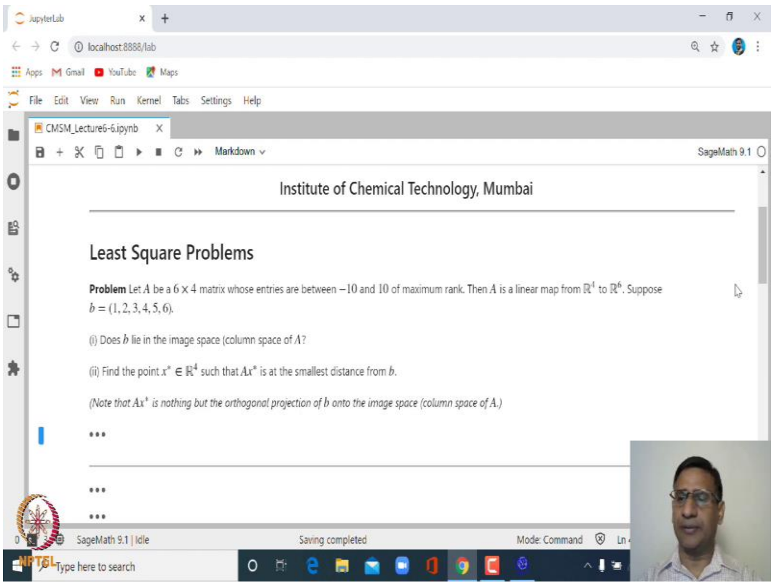
Computational Mathematics with SageMath
Prof. Ajit Kumar
Department of Mathematics
Institute of Chemical Technology, Mumbai

Least Square Problems
Lecture – 40
Least Square Solution with SageMath

Welcome to the 40th lecture on Computational Mathematics with SageMath.In this lecture we will look at Least Square Problems as an application to linear algebra. We have already seen how to solve least square problems using calculus. Let us first start with a problem.

(Refer Slide Time: 00:38)

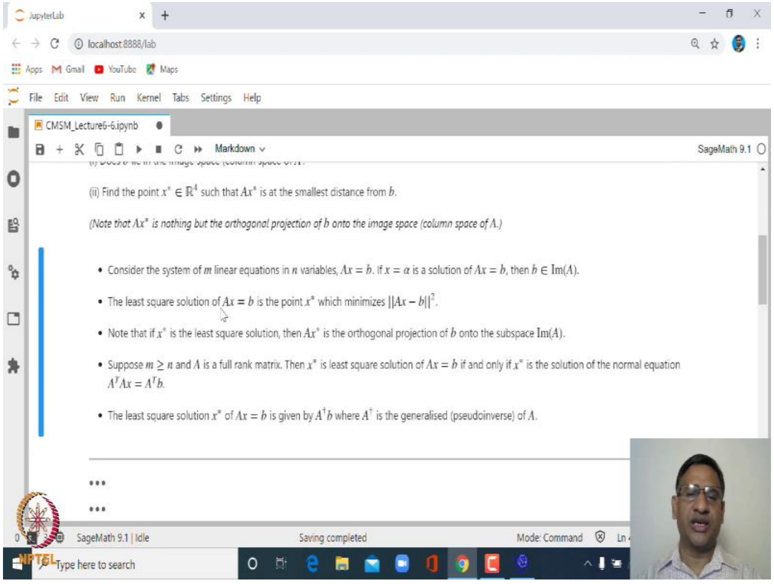


Suppose you have a random 6 cross 4 matrix whose entries are let us say integer from minus 10 to 10. Then A can be thought of as a linear map from R4 to R6. Now suppose there is a vector b in the codomain, which is 1, 2, 3, 4, 5, 6. Now, the question is, does b lie in this image of A, that is column space of A? If it lies, that is same as saying that Ax equal to b, has a solution.

However if it does not lie in the image space of A, in case b does not lie in the image space of A, then then we can ask, what is the point x star let us say in R 4, such that the image of x star under A, that is Ax star is at the smallest distance from b in the codomain. Such a x star is known as least square solution of this system of linear equation Ax equal to b.

