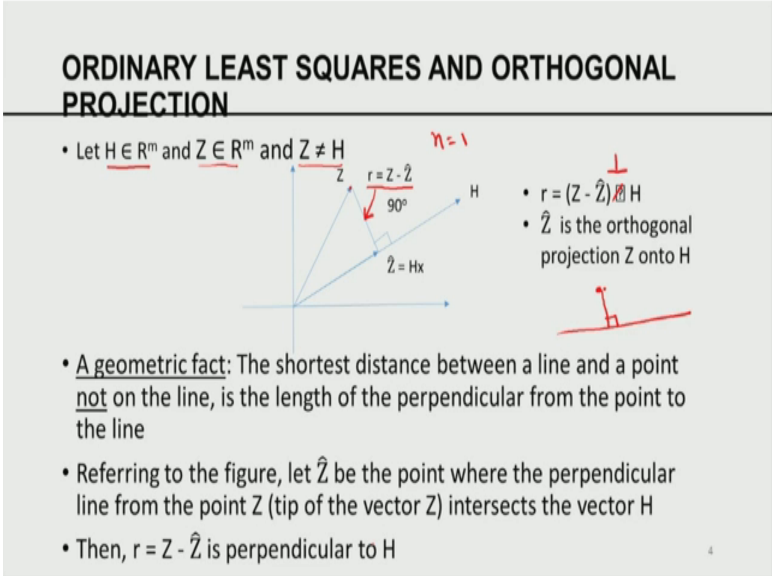
Projection as matrices; so in the last slide we saw matrices P 1 and P 2; P 1 is called an orthogonal projection matrix, P 2 is called the oblique projection matrix. Every projection matrix has a fundamental property that is idempotent by idempotent is I mean P 1 square is equal to P 1, P 2 square is equal to P 2. So, here I am looking for a matrix whose power is equal to itself.

Let us recall in terms of numbers; if I have a number a, if I want a square to be equal to a; to do that I have to solve this equation. This equation essentially tells you a times a minus 1 must be equal to 0; that essentially gives you either a is equal to 0 or a is equal to 1. So, there are only 2 numbers which when squared is equal to itself 0 and 1, but in the case of matrices P 1 square is equal to P 1, it can be solved and the solution we saw is given by the matrix in the previous slide.

Likewise P 2 square is equal to P 2 can be solved one of the solutions for that is the matrix P 2 given in the previous slide. What is the difference between P 1 and P 2? P 1 is symmetric, but P 2 is not symmetric. Now, I am going to state a very general property of orthogonal a projection matrix. An orthogonal projection matrix is idempotent and symmetric; an oblique projection matrix is idempotent, but not symmetric.

So, every projection matrix is idempotent; it is a symmetry or not symmetric nature of the idempotent operator is going to decide, whether the resulting projection is going to be an orthogonal projection or an oblique projection. It can be shown every projection matrix is singular that is it is rank deficient; if it is rank deficient the determinant is 0. Please verify that the determinant of P 1 in our area slide is 0, the determinant of P 2 in our area slide is also 0. So, in this slide we are summarizing the general properties of projection matrices; projections are of 2 types orthogonal or oblique. Every projection matrices must be idempotent; in addition if the projection matrix are symmetric, it is orthogonal projection; if it is not symmetric, it corresponds to oblique projection.

(Refer Slide Time: 07:41)



Now, ordinary least square solutions can be viewed from an orthogonal projection point of view. Let H; so I am considering a very special case, where H is a column vector; that means, n is equal to 1; that vector H is given by this line H; Z is the vector in R m. So, both H and Z are vectors in R m; I am giving an example of a 2 dimensional representation, assuming m is equal to 2, but the whole analysis holds for any m. let Z be not equal to H therefore, if I draw the vector Z this represents the vector Z, this represent the vector H.

