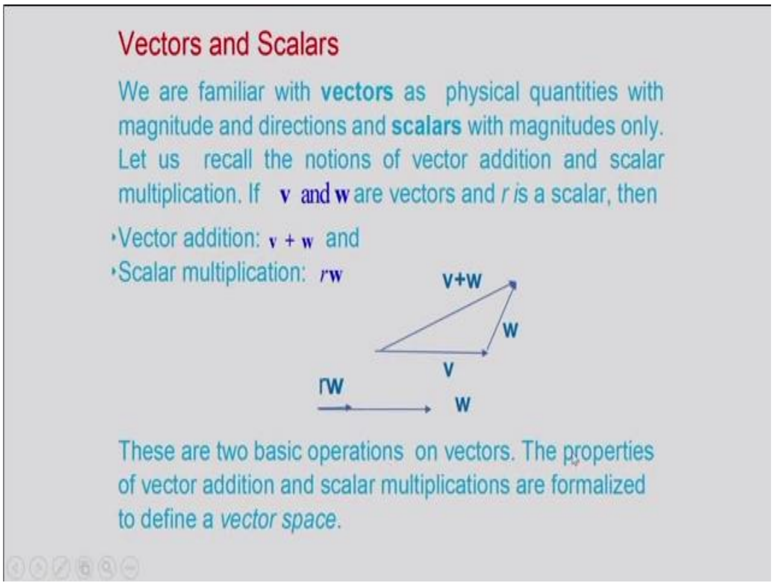
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture No 1
Linear Algebra of Random Variables

Hello students, in the last lecture we introduced the basic concepts of probability and random variables. We noted that, random variables are functions from the sample space to the set of real numbers. Like other functions random variables also can be interpreted as members of a vector space and many results of linear algebra can be extended to random variable theories. Such interpretations are exploited in statistical signal processing.

In this lecture we will introduce the basic concepts of linear algebra and their extension tool SSP. We will start with the concept of vectors and scalar.

(Refer Slide Time: 01:24)

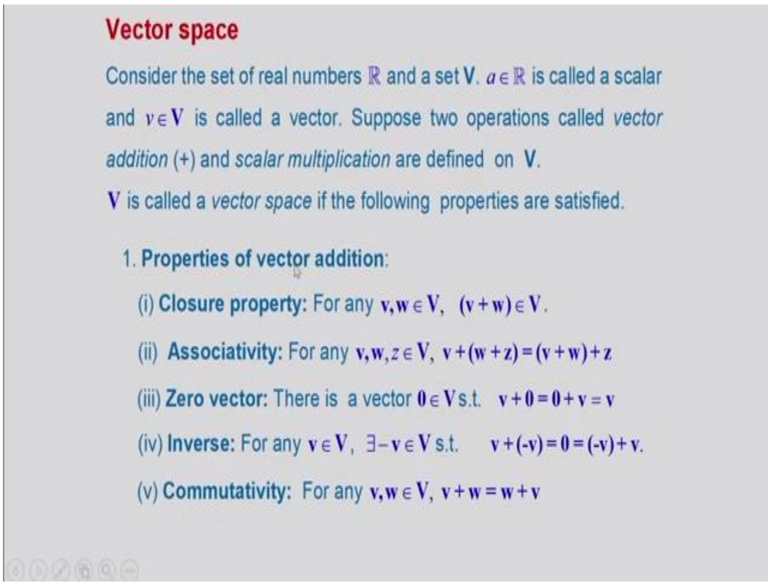


We are familiar with vectors as physical quantities with magnitude and directions and scalars with magnitudes only. We are familiar with those concepts in physics. Let us recall the notions of vector addition and scalar multiplication. If v and w are vectors and r is a scalar, then vector addition is v + w and scalar multiplication is rw, now in a two dimensional plane.

If we write this is v this one is w then v+w is given by this, this is the vector addition. Similarly if w is a vector then we can scale it by a vector r. So that rw this is the new vector rw is a scaled version of w and r can be bigger than 1 then this vector is extended this side.

These are the vector addition and scalar multiplications are two basic operations on vectors. The properties of vector addition and scalar multiplications are formalized to define a vector space. So we will go to vector space now.

(Refer Slide Time: 02:40)



Vector space consider the set of real numbers R and they set V, a belong to R is called a scalar any real number we will call it as a scalar. It may be a complex number but we are considering the scalar of real numbers only, v belonging to V is called a vector. The elements of this space this set V in called vectors. Suppose two operations called vector addition, that is denoted by + and scalar multiplication are defined on V. V is called a vector space.

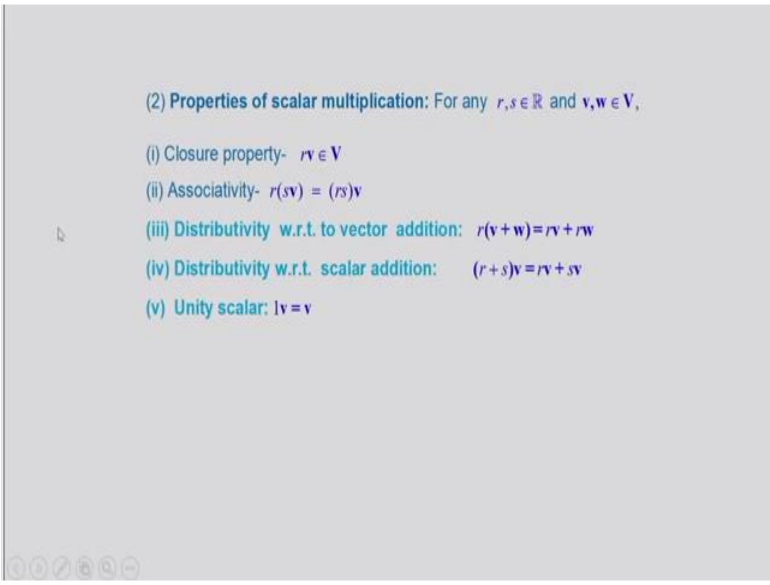
If the following properties are satisfied, there are a number of properties. First properties are properties of vector addition. So first one is known as the closure property. So for any vector v,w belonging to V, v+w vector addition of v and w is also a member of V. So that way vector addition is closed or this set of vector is closed under vector addition. Now vector addition is associative, what does it mean for any vectors v and w, v w and z, v + sum of w + z is equal to sum of v + w + z.