You can note note that this Ax star is actually, nothing but but the orthogonal projection of b onto the image space of A. So, that is what we will obtain. We will obtain the least square solution as an orthogonal projection of b onto image space of A.

(Refer Slide Time: 02:14)



Let us see what we are saying here. In this case, we have a system of m linear equations in n variables. Let us say that is Ax equal to b.

If x equal to alpha is a solution of Ax equal to b, then b lies in the image space of A, and the least square solution of Ax equal to b is x star which minimizes this distance of Ax from b for all x. This is a minimization problem. That is how we have seen as a solution to this problem in calculus.

Note that if x star is a least square solution of least square solution, then Ax star is the orthogonal projection of b onto image space of A, that is what I already said. Now, suppose this m is bigger than equal to n, that is, you have more equations than the number of variables, and let us assume that A is a full rank. That means, the rank of a will be n here.

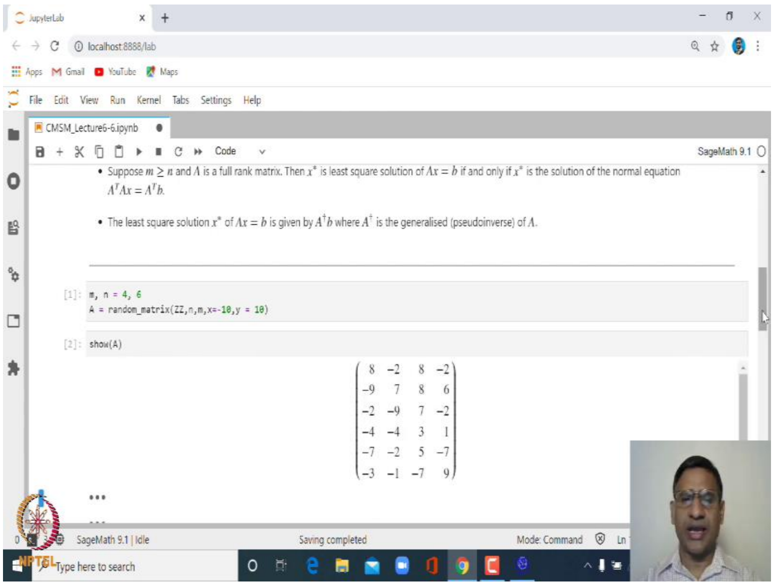
In this case solving this A x equal to b, or finding the least square solution of this A x equal to b, is same as, you just multiply both sides of Ax equal to b by A transpose. So, what you will get? A transpose A x is equal to A transpose b. This is what is called the normal equation of this least square problem.

Now, the least square solution of Ax equal to b, is nothing but the solution of this normal equation A transpose Ax is equal to a transpose b. Since A is of full rank, that is rank of A is n, A transpose A, which is n by n matrix will be invertible matrix and hence the inverse will exist. So, finding least squares solution of Ax equal to b is nothing, but solving this normal equation for x.

One can also obtain this least square solution namely, x star as a solution to this Ax equal to b, But since A may not be invertible matrix, what we need to do is, we need to find the generalized inverse of A and then multiplied by b. Notice that if A is invertible, the solution x is A inverse b, but if even if A is not invertible, we can find its generalized inverse.

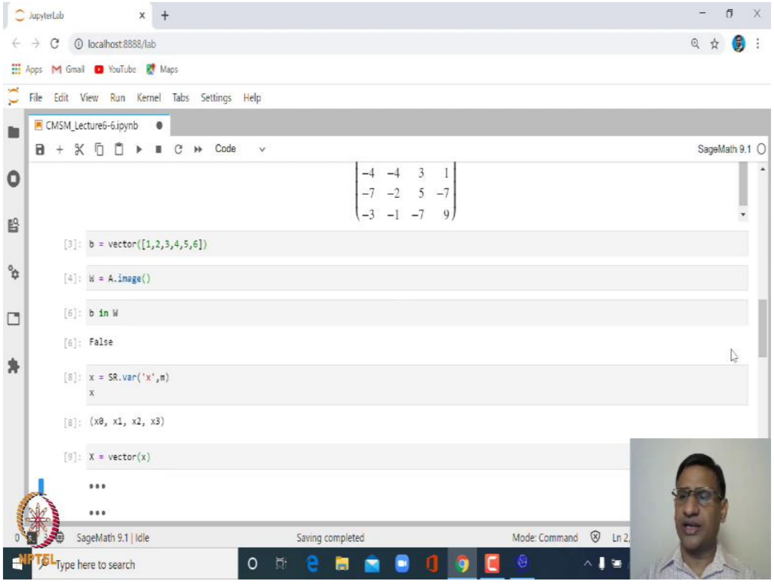
We will look at this, how we can solve this in using singular value decomposition, but Sage also has an inbuilt function to find generalized inverse of a matrix. So, in this case what will you get? You will get least square solution x star, as generalized inverse of A times b. This is how we denote generalized inverse of A. Now, let us look at how we can solve this problem, which we have stated.