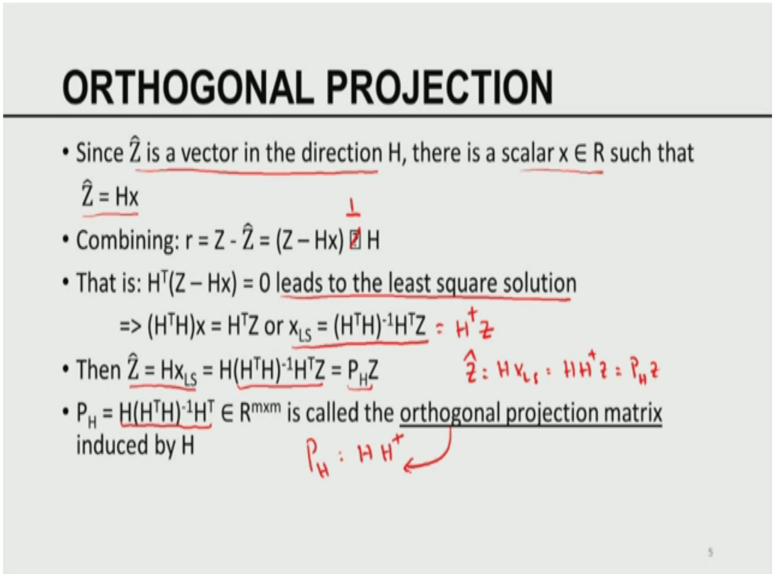
Now, I can project the vector Z onto H to get Z hat. Z hat is a vector that is along the direction of the vector H. So, I should be able to get Z hat as H times x; where x is a scalar. So, the question in the least squares projection is such that, I would like to be able to find the constant x such that the projection Z hat is an orthogonal projection of Z on to H; that is a geometric point of view. To do that; I am going to consider the difference of the vector Z minus Z hat and that is this vector, this vector is Z minus Z hat. I would like my Z minus Z hat to be perpendicular to Z hat, but perpendicular to Z hat is also equal to saying perpendicular to H. Therefore, in here I m requiring my r which is the residual; you remember the residual when we talked about least square solution; Z minus Z hat is the residual or the error in the projection must be perpendicular or orthogonal to H, Z hat is the orthogonal projection of Z onto H.

Now, I would like to relate another geometric fact; it is well known that if I have a line and if I have a point not on the line, the shortest distance from the point to the line is the length of the perpendicular from the point to the line; that is a very well known fact in basic geometry. So, you can think of the line to be my line H; you can think of my point to be the tip of Z; I am trying to draw the perpendicular from Z to H.

So, if it is perpendicular the angle is 90 the angle between r and H is 90; therefore, all when the angle is 90 is the shortest distance, the residual is the shortest length. So, the shortest distance between a line and a point not on the line; is the length of the perpendicular from the point to the line. So, referring to the figure Z hat, the point where the perpendicular line from the point Z the tip of the vector Z intersects the vector H; therefore, all it is Z minus Z H is perpendicular to H.

So, that is the simple geometric fact where the minimum of the residual essentially comes from the simple fact that the distance between the line and the point is shortest when the line is perpendicular; that is the geometric fact we are trying to use.

(Refer Slide Time: 11:34)