So if we do this sum first and then, it will be same as you first take the sum of v+w and then v+z. so that way this property is known as the associative property of vector addition. Similarly there is a vector 0; it is a member of the vector space. Such that v + 0 is equal to 0 + v. v+ 0 is equal to 0+ v is equal to v. So that way if we add 0 to the vector any vector it will remain the same vector.

And v has also an additive inverse. So for any v belonging to V, they are exist a– v. That is the negative of the vectors. Such that v+ -v is equal to 0 and finally vector addition is commutative. What does it mean? v+w is equal to w+v. So if these properties are satisfied, these five properties are satisfied. Then V is V+ vector space V and that operation plus R known as a commutative group. These properties are known as the group properties.

So that way vector addition forms a group on the vector space. Okay, so that way we see that to define vector space we need vector addition and vector addition satisfy these five properties.

(Refer Slide Time: 05:47)

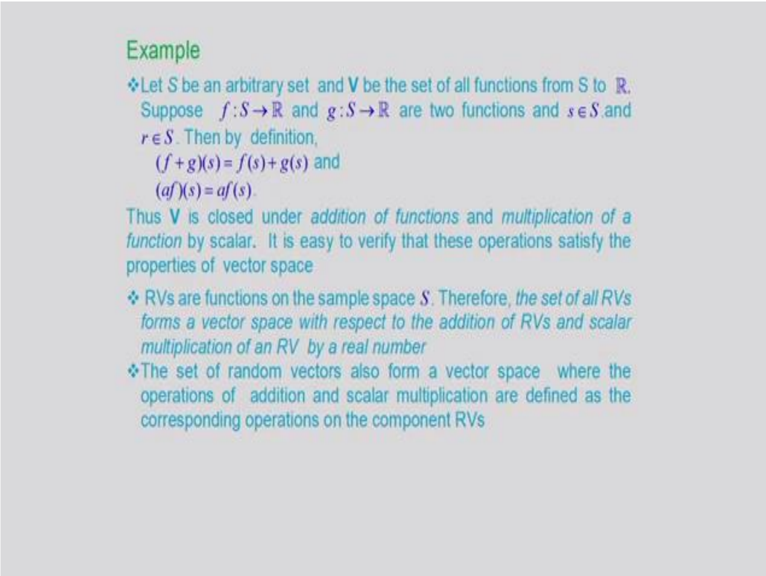


Similarly properties of scalar multiplication, we have the following properties. What are those properties? For any r,s belonging to R. r and s are scalars and v,w belonging to V, v and w are vector. Suppose closer property that is scalar multiplication of V, which r,v belongs to V. if we multiply a vector by a scalar we will get another vector associativity. So r into s v is equal to r s into v. So that way first we can scale v then we can scale s v. This is same as scaling v by vector r s.

Distributivity with respect to vector addition; so here two operations are there, vector addition and scalar multiplication. So we have to consider distributivity with respect to vector addition. That is r of v+w is equal to r v + r w. similarly with respect to scalar addition also scalar multiplication is distributive. So that we can write as r+ s into v, that is same as rv + sv. So distributivity with respect to scalar addition.

So if we have two scalar r+ s into v, that will be same as rv + sv. So this is the distribution with respect to scalar addition and also we have the unity scalar 1 into v is equal to v. so therefore in a vector space two operations are there one is scalar multiplication, one is vector addition and vector addition satisfied five properties scalar multiplication also satisfies these five properties. So we have defined a vector space.

(Refer Slide Time: 07:51)



Now we can give some example, simple examples are there for example: if we consider suppose normal vectors in a three dimensional space and then the collection of those vectors will form a vector space with respect to vector addition and scalar multiplication. I will give another example: suppose, this example is important for us, let S be an arbitrary set and V be the set of all functions from S to R.

Suppose f from S to R and g from S to R are two functions and s and r are two elements of S. Then by definition f+g at the point s is equal to fs + gs and also af of s, af at point s is equal to a times f of s. So that way since we are considering all function this is also a function, this is also a function. Thus V is closed under addition of functions and multiplication of function by a scalar. You are multiplying a function by a scalar you are getting another function.

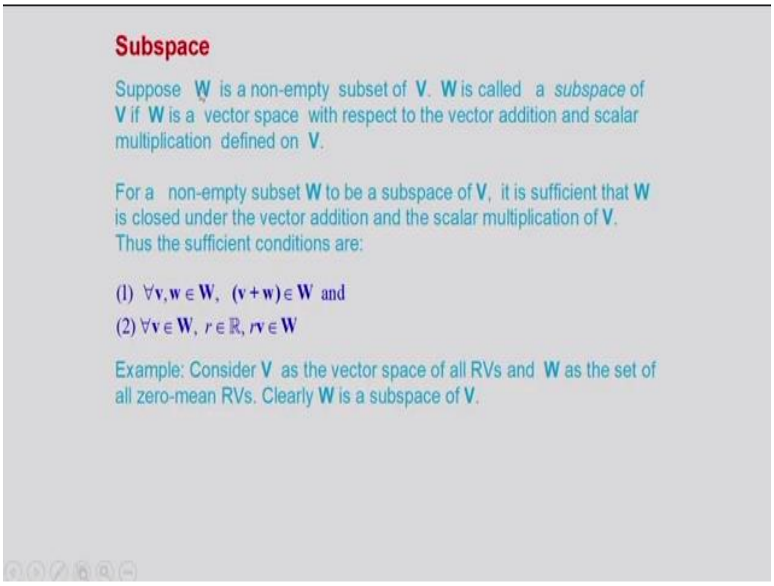
And similarly you are adding two function, you are getting another function. That way, these spaces of all functions are closed under addition function and scalar multiplication of a function by scalar. Thus these closed under addition of functions and multiplication of a function by a scalar. It is easy to verify that these operations satisfy the properties of the vector space. So that addition of function is associative.