(Refer Slide Time: 05:23)



Let us start with m and n as 4 and 6 and define a random matrix over integers whose entries are coming from minus 10 to 10 right. Now, let us see what this matrix is. This is how this matrix looks like.

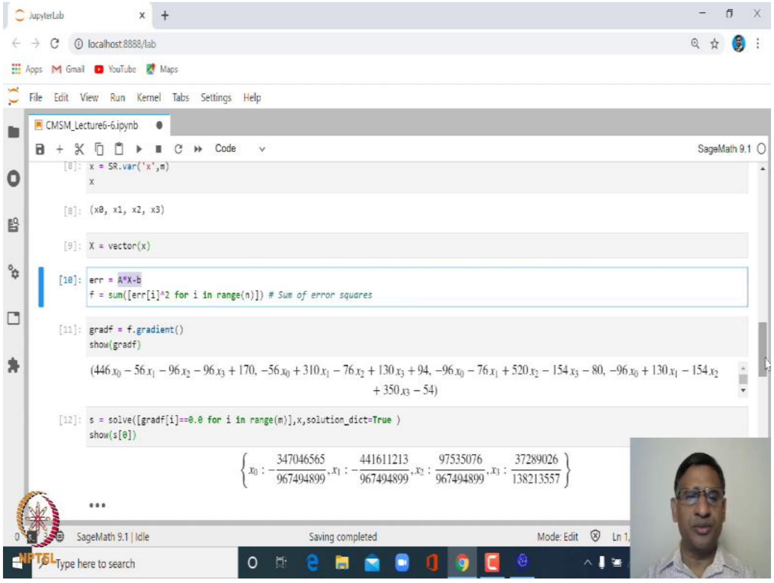
(Refer Slide Time: 05:44)



Next let us let us define a vector b which is 1, 2, 3, 4, 5, 6. Now, we can ask whether b lies in the image space of A. How do I how do check? You can define, let us say W, equals to A dot image, which is same as saying column space of A. Now, we can check if b in W? The answer is not true. So, b does not lie in W.

Now we can ask what is the least square solution of this Ax equal to b. Let us define x as variable x0, x1, x2, x3, x4, x5. You can check, what is this x here; x0, x1, x2, x3. Since m is 4, this is x0, x1, x2, x3, we have seen this earlier. Now, let us define capital X as a vector which is going to be column vector x0, x1, x2, x3.

(Refer Slide Time: 06:59)

