**Neural Networks for Signal Processing-I**

**Prof. Shayan Srinivasa Garani**

**Department of Electronic System Engineering**

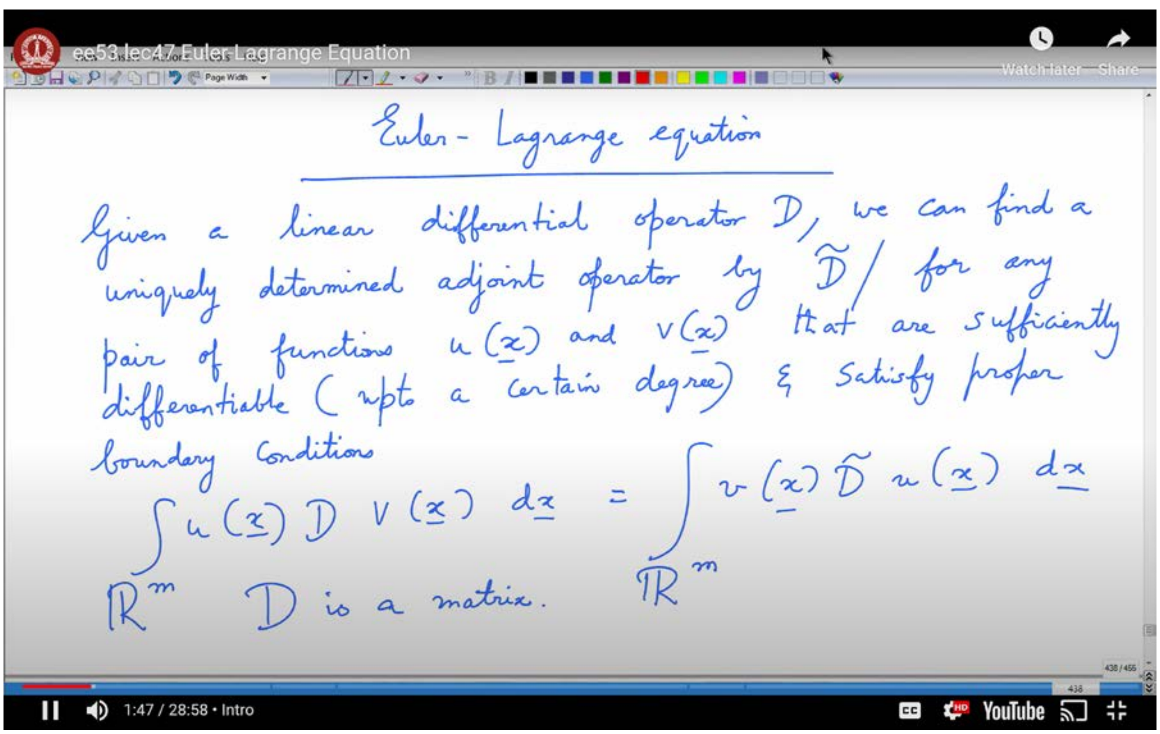
**Indian Institute of Science, Bengaluru**

**Lecture – 47**

**Euler-Lagrange Equation**

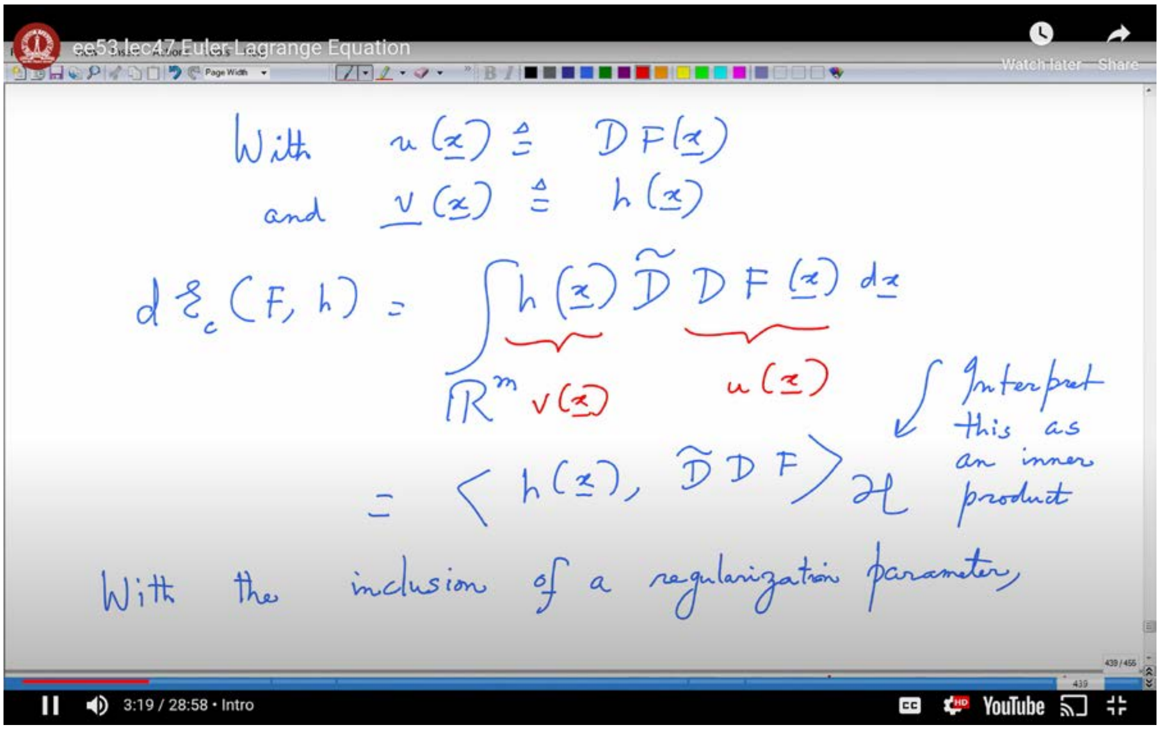
Let's delve into the Euler-Lagrange equation after having studied the Trichinow functional. We've computed the differentials of this functional separately over the standard error term and the regularization term, and these are referred to as the Fréchet differentials. Given a linear differential operator, denoted by $\mathcal{D}$, we can uniquely determine an adjoint operator, denoted by $\widetilde{\mathcal{D}}$, for any pair of functions u(x) and v(x) acting on data points x. This is subject to certain boundary conditions and under the assumption that these functions are sufficiently differentiable to a certain degree.

(Refer Slide Time: 01:47)



Now, consider an integral in an m-dimensional space of u(x) multiplied by this operator $\mathcal{D}$ acting on v(x), integrated over dx. This is equivalent to taking v(x) and applying the adjoint operator $\widetilde{\mathcal{D}}$ to u(x).