Similarly we can define a 0 function and other properties of scalar multiplication also we can establish. So therefore what we say that, the set of all functions from any set S to R is a vector space with respect to addition of function and scalar multiplication of function. No random variables are functions on the sample space S. Therefore the set of all random variables form a vector space with respect to addition of random variables.

And scalar multiplications of a random variable are real number. So that way this is one important result, what we say that random variables form a vector space with respect to the addition of random variables and scalar multiplication of a random variable by a real number. The set of random vectors also form a vector space.

Where the operations of addition and scalar multiplications are defined as the corresponding operations on the component random variables. So that way we can define a vector space of random vectors also. Let us define a subspace,

(Refer Slide Time: 11:04)

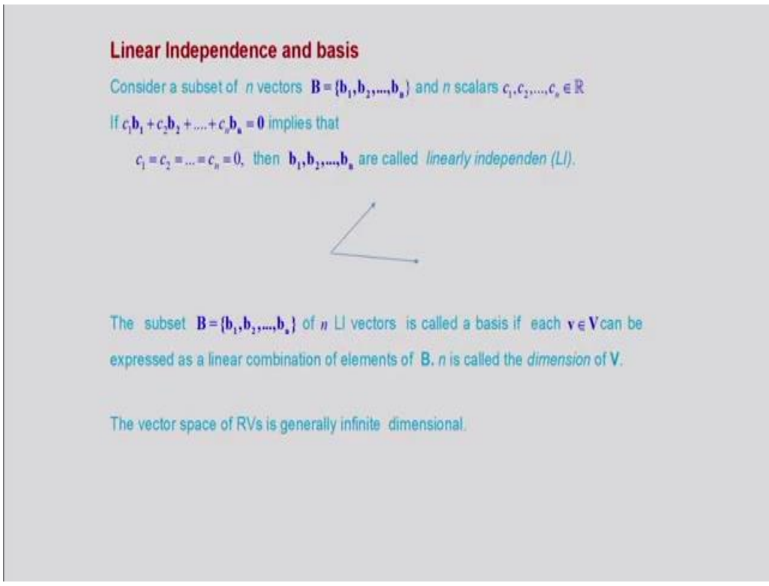


Suppose W is a non-empty subset of V. W is called a subspace of V. if W is a vector space, it itself is a vector space with respect to these same operations of vector addition and scalar multiplication defined on V. so W is a subset and it itself is a vector space with respect to the operations of the original vector space. For a non-empty subset W to be a subspace of V, it is sufficient that W is closed under vector addition and scalar multiplication of V.

Thus the sufficient conditions are for all v w belonging to be v+w soon also belong to W. For all V belong to W and r belonging to R, r v also belong to W. So these are the properties needed for a subset W to be a subspace. We can show that 0 is an element of this subspace and other properties of vector space are satisfied by subset a subspace also. Example we will consider one example:

Suppose V as the vector space of all random variables and W as the set of all 0 mean random variables. If we aim to determine random variable you will get a 0 mean random variable, if we multiply a 0 mean random variable by a scalar you will get a 0 mean random variable. Therefore W is a subspace of V. so the case of the vector space of random variables, we can define another subspace that is this subspace of the 0 mean random variables.

(Refer Slide Time: 13:02)



Now for the vector space linear independence and basis are two important concepts. Consider a subset of n vectors that is B is equal to b1, b2 up to bn and n scalars c1, c2 up to cn. Now if c1 b1+ c2 b2+ up to cn bn is equal to 0 vector implies that, c1 equal to 0. If this is equal to 0, then c1 must be equal to 0. c2 must be equal to 0 and so on. Then b1, b2, bn are called linearly independents. This we abbreviate as a LI.

So if this left hand side is equal to 0, then all the coefficients must be identically 0. Then we will say that these factors are b1, b2… bn are linearly independent. Linear independence says that suppose any vector if we consider in this set b1, b2 upto bn, you cannot represent one vector in terms of the list one vector as a linear combination of the list. For example; let us consider, this two-dimensional vector.