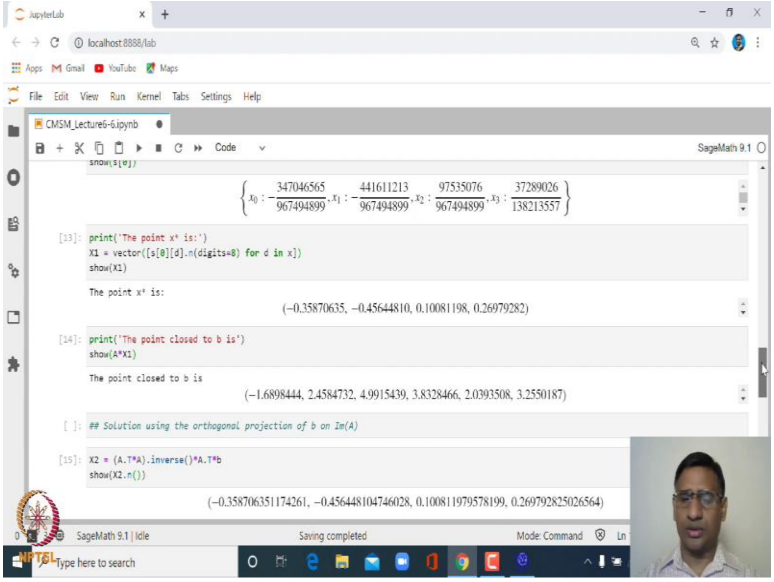


Now, if you recall in calculus, we solved this problem. First we define f as sum of the error squares and error in this case is nothing, but AX minus b. So, for any for any x0, x1, x2, x3, AX is in the image and its error is going to be AX minus b. Then find the sum of the error square, that is, just look at the ith coordinate of this AX minus b. This will be a column vector, look at the ith coordinate of this. This is going to be the error, take the error square and take the sum, that is your f. So, f is going to be a function of x0, x1, x2, x3 and we want to minimize this function.

In calculus, how did we do? Let us quickly do that. We obtained the gradient of f with respect to x0, x1, x2, x3 and this is what we get. Then we solve this for each each coordinate of the gradients would be equal to 0. That is, the critical point and when you solve this for x, the solution is going to be given by, let us just wait it is taking time.

So, x0 is this one, x1 is this, x2 is this, x3 is this. This is how, we have already obtained using calculus. In calculus we we wrote this as each coordinate, here we are writing this as a matrix, as a distance from b to AX.

(Refer Slide Time: 08:44)

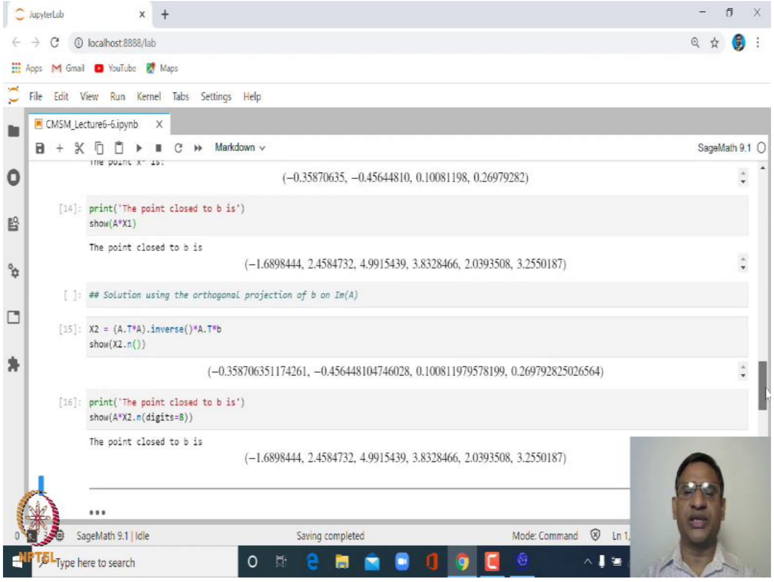


Now, in this case if you want to print the solution as a decimal value, this is what you get, right.

Next let us see the point which is closest to b, is A star X1, where X1 is this solution vector or the least square solution of A x equal to b.