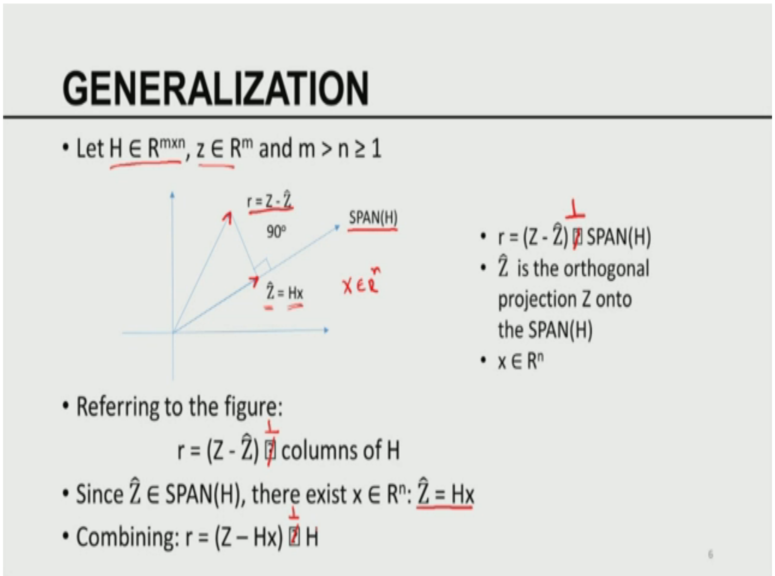
Since Z hat is a vector in the direction of H; there is a scalar such that Z hat is equal to H of x; that is a very well known fact. Because, if I give a direction any segment of the vector can be obtained by multiplying the direction by a constant; so x is a scalar, so by combining the fact that Z minus H x is equal to Z H, which is perpendicular to H.

Now, this perpendicular condition will be implied if the inner product of the 2 vectors are 0, so H is perpendicular to Z minus H. Therefore, H transpose Z minus H must be equal to 0 and that naturally leads to least square solution. So, you can multiply both sides; you get H transpose H x is equal to H transpose Z or X LS is equal to H transpose H inverse H transpose Z. So, Z hat, so this we have already seen to be H plus Z; where H plus is the generalized inverse.

Now, once I know X LS; I can get Z hat to be H times X LS; H times this is X LS I substitute that fact in here. So, I have now a matrix H times H transpose; H inverse H transpose the matrix operates on Z, this matrix I am going to call it P of H. This P of H is called the orthogonal projection matrix induced by H. Therefore, the least square solution Z hat is equal to H times X LS is equal to H plus H, therefore, H times H plus Z and that is equal to P H Z, P H is equal to H H plus and that is the orthogonal projection matrix that we are interested in.

So, you can readily see I get the same formula that I obtain by minimizing f of x, which is the square of the norm of the residual. The same result that we did analytically by minimizing f of x is obtained by a very simple geometric fact; which states that the perpendicular from a point; not on the line, to the line gives the shortest distance from the point to the line. It is a very simple geometric fact we all learn, when we are first introduced to geometry in high school.

(Refer Slide Time: 14:39)

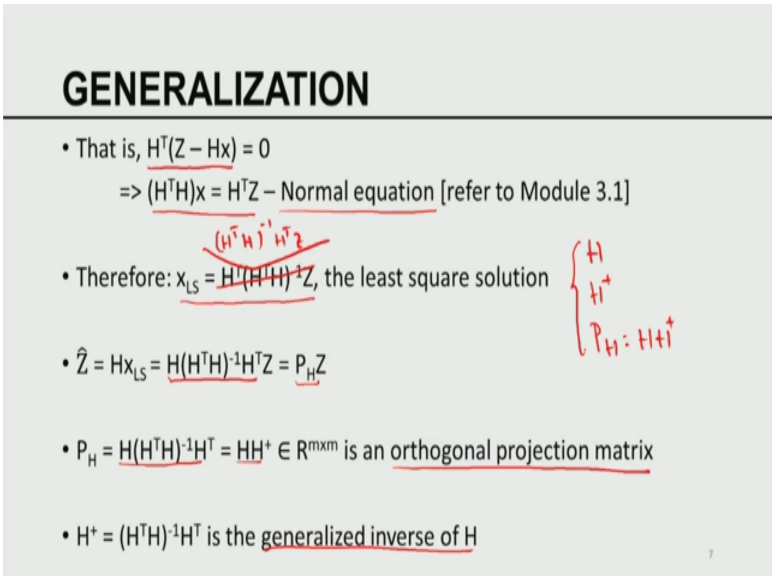


So, what is the generalization of this? Now, let us consider in this case H is m n; a matrix m rows and n columns. So, previous analysis was a special case when n is 1; now m is greater than n, which is greater than equal to 1 is the generalization of that. Let Z be a vector in R m; now since H has n columns, I am now going to consider a subspace kind by the columns of H; that is the subspace onto which I would like to be able to project my vector Z, that is the vector Z; if I project that vector Z hat is given by this.

