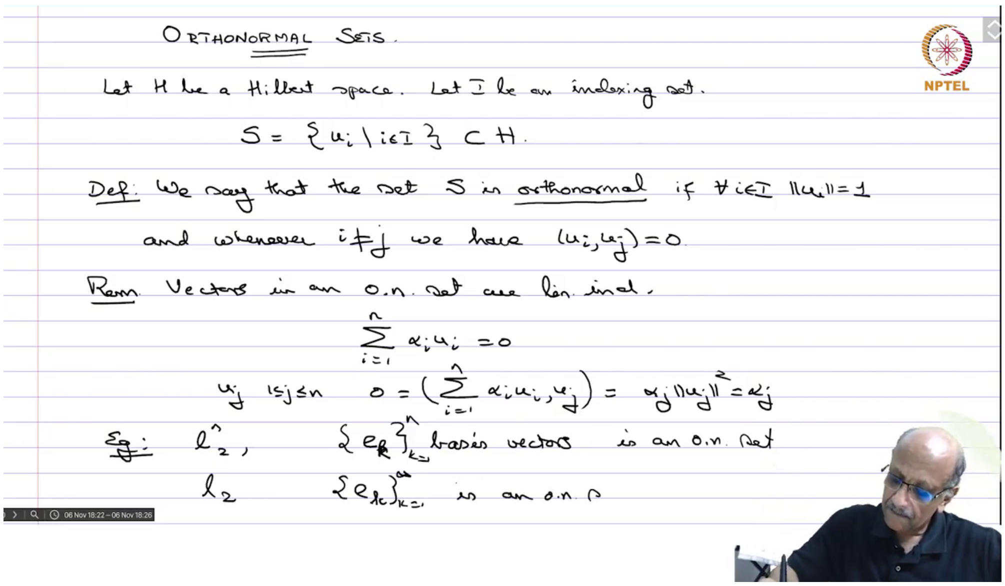
Functional Analysis
Professor. S. Kesavan
Department of Mathematics
The Institute of Mathematical Sciences
Lecture No. 54
Orthonormal Sets

(Refer Slide Time: 0:16)



We said that the distinguishing feature of Hilbert spaces over Banach spaces was the orthogonality of vectors. And in fact, so far we have seen one important result that every closed subspace of a Hilbert space, due to the orthogonal decomposition $M + M^{\perp}$ is complemented. So, every closed subspace is complemented unlike a Banach space where it cannot happen. So, now, it is time for us to study this orthogonality in a little more detail and so we come to the most important section in this chapter.

Orthonormal Sets

Let H be a Hilbert space. Let I be an indexing set. It could be empty, it could be finite, it could be countable, it could be uncountable. So, let $S = \{u_i / i \in I\} \subset H$

Definition. We say that the set S is orthonormal if $\forall\, i \in I \;\; ||u_i|| = 1$. That is the normal part. Every vector is normalized. So, every vector has unit norm and whenever $i \neq j$, we have $(u_i,$

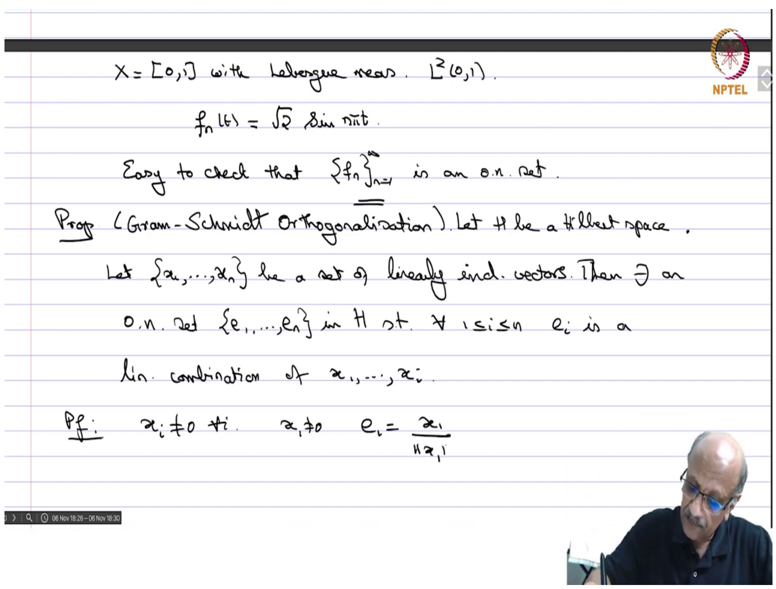
u_i)=0. So, they are all orthogonal to each other and they are all normalized vectors, namely, they have norm equal to 1 and therefore, such a set is called orthonormal.