(Refer Slide Time: 03:19)



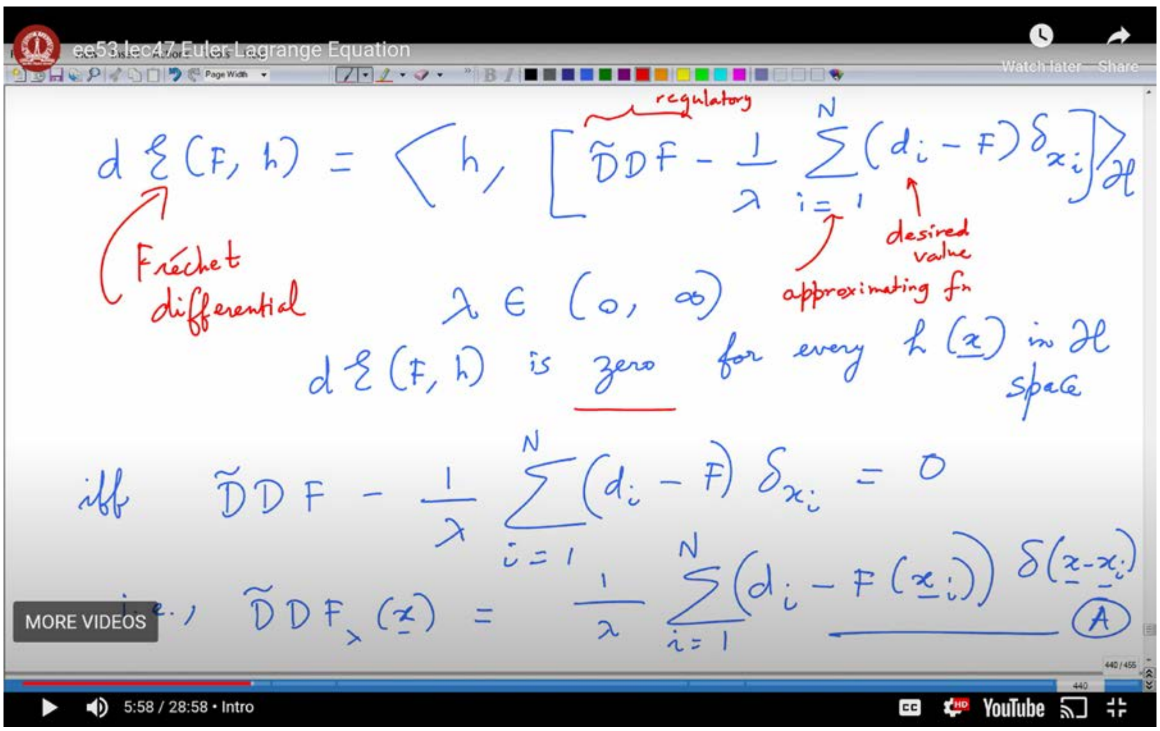

Let's make the following definitions: if we let u(x) represent a differential operator acting on a function f(x), and v(x) as h(x), then this differential, DEC$(f, h)$, can be expressed in the following form:

$$\mathrm{DEC}(f, h) = \int_{R^m} h(x)\mathcal{D}\, f(x)\, dx$$

Here, h(x) corresponds to v(x), and $\mathcal{D}$f(x) corresponds to u(x) based on our definition. We also have an adjoint operator linking v and u, similar to what we discussed earlier. This can be interpreted as an inner product between h and $\widetilde{\mathcal{D}}f$ in the space where this inner product is defined.

With the inclusion of a regularization term, we form our Fréchet differential. Here, we introduce an approximation error term and compute the inner product between h and the operator. However, it’s important to note that the arrows indicating the relationships might seem reversed, but this is due to the nature of our regulatory constraints. The approximation error pertains to the difference between the desired response and the approximating function.

(Refer Slide Time: 05:58)




Now, the Fréchet differential becomes zero for every h in this inner product space if the second term is set to zero. This implies that:

$$\tilde{\mathcal{D}}\mathcal{D}f = \frac{1}{\lambda}\sum_{i=1}^{n}\left(d_i - f(x_i)\right)\delta(x - x_i)$$

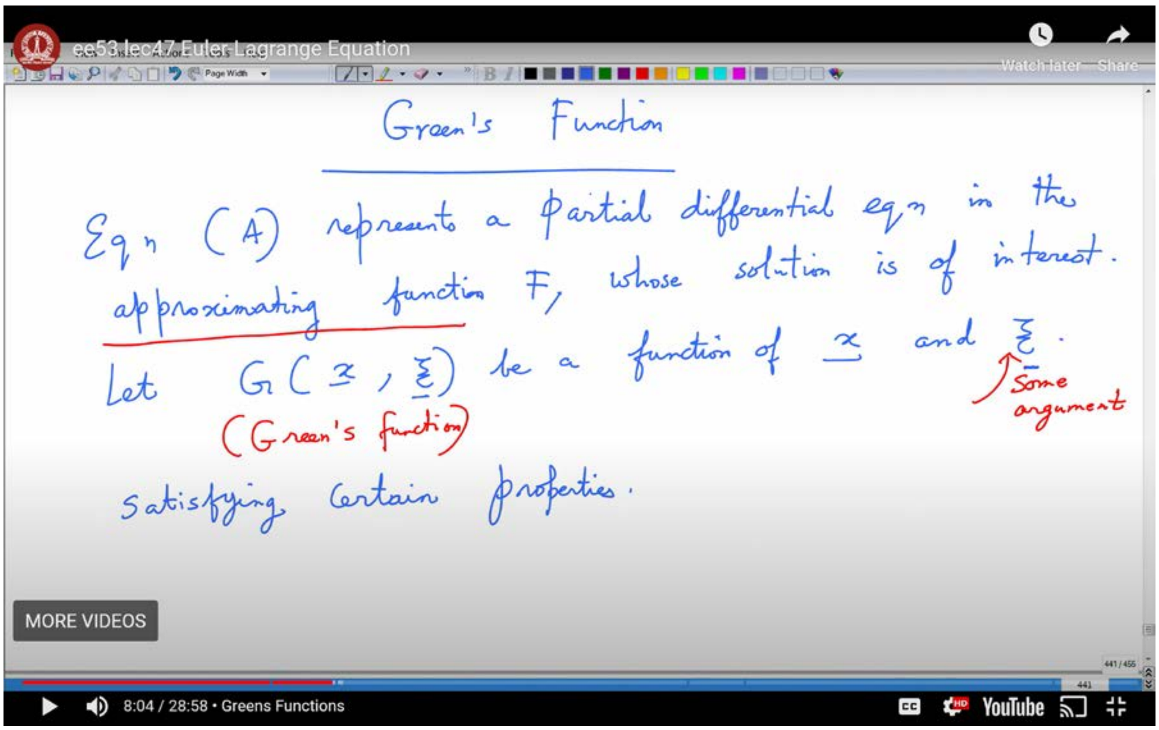
Here, $\delta(x - x_i)$ is the Dirac delta function, indicating that the function f is evaluated at the point $x = x_i$ because f(x) is a general hypersurface over the space, but we are sampling a discrete set of data points from this surface, specifically $f(x_i)$, where $x_i$ is the selected point and $f(x_i)$ is the desired response $d_i$. This is the reason we need to incorporate $\delta(x - x_i)$.

Upon simplification, we obtain:

$$\widetilde{\mathcal{D}}\mathcal{D}f_{\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n}\big(d_i - f(x_i)\big)\delta(x - x_i)$$

indicating that we are evaluating at $x = x_i$.

(Refer Slide Time: 08:04)

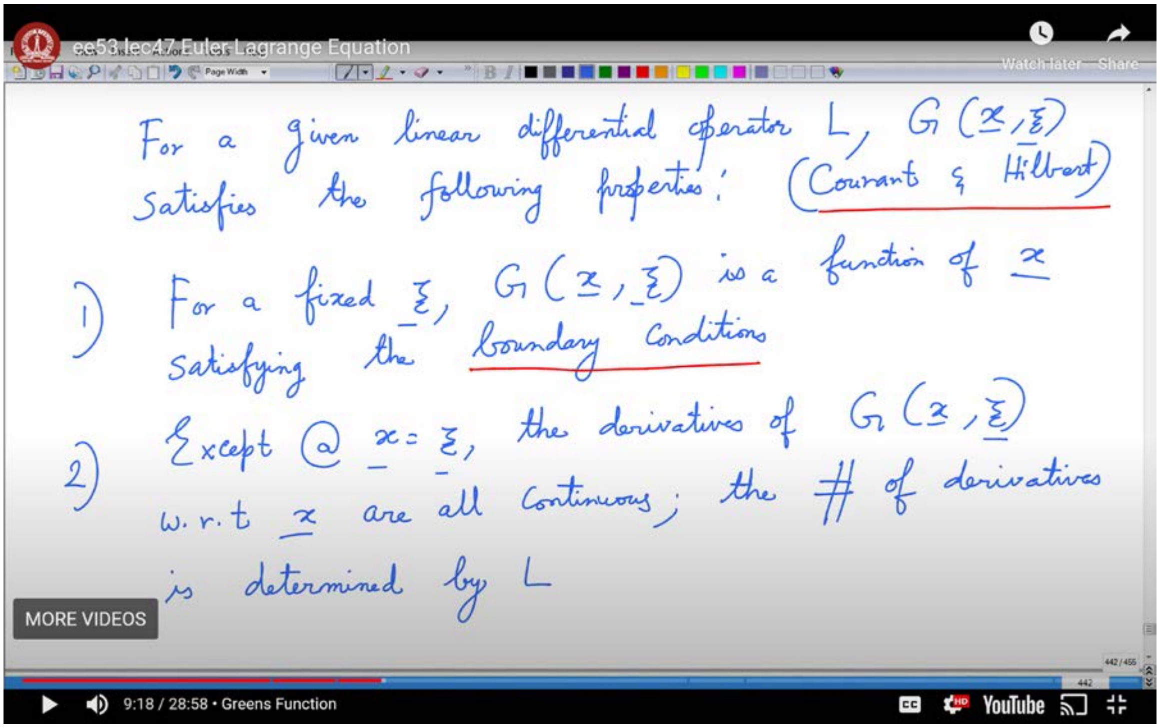



Now, let’s introduce the concept of Green's functions to appropriately link these differentials. This requires us to take a slight detour to discuss the fundamental aspects of Green's functions and their application in this context.

The equation I wrote earlier, $\widetilde{\mathcal{D}}\mathcal{D}f_{\lambda}$, represents a partial differential equation in the approximating function F whose solution we are interested in. To better understand this, let's delve into Green's functions and how they relate to our problem, specifically the solution of the approximating function problem in this mapping context.

Consider g(x, ζ) as a function of x and ζ, where ζ is a parameter, which could be a vector. This function is known as the Green's function and is characterized by certain properties established through the work of mathematicians like Courant and Hilbert.

(Refer Slide Time: 09:18)

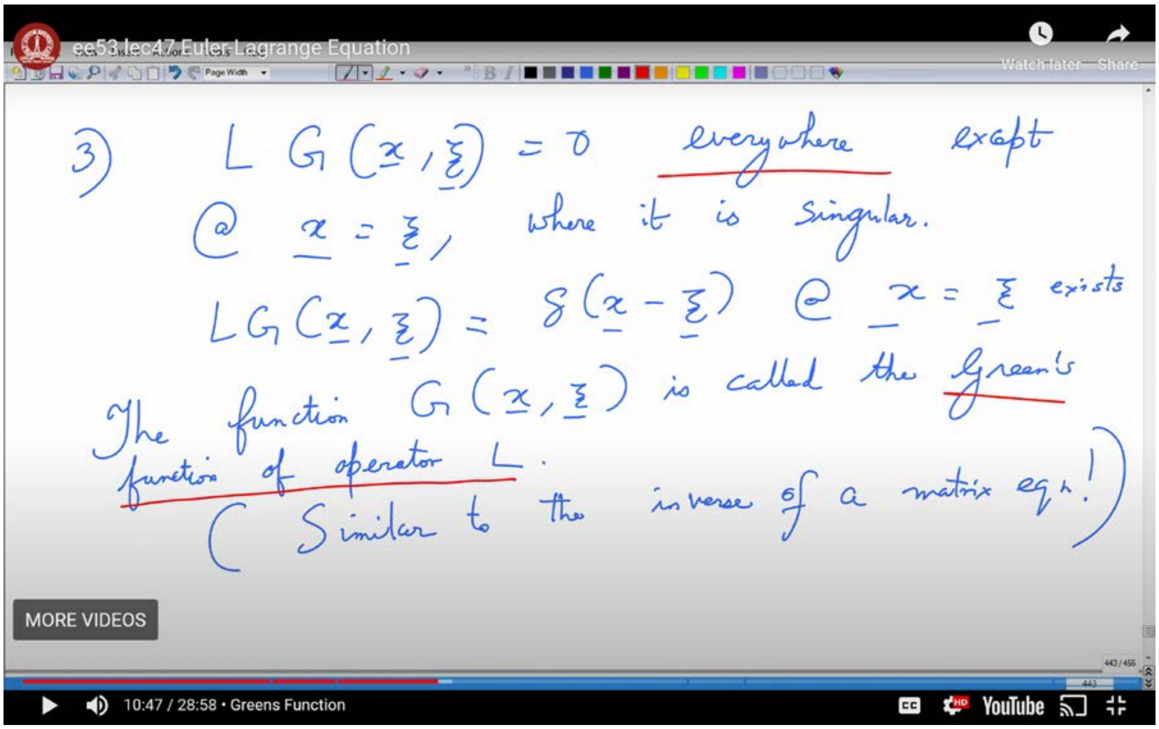



Given a linear differential operator $\mathcal{L}$, the Green's function g(x, ζ) satisfies the following key properties:

- For a fixed ζ, g(x, ζ) is a function of x that meets the boundary conditions specific to the problem.
- The derivatives of g(x, ζ) with respect to x are continuous except at x = ζ, where the function becomes singular. The number of derivatives is dictated by the linear differential operator $\mathcal{L}$.
- The operator $\mathcal{L}$ acting on g(x, ζ) yields zero everywhere except at $x = \zeta$, where it results in a singularity. Specifically, $\mathcal{L}g(x, \zeta) = \delta(x - \zeta)$, where $\delta(x - \zeta)$ is the Dirac delta function, which is non-zero only at x = ζ and zero otherwise.

The Green's function can be thought of as analogous to the inverse of a matrix in linear algebra. Just as $A \times A^{-1} = I$ for a matrix A, the Green's function satisfies $\mathcal{L}g(x,\zeta) = \delta(x-\zeta)$.

(Refer Slide Time: 10:47)




In the context of our study, let φ(x) be a continuous or piecewise continuous function in an m-dimensional space. The function f(x) can be expressed as an integral over this space:

$$f(x) = \int_{R^m} g(x,\zeta)\phi(\zeta)\ d$$

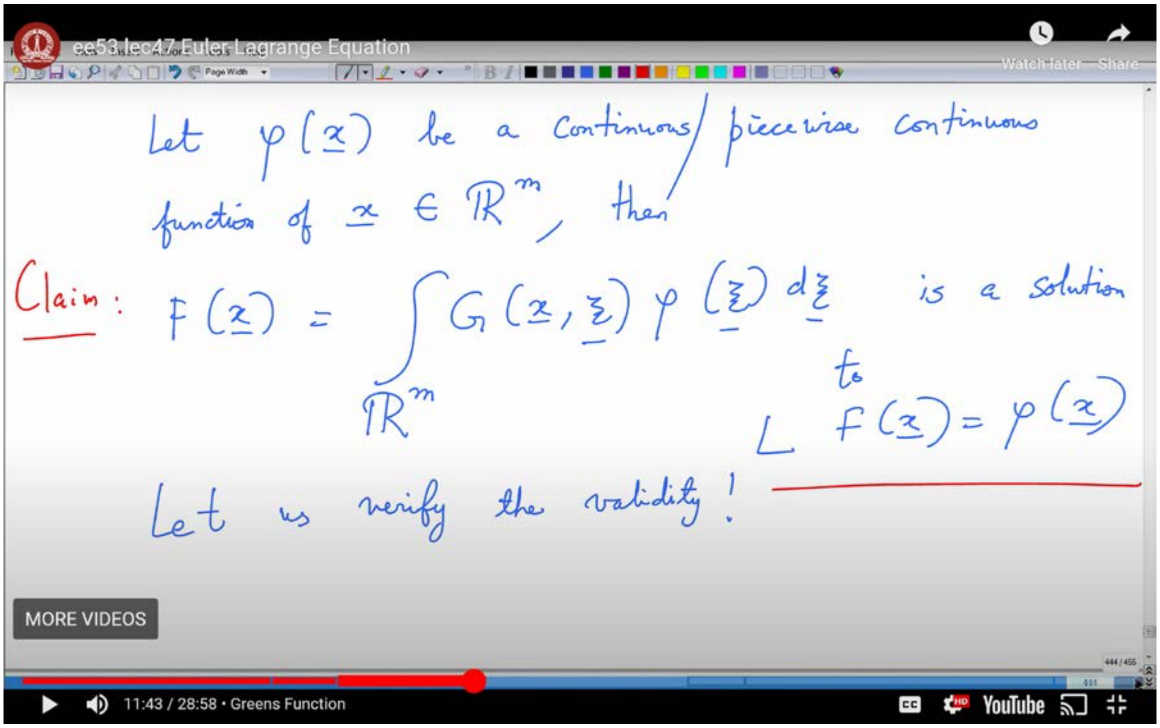
We claim that this integral represents a solution to the equation $\mathcal{L}f(x) = \phi(x)$. To verify this claim, consider applying $\mathcal{L}$ to f(x):

$$\mathcal{L}f(x) = \mathcal{L}\left(\int_{R^m} g(x,\zeta)\phi(\zeta)\ d\zeta\right)$$

Since $\mathcal{L}$ is a linear operator, it can be moved inside the integral:

$$\mathcal{L}f(x) = \int_{R^m} \mathcal{L}\, g(x,\zeta)\phi(\zeta)\, d$$

(Refer Slide Time: 11:43)




Given the property of the Green's function, $\mathcal{L}g(x,\zeta) = \delta(x-\zeta)$. Thus, the integral simplifies to:

$$\mathcal{L}f(x) = \int_{R^m} \delta(x-\zeta)\phi(\zeta)\ d$$

The Dirac delta function $\delta(x-\zeta)$ is zero except when x = ζ, so the integral evaluates to:

$$\mathcal{L}f(x) = \phi(x)$$

This confirms that f(x) is indeed a solution to $\mathcal{L}f(x) = \phi(x)$.

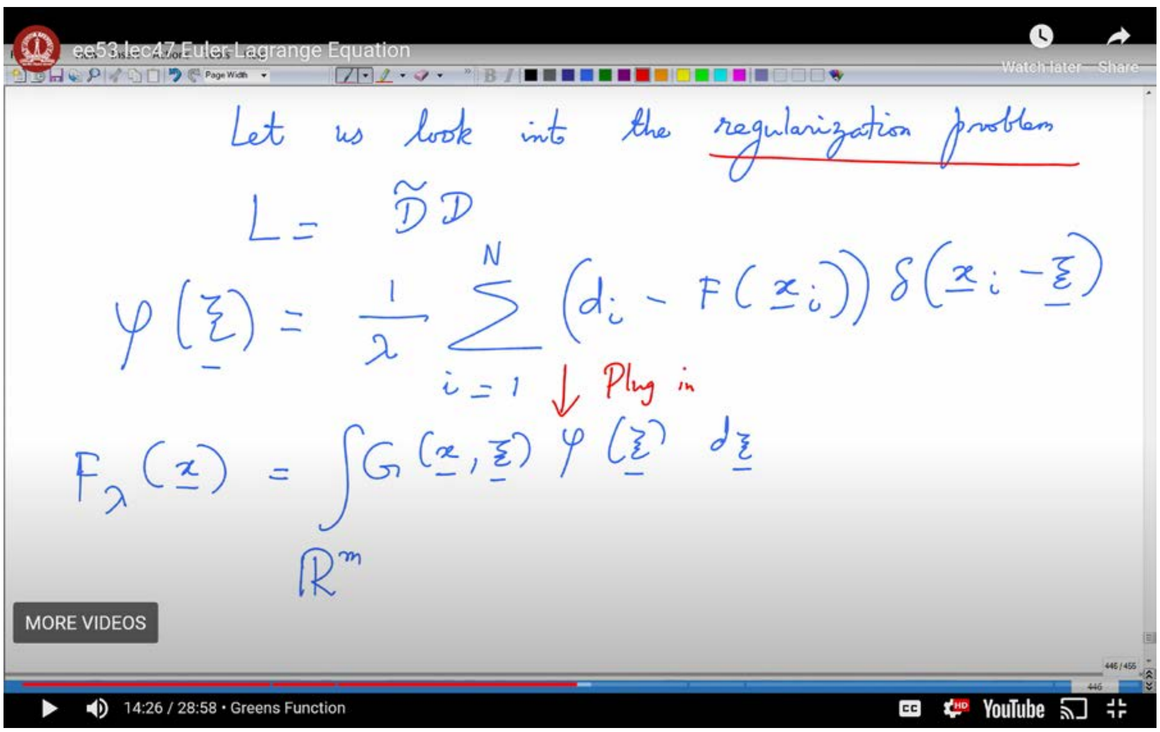
In the context of our regularization problem, let's express the linear differential operator $\mathcal{L}$ as $\widetilde{\mathcal{D}}\ \mathcal{D}$, and let φ(ζ) be given by:

$$\phi(\zeta) = \frac{1}{\lambda}\sum_{i=1}^{n}\big(d_i - f(x_i)\big)\delta(x - x_i)$$

where $d_i$ represents the desired values at data points $x_i$. Substituting this expression into our integral for $f_\lambda$, we get:

$$f_\lambda = \int_{R^m} g(x,\zeta)\phi(\zeta)\ d$$

(Refer Slide Time: 14:26)




This setup allows us to further simplify and explore the solution to our regularization problem using Green's functions.

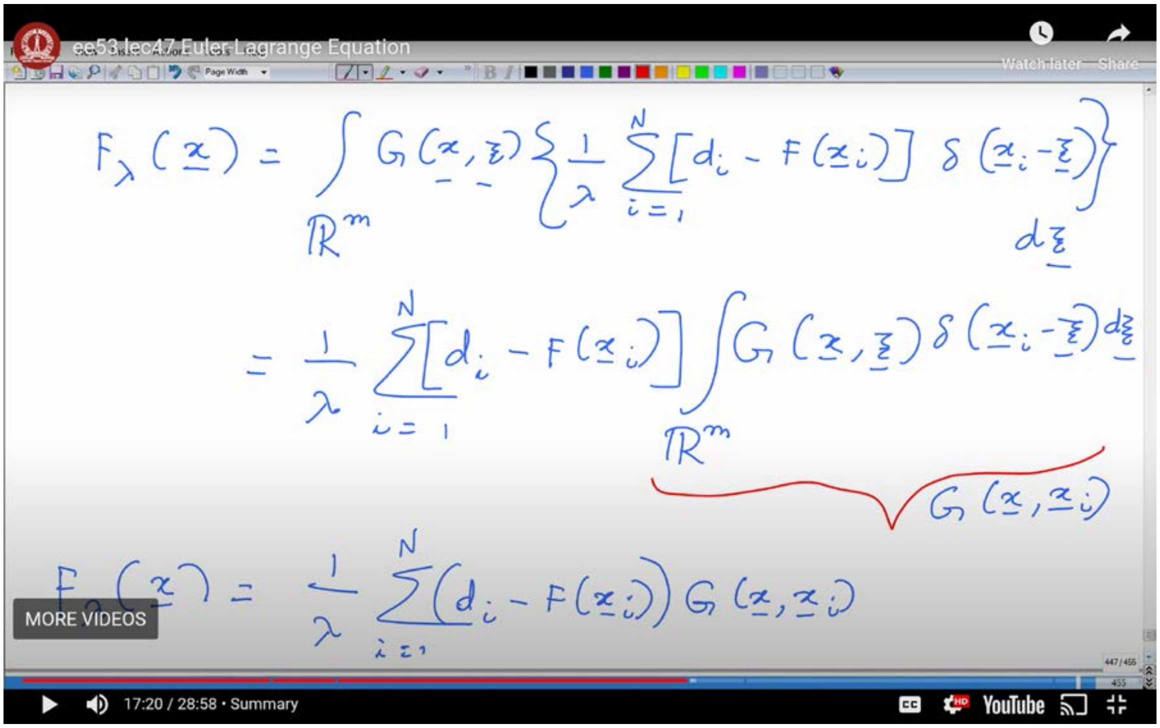
What I find is as follows: I substitute $\varphi(\zeta)$ directly into f and perform the necessary simplifications. In this integral, which is over $d\zeta$, I pull out the terms involving $d\zeta$ and leave the remaining terms outside. The integral now takes the form:

$$\int g(x,\zeta)\delta(x_i - \zeta)\ d$$

Outside the integral, I have:

$$\frac{1}{\lambda}\sum_{i=1}^{n}\left(d_i - f(x_i)\right)$$

(Refer Slide Time: 17:20)



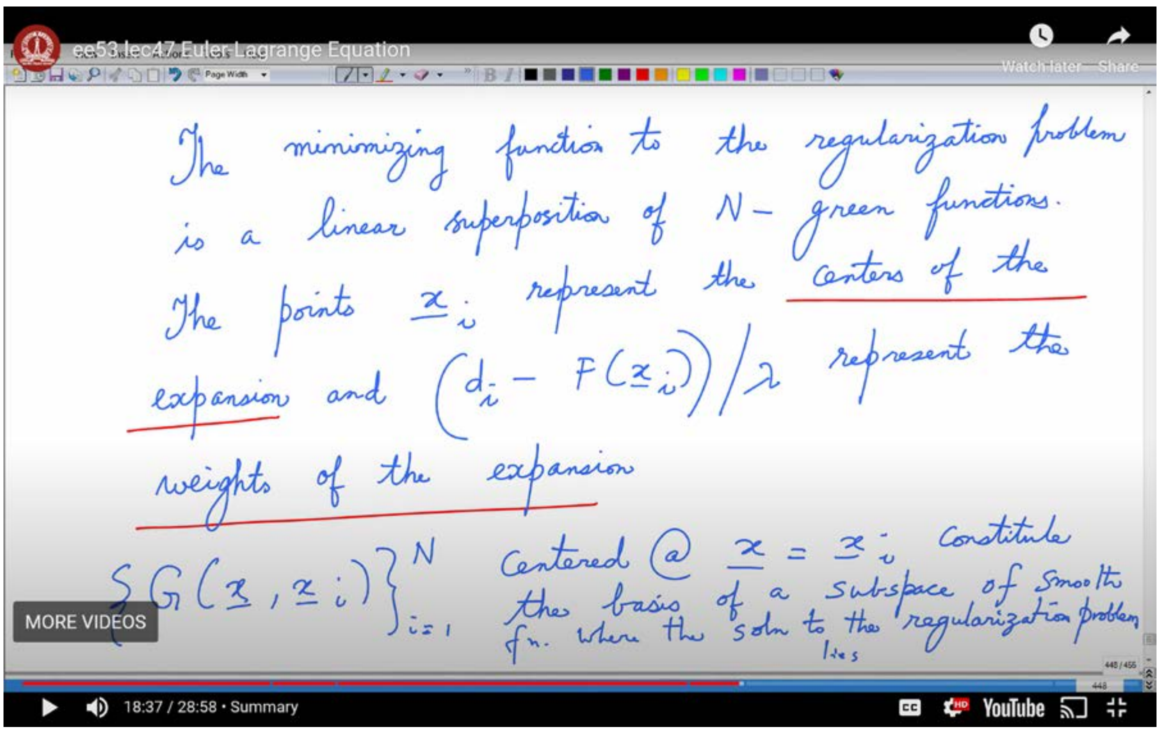

The Dirac delta function $\delta(x_i - \zeta)$ is non-zero only when $\zeta = x_i$. Consequently, the integral collapses to $g(x, x_i)$ because for $\zeta \neq x_i$, the integral evaluates to zero. Therefore, the simplified form of our approximating function $f_\lambda$ is:

$$f_\lambda(x) = \frac{1}{\lambda}\sum_{i=1}^{n}\left(d_i - f(x_i)\right)g(x, x_i)$$

Here, x is a continuous variable representing the surface, while $x_i$ are discrete points. The term $\frac{d_i - f(x_i)}{\lambda}$ acts as the weight for each Green's function $g(x, x_i)$, which depends on the data point $x_i$. We have n Green's functions, each corresponding to a data point $x_i$, and we are essentially taking a linear combination of these Green's functions.

To summarize, the function $f_\lambda$ that minimizes the regularization problem is a linear superposition of n Green's functions, where $x_i$ are the centers of the expansion and the coefficients $\frac{d_i - f(x_i)}{\lambda}$ represent the weights of this expansion. These Green's functions, centered at $x = x_i$, form a basis for a subspace of smooth functions where the solution to the regularization problem lies. Thus, $f_\lambda$, the surface as a function of x, is represented as a linear combination of these Green's functions. The weights associated with each Green's function depend on the data pairs ($x_i$, $d_i$), with each Green's function having an argument dependent on the data point $x_i$. This interpretation is crucial.

(Refer Slide Time: 18:37)




Now, you might wonder how to choose the weights. The weight $w(x_i)$ is defined as:

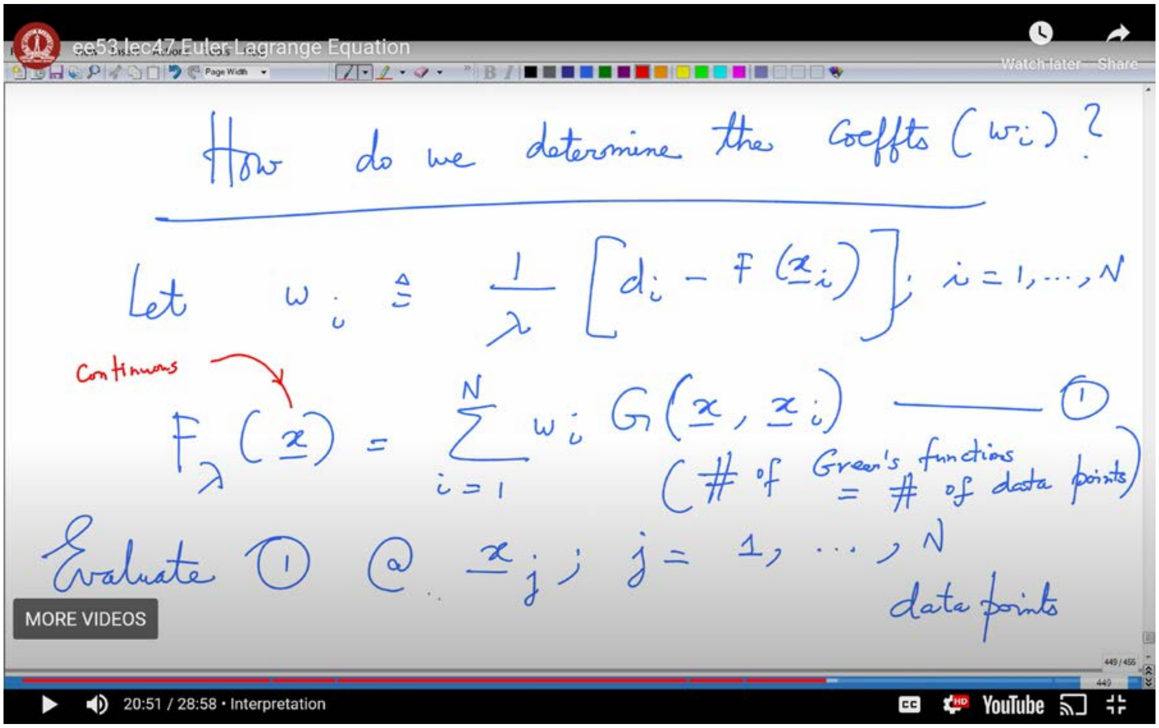
$$w(x_i) = \frac{1}{\lambda}\left(d_i - f(x_i)\right)$$

Substituting this into our expression for $f_\lambda(x)$, we get:

$$f_\lambda(x) = \sum_{i=1}^{n} w(x_i) g(x, x_i)$$

where x is continuous and $x_i$ are discrete data points. We then evaluate this equation at every sample point $x_j$, for j = 1, 2, …, n.

(Refer Slide Time: 20:51)



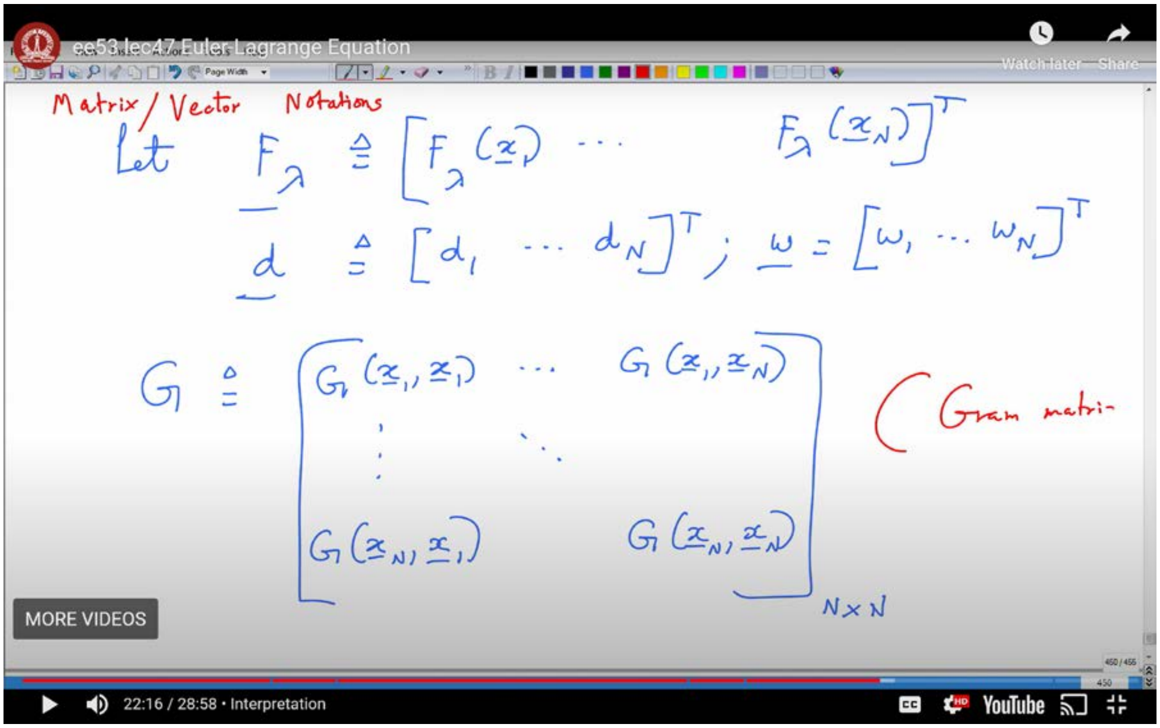

To clarify, x is a continuous variable, but when x is replaced by the sample points $x_j$, we can link the left-hand side and the right-hand side of the equation. By evaluating at all these n data points, we introduce some matrix/vector terminology. Define $f_\lambda$ as the vector:

$$\mathbf{f}_\lambda = \begin{bmatrix} f_\lambda(x_1) \\ f_\lambda(x_2) \\ \vdots \\ f_\lambda(x_n) \end{bmatrix}$$

This results in an n-dimensional column vector. Similarly, the desired responses are $d_1$, $d_2$, …, $d_n$, and the weights are given by:

$$\mathbf{w} = \begin{bmatrix} w(x_1) \\ w(x_2) \\ \vdots \\ w(x_n) \end{bmatrix}$$

(Refer Slide Time: 22:16)




Define $G$ as the matrix whose entries are the evaluations of the Green's function at pairs $(x_i, x_j)$. This matrix $\boldsymbol{G}$ is n × n.

In matrix form, we have:

$$w = \frac{1}{\lambda}(d - f_\lambda)$$

where $\boldsymbol{d}$ is the vector of desired responses, and $f_\lambda$ is the vector of approximating function evaluations. Therefore:

$$f_\lambda = Gw$$

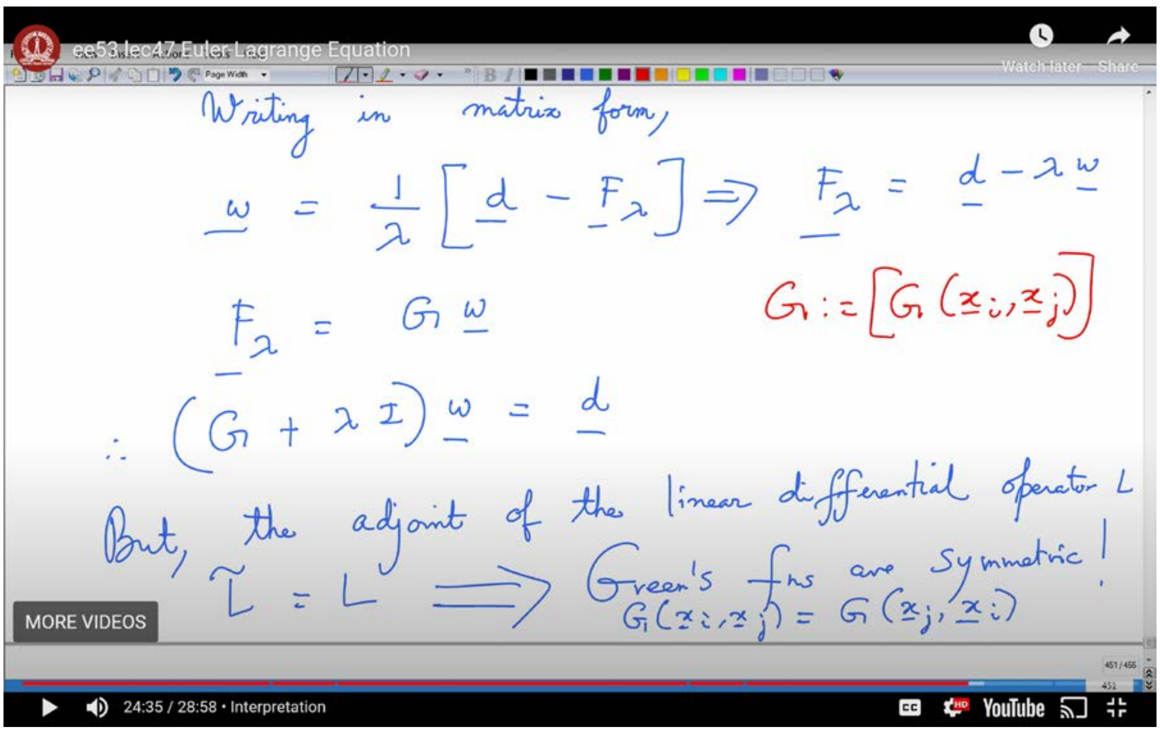
Plugging everything together, if we need $d$, we focus on this equation:

$$f_\lambda = Gw$$

To isolate D on one side of the equation, we start with the expression $f_\lambda + \lambda\, w$. The weight vector $\boldsymbol{w}$ can be written as $I\ \boldsymbol{w}$, where $\boldsymbol{I}$ is the identity matrix. By simplifying, we obtain the equation:

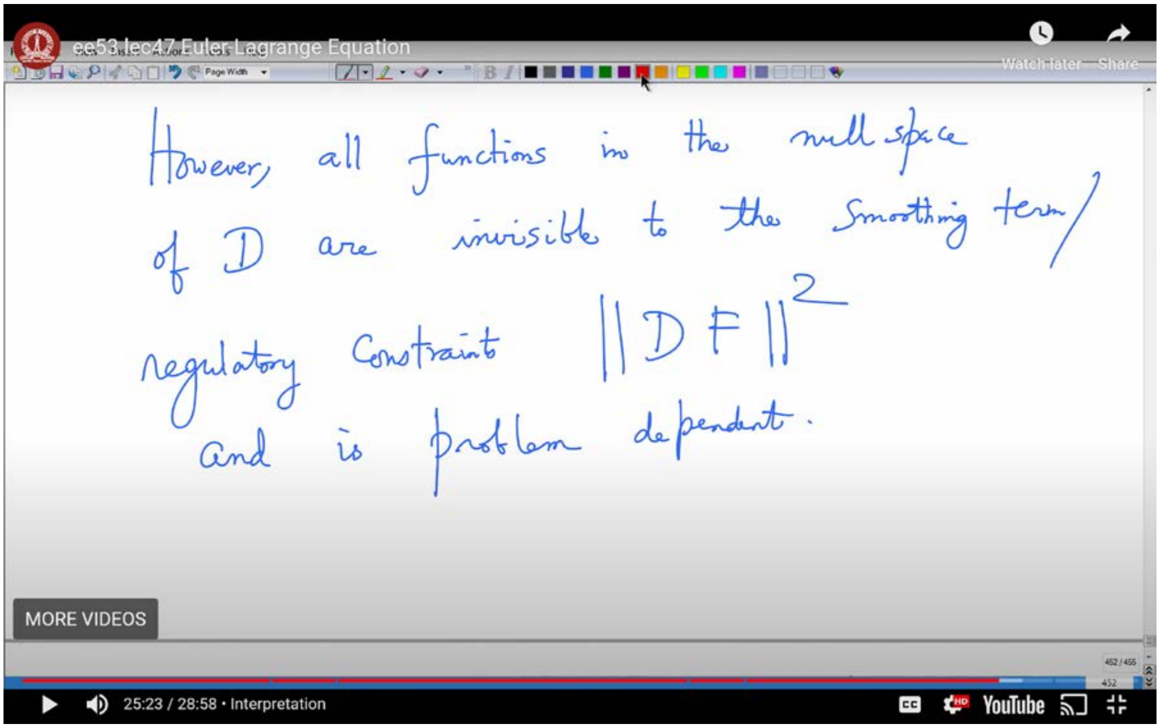
$$G + \lambda I w = \text{desired response}$$

(Refer Slide Time: 24:35)

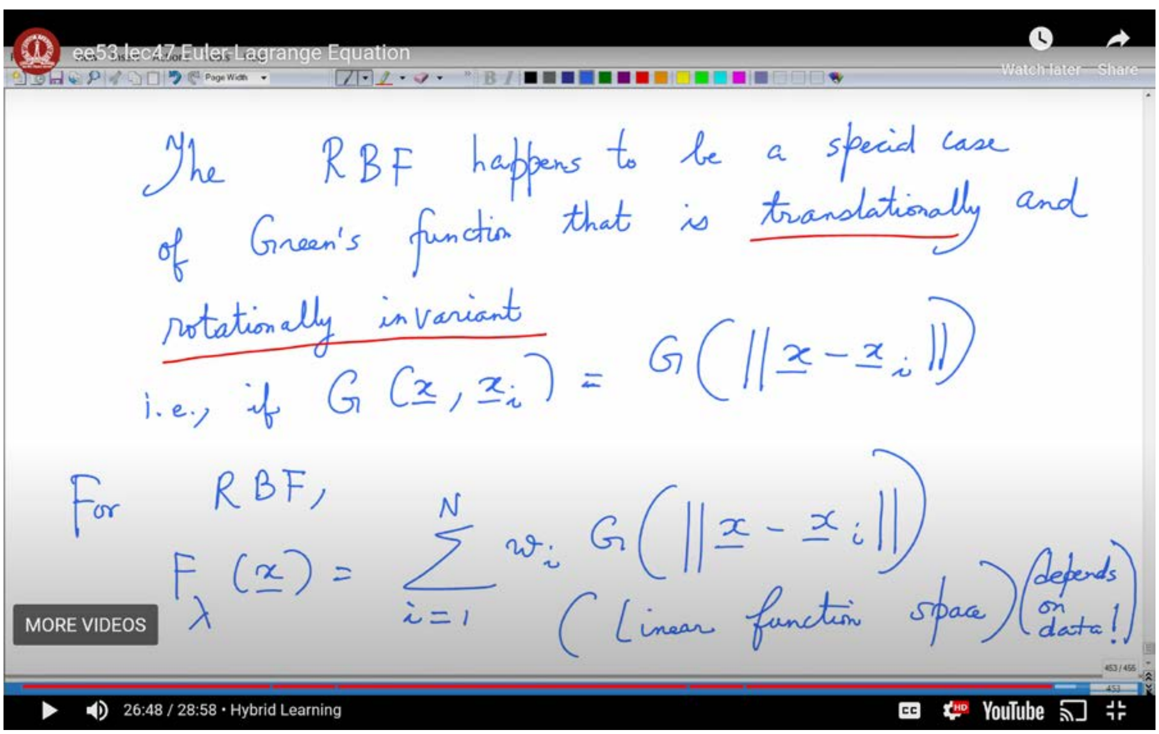



Given that $\tilde{L} = L$, which means the adjoint of the linear differential operator is the same as the operator itself, this implies that the Green's functions are symmetric. Thus, the matrix $\boldsymbol{G