So, Z hat in this case is the projection; it is still H times x, but in this case x is a vector belonging to R n. When n is 1, x was a scalar that is what we had gotten earlier. Again Z hat minus Z is r the residual is Z hat minus Z and we would like r to be perpendicular orthogonal to the span of H. We all know span of H is the linear combinations in the columns of H; that means, Z minus H of x must be perpendicular to every column in H.

So, referring to the figure we are going to get r is equal to Z minus H must be perpendicular to the columns of H; since Z hat belongs to the span of H that exists a vector x such that Z hat is equal to H of x; we have already talked about that earlier. So, combining this we now know r; which is equal to Z minus H is perpendicular to H; same argument. If H has only one column, if we project it onto that vector, if H has multiple columns; we project it onto the span of H, which is the set of all linear combinations of the columns of H.

(Refer Slide Time: 16:53)



Therefore, we would like to enforce this condition; which is H transpose Z minus H must be equal to 0; that essentially gives you the normal equation; which is H transpose. H x equal to

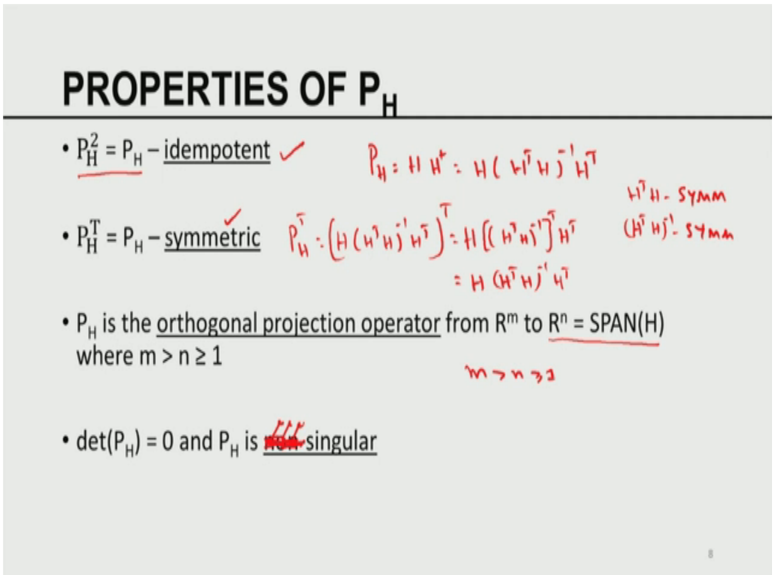
H transpose Z; this is called the normal equation. The solution to this normal equation is given by x least square is equal to H transpose.

So, the least square solution is given by this solution; the expression given there. So, Z hat is equal to H of X LS. So, which is equal to; so X LS is given by the solution of this, which is H transpose; H inverse H transpose Z, that is the correct solution, that is the least square solution; that solution we have already seen in the previous module.

Therefore, Z hat is equal to H of X LS, which is again if I substitute X LS; by this, I get this operator operating on H that operator is called P of H. So, P of H is given by this which is H plus; that is called the orthogonal projection matrix, induced by the given matrix H. And please remember H plus is the generalized inverse, we have already seen.

So, now we are seeing very many different types of matrices that come into play. We have the matrix H, we have the matrix H plus which is the generalized inverse. We have the matrix P H; which is equal to H; H plus, so this is projection that is the generalized inverse that is the given matrix. So, all these three matrices are naturally associated with the notion of least squares.

(Refer Slide Time: 18:57)



Now, I am going to talk about the general properties and verify that it is going to be an orthogonal projection; much of it is going to be left as a homework problem; we already know P H equal to H; H plus that is equal to H; H transpose; H inverse, H transpose that is the P of H. Now, I would like to ask you to verify this is idempotent; so what does it mean?