Remark. Vectors in an orthonormal set, I will write o n for orthonormal, so orthonormal sets are linearly independent. So, if you have $\sum_{i=1}^{n} \alpha_i u_i = 0$, then you take any one u_i, $1 \leq j \leq n$, , so you have $\sum_{i=1}^{n} \alpha_i (u_i, u_j) = 0 = \alpha_j ||u_j||^2 = \alpha_j$ as $\alpha_i (u_i, u_j) = 0\ \forall\, i \neq j$. So, each $\alpha_j = 0$ and therefore, the vectors are linearly independent.

Example. So, if you take l_n^2 or l_2 . Let me write separately the basis vectors.

$\{e_k\}_{k=0}^{n}$ is a basis of l_n^2. $\{e_k\}_{k=0}^{\infty}$ is a basis of l_2. These are all sequences. 1 in the k-th place, 0 elsewhere, is an orthonormal set.

(Refer Slide Time: 4:47)



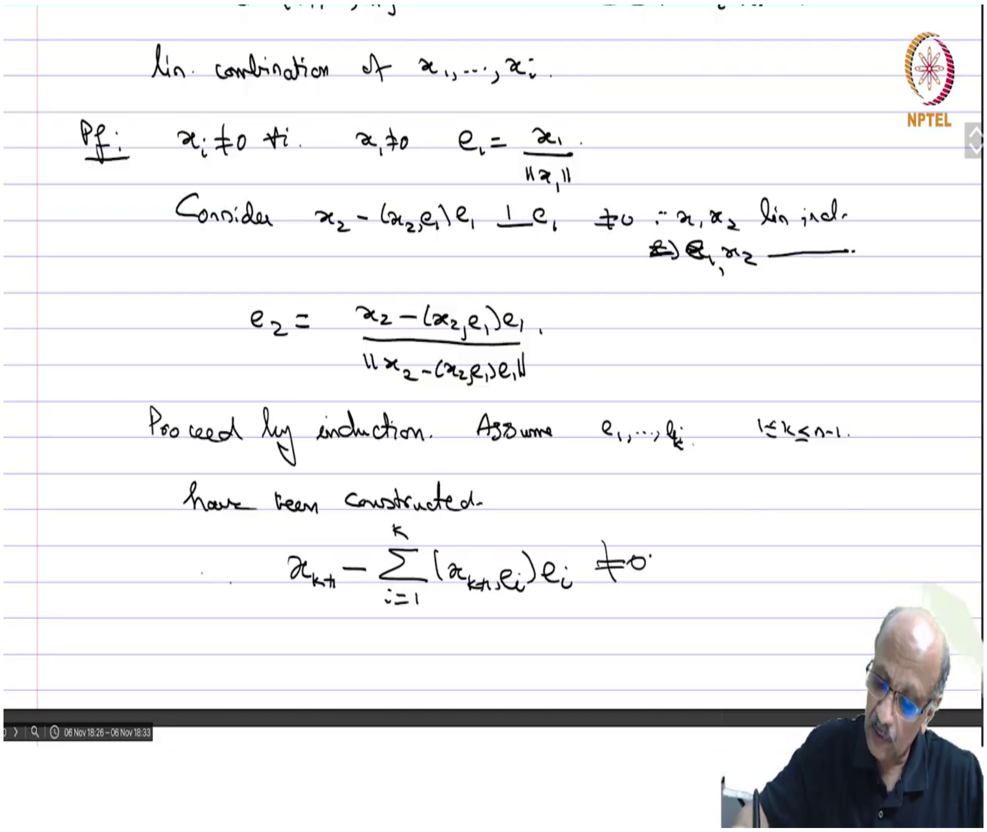
Then, let us take $X = [0, 1]$ with the Lebesgue measure. And then you look at $L^2(0, 1)$. So, then if you write $f_n(t) = \sqrt{2}\ \sin\ (n\pi t)$, then easy to check that $\{f_n\}_{n=1}^{\infty}$ is an orthonormal set. So,

$\int_0^1 2\sin^2(n\pi t)dt = 2\frac{1}{2} = 1$ and if you have, if $n \neq m$, $\int_0^1 \sin(m\pi t)\ \sin(n\pi t)dt = 0$ because that will be defined as the difference of some cosines and then you integrate, you will get 0. So, $\{f_n\}_{n=1}^{\infty}$ is an example of orthonormal sets. So, now, we come to an important proposition.

Proposition: (Gram-Schmidt Orthogonalisation.) Let H be a Hilbert space. Let $\{x_1, x_2, \cdots, x_n\}$ be a set of linearly independent vectors. Then, there exists an orthonormal set $\{e_1, e_2, \cdots, e_n\}$ in H such that for every $1 \leq i \leq n$, e_i is a linear combination of $\{x_1, x_2, \cdots, x_i\}$. So, you take x_1, then you get e_1. Take $\{x_1, x_2\}$, you get e_2 which is a linear combination of x_1& x_2 and so on. So, you build progressively this orthonormal set. So, the span of $\{e_1, e_2, \cdots, e_n\}$ is the same as the span of $\{x_1, x_2, \cdots, x_n\}$ because each e_i is a linear combination of the $x_1, x_2, \cdots, x_n$'s and therefore, they span the same space. It is orthonormal so it is also linearly independent.

Proof. $x_i \neq 0\ \forall\, i$. So, $x_1 \neq 0$. So, you put $e_1 = \frac{x_1}{||x_1||}$. So, this automatically is a vector which is a unit vector.

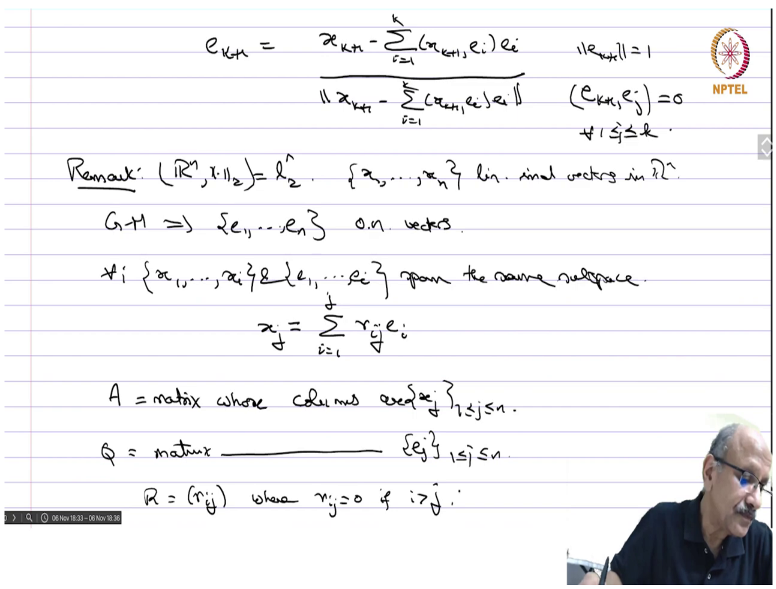
(Refer Slide Time: 8:23)



Now, you take e_2, so you consider $x_2 - (x_2, e_1)e_1$. Now, this vector is orthogonal to e_1. Because if you take the inner product you get $(x_2, e_1) - (x_2, e_1)(e_1, e_1) = 0$, as $(e_1, e_1) = 1$. So, this is orthogonal to e_1. And $x_2 - (x_2, e_1)e_1 \neq 0$ since x_1, x_2 are linearly independent and this implies that e_1 and x_2 are also linearly independent. Because e_1 is just a multiple of x_1. And here, you have a non-zero coefficient for x_2 so this cannot be the 0 vector. Therefore, I can divide by its norm. So, I define $e_2 = \frac{x_2 - (x_2, e_1)e_1}{||x_2 - (x_2, e_1)e_1||}$. So again, e_2 is therefore just a multiple of x_2 for the first one. So, e_1 and e_2 are orthogonal to each other and e_2 is a linear combination of x_2 and e_1 and that means it is a linear combination of x_2 and x_1. Now, you proceed by induction. So you assume, we have constructed $\{e_1, ..., e_{ki-1}\}$, for $1 \leq k \leq n - 1$.

So, we are going to define e_{k+1}. So, we now look at $x_{k+1} - \sum_{i=1}^{k} (x_{k+1}, e_i)e_i$. Now, this again, cannot be 0 because it is a linear combination of x_{k+1} and $\{e_1, e_2, \cdots, e_k\}$, but $\{e_1, e_2, \cdots, e_k\}$, by induction hypothesis, are a linear combination of $\{x_1, x_2, \cdots, x_k\}$ so the whole thing is a linear combination of $x_1, x_2, \cdots, x_{k+1}$ and the coefficient of x_{k+1} is 1, which is not 0, so $x_{k+1} - \sum_{i=1}^{k} (x_{k+1}, e_i)e_i \neq 0$.

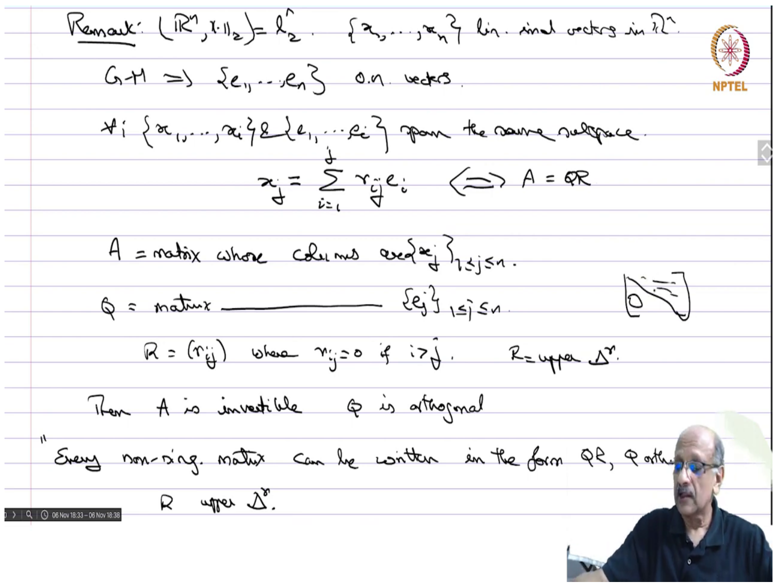
(Refer Slide Time: 11:41)



Therefore, I can now define $e_{k+1} = \frac{x_{k+1} - \sum_{i=1}^{k} (x_{k+1}, e_i)e_i}{||x_{k+1} - \sum_{i=1}^{k} (x_{k+1}, e_i)e_i||}$. Then, $||e_{k+1}|| = 1$ and $(e_{k+1}, e_j) = 0, \forall 1 \leq j \leq k$. So, then, the induction hypothesis is complete and therefore you can go on. So, this can be done for any n and you can construct the orthonormal base vectors.

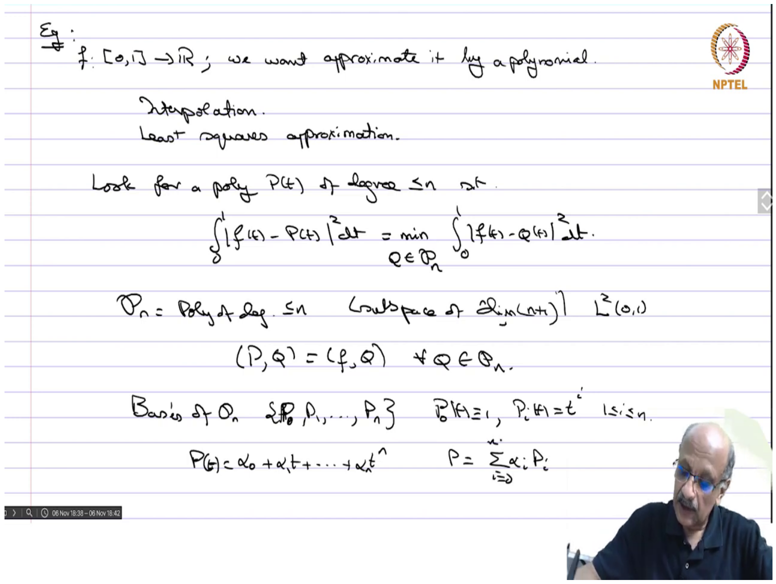
Remark. Let us take $(\mathbb{R}^N, ||.||_2)$, that is, l_n^2. Then, you take $\{x_1, x_2, \cdots, x_n\}$ linearly independent vectors in $\mathbb{R}^N$. And then we apply the Gram-Schmidt process to give you $\{e_1, e_2, \cdots, e_n\}$ which are orthonormal vectors. So, now, what, so we can, they span, for any i, $\{x_1, x_2, \cdots, x_i\}$and $\{e_1, e_2, \cdots, e_i\}$, so, $\forall i$ $\{x_1, x_2, \cdots, x_i\}$ and $\{e_1, e_2, \cdots, e_i\}$ span the same subspace. So, let us write $x_j = \sum_{i=1}^{j} r_{ij} e_i$. Now, let A be the matrix whose columns are x_j. Q be the matrix whose columns are e_j. And R be the matrix (r_{ij}), where you put $r_{ij} = 0, if\ i > j$.

(Refer Slide Time: 14:56)



Then, A is invertible because the columns of the matrix are linearly independent. Now, Q is a matrix whose columns are all orthogonal to each other and therefore matrix Q is orthogonal. $Q\,Q^T = Q^T Q = I$. Because that is just the definition of an orthogonal matrix so the columns of the orthogonal matrix are orthogonal vectors. And therefore, $x_j = \sum_{i=1}^{j} r_{ij} e_i$ is the same as $A = QR$. This is a theorem in matrix theory which we have proved. Every non singular matrix can be written in the form QR, Q orthogonal and R upper triangle. Namely, R is a matrix where below the diagonal everything is 0. So, this is a theorem in matrix theory which we have now proved using the Gram-Schmidt Orthogonalisation process.

(Refer Slide Time: 16:37)



So, what is the use of these orthogonal matrices? They are very useful. So, let us give an example. So, you have a function.

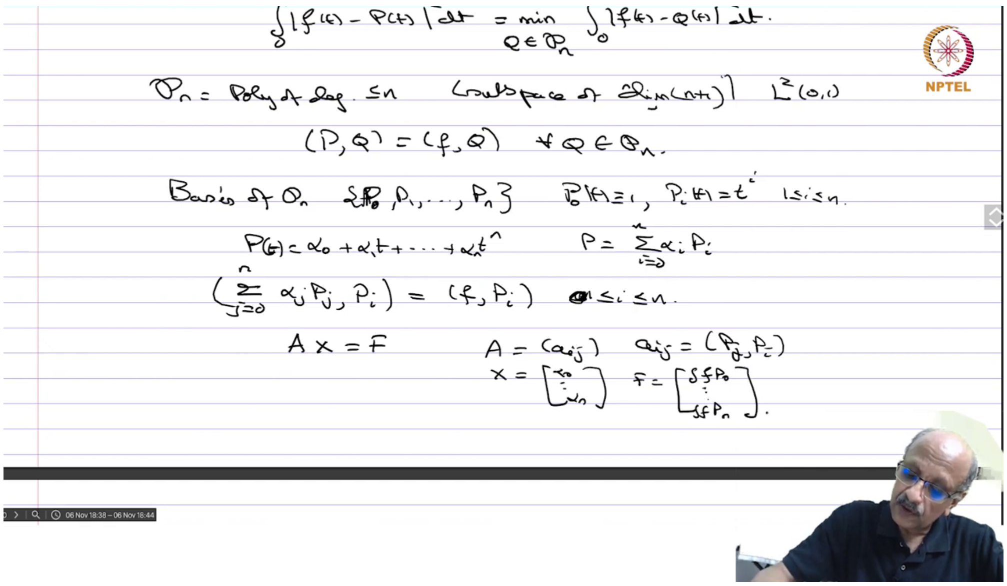
Example. Suppose I have a continuous function $f\colon [0,\ 1] \to \mathbb{R}$. Then we want to approximate it by a polynomial. There are various ways, Lagrange interpolation is one way. Another method is least squares approximation. What is a least squares approximation?

You look for a polynomial $P(t)$ of degree, let us say $\leq n$ such that

$\int_0^1 |f(t) - P(t)|^2 dt = \min_{Q \in \wp_n} \int_0^1 |f(t) - Q(t)|^2 dt$. Where, so what is this? $\wp_n$ is this polynomials of degree $\leq n$. So, this is a subspace of continuous functions of dim $(n + 1)$. We are going to look in the space $L^2(0, 1)$ because we are taking integrals here and here. So, $L^2(0, 1)$ is a Hilbert space. You have finite dimensional subspace therefore, it is a closed subspace. Then we know that you can always find a proximinial point, that means there exists a unique polynomial P which satisfies this condition and P is called the least square approximation of f. Now, how do we do this? So, this polynomial P, what is it characterized by? So, this characterization is, $(P,\ Q) = (f,\ Q)\ \ \forall\, Q\ \in \wp_n$. So, P is characterized by these equations, namely the inner product; we have seen this when proving the Lax-Milgram Lemma, even before that. So, now, since this is a linear space and this is a linear relationship in P and Q, therefore it is enough to check it only for the basis elements. So, basis of $\wp_n$, standard basis are the functions $\{P_0, P_1, \ldots, P_n\}$ where $P_0(t) = 1$, $P_i(t) = t^i$, $1 \le i \le n$, so $\{1, t, t^2, \ldots\}$ forms a basis.

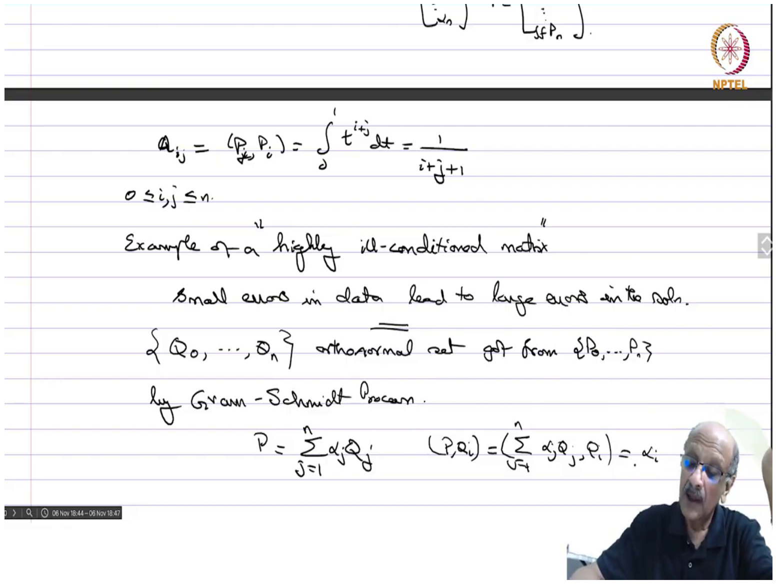
So, if you take the polynomial $P(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n$. So, $P(t) = \sum_{i=0}^{n} \alpha_i P_i(t)$.

(Refer Slide Time: 20:53)



So, then if you substitute that in equation $(P, Q) = (f, Q)$,, you will get following linear system. So, you substitute here. So, we are going to write, $(\sum_{j=0}^{n} \alpha_j \; P_j, P_i) = (f, P_i)$ for $1 \le i \le n$. So, this gives you $n + 1$ linear equations in $n + 1$ unknowns $\alpha_0, \alpha_1, \cdots, \alpha_n$. So, once you find the α's you know P and therefore, you have found the P^2. Now, what is this thing? This can be written as $Ax = F$, where $A = (a_{ij})$, $a_{ij} = (P_j, P_i)$. And X is a vector $X = \{\alpha_0, \cdots, \alpha_n\}$ and F is a vector which is $F = \{\int f P_0, \ldots, \int f P_n\}$.

(Refer Slide Time: 22:10)

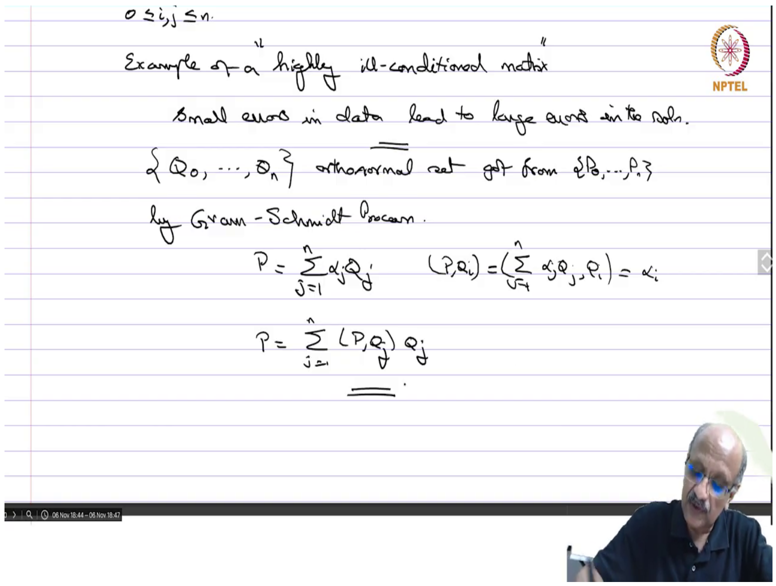


And this matrix A, $a_{ij} = (P_i, P_j) = \int_0^1 t^{i+j} dt = \frac{1}{i+j+1}$. So, this is a nice, symmetric matrix with, for $0 \le i, j \le n$. So, this is a $(n+1) \times (n+1)$ matrix and therefore it is technically a nice, in fact it is positive definite also, you can check that because these are basis vectors. So, it can be solved. But when n is large, this is a very bad matrix. In fact, this is an example of a highly ill-conditioned matrix. So, this matrix whose entries are $\frac{1}{n+j+1}$ is an example of a highly ill-conditioned matrix. This is a terminology in numerical analysis which means small errors in data lead to large errors in the solution. That means round of errors which are small, which are natural to occur will blow up when you come to, the solution which you get by inverting this matrix will be nowhere the real solution because there will be, computers will make a lot of errors and this matrix of this nature will completely destroy it.

So, it is at this situation that the orthonormal basis, so suppose we do the Gram-Schmidt Orthogonalisation and then we get $\{Q_0, Q_1, \ldots, Q_n\}$ is orthonormal set got from $\{P_0, P_1, \ldots, P_n\}$ by Gram-Schmidt. So, we get the Gram-Schmidt process and then, so let me write Gram-Schmidt. So, then, $P = \sum_{j=1}^{n} \alpha_j Q_j$ and then, it is very easy now.

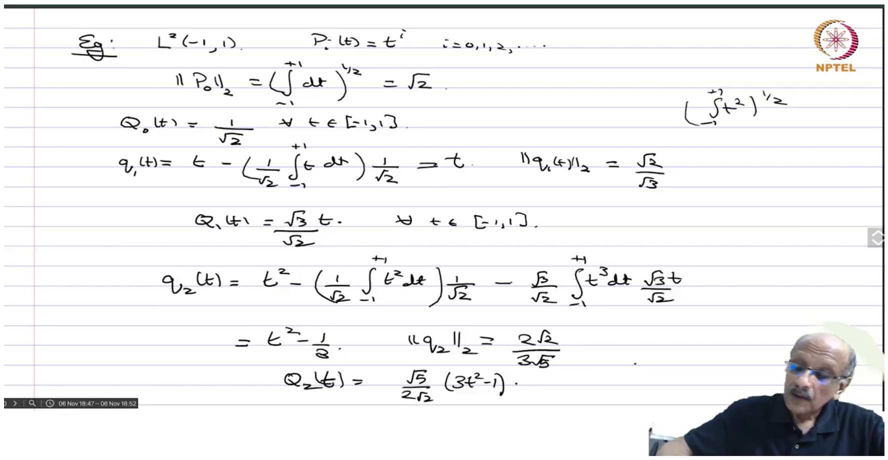
If I take $(P, Q_i) = \sum_{j=1}^{n} \alpha_j (Q_j, Q_i) = \alpha_i$, because only one of the terms is going to survive, all others will disappear because of the orthogonality conditions. Therefore there is nothing to, no need to solve anything.

(Refer Slide Time: 25:26)



I, therefore can write, immediately I know what to, so I can write $P = \sum_{j=1}^{n} (P, Q_j) Q_j$. That is all. So, I have got a polynomial without any difficulty. So, this is one of the uses of Gram-Schmidt Orthogonalisation. It helps you to immediately write down the solution.

(Refer Slide Time: 25:52)



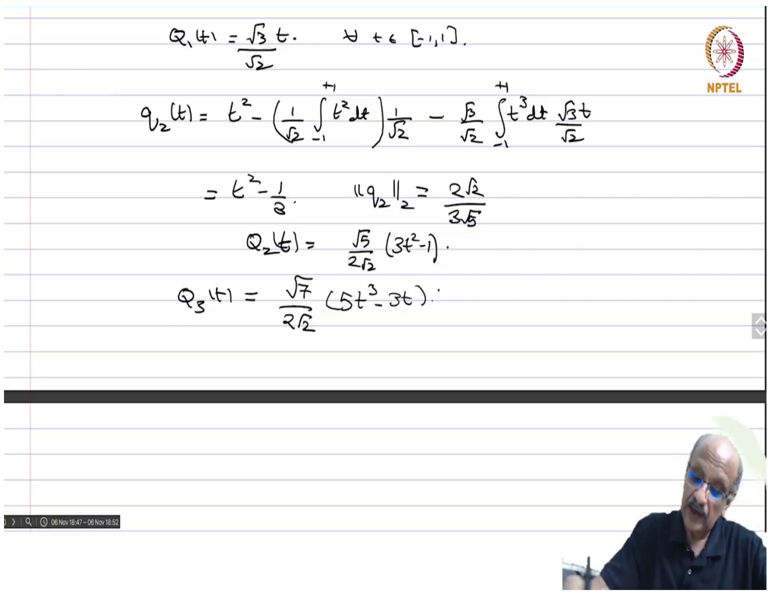
Example of this Gram-Schmidt process.

So, let us take the space $L^2(-1, 1)$ and let us take $P_i(t) = t^i, \; i = 0, 1, 2, \cdots$. We want to construct the Qs. So, what is P_0? So, we have to take $||P_0||_2$. $||P_0||_2 = (\int_{-1}^{1} dt)^{1/2} = \sqrt{2}$. P_0 is constant. So, we write $Q_0(t) = \frac{1}{\sqrt{2}}, \forall t \in [-1, 1]$. Now, we have to look at the function. So now, consider the following function., Note, $P_1(t) = t$. So, $q_1(t) = t - (t, Q_0) = t - \frac{1}{\sqrt{2}}\int_{-1}^{1} t\, dt = t$. So, that is just the inner product of q, then, multiply it by $Q_0(t) = \frac{1}{\sqrt{2}}$.

Now, $\int t\, dt$, t is an odd function. So, if you integrate, you would have got t^2, it is the same at +1 and -1, so we will get $\int_{-1}^{1} t\, dt = 0$, so this term will not come. So, this is just t. And then $||q_1(t)||_2 = (\int_{-1}^{1} t^2\, dt)^{1/2} = \frac{\sqrt{2}}{\sqrt{3}}$. And therefore, you write $Q_1(t) = \frac{\sqrt{3}}{\sqrt{2}}t \quad \forall\, t \in [-1, 1]$. So, now, let us take $q_2(t) = t^2 - (\frac{1}{\sqrt{2}}\int_{-1}^{1} t^2 dt)\frac{1}{\sqrt{2}} - (\frac{\sqrt{3}}{\sqrt{2}}\int_{-1}^{1} t^3 dt)\frac{\sqrt{3}}{\sqrt{2}}t = t^2 - \frac{1}{3}$. Note

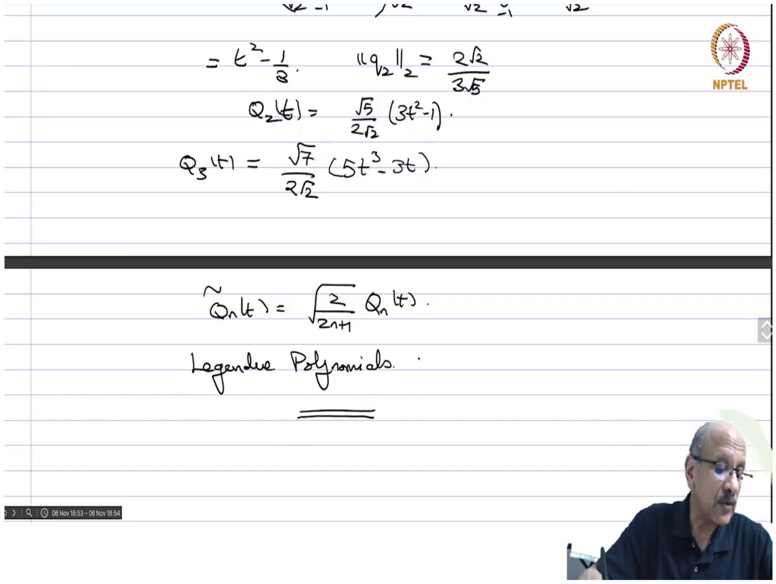
$\int_{-1}^{1} t^3 dt = 0$ as t^3 is an odd function. And then $||q_2(t)||_2 = \frac{2\sqrt{2}}{3\sqrt{5}}$. So then, you have $Q_2(t) = \frac{\sqrt{5}}{2\sqrt{2}}(3t^2 - 1)$.

(Refer Slide Time: 30:37)



And similarly, you can proceed like this and you can show for instance that $Q_3(t) = \frac{\sqrt{7}}{2\sqrt{2}}(5\,t^3 - 3\,t)$. So here, you notice something. Even polynomials which you get are only involving the even powers of t and the odd polynomials are involving only the odd powers of t. That is just because of the space, $[-1, 1]$ and the inner product which we have here.

(Refer Slide Time: 31:18)



Remark. If you have $\tilde{Q}_n(t) = \sqrt{\frac{2}{2n+1}}\, Q_n(t)$, these are nothing but the famous Legendre Polynomials. So, the Legendre Polynomials in applied mathematics are very famous in numerical analysis also, so these are orthogonal polynomials. They also come from differential equations, you might have studied. And these are nothing but, the, you take $\{1,\ t,\ t^2, ..\}$ etc., apply the Gram-Schmidt Orthogonalisation in $L^2(-1,\ 1)$ and then you get them with the scaling factor in the front. So, that is an example of this.

So now, we want to study in detail the properties of Orthonormal sets. So, we have seen examples of Orthonormal sets, we have seen examples of the Gram-Schmidt Orthogonalization process, we have also seen an application of why such orthogonalization is desirable and now, we will now study the important properties of this.