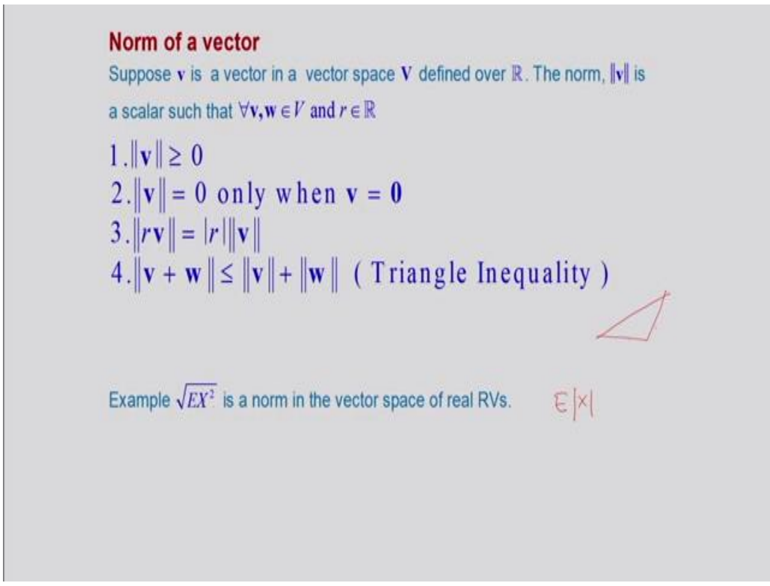
Suppose this is my b1 and this is my b2 normally when we write by hand we give a bar to denote a vector. Now this b1 cannot be written in terms of b2 or we cannot write b1 as a scaled version of b2. Therefore b1 and b2 are independent. However if we have another vector suppose b3, now this b3 can be represented in terms of as a linear combination of b1 and b2. Therefore b1, b2 and b3 in this case will not be linear independent.

So we introduce the concept of linear independence, linear independence basically means that in those collection of vectors one vector cannot be expressed as a linear combination of the list. it is now we will tell what is their basis this subset B is equal to b1, b2 upto bn of n LI vectors is called a basis. If v belonging to V can be express as a linear combination of elements of B. so there are n elements.

And we can express any vector in V as a linear combination of these n vectors and in that case n is called the dimension of V. therefore what is a basis then basis are those vectors which are linearly independent and their linear combination can generate any vector V belonging to the vector space. Now we told that, random variables also form a vector space but unfortunately the dimension of this vector space is infinite and finding a basis for this vector space is a complicated task.

Norm of a vector, we know that suppose in a two dimensional vector. I have the length of the vector and the concept of length can be generalized for a vector space.

(Refer Slide Time: 16:47)



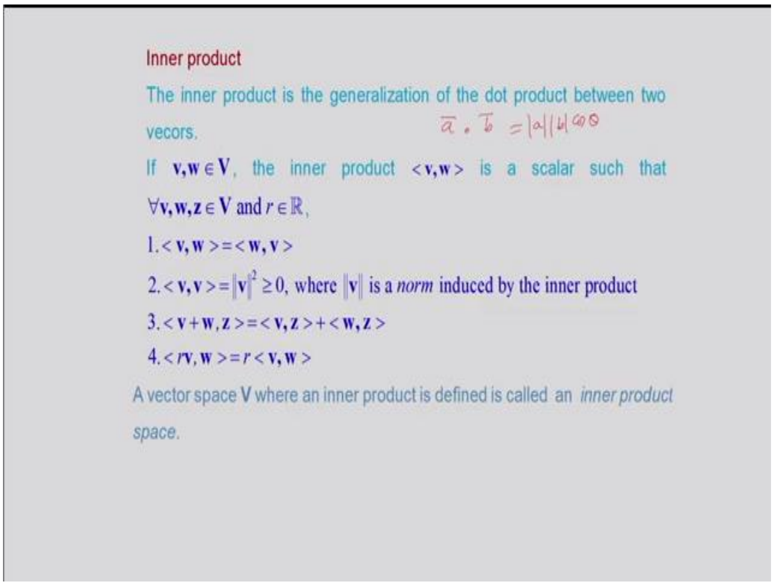
Suppose v is a vector in a vector space V defined over R. the norm denoted by this notation, so norm of v is a scalar. Such that for all v,w belonging to be r belonging to R. these four properties are satisfied. So number one is norm is non-negative. So norm of v is always greater than or equal to zero. Then when norm will be equal to zero normally it will be equal to zero. Only when v is equal to zero that vector itself is zero vectors.

Then the normal way the scalar multiplication of a vector is equal to norm of the scalar that is magnitude of the scalar because scalar is a real number, so it is a magnitude of the scalar multiplied by the norm of the vector. Now fourth property is important that is there we know this is the triangular inequality suppose in the case of length. We know that, suppose if you in any triangle if you consider this third side will be less than equal to sum of these two sides.

so this result is that is norm of v+w is less than or equal to norm of v+ norm of w, this is the triangular inequality. So any major that satisfied these four properties qualifies to be a norm. For example; in the case of vector space of random variables root over E of X square is a norm. Similarly we can have, suppose your mode of X that is also a norm. So here you see that since it is a random variable that to define norm.

We are taking the help of expectation operation because this expectation operation will convert it into a scalar.

(Refer Slide Time: 18:49)



So we have defined norm now, we will define what is an inner product between two vectors? The inner product is the generalization of the dot product between two vectors. Suppose I know that, if a is vector suppose b is a vector a dot b is equal to many dot of a into many dot of b into cos of angle between them cos of theta. So now this dot product vector dot product satisfies certain properties and these properties are as generalized to define the inner product of two vectors.

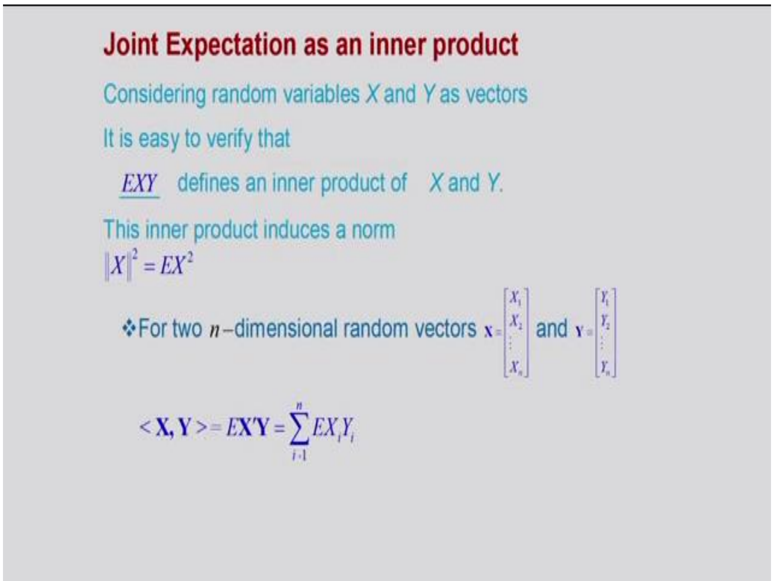
If v,w belong to capital V, the inner product this is denoted by that is the angular bracket v,w is a scalar. So inner product is a scalar, such that for all v,w,z belong to V. So any three members of V and r belonging to R. what will happen that inner product of v, w is equal to inner product of w,v. so inner product is commutative. This property is known as symmetry property. Inner product of v,w is equal to inner product of w,v.