Please if you multiply this matrix by itself is equal to the matrix itself. So, P H square is equal to P H; therefore, it is idempotent. Please verify this; it can be also verified P H transpose is P H, why P H transpose is P H? Let me quickly illustrate that, P H transpose is equal to H; H transpose; H inverse, H transpose; transpose. From matrices, we already know the transpose of the product, of the transpose has taken in the reverse order. So, this is equal to H transpose; that is equal to H transpose, H inverse transpose and H transpose.

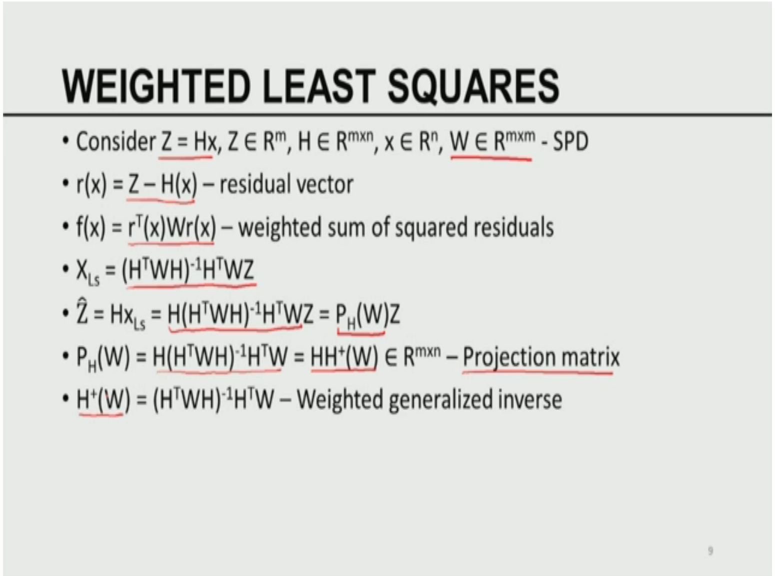
Now, H transpose H is symmetric, if H that is a (Refer Time: 20:39) is a symmetric matrix; H H transpose inverse is also symmetric. There is a general theorem that says that the inverse of a symmetric matrix is also symmetric. If the inverse of a symmetric matrix is symmetric, its transpose is equal to itself. Therefore, this is equal to H transpose; H transpose, H inverse; I am sorry I made a mistake, I would like to be able to correct myself; once I eraser that is correct.

So, that is the correct way of doing it. So, the transpose of the product is the product of the inverse transpose has taken the reverse order. So, H transpose transposes H; this is the transpose of the inverse, this is H transpose. Therefore, I get the correct formula; the correct formula is given by H transpose therefore, it is symmetric; that is verified.

So, P H is idempotent and symmetric; so, by definition is an orthogonal projection from R m to R n; which is the span of H. So, please look at this now; m; m is greater than n is greater than equal to n to 1. So, R m is a larger space, R n is a subspace; the span of H because H is a full rank, it generates the subspace R of m. You can also verify the determinant of P H is 0 and hence P n is singular, it is singular and hence P n is singular that is the property we can easily verify and the determinant is (Refer Time: 22:38) means it must be singular; that is the way things are ok good.

Now, I would like you to be able to verify that particular property.

