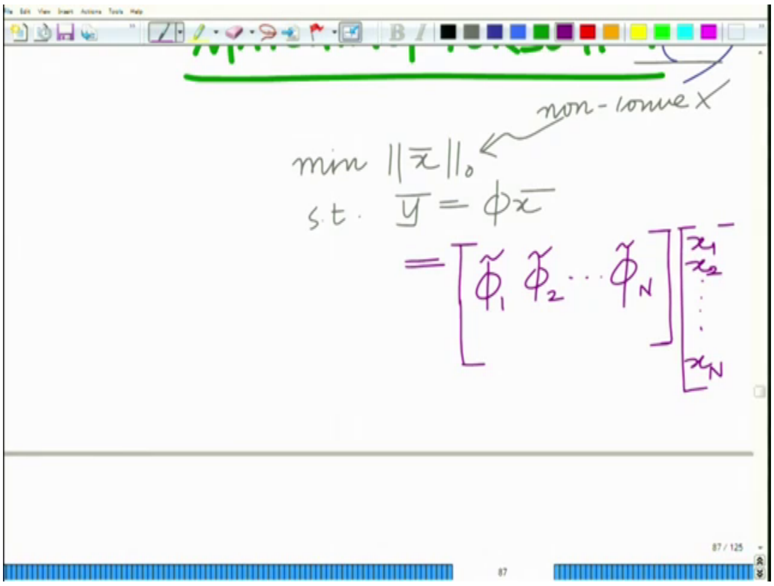
Applied Optimization for Wireless, Machine Learning, Big Data

Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 57
Practical Application: Orthogonal Matching Pursuit (OMP) algorithm for Compressive Sensing

Hello, welcome to another module in this massive open online course. So, we are looking at compressive sensing and we are seen that the cost function for the compressive sensing problem that is minimizing the l 0 norm is highly non convex and therefore, we have to invent or come up with intelligent techniques to solve this thing and in this module, we are going to look at one such technique which is termed as the orthogonal matching pursuit ok.

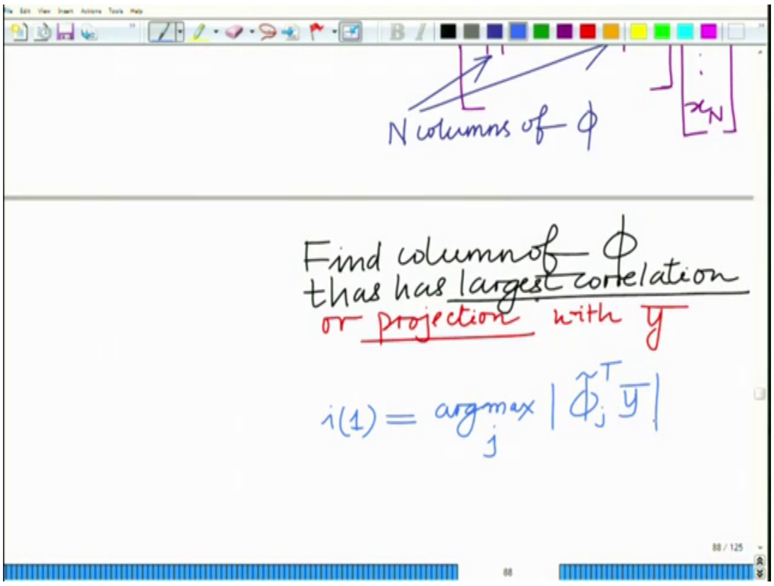
(Refer Slide Time: 00:38)



So, we want to start looking at one of the seams schemes for solving for sparse signal recovery, this is known as orthogonal matching pursuit what happens? And there is also abbreviated as OMP, I remember the genesis of problem, this problem is something like this, we want to enforce the sparsity, that is we want to minimize the number of non 0 entries given by the l 0 norm subject to the constraint y bar equals phi x bar, this is non convex. So, we want to come up with a scheme to solve this and that particular scheme is OMP orthogonal matching pursuit.