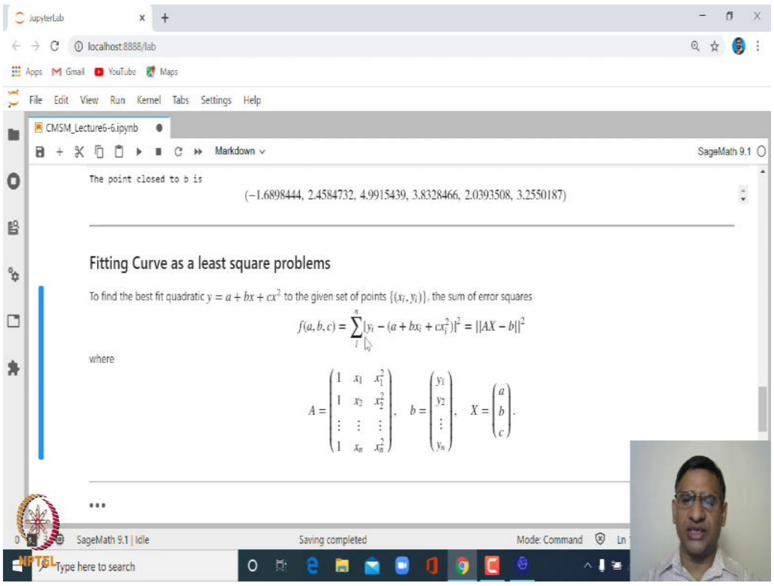
Next this point AX1, which we obtained using calculus, is nothing but orthogonal projection of b onto image space of A. So, how does one find orthogonal projection of b onto image space of A? We have already seen that, this is nothing but take A transpose A, inverse of this and then multiply by A transpose b. That is your point X2, which is same as, what you have obtained using calculus.

(Refer Slide Time: 09:53)



Now, multiply this by AX2, this is what you get, as the least square solution. So, we have seen that orthogonal projection of b onto subspace W, which is obtained by the image space, or column vectors of basis of W, is given by A into A transpose A inverse times A transpose b. And that is what we have got. So, least squares solution can be thought of as an application to this orthogonal projection and as an application to linear algebra.

(Refer Slide Time: 10:36)



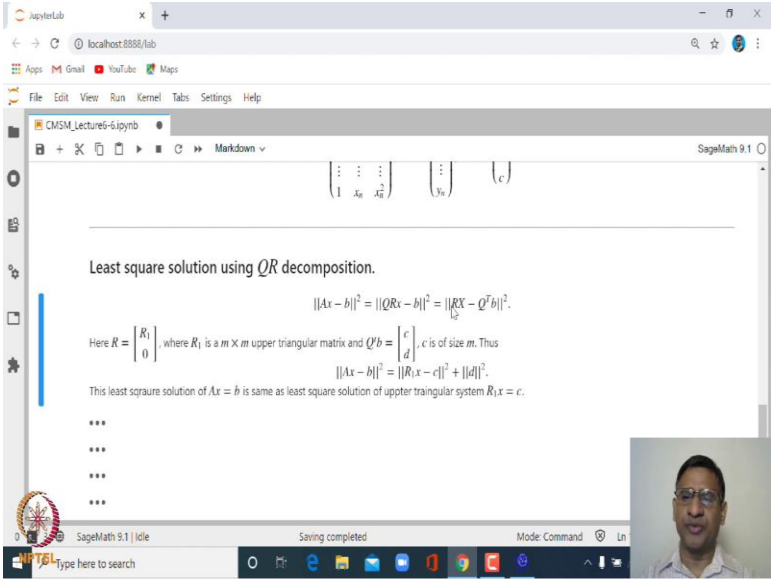
Now, this again we have seen that you can fit a curve to a given set of points in the plane or also fit some kind of plane in this space and this can be generalized.

If you have to find, let us say best fit quadratic y equal to a plus bx plus cx square to a given set of points xi, yi, i going from 1 to n, then this problem, we define, what is the sum of the errors square as a function of a, b, c and this is what we got.

This is summation yi minus a plus b xi plus c xi square, the whole square of this, that is the error of any point on this curve from this yi. And this can be expressed as matrix norm square of AX minus b, where what is capital A? Capital A is the column vector 1 1 1 and the next column is x1, x2, x3 up to xn. And the next column will be x1 square x2 square, x3 square up to xn square. And then b is y1, y2, y n. So, if you take A as this and capital X is a, b, c, you take AX, what will you get? The first entry of AX will be a plus b x1 plus c x1 square, the second row is going to be a plus bx 2 plus cx 2 square and so on. And this minus b, the first Ax minus b, the norm squared is nothing but the first entry in this summand. So, that is how you are converting this best fit problem to a least square problem. That is minimizing this sum of the error square. So, I will not solve any problem using this, I will just leave it an exercise. We already know how to obtain this solution of this least square problem.

And the solution in this case is going to be X star is going to be A transpose A inverse of that into A transpose b and then you multiply, this what you get X star by A you get the least square solution.