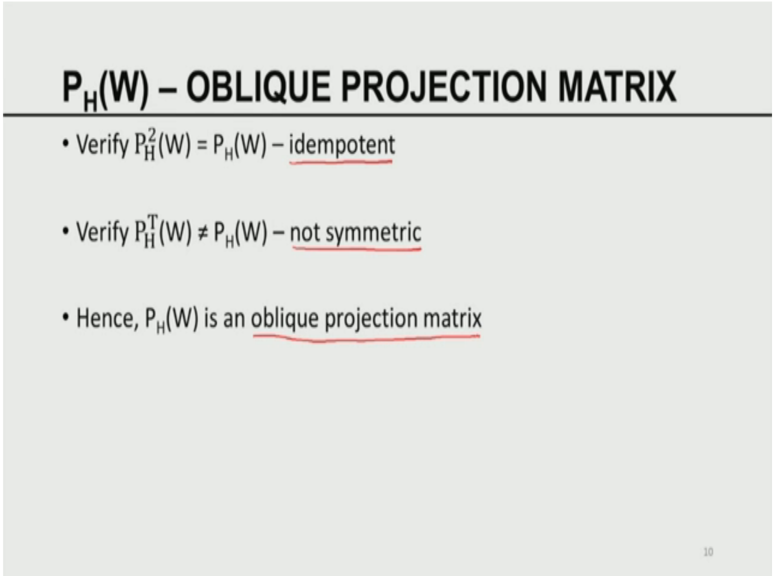
(Refer Slide Time: 22:49)



So, I am going to now go to the least squares with the weight present. So, in the case of weighted least squares consider Z is equal to H of x; there is a weight matrix therefore, I am going to be concerned with the residual Z minus H of x; that is the residual vector. My f of x is r transpose W of x; r transpose W of r; that is a weighted sum of square residual. We have already seen the least square solution is given by that, we have already seen in the previous slides and previous lectures, so Z hat is H of Z L S.

So, this is going to be the solution least square solution; I am providing a summary. So, P H W is given by the matrix that is the underlined. So, P H matrix is this that can be thought of at H; H plus W. H H plus W is the inverse in the weighted case generalized inverse in the weighted case; that is the so called projection matrix and H plus W is called the weighted generalized inverse. So, these are all simply a summary of all the quantities that we have considered so far.

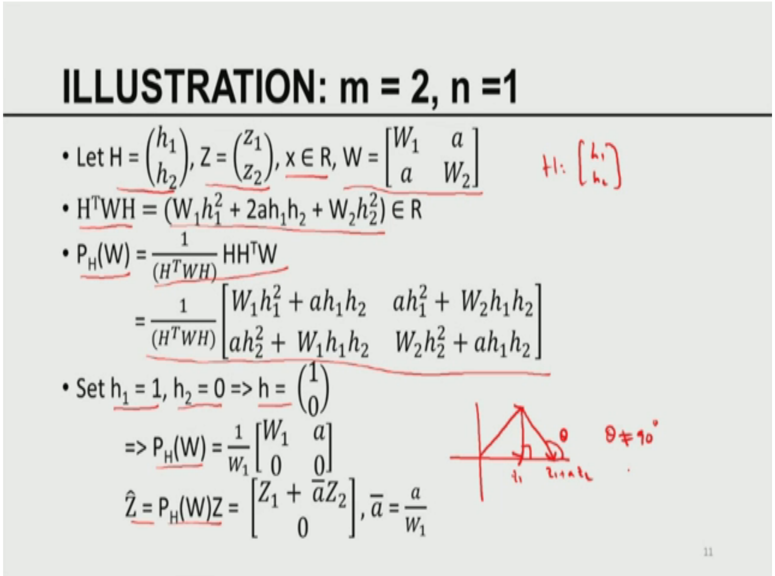
(Refer Slide Time: 24:15)



Again it is a very simple exercise to verify that this matrix is idempotent; this matrix also idempotent, this matrix is not symmetric and hence an idempotent matrix that is not symmetric has to be an oblique projection matrix; that is the general conclusion. What does this mean? When you do problems in 3D war, we always consider your weighted sum of squared errors. Therefore, the solution to the 3D war problems; with weight matrix in those cases, the weight matrix is simply the inverse of the covariance matrices observation covariance or background covariance.

So given the observation covariance matrix and the background covariance matrix, so, long as they are not identity matrices; we are always dealing with weighted least squares within the context of 3D war. So, almost 3D war solutions are giving raise to the so called oblique projection; only in the unweighted case, do we have an orthogonal projection; so that is the beautiful geometric view of things one has to remember.