What happens in orthogonal matching pursuit? You write this matrix phi as a set of columns it is corresponding columns ok. So, you have phi 1 tilde, phi 2 tilde. Remember each is a column ok. So, the set of columns are since, we call is the rows as phi 1 bar, I am calling the columns as phi 1 tilde, phi 2 tilde since, this is a N m cross N matrix there are N columns ok. So, this is phi N tilde times x 1 x 2 up to x N.

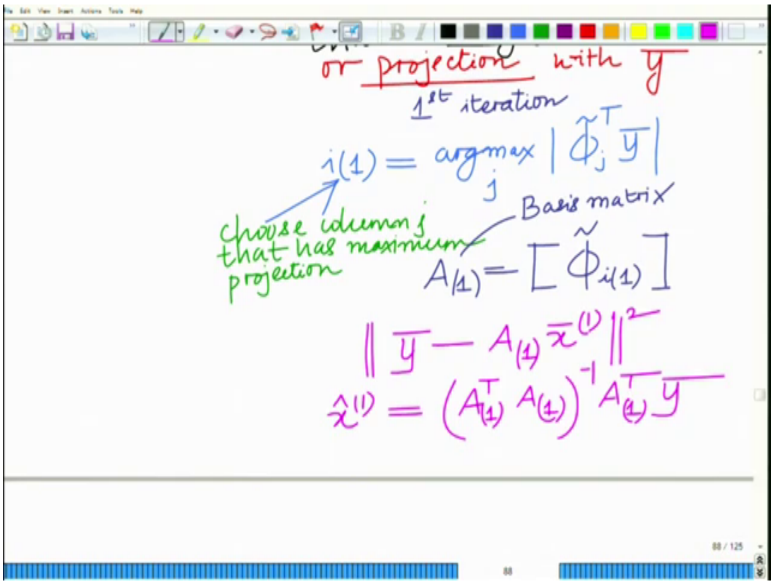
(Refer Slide Time: 02:24)



And so these are N columns, N columns of the matrix phi, now the orthogonal matching pursuit remember the name itself implies matching alright. So, you are looking for something that matches what is the thing that matches? You are looking for the column that closely matches the vector x bar, which means basically you have to find the projection of x bar on each of these columns of you have to find the projection of y bar on each of these columns of the matrix phi and choose the one that has the maximum.

So, you are basically trying to find, which of these columns is probably there in the linear combination alright and therefore, you try to find the one which closely matches y bar which means basically, the column which has the largest projection on y bar or the column which basically, yields the largest projection of y bar, so the way to do that is. So, we find the column of phi? Find the column phi? That will has largest correlation or basically projection with y bar. So, the way to do that is we have i 1 equals at the first stage, what we do at the first stage? Is basically we choose i 1 the index, one amongst all the j columns.

(Refer Slide Time: 05:02)



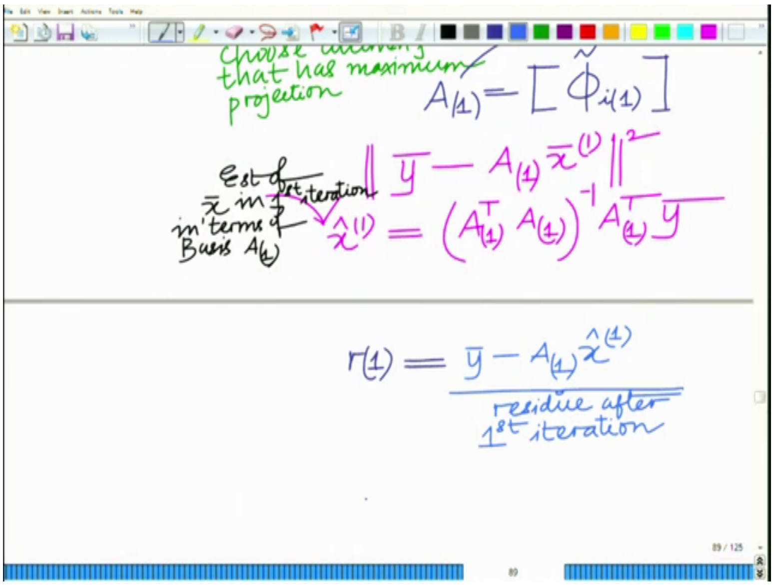
Let me write it clearly as the 1 that maximizes the as a column j at this where is, you look at all the columns phi tilde j take the projection phi tilde j transpose y bar and choose the one that has the maximum projection. So, choose the 1 right choose the column j that has the maximum projection ok.