Now what happen if we take the inner product with itself, so inner product of v,v is equal to norm of v square, which is greater than equal to 0. Now where now this norm of v is a norm induces by the inner product from inner product itself we can induce a norm. What is that norm? That is the inner product of v with itself. So that way inner product will define a norm. Third property is inner product of v+w,z is equal to inner product of vz plus inner product of wz.

This is the third property and fourth property is, if we scale one vector suppose r, inner product of r,v,w is same as r times inner product of v,w. so these are the properties also satisfied by dot products. It is very easy to verify and here we'll be using these properties as the definition of inner product. A vector space V where an inner product is defined is called an inner product space. Now from vector space we have come to inner product space.

In fact when norm is defined that vector space is known as the normed linear space or normed vector space. Now we have defined normed inner product space.

(Refer Slide Time: 21:41)

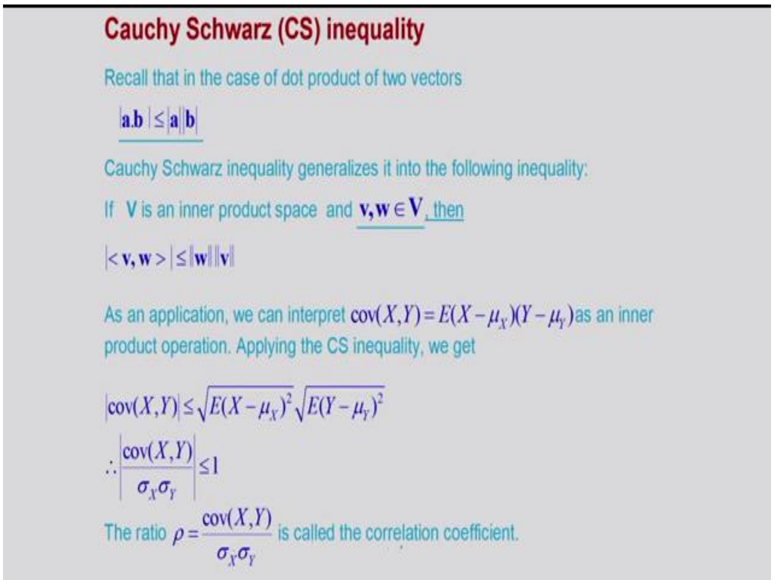


The joint expectation as an inner product operation considering random variables X and Y as vectors. It is easy to verify that, E of XY. Because we know that, E of XY is a real number. It is a scalar defines an inner product of X and Y. For example E of XY is equal to E of Y X. Now E of X square will be always greater than equal to 0, like that we allow all other properties we can solve to be satisfied therefore this joint expectation of XY is an inner product operation.

So that is a very important conclusion now E of XY defines an inner product of random variable X and Y. when they are represented as vectors and this inner product induces a norm. What is that norm? Norm of X square that is equal to E of X square. So we can define inner product of two random variables. Similarly for two random vectors, also a random vector X is equal to suppose this and Y is equal to this.

Then inner product, we can define as E of X transpose Y, X transpose Y will convert these two vectors into scalar and then this scalar when we take the expectation we get the inner product. Okay, so X transpose Y that is a scalar and when we take the expectation we get the inner product. So we have defined inner product.

(Refer Slide Time: 23:21)



In the case of inner product there is a very important inequality what is known as Cauchy Schwarz inequality. recall that in the case of dot product of two vectors and that is magnitude of a dot b is less than equal to magnitude of a into magnitude of b. this is because cos theta is magnitude of cos theta is less than equal to one. So from the definition of dot product this result follows.

CS inequality, Cauchy Schwarz inequality generalizes this inequality into the following inequality. How it is generalize, if V is an inner product space vector V, this be a inner product space and small v,w belong to V. If V is an inner product space and small v,w are members of V. then magnitude of the inner product of v and W is less than equal to norm of w into norm of v.

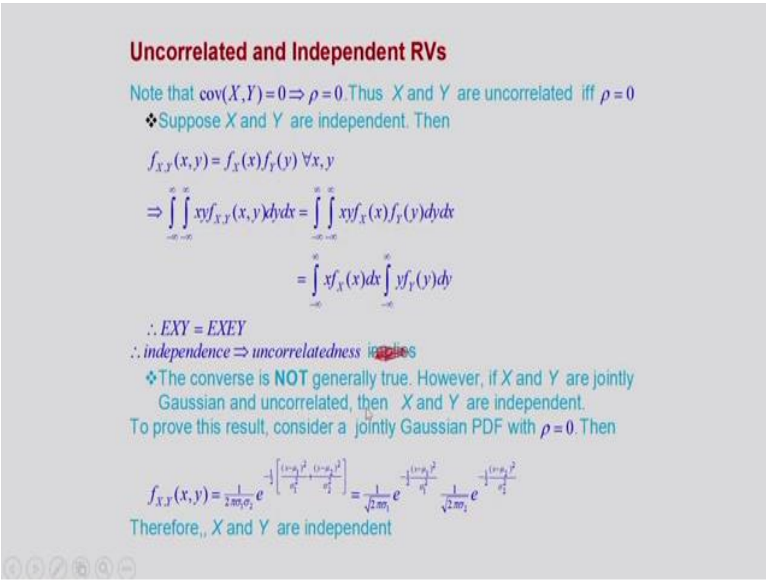
So this is a very important result Cauchy Schwarz inequality will not prove this but we will assume that it is satisfied. This CS inequality will be used to derive an important result of joint random variables. So we can interpret covariance of X, Y. suppose covariance of X, Y is equal to by definition E of X - mu x into Y – mu y. So this is an inner product on these two random variable, X - mu x into Y - mu Y.