(Refer Slide Time: 25:31)



Now, going to quickly illustrate by an example n is 1; that means, there is only one unknown, m is 2; that means, H is a matrix. So, now you can readily see H is a vector which is equal to h 1, h 2; that is given by here. Z is a vector because m is 2, x is a real number, I am now going to conjure up a simple matrix; which is symmetric; the weight matrix is always symmetric. Even though the weight matrix is symmetric, the projection matrix resulting from the weight matrix is not symmetric; that is something one needs to keep in mind.

So, W 1, W 2 are 2 diagonal elements; a is the off diagonal elements. So, h transpose W H; if you do the multiplication, you will get this quantity; which is the real number. Now, we have already seen the expression for P H W in the previous slide; I am now going to substitute all this and that takes this form. So, if you do the simplification; the projection matrix becomes that particular matrix which is a; each of the elements have 2 terms; it is an addition of 2 terms.

Now, I am going to set a special case h 1 is 1, h 2 is equal to 0; that means, h is equal to 0, 1. In this case, your projection matrix becomes the one that is given here or the projection Z hat is given by P H; W Z, which is given by Z 1 plus a bar Z 2; a bar is a by W 1 and this is something, we have already seen. The very first opening example of an oblique projection, so what is that we have shown? If there is a weight matrix, the resulting projection is not an orthogonal projection and that is the conclusion that we have; why is this is not an orthogonal projection?

If you have this; if this is Z, that is an orthogonal projection; now if you get this, so this is Z 1, this is going to be Z 1 plus a times Z 2. Z 1 plus a times Z 2 is not equal to Z 1 unless; a is equal to 0. So, if you assume; a is not equal to 0, the angle here is 90; here is theta, theta in general is not equal to 90. Therefore, when a is not equal to 0, the projection is an oblique projection; that is the important thing that one has to keep in mind.

(Refer Slide Time: 28:27)

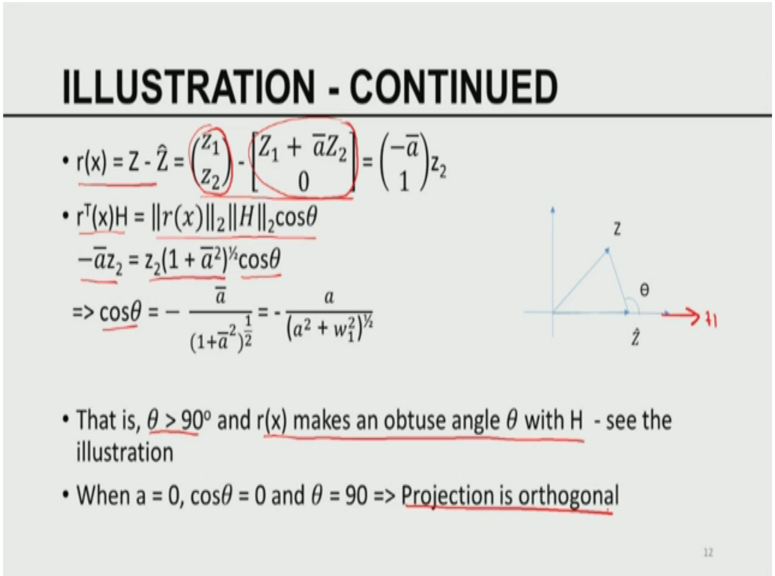
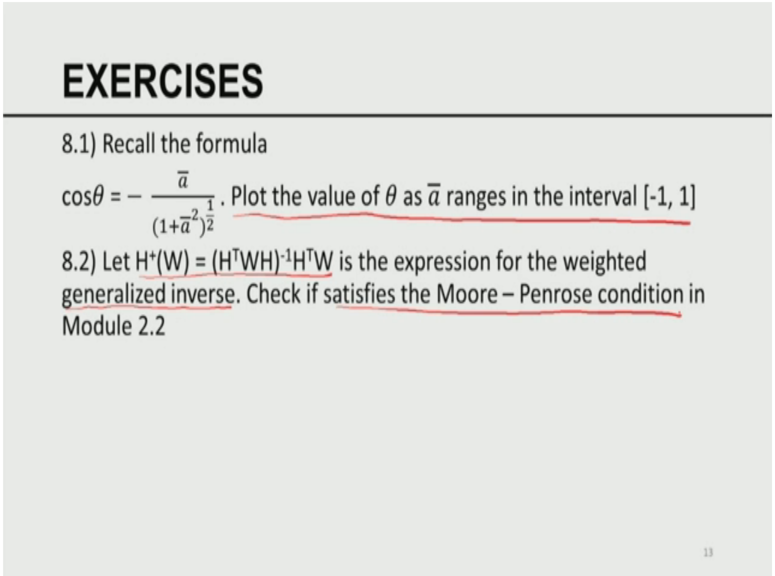