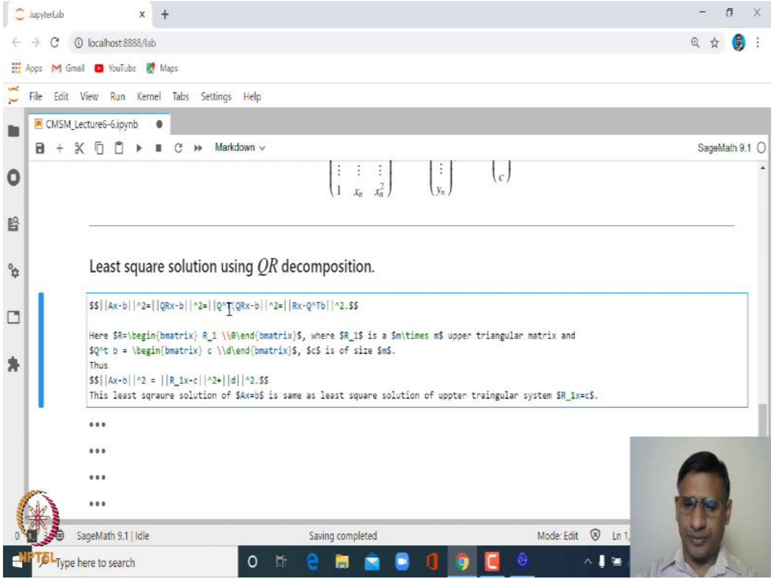
(Refer Slide Time: 12:55)



This least square solution, we can also obtain using QR factorization. So, how do we do that? Let us see this. We want to minimize, Ax minus b the whole square. Now suppose A has QR factorization ,A is equal to Q times R. Then this norm of Ax minus b whole square is QR minus b the norm square.

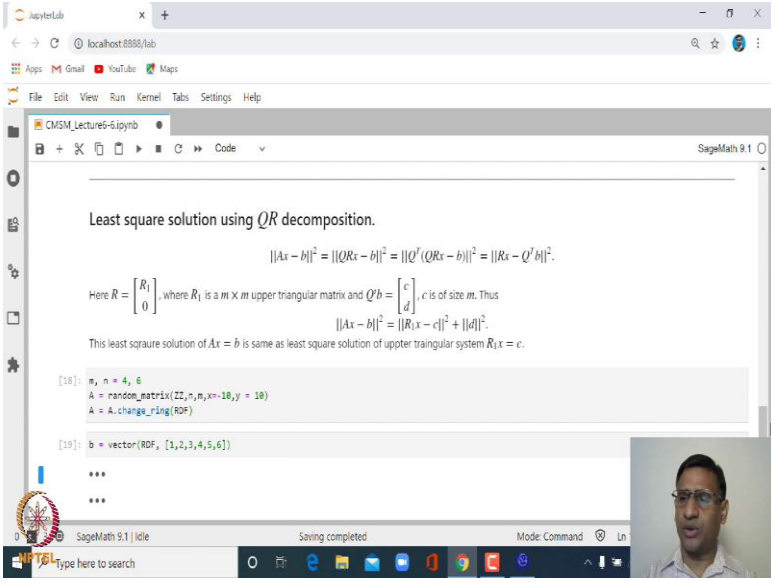
Now, we know that, since Q is orthogonal matrix the norm of Qx will be same as norm of x. So, if I take norm of this square, that is same as multiply this whole thing by Q transpose. Q is orthogonal therefore, Q transpose will also be orthogonal. But on the left hand side Q transpose Q is identity. So, this terms reduces to R x this would be a small x and minus Q transpose b.

(Refer Slide Time: 14:02)



This let me change this to small x and let me also add this intermediate step here. This is equal to, we are multiplying this by Q transpose.