Now since, it is an inner product operation we can apply Cauchy Schwarz inequality. What we will get? That is, if we take the magnitude of this that is same as magnitude of this quantity. And that must be less than equal to now inner product is less than equal to norm. Here what is the norm? Norm is X minus mu x, expectation of X - mu x whole square and square root of that. Similarly norm of this quantity is expected sum of Y - mu y whole square and square root of that. So what we get therefore applying the Cauchy Schwarz inequality, the covariance of X Y is less than equal to square root of E of X - mu x Whole Square into square root of E of Y- mu y whole square.

And these quantities are nothing but variance, so and if we take this square root then it will get Sigma x, if we take the square root of this we will get Sigma y. so therefore covariance of XY divided by Sigma x into Sigma y its magnitude. Because, I am taking to the denominator of this must be less than equal to 1 and this quantity is the correlation vector. So covariance of XY divided by Sigma x into Sigma y, it is known as the correlation coefficient.

And we see that magnitude of rho is always less than equal to 1. So this is one important result.

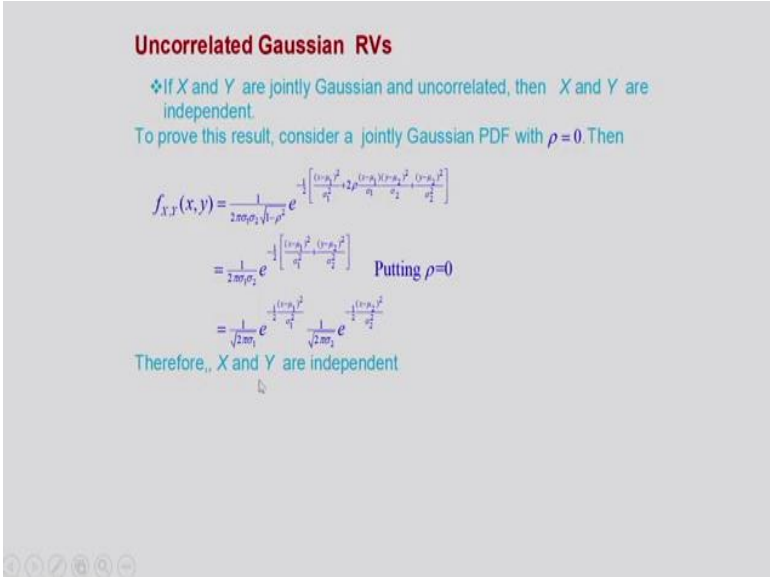
(Refer Slide Time: 26:48)



Now let us see, relation between uncorrelated and independent random variables. Note that, covariance of XY equal to 0 implies that rho is equal to 0. Therefore X and Y are uncorrelated. If rho is equal to 0, this can be taken as a definition. Now suppose X and Y are independent, then by definition joint density is product of the marginal density for all x,y. Now you multiply both side by x,y and take the double integral which respect to dy dx.

And now using the separability, we get that this is equal to integration of x fx of x dx into integration of y fy of y dy. So and this is E of X and this is E of Y. So E of XY is equal to dx into dy. That is, if x and y are independent then they are uncorrelated. Independence implies uncorrelatedness. The converse is not generally true. So uncorrelatedness does not imply independence in general. However if X and Y are jointly Gaussian and uncorrelated then X and Y are independent.

(Refer Slide Time: 28:13)

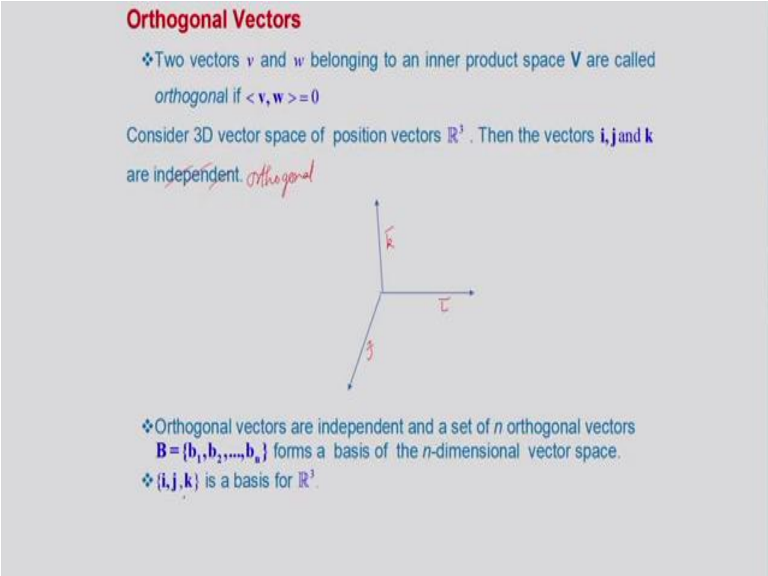


We can see this if X and Y are jointly Gaussian and uncorrelated, and then X and Y are independent. To prove this result, consider a jointly Gaussian PDF with rho is equal to 0. we know what is the jointly Gaussian this density function is given by 1 by 2 Pi Sigma1, Sigma2 into root over 1 – rho square into e to the power minus half of these quantities. Okay, now uncorrelated X and Y are uncorrelated Gaussian. Therefore Y is equal 0.