And now, what we do? Is we compute, we start building the basis matrix that is A 1, which basically comprises of this first column, phi tilde of i 1 ok. So, this is at this stage, this is a single column matrix ok. So, this is A 1. So, we think of this as the basis matrix, this is by the way the first iteration of the algorithm and now what we do? Is given, this basis matrix we find try to find the best estimate of the vector x bar? So, we minimize. So, it is like we are finding this basis alright, this is the basis for the linear combination A 1 and now we are trying to find minimize the least squares norm, such that you find the best vector x bar 1, in the first iteration that minimizes, this error this error corresponding to y bar.

And therefore, the estimate x hat 1, corresponding to our basis A 1 is given as A 1 transpose A 1 inverse A 1 transpose y bar, this is basically the first iteration ok. So, what we are doing is you are trying to estimate the basis ok. That is estimate the columns of phi, which are present in the linear combination to give rise to remember only few elements of x bar or non 0, which means only few a few columns of phi are present in the linear combination. So, which are those columns? That is what we are trying to find by this orthogonal matching pursuit. So, we take the projection of each column one y bar, finding the one that has a maximum project, choosing that column as the basis, then finding the best type for optimization to the y bar based on that basis that is what we are doing here by solving this least squares problem.

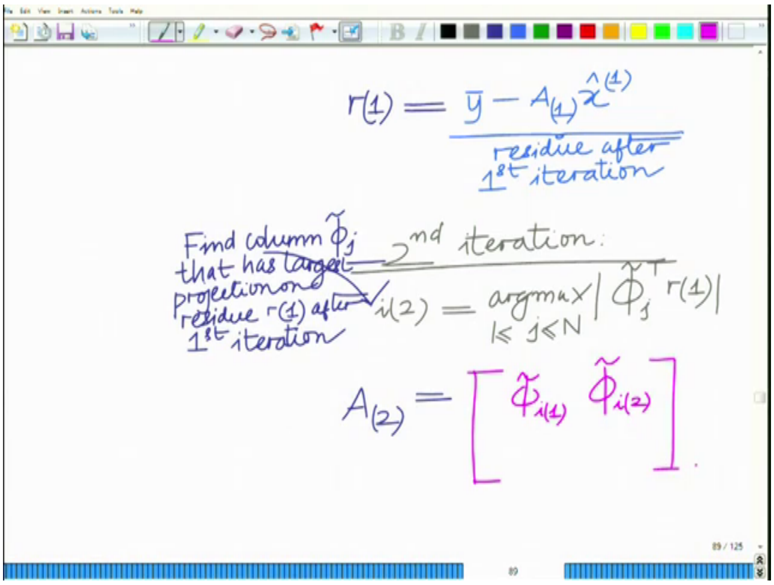
Now, what we find is for find the residue, the residue that is left after getting this best possible approximation.

(Refer Slide Time: 08:00)



So, the residue. So, this is basically Estimate of x bar in remember, but first iteration in terms of the basis A 1 and now the residue after the first iteration that is after your approximation, that is remember y bar minus the basis times x hat of 1, that is basically you have estimated x hat of 1, whatever is remaining that is y bar minus the basis times x hat of 1.

(Refer Slide Time: 09:08)

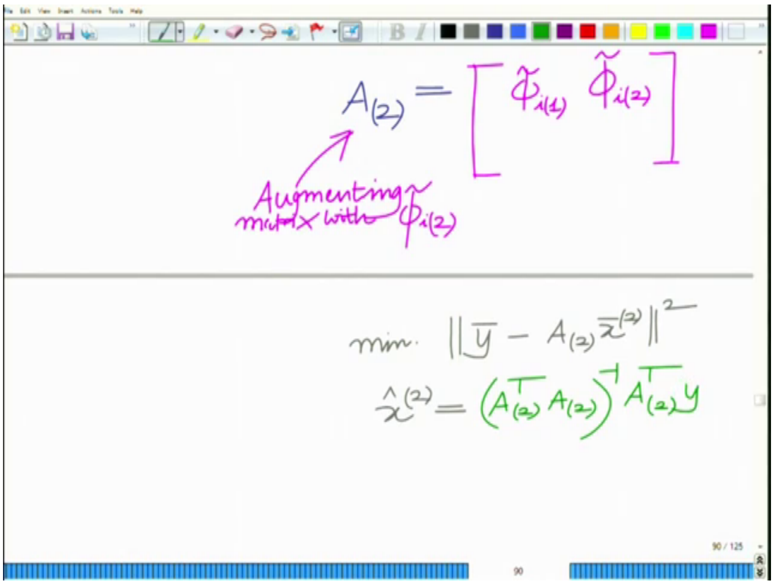


This is your residue after the this is your residue after the first iteration ok. So, this is the first iteration in the orthogonal matching pursuit.

What do we do in the second iteration? Now in the second iteration now, we start. So, this completes the first iteration now, let us do the second iteration ok, I hope this is clear. In the second iteration, we do something very simple; we take the projection of this on the residue. So we find the column, which has the maximum projection on the residue after the first iteration.

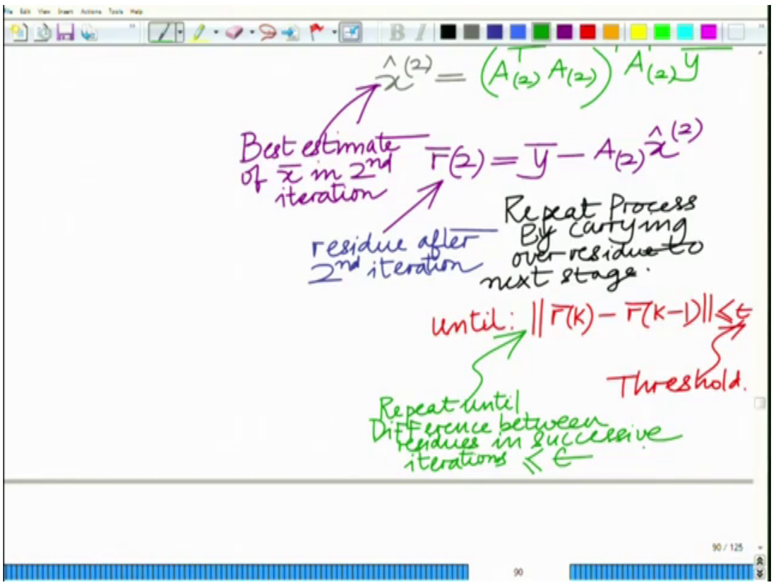
So, what we are doing here and this is the key step, find column phi tilde j that has the largest projection on residue r 1, after the after the first iteration? So, after the first iteration you are find we have left with the residue r 1 ok, this still has to be approximated. So, you find the projection of each column of phi on this residue and choose the column, which now has the maximum projection on this residue that is what we are doing that is your index item.

(Refer Slide Time: 11:17)



Now, you augment your basis matrix your basis matrix, now become previously you have remember, you have A 1 column matrix phi tilde i 1. Now you augment it with phi tilde i 2 ok. So, you are augmenting the basis matrix A 2, you are augmenting this with the column phi tilde i 2. Once again, you find the best estimate x bar via least squares, now against again you find you have the basis matrix, you find the best estimate x of x bar at the second iteration. The least squares estimate x hat 2, that will be given as A 2 transpose, we know the least squares procedure A 2 transpose A 2 inverse A 2 transpose

(Refer Slide Time: 12:14)



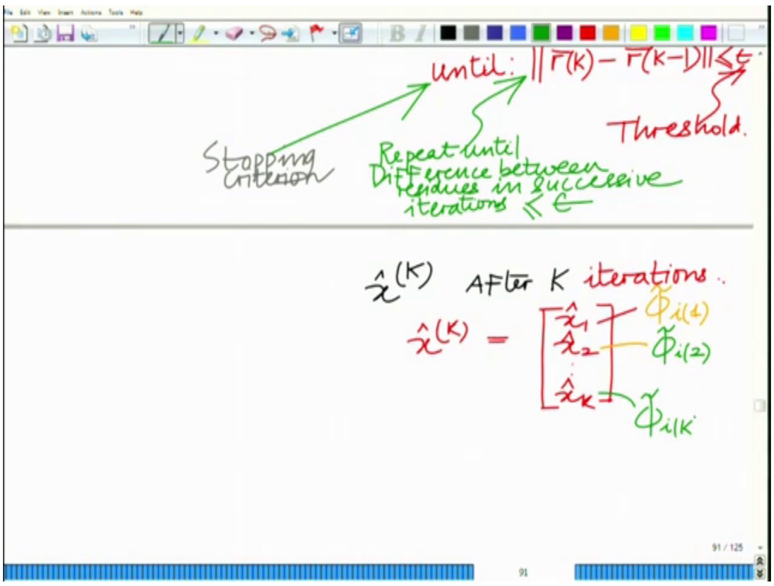
So, what is this x hat? This is the estimate of x best estimate of x hat 2 x bar best estimate of x bar in the second iteration.

Now, you find the residue r bar 2, which is y bar minus the basics matrix times x hat and what is this? This is the residue after the second iteration; this is the residue after the second iteration and then subsequently, now what you do in the third iteration? Take this residue, find again the projection of each column of the matrix phi, choose the one that maximizes, there has a maximum projection or let expand your basis or augment the basis least squares estimate, alright find the residue and alright.

So, keep repeating this process by projecting on their projecting at each iteration start by projecting the residue from the previous iteration onto each column of phi, choose the one that maximizes, choose the one that has a maximize maximum projection and then repeat the process ok.

So, repeat process by carrying over residue to the next stage so each stage the residue carries over to the next stage. Now till when to the you repeat there can be several criteria for instance, you might have a fixed number of iterations or typically you repeat until this residue, the difference between the residue, that is a residue stops decreasing, that is you repeat until such stage that your residue that is let us say, you have K iterations, that is this epsilon is some threshold.

(Refer Slide Time: 15:54)

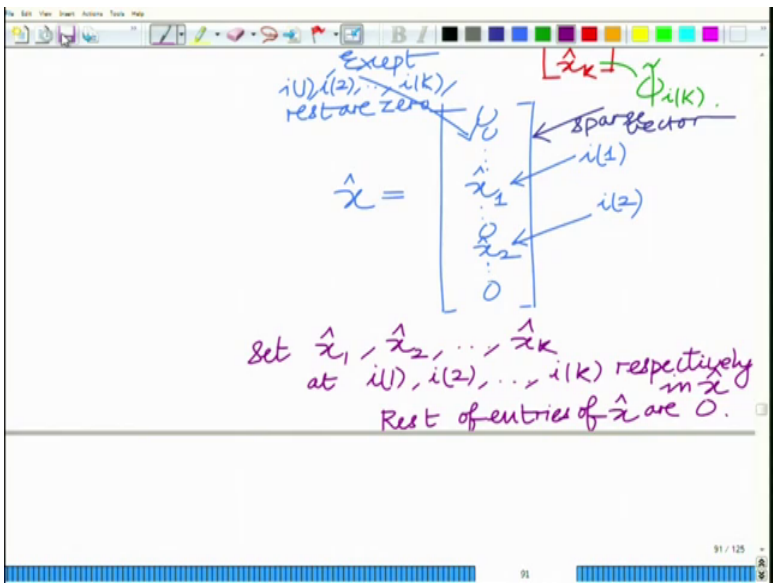


So, what we are saying is repeat until the difference between residues is smaller than epsilon residue does not decrease any further ok, difference between residues in successive iterations is less than or equal to epsilon that is the residue does not decrease and if the epsilon is the threshold, this is termed as the stopping criteria, every iterative algorithm you have to choose a suitable stopping criteria ok.

So, this is termed as this is termed as the stopping criteria, there can be many stopping criteria for instance, you can have a fixed number of iterations, residue until there is a difference between the successive residues. You can iterate until the difference between the the residues stops decreasing any further, there is the difference between the successive residues is less than some predefined quantity epsilon or so on there can be various different stopping criteria and now after the stopping criterion. Let us say, you stop after K iterations x hat. So, you have after K iterations you have x hat K, which means you basically obtained a sufficiently fairly good up estimate of approximation to y bar and the residue is not decreasing any further after K iterations, let us say you have x hat K.

And naturally if you look at x hat K, that will have elements remember after K iterations the augmented matrix is of size x K columns. So, you will have x hat 1, you will have the estimated vector has K elements x hat 1, x hat 2 up to x hat k, but remember these correspond to columns for instance, this corresponds to column phi tilde i 1, first one corresponds to column, first one corresponds to column phi tilde i 1. Second one corresponds to column phi tilde i 2 and last one corresponds to column phi tilde i K and so on.

(Refer Slide Time: 18:17)



And then therefore, how do you finally, estimate the vector x? Now x hat will simply be a vector that mostly contains zeros, except corresponding to the location i 1, you have x hat over, mostly zeros corresponding to the location i 2, you have x hat 2 ok.

So, this is at location i 1, this is at location i 2 and the rest are zeros or rest zeros. Except your locations, these are known as the sparse locations rest except i 1. So, we say except i 1, i 2, i K rest are 0. So, this is a sparse vector basically. This is a sparse vector and. So, finally, to conclude what we have is set x hat 1, x hat 2 up to x hat K at i 1, i 2 i K respectively, in x bar or in x hat and rest of the entries of x hat are 0. Remember all this is saying is basically, through these various iterations you have estimated the basis columns, which is which are phi tilde i 1 phi tilde i 2 at the second iteration so on, until that the Kth iteration, you get phi tilde i K, which gives you the K largest projection on the residue, at the Kth iteration, you can say that and therefore what you have is basically you have this set of columns i phi tilde i 1 phi tilde i 2 phi tilde i K and therefore, the x hat vector that you estimate naturally we have K entries ok, x hat 1, x hat 2 up to x hat K.

So, each of these entries corresponds to that particular column alright and therefore, in the original vector x bar, you have to set these entries x hat 1, x hat 2, x hat K add the corresponding columns that is i 1, i 2, i K rest of the entries of x bar will be 0, that is what gives you the sparse vector x bar alright that is the vector, that is estimated using the orthogonal matching pursuit ok.

So, it is a very simple algorithm, it is a very simple, it is a very intuitive that is at each stage you are finding, the column of phi which has which has the largest projection on the residue on which, the residue has the largest projection ok. So, that is what you are finding degeneration and you are expanding the basis and at the same point, once you have the basis, you are trying to find the best vector x bar that minimizes the residue and you are computing that residue, taking it is projection in the next stage, next iteration repeating that way alright. So, we will stop here, we look at an example in the subsequent module, but we will stop here.

Thank you very much.