Illustration continued; so, I have already talked about this. So, r x is the error; the projection is given by Z minus Z hat, the actual error can be computed by this. So, this is the actual vector; this is the projection, the difference between the 2 is given by minus a bar 1, multiplied by Z 2. So, if I consider r transpose H; I get this and that is exactly that is essentially given by the; so I would like you to see this.

So, this is the vector H, so if you project that Z hat is the projection; the angle is theta that. So, by Schwarz inequality; the inner product is equal to the norm of r x times; the norm of H times cosine theta. I can compute each of these quantities explicitly; this is the inner product, this is the norm; that is a cosine of the theta. So, cosine of theta is equal to given by this ratio and that ratio essentially tells you the angle is not 90 degrees; that is theta is greater than 90 and r x makes an obtuse angle with H; see the illustration. So, when a is 0, cosine is 0 in which case theta is 90 the projection becomes orthogonal.

So, this is a very simple graphical illustration using a 2 dimensional case, where we can readily see; you can have weights. But for certain sets of weights, the projection is orthogonal for certain types of weights, the projection can be non orthogonal projection, so that is the important conclusion you coming out of this exercise.

(Refer Slide Time: 30:29)



So, in summary what is that we have accomplished in this small module? We have seen the importance of least square solution, within the context of data simulation, within the context of solving inverse problems. We are simply trying to embellish the character of the least square solution by relating the properties of the least square solution; to a very simple geometric fact, which we all have learnt in our first course; in geometry that of orthogonal projection and oblique projections.

So, what is the conclusion? You have a static deterministic inverse problem and you formulate it as a unweighted problem. The least square solution is given by orthogonal projection; if you formulate it as a weighted least square, the solution is given by an oblique projection. This essential difference between orthogonal oblique is essentially coming out of whether there is weight; whether there is not weight, recall the formula. So, I have couple of exercises; recall the formula that we have already seen. So, I would like you to be able to plot the value of theta as a bar ranges in the interval minus 1 to plus 1.

So, this exercise essentially tells you how the angle theta varies with the choice of a bar. And please recall from couple of slides earlier, a bar depends on a. So, a essentially controls a bar and as a bar ranges within minus 1 to plus 1; theta sweeps through a particular range. And I would like you to be able to plot this perhaps using MATLAB and convince yourself; what is the range of rotation angle theta 1 gets with a? The second exercise relates to the expression for the weighted generalized inverse check, if it satisfies; so, what is the idea here?

Now, any generalized inverse must be able to satisfy the Moore Penrose condition in module relating to matrices. So, here is an exercise that I would like to revisit; the Moore Penrose condition; that defines a generalized inverse and check to see, whether this expression for the generalized inverse with the weight satisfies the Moore Penrose. I think it will be a very worthwhile exercise to do, with this we come to the end of the discussion relating to the geometric facts and geometry interpretations of least square solutions.

Thank you.