If I put rho is equal 0 here, I will get this expression and this rho will be equal 0. Therefore this will be 1 only so I will get this expression. Now I can separate out these two vectors. So that I will get this one PDF here and this is another PDF. This is the PDF of X and this is the PDF of Y. so in this case joint PDF is equal to the product of the marginal PDF. We should also recall that if x and y are jointly Gaussian then they are individually also Gaussian.

Therefore, this is the PDF of X and this is the PDF of Y and the joint density is product of marginal densities. Therefore x and y are independent. So the concept of inner product can be used to define orthogonal vectors.

(Refer Slide Time: 29:49)



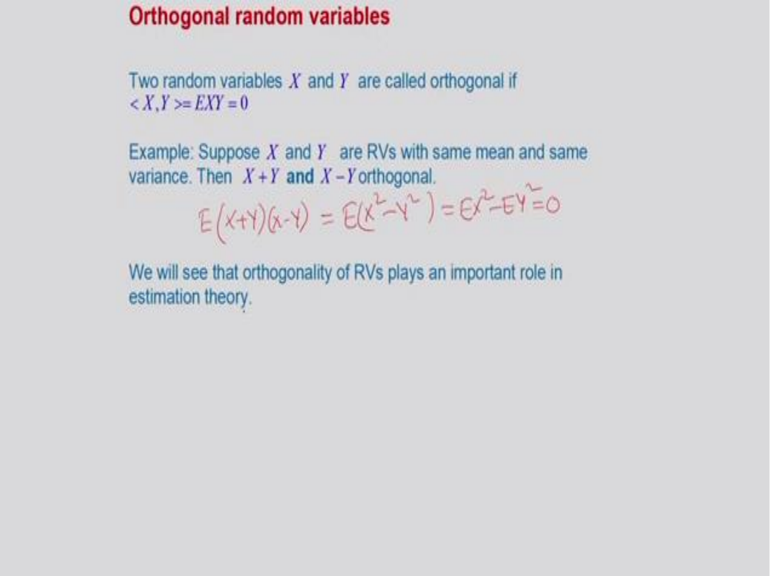
Two vectors v and w belonging to an inner product space V are called orthogonal. If the inner product of v and w is equal to 0. Just like in the case of three-dimensional space, suppose when the vectors are perpendicular then inner dot product is 0. The same concept is use here, two vectors will be orthogonal when they are inner product is equal to 0. Now consider 3D vector space of position vector R 3 that we generally do not by R 3.

Then the vectors i,j,k this is the suppose i vector this is the j vector and this is the k vector. Suppose and they are orthogonal, so these vectors are orthogonal. so the vectors i,j,k, are orthogonal because, here inner product is the dot product, dot product between i and k will be equal to 0. Dot product between i and j will be equal to 0. Like that now orthogonal vectors are independent.

That is one important result that orthogonal vectors are independent and i,j,k are orthogonal. Therefore i,j,k also will be independent and therefore now we know that this n independent orthogonal vectors because they are independent they will form a basis of the N dimensional vector space. So that way if we have orthogonal vectors and n orthogonal vectors then those n orthogonal vectors R basis of an n-dimensional vector space.

And therefore this i,j,k is a basis for R3. Because i,j,k are independent and this is three dimensional space therefore this is a basis. So now this concept of orthogonal vectors can be generalized to orthogonal random variables.

(Refer Slide Time: 31:51)

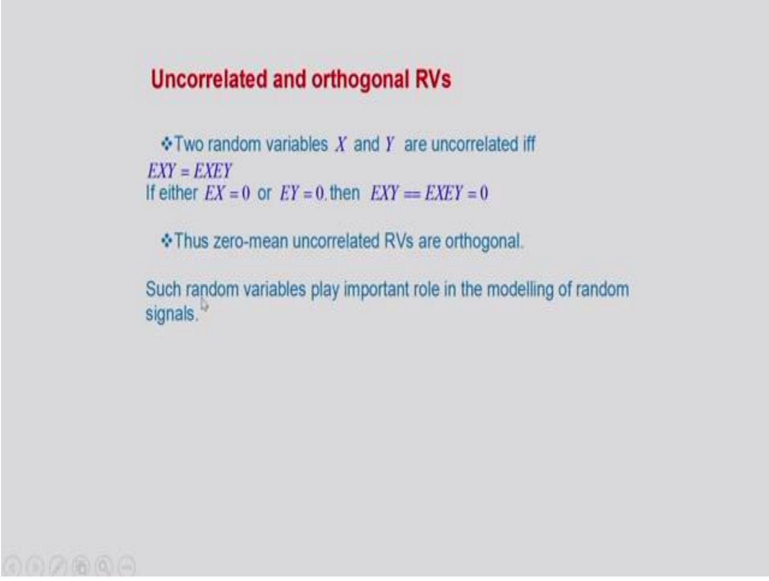


Now orthogonality of vectors can be extended to random variables. Two random variables X and Y are called orthogonal. If inner product of X Y that is equal to E of X Y equal to 0, If E of X Y is equal to 0, then X and Y are called orthogonal. Example, suppose X and Y are random variables with same mean and same variance. So both mean and variance of X and Y are equal. So in that case X +Y and X- Y are orthogonal.

How suppose X + Y, E of into X -Y that is equal to E of X square- Y square. That is equal to E of X square - E of Y square. Now since both are of the same mean and variance therefore E of X square and E of Y square are equal, that will be equal to 0. We will see that orthogonality of random variables plays an important role in estimation theory. Particularly we will see that it for optimal estimation ER is orthogonal to data.

This result we will be using. So we have defined orthogonal random variables and uncorrelated random variables. But how they are related?

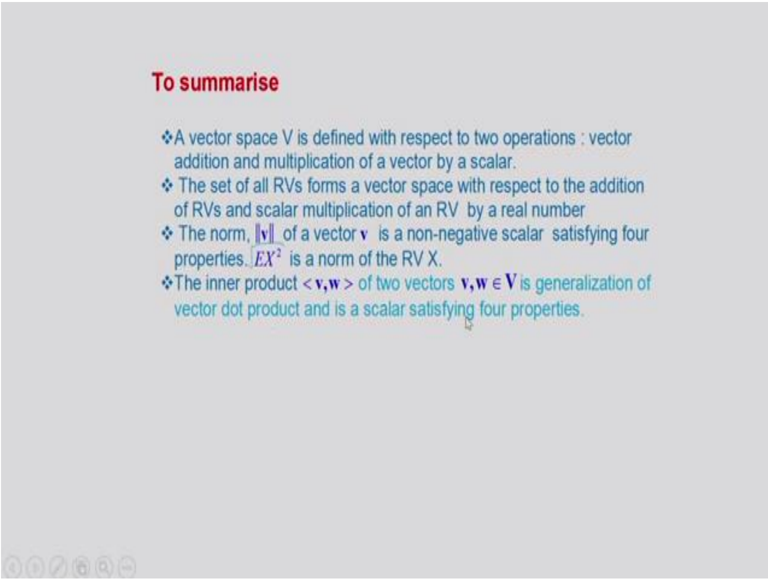
(Refer Slide Time: 33:23)



Two random variable X and Y are uncorrelated, if E of XY is equal to EX into EY. That is there we covariance of X Y is equal to 0 from that we get that E of X Y is equal to EX into EY. Now if either EX is equal to 0 or EY is equal to 0, then E of XY is equal to and now since one of the EX or EY is 0. Therefore this product EX into EY will be equal to 0. So therefore 0 means uncorrelated random variables are orthogonal.

Such random variables play important role in modelling of random signals. For modelling random signals we need a sequence of zero mean uncorrelated random variables, in fact they are orthogonal.

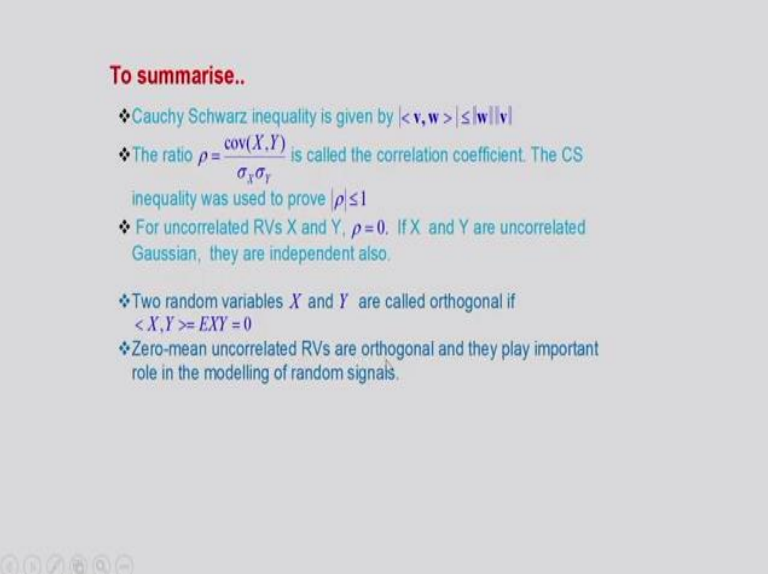
(Refer Slide Time: 34:18)



Let us summarize the lecture today, a vector space V is defined with respect to two operations: vector addition and multiplication of a vector by a scalar and those properties of vector addition and scalar multiplications must be satisfied. The set of all random variables forms a vector space with respect to the addition of random variables and scalar multiplication of a random variable by real number.

The norm v of a vector v is a non-negative scalar. so we define norm is a non-negative scalar satisfying four properties and square root of EX square is a norm of the random variable X. The inner product v,w of two vectors v and w is generalization of the vector dot product. Inner product is a generalization of vector dot product and it is scalar inner product is a scalar satisfying four properties. Those properties we stated,

(Refer Slide Time: 35:30)



Then we introduced the Cauchy Schwarz inequality is given by magnitude of inner product of v,w is less than equal to norm of w into norm of v and we defined it ratio rho is equal to covariance of XY divided by Sigma X into Sigma Y this is called the correlation coefficient. The Cauchy Schwarz inequality was used to protect magnitude of rho is less than equal to 1. We used Cauchy Schwarz inequality so that the magnitude of the correlation ratio is less than equal to 1.

For uncorrelated random variables X and Y, rho is equal to 0. So since they are uncorrelated rho is equal to 0. If X and Y are uncorrelated Gaussian, then they are independent also using the joint PDF of the Gaussian random variables. We saw that if rho is equal to 0 then X and Y are independent. Also generally independent random variables are uncorrelated but uncorrelated does not imply independence, only in the case of Gaussian.

If they are uncorrelated then they are independent also two random variables X and Y are called orthogonal. If inner product of XY that is equal to E of X Y equal to 0. Zero mean uncorrelated random variables are orthogonal and they play important role in the modelling of random signals. That we will be using in a future lecture. Statistical signal processing also uses matrix theory extensively.

We will not cover matrix theory in this lecture but I will be introducing it whenever application comes. In the next lecture we will give the foundation of random process. Thank you.