(Refer Slide Time: 14:23)

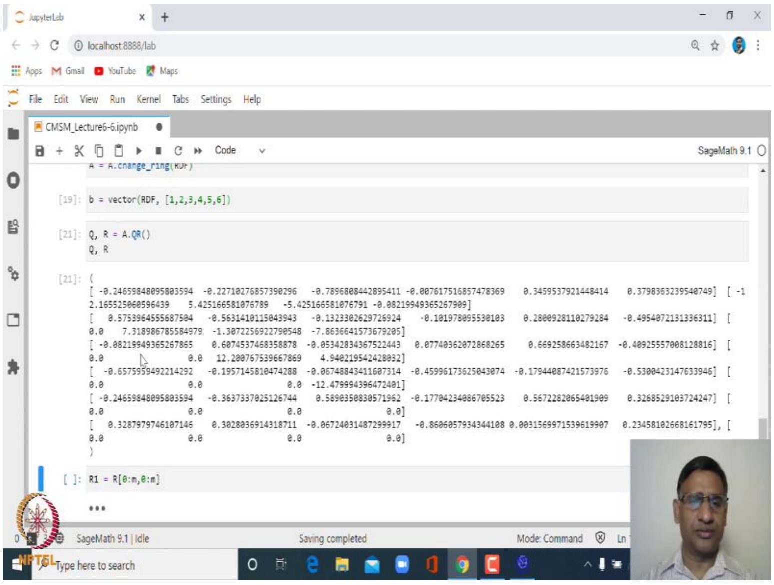


So, norm of Ax minus b whole square is same as norm of Q transpose Ax minus b, but A you replace it by QR. So, this now reduces to norm of Ax minus b whole square reduces, to norm of Rx minus Q transpose b the whole square. Now, this least square problem Ax equal to b has reduced to solving Rx is equal to Q transpose b. Since R is in an upper triangular matrix this is the same as solving this upper triangular system which is much easier. Here this R is upper triangular, you split this R as R1 and 0, 0 is a 0 matrix, where R1 is m cross m upper triangular matrix and Q transpose b you also split as c, d, where c is of size m and the d is going to be the size n minus m.

Therefore the norm of Ax minus b whole square, which is norm of Rx minus Q transpose b whole square will reduce to R1 x minus c, the norm square plus norm of d squared. But norm d is actually a fixed constant. So, minimizing norm of Ax minus b whole square is same as saying minimize this R1 x minus c whole square. So, the least solution of Ax equal to b is same as least square solution of R1 x equal to c.

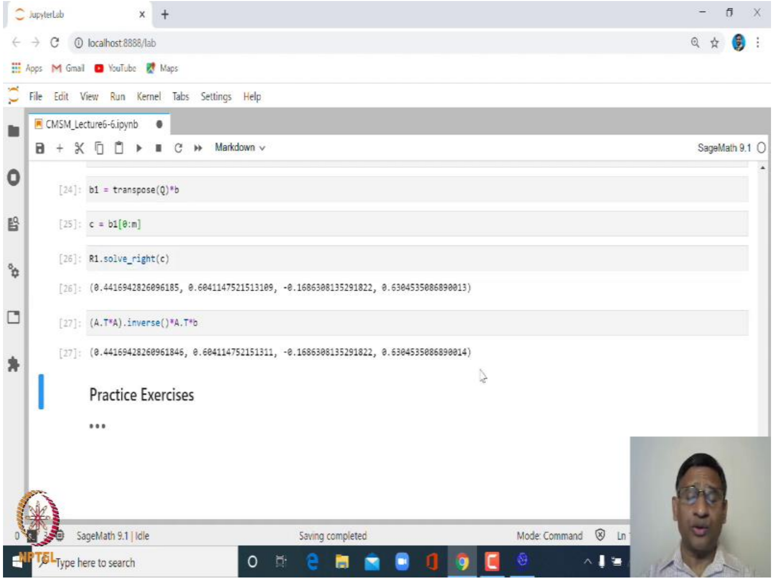
Let us see how we can solve this. Again let us take a random matrix 4 cross 6 random matrix and over integer. Since we want to find QR factorization, you may have to change the underlying domain from which the entries are taken. So, let us change this ring to RDF, that is extended real field. The vector b is again 1, 2, 3, 4, 5, 6. Then what do we do? We define QR factorization of A, Q, R is equal to A dot QR.

