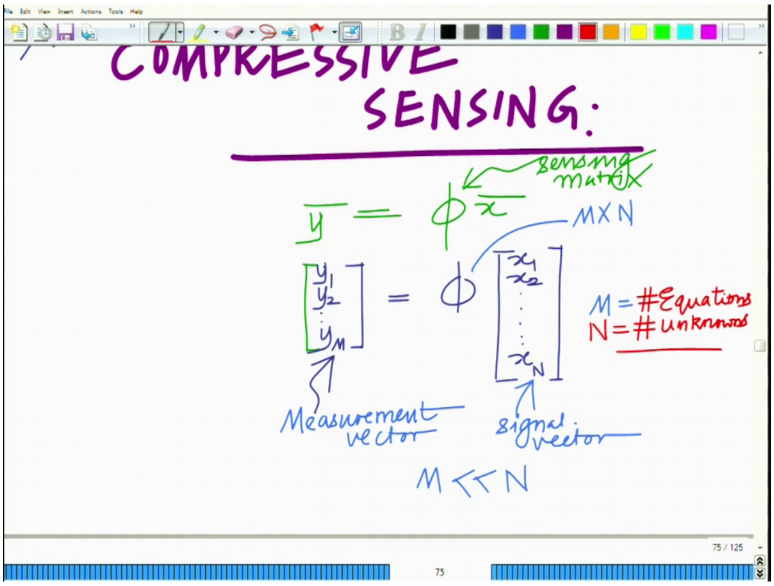
Applied Optimization for Wireless, Machine Learning Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 56
Practical Application

Hello welcome to another module in this Massive Open Online Course. So, we are looking at compressive sensing all right where you try it to compress not after the sensing process all right. So, you try to compress during the sensing process itself by making much fewer number of measurements in comparison to the dimension of the signal. And then later try to reconstruct the signal from the very few measurements may all right.

(Refer Slide Time: 00:41)

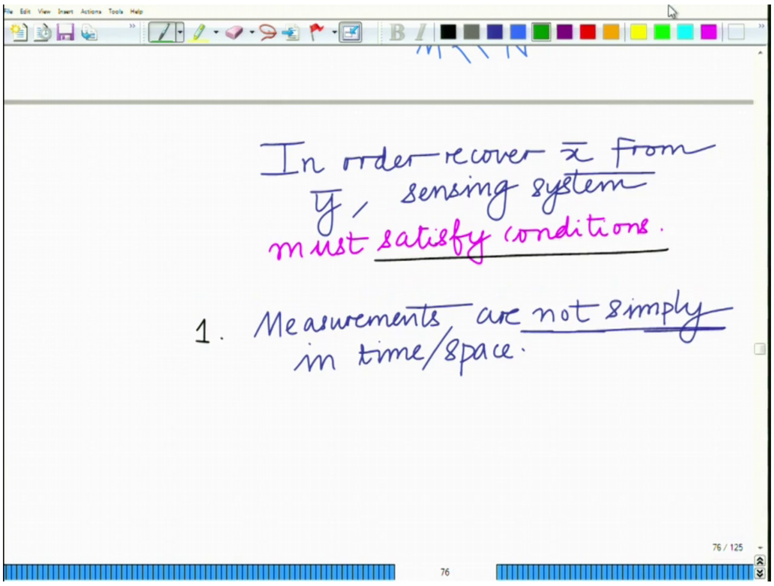


So, let us continue our discussion on compressive sensing. Let us see what are the necessary conditions for this. So, you are looking at compressive sensing and what we have seen is 1 thing that has to be done we are doing is essentially you have a measurement vector y of equals phi times x bar. And this is you are sensing matrix. And you have M measurements y 1 these are your M measurements you have your matrix phi. And I am going to draw this vector much larger than y bar purposefully because you are making much fewer measurements than the length of the signal x.

So, this is your measurement vector. And this is your signal and this is your sensing matrix and remember this is an M cross N matrix. And what we have is at M is less than or equal to and that is we make significantly fewer measurements or let us say M is significantly lower than l all right. Now if you view this as a system of equation equations then we have M as a number of equations N equals number of unknowns or number of equations is less significantly lower than the number of unknowns.

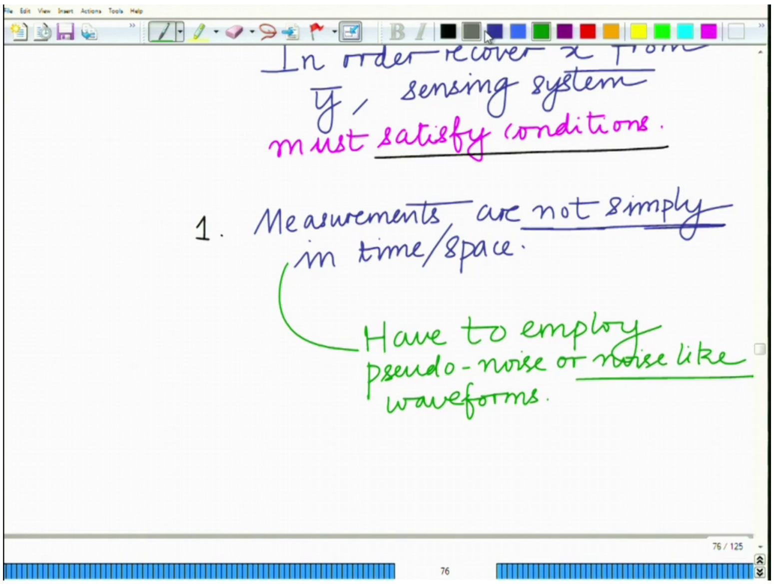
So, simple linear algebra tells us, high school level linear algebra tells us that 1 cannot reconstruct the vector x bar of length N which basically comprises of N unknowns from simply M equation since the number of equations is much lower than the number of unknowns. So, this is an underdetermined system all right so, our systems. So, one cannot uniquely determine x bar. Therefore this system, the sensing system has to satisfy certain special properties In order to recover x bar from the observation.

(Refer Slide Time: 03:37)



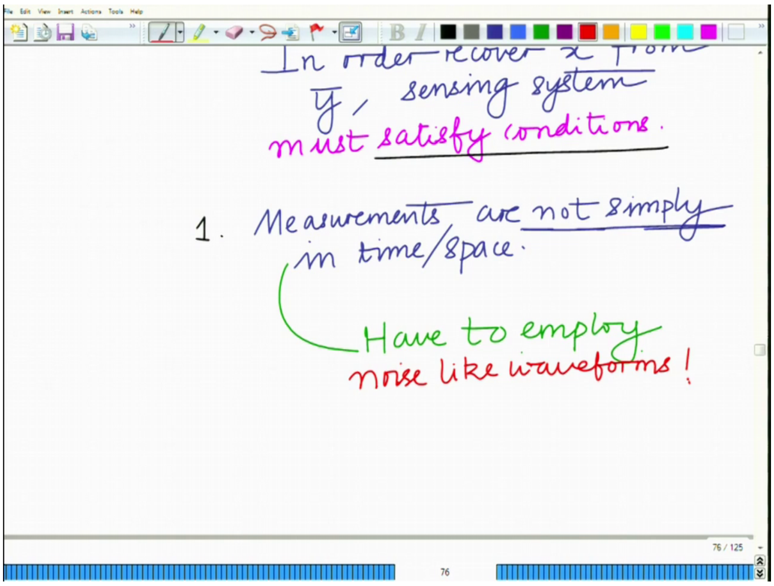
Now what are those conditions or what is what the properties the system has to satisfy that is what to look at. So, in order to recover or reconstruct; in order to recover x bar from y bar, the sensing system must satisfy some conditions. What are those conditions for instance. Let us look at the first condition. The first condition states that this must satisfy some kind the first condition states that measurements are not simply in the time or space. These are not simply in time or space. These are not simply in time or space rather.

(Refer Slide Time: 04:59)



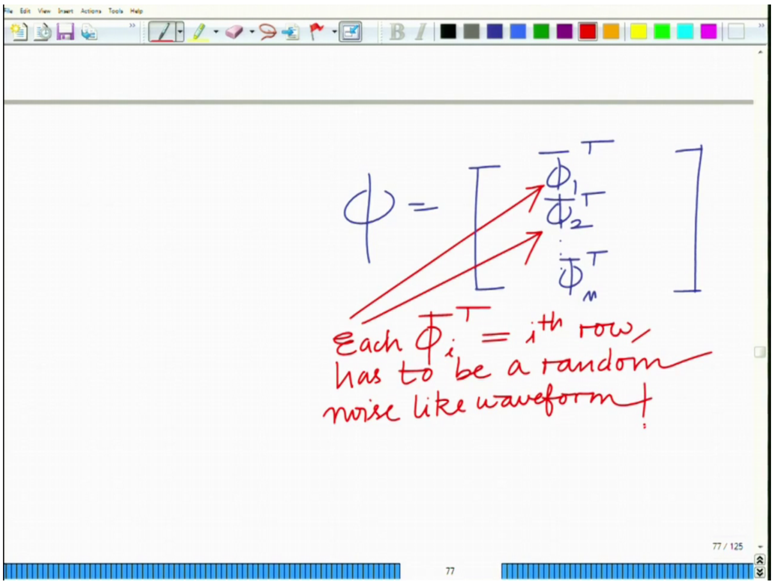
They have to employ noise like measurements have to employ noise like or one can say pseudo noise or noise like these have to employ pseudo noise or noise like waveform. Or let us just say noise like waveforms measurements have to employ.

(Refer Slide Time: 05:35)



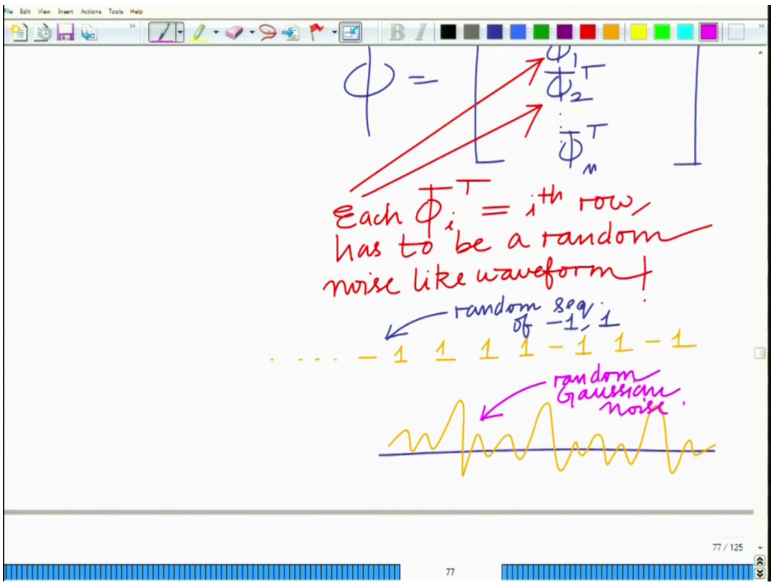
What is the meaning of that the meaning of that is if you look at this sensing matrix.

(Refer Slide Time: 05:55)



If you look at remember we said we have the sensing matrix which is M cross N which means there is M rows and M columns. If you look at each row which we are denoting by phi 1 bar transpose phi 2 bar transpose phi M bar transpose. So, these are the rows. So, each row of the sensing matrix each phi i bar transpose which is the ith row has to be a random noise. This has to be a noise like wave form which means that it has to be something very random, it has to look like a random.

(Refer Slide Time: 06:57)

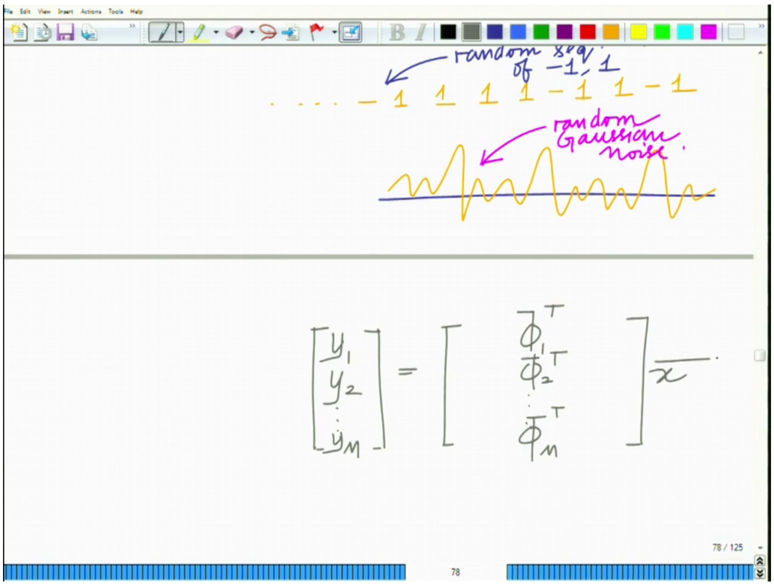


Or pseudo random sequence for instance like some random sequence of, either can be a random sequence a random sequence of minus 1 comma 1. Or it has to be some random noise like waveform such as Gaussian so on.

So, it has to be either a random noise like or a random sequence ones minus one so on. So, each row that is we denote the ith row by phi i bar transpose of the sensing matrix phi, this has to look like noise just to look purely like noise it cannot be one followed it not.

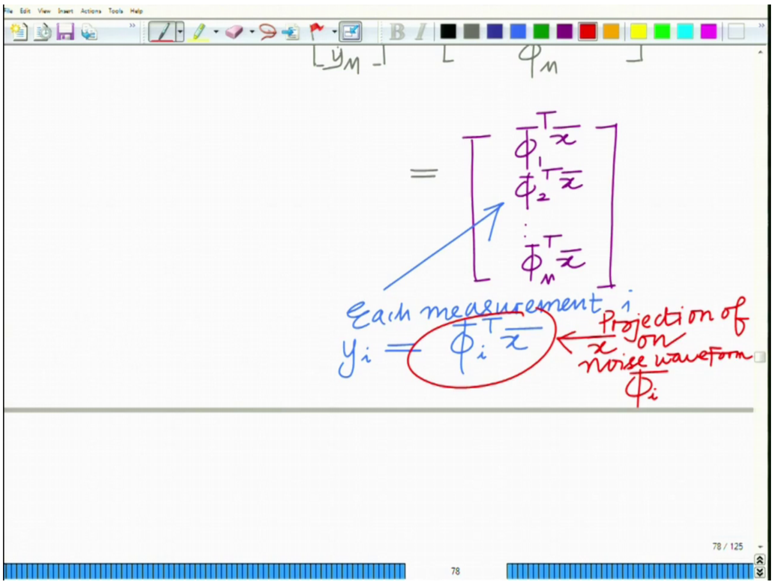
So, the matrix phi cannot look like an identity matrix. 1 followed by 0 0 1 0. It cannot look like an identity each row has to look like and these rows have to look like independent realization of the noise waveforms alright. And thereby when you are making the measurement what you are doing is you are taking the projections of the signal on this noise like waveform.

(Refer Slide Time: 08:19)



So, when you are making the measurements, you have your y 1 y 2 y M which is equal to y 1 bar transpose phi 2 bar transpose phi M bar transpose times x bar.

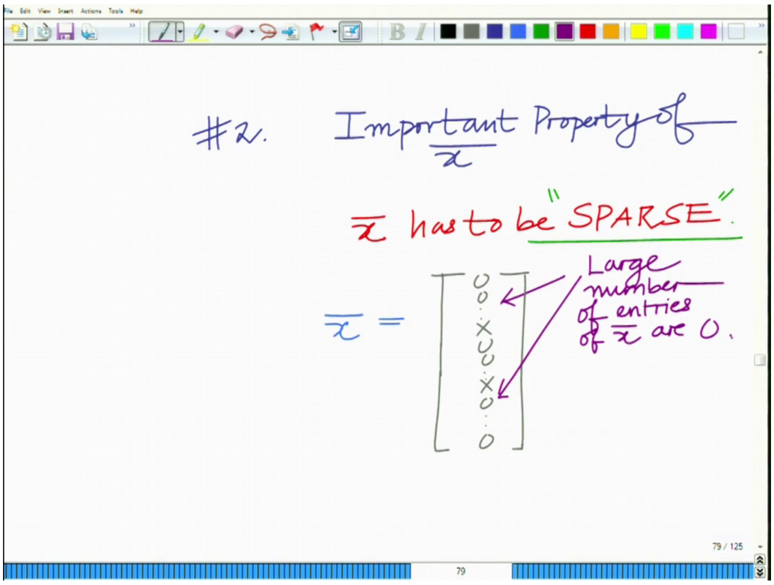
(Refer Slide Time: 08:53)



So, this is equal to phi 1 bar transpose x bar phi 2 bar transpose x bar phi M bar transpose x bar. So, each is a projection. So, each measurement if you look at this; each measurement i which is phi i bar transpose x bar the measurement y i is phi i bar transpose x bar. This is the projection of x bar. This is the projection of x bar on the random noise like waveform phi i bar. This is the projection of x bar.

So, what you doing when you are taking these measurements which is each row of phi the sensing matrix is a noise like waveform and when you take in the measurement what your doing is nothing, but taking the projection of this x bar a linear combination of x bar using this noise like waveform which is the row phi i bar. So, the ith measurement is phi i bar transpose x bar all right. So, that is an important property of the sensing matrix.

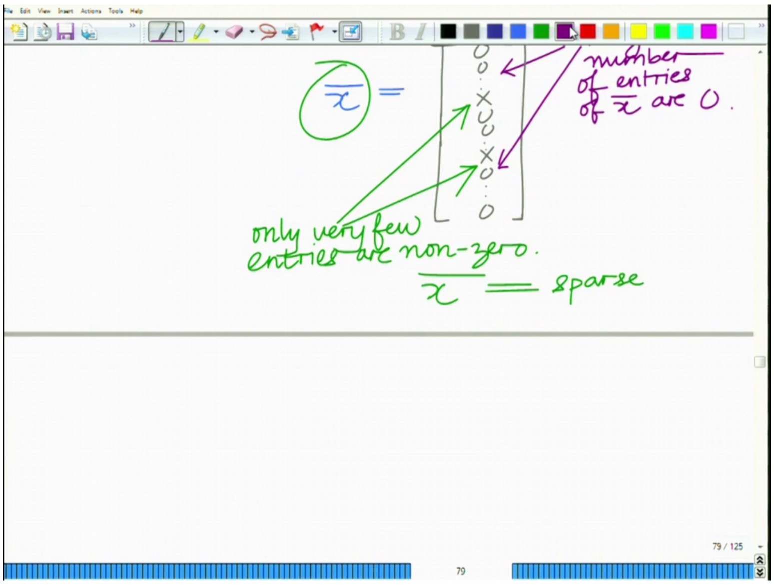
(Refer Slide Time: 10:31)



Now, let us talk about the say vector x bar. Now the vector x bar itself has to satisfy an important property. And this I would say is the important property of most important property of x bar the vector x bar. And I am going to introduce this terminology which we are going to use very frequently x bar has to be sparse and this is a very important property; x bar has to be sparse what is the meaning of this. Remember when we say something is sparse when we say an area is sparsely populated which it means that there are very few people in that all right. This is sparse implies that it is very few some it is sparse and some object implies there is very few numbers of that sparse object in particular if it is sparsely populated a country is sparsely populated implies that there are very few people in there.

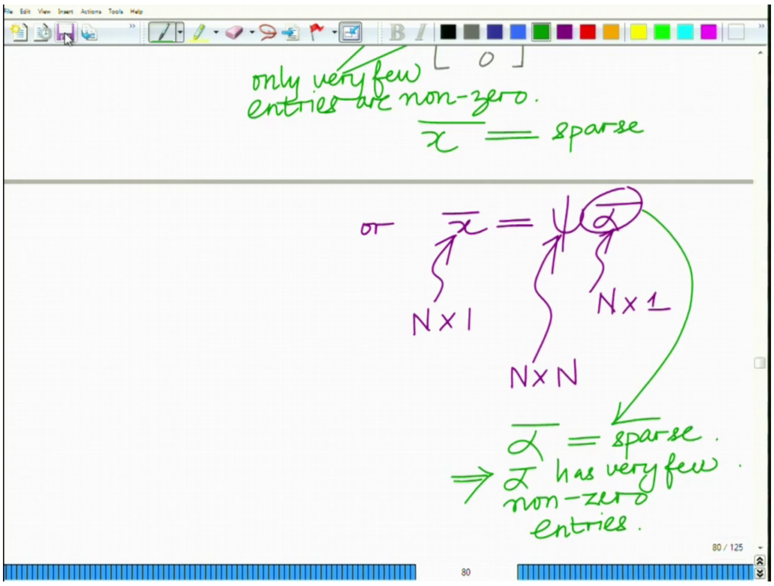
Now, x bar is sparse implies that very simply it implies that a large number of entries of x bar are 0 only very few entries are non 0. So, this implies that basically what this implies is that large number of entries. So, sparse implies large number of are 0 only very few entries which are marking by these x's.

(Refer Slide Time: 12:13)



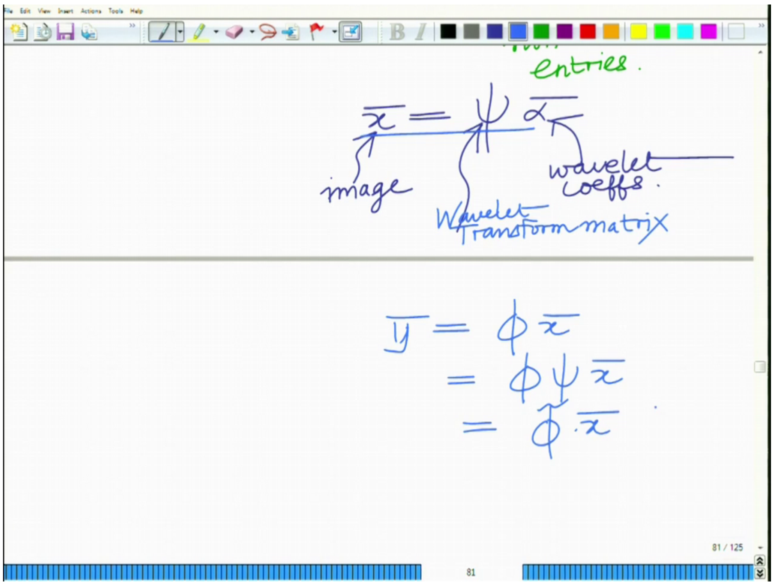
For instance only very few entries are non 0 only very few entries are non 0. So, a large number of entries of x bar are 0 very few entries. So, x bar is a sparse vector or so, x bar is sparse or alternatively and this is typically what happens or x bar equals psi some matrix psi times alpha bar.

(Refer Slide Time: 13:15)



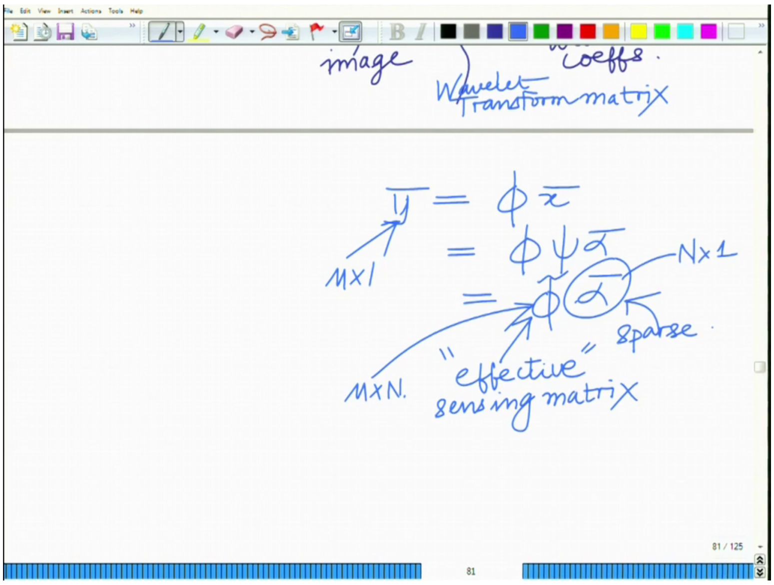
So, x bar is N cross 1. This alpha bar is N cross 1 and psi is an N cross N basis such that alpha bar is a sparse vector that is alpha bar equal sparse implies alpha bar has very few, alpha bar has very. So, either x bar is SPARSE. So, vector that you are trying to recover or more importantly and this is what happens more frequently like an image x bar can be expressed as a linear transformation of a sparse vector. For instance we take an image. Image if you look at it in the special domain it is not possible when you take it for instance the wavelet transform it is an excellent example. If you look at the wavelet coefficients of an image then they are sparse very few coefficients are non 0.

(Refer Slide Time: 14:43)



So, therefore, I can have an image x bar which is psi times alpha bar. So, let us say this is image these are the wavelet coefficient. And this is your wavelet transform matrix linear transform. So, x bar is sparse or x bar can be expressed in terms of alpha bar which is sparse and therefore, now if you substitute this the sensing model becomes y bar equals phi x bar equals phi psi times x bar equals phi tilde times x bar.

(Refer Slide Time: 15:43)



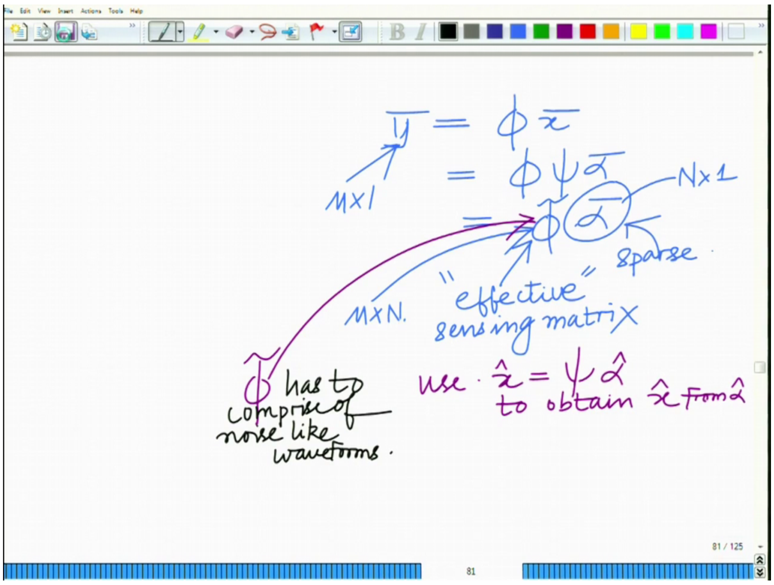
I am sorry phi psi times alpha bar. And if you do note phi psi as y phi tilde the effective sensing matrix this phi tilde now becomes your effective sensing matrix. And now alpha bar is SPARSE. And you still have the same relation this is M cross 1 this is M cross N and this is N cross 1

Now, in this model it is as if you are trying to you are trying to sense the vector alpha bar which is in the wavelet domain. Now once you get the wavelet coefficients naturally you can reconstruct the image right because image and wavelet they have a 1 to 1 correspondence. If you have an image you can represent in terms of wavelet transform right. If you have the wavelet transpose the wavelet coefficients they can reconstruct the image all right.

But the wavelet coefficients satisfy a very important property which is that this coefficients the wavelet vector of wavelet coefficients is sparse. And that is very amenable to compressive sensing and this is what is preferred frequently happens and the model can be extended right.

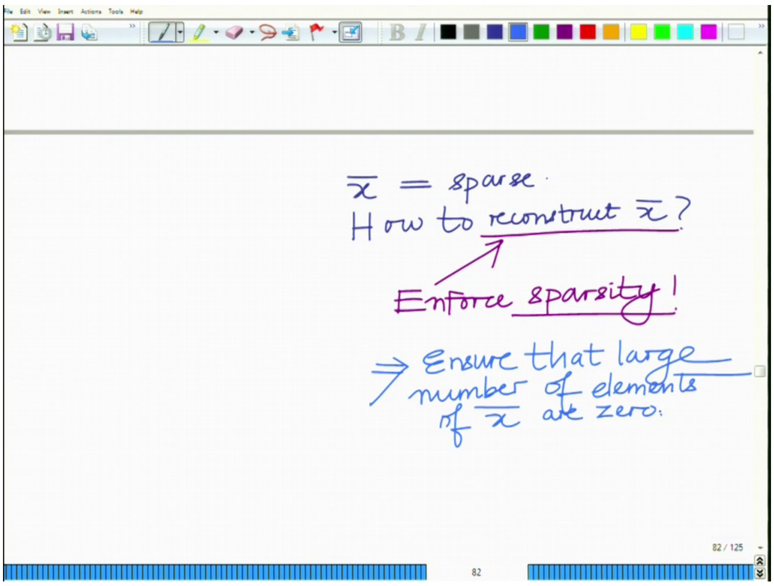
So, your y is equal to phi times x bar you substitute for x bar psi times alpha bar. So, it becomes phi psi into alpha bar and now you reconstruct alpha bar; from alpha bar which is sparse vector sparse signal recovery or compressive sensing and then you get the image from the wavelet coefficients all right or x bar from alpha bar using x bar equal to psi alpha bar.

(Refer Slide Time: 17:13)



So, once you get alpha bar use x hat equals psi alpha hat to obtain estimate x hat from alpha hat so, that is what you do. And now if you look at this phi tilde matrix this has to contain noise like waveforms. So, phi tilde which is your effective sensing matrix has to comprise of this has to comprise of noise like waveforms or contains noise like waveforms. Now how to reconstruct this alpha bar or x bar if x bar is sparse how to reconstruct. Let us assume that x bar is sparse and phi contains noise like waveforms.

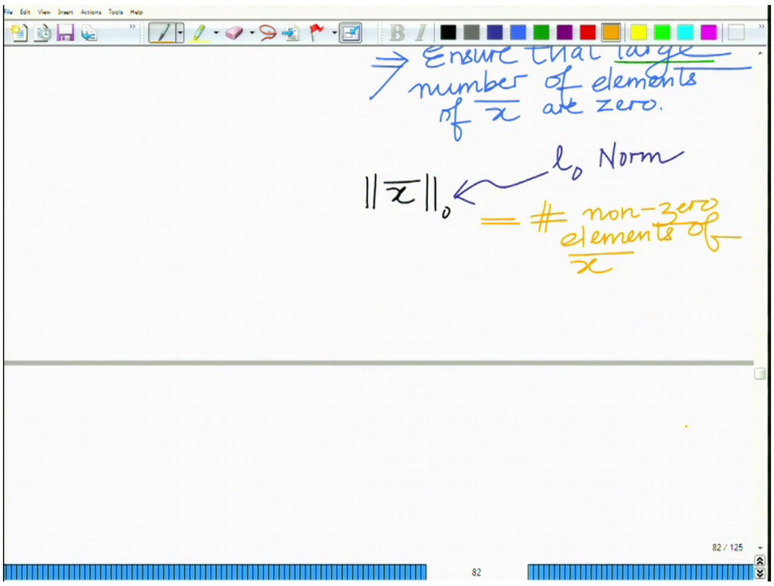
(Refer Slide Time: 18:13)



Now, the question that we want to ask is how to reconstruct x bar. To reconstruct x bar what we have to do is we have to enforce the sparse it remember x bar is a sparse vector. So, we have 2 enforce sparsity so, how to reconstruct x bar the answer is enforce sparsity. What is a mean to say enforce sparsity implies, we have to somehow impose a condition that x bar contains a large number of remember we said x bar is a sparse vector.

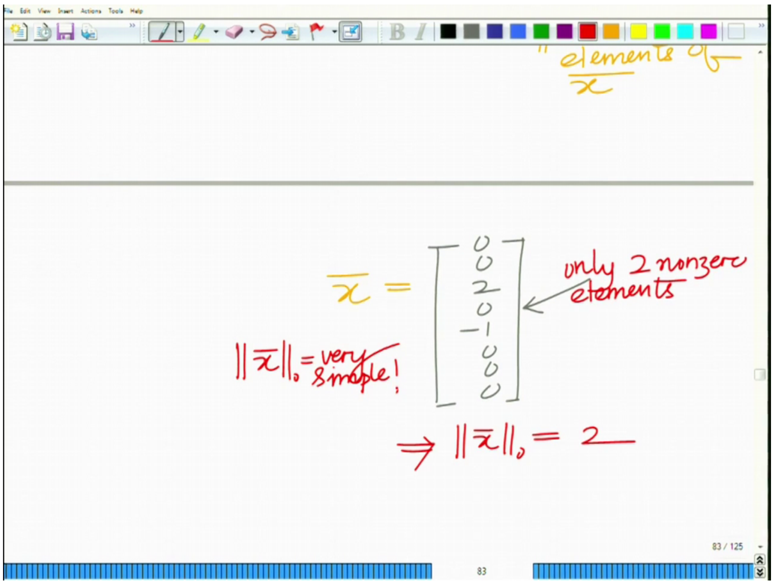
So, we have to enforce this sparse this sparsity which implies that we have to ensure that the reconstructed vector x bar is such that a large number of elements are 0's only some elements are non 0. So, which implies we have to ensure that large number of elements of x bar are 0; we have to ensure that large number of elements of x bar are 0 this is what. In fact, ensure that large number of elements of x bar; large number of elements of x bar are 0 all right. So, that x bar is sparse and this is precisely what is enforced by what we call the l 0 norm.

(Refer Slide Time: 20:05)



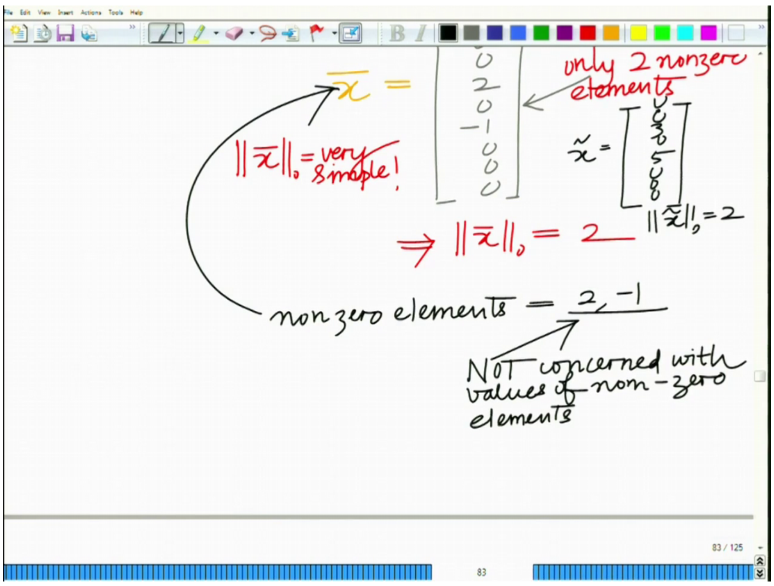
That is if you denote the l 0 norm of a vector this is l 0 norm like similar to l 1 l infinity and l 2 norm; this is the l 0 norm this l 0 norm and l 0 norm this equals the number of non 0 element of x bar. So, this is precisely what you want to minimize, the number of non 0 elements of x bar which is l 0 norm.

(Refer Slide Time: 20:51)



For instance let us say you have the vector x bar equals 0 0 2 0 minus 1 0 0 0. Now, l 0 now look at this is a 8 elements, but only 2 non 0 elements. In which implies if you look at l 0 norm of x bar that is simply 2 alright. So, very simple all you have to do is count the number of non 0 elements of x bar all right. Whatever is the number of non 0 elements that is the l 0 nothing else to be done. So, it is very simple their concept of l 0 norm itself is very simple. So, this l 0 norm concept; this l 0 norm concept is very simple, but there is a problem so in fact, the non 0 elements.

(Refer Slide Time: 22:01)

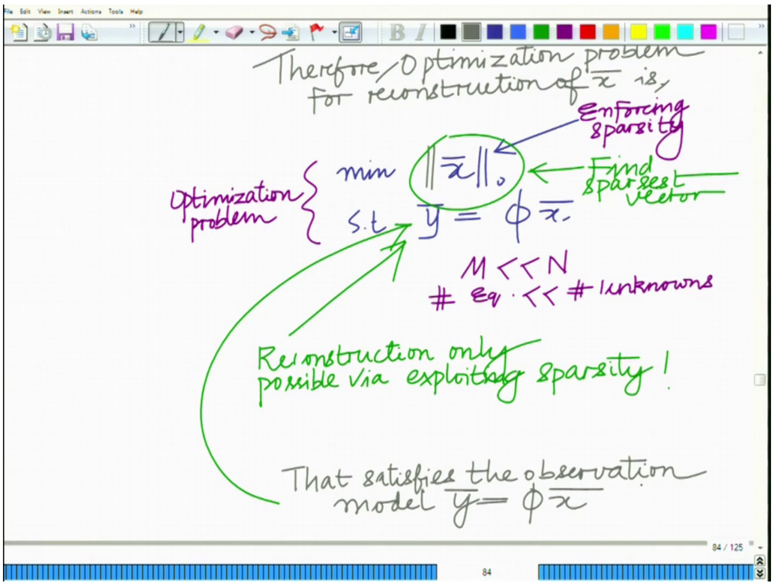


We are not concerned with the values of the non 0 element we just have the number of non 0 elements. We are not concerned with observed here unlike the other norms not at all concerned with values of something that is strange.

For instance this could have been 2 comma 1 and in fact, if you have another x tilde which is 0 0 3 0 5 0 0 0 non 0 elements are 3 comma 5 and if you look at l 0 norm of x tilde that is also 2 alright. It is the interesting part is we are not at all concerned with what are the values of these non 0 elements. We are only concerned with the number of non 0 elements that is the most interesting about the l 0 norm and that is what. In fact, makes it very complex this problem and therefore.

So, the optimization problem for now reconstruction and therefore, at this is very interesting.

(Refer Slide Time: 23:17)



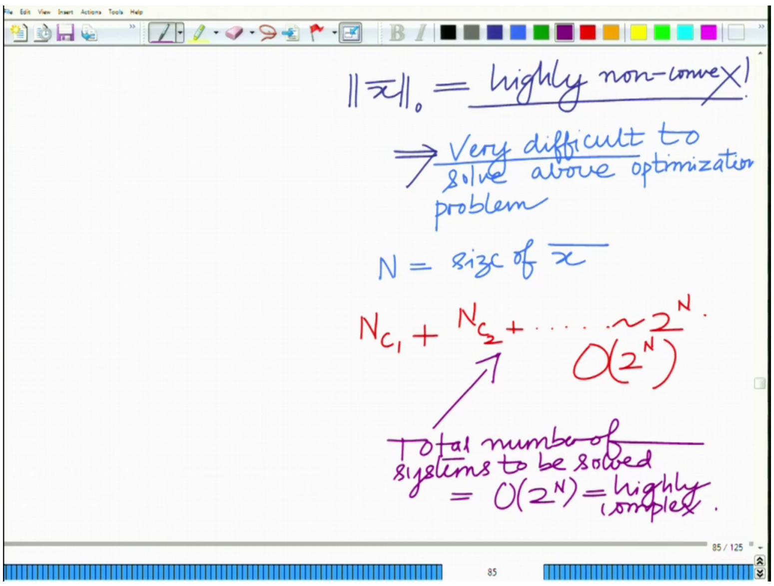
Therefore optimization problem for reconstruction of x bar is the same thing that we talked about. We have the sensing model we want to minimize the l 0 norm. And now subject to the constraint that y bar equals phi times x bar, what is y? Y bar is observation vector that is what something that we have seen already seen y bar phi sensing matrix x bar is a vector that we are trying to estimate x bar.

So, we minimize. So, this is basically enforces this is what we mean by enforcing sparsity. This is enforcing sparsity because we are minimizing the l 0 norm that is minimizing the number of non 0 elements. So, this is the relevant optimization problem. And why do we need this we need this because M number of equations, M is less than or equal to N. Which we implies the number of equations is less than or equal to number of unknowns and therefore, one has to exploit sparsity.

So, reconstruction is only possible exploiting because we said if you treat this reconstruction only possible via reconstruction, only possible via exploiting sparsity. So, what does this mean? Your enforcing sparsity which means you are trying to find here finding the sparsest vector; find the Sparsest vector which satisfies that satisfies the observation model. So, that is what the optimization problem is. That satisfies observation model y bar equal to y bar equal to phi x bar. So, if you trying to find this Sparsest vector.

Now what is, now we have written an optimization problem like similar to several optimization problems that we have seen before. We have an optimization problem you might think that you can solve it similar to fashion that we have done before an optimization objective plus Lagrange multiplier. The problem with this is this optimization problem not only is the objective non differentiable this optimization problem is highly non convex.

(Refer Slide Time: 27:07)

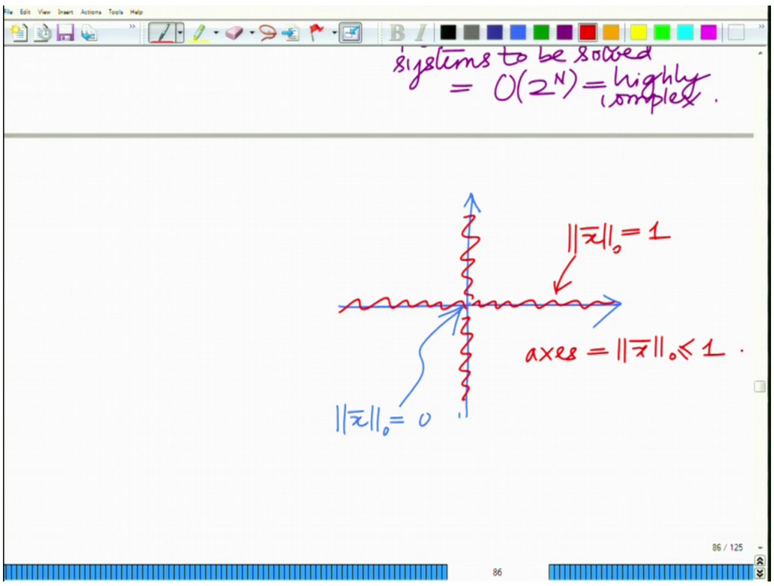


If you look at the l 0 norm; norm x bar is 0 l 0 norm this is highly convex. This is highly non convex implies very difficult to solve above optimization problem is very difficult to solve over. Consider a simple technique what you can do is for instance let us say you have M elements or M N equals size of x bar, what you can do is set only one non 0 element of x bar solve it already you will have one; you will have only one unknown, but the non 0 unknown can be in any of the N positions.

So, we will have to solve N systems and then similarly you can have 2 non 0 elements the 2 non 0 elements can you can be in any of the 2 locations. So, you have N c 2 such combinations. So, you have to solve N c 2 system. So, so on and so, forth if you look at this total number of total number of systems that have to be solved is N choose 1 plus N choose 2 plus so on and this will be 2 to the power of N or 2 the approximately you do not need to get it the exact number what you can see is this rises exponentially.

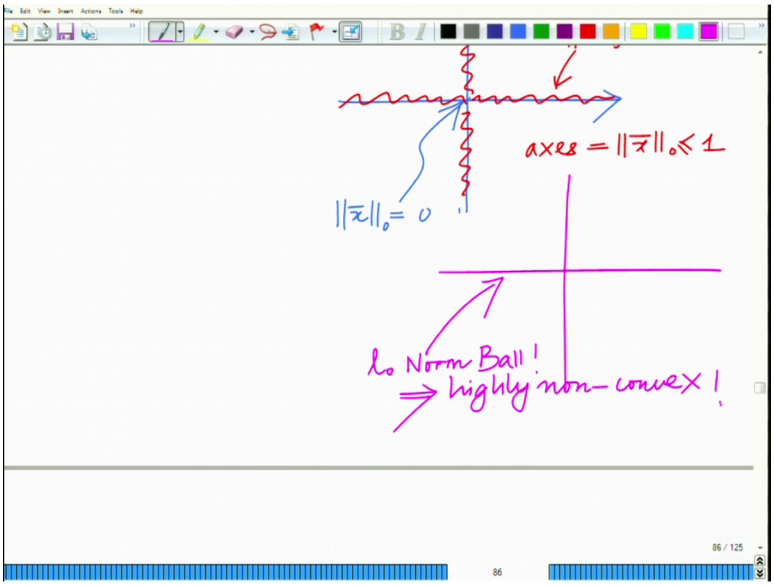
So, this is of the order 2 to the power or you can write order 2 to the power of N so, total number of systems, which is highly complex for instance. If we have a signal vector x bar of dimension 100 then you have to solve 2 to the power of something of the order of 2 to the power 100 which is impossible alright that cannot be solved by any known computer either now or in a very far right. So, the point is that this problem although it is very simple to state the optimization problem it is an extremely complicated optimization formula.

(Refer Slide Time: 30:13)



Now, if you look at for instance the l 0 norm, what is the l 0 norm? If you look at for instance l 0 norm equal to l 0 norm if you look at a 2 dimensional plane. This one point has l 0 norm equal to 0 because that is the origin. Now, this axis other than origin. If you look at this axis that has l 0 norm because on the axis you have only ones non 0 element. So, this is rest all of the plane other than axis and origin this. So, if you look at l 0 norm less than equal to 1 that comprises only of the axis. So, if you look at this object it is highly non convex. So, if you look at the l 0 norm ball.

(Refer Slide Time: 31:17)



The l 0 norm ball this is the l this is literally the l 0 norm ball. Which is basically highly this is highly non convex and that is the problem in solving this optimization problem all right. So, we would like to make very few measurements agreed projections or noise like waveforms agreed, the vector x bar is sparse or it is sparse in some appropriate domain x bar equals psi times alpha bar, but the problem is that the eventually the problem that has to be solved in force sparsity.

Which is basically minimizing the l 0 norm that has an exponential complexity it is an N p hard problem. And which is therefore, practically impossible to solve for large signals x bar that is signals of the size let us say even 20 samples 20 samples or 25 becomes very difficult exceedingly difficult and if with 100 samples it is impossible to solve and therefore, 1 has to come up with other engineers techniques to solve this optimization problem and that is basically at the heart of compressive sensing let us say. What are the other techniques that can be used to solve this optimization problem which we are going to look at in subsequent modules that forms the basis for compressive sensing and that is what in fact; these techniques have revolutionized the field of compressive sensing all right. So, let us stop here and continue in the subsequent modules.

Thank you very much.