(Refer Slide Time: 16:42)



So, we can print what are Q and R. So, this is the two matrices Q and R. It looks slightly complicated, but since this is a 4 cross 6 matrix it looks like this.

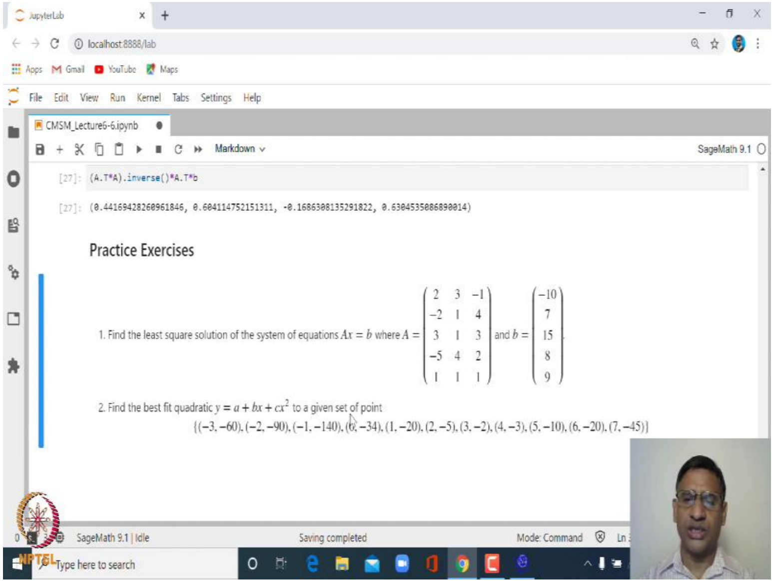
(Refer Slide Time: 17:15)



Now, let us define R1. R1 is, from R you take out the first m rows and m columns that is your R1, and then define b1 as Q transpose b ,that is b1. Again from b1 you define this as c and d. So for c, take the first m component of b1 that is your c. Now remember we need to solve R1 x is equal to c, that is what we need to do. So, what we have to do? We need to solve R1 dot solve right and bracket c. That is the solution you are getting, so that is your x star. Now if you look at what is the solution using orthogonal projection.

So, it is A transpose A inverse of this into A transpose b. What you get here is same as what you get in this case. So, solving this least square problem is same as solving this system of linear equation R1 x is equal to c. Since R 1 is invertible matrix, solving it is quite easy you can obtain this using Gaussian elimination method or simply take inverse of R1 multiplied by c. So, that how you can solve this least square problem using QR factorization except that you need to do this computation of finding QR factorization of A. But once you have done this, then it is quite easy.

(Refer Slide Time: 18:59)



Now let us look at some exercises. Just two exercises here.

Find least squares solution of this system of linear equations which is 5 equations in 3 variables and b is this vector.

And find the best fit quadratic y equals to a plus bx plus cx square to a given set of points this. This problem, solve using minimizing norm of AX minus b whole square. We have already seen what should be A, what should be X what should be b in this case. So, that is what you have to use and once you have a b and capital X, then the least square solution will be A transpose A inverse of that into A transpose b that is it.

Instead of quadratic, you can try to fit cubic, you can try to fit nth degree polynomial. So, if you want to fit nth degree polynomial, this matrix a will have first column as all 1, 1, 1, second column will be x1, x2, xn, third column will be x1 square x2 square x3 square and so on. Last column will be x1 to the power n, x2 to the power n, dot dot xn to the power n. So, fitting nth degree polynomial to a given set of points is quite easy.

Similarly, if you have to fit a plane to a given set of points in R 3, then plane can be written as z is equal to a plus by plus c. So, in this case the matrix A is going to be the first column will be all 1, 1, 1; second column will be coefficient of x, third column will be coefficient of y. And b is going to be the coefficient for z. So, z1, z2, zn will be the the column vector will be b, and x1, x2, xn will be the second column of A, y1, y2, yn will be hm the third column of the matrix A. That is how you can solve this least square problem and as an application to linear algebra.

There are other methods also to solve these least square problems, but we will not get into all these things.

Thank you very much, we will see you in the next class. In the next class we will look at singular value decomposition which is again a concept which has several applications.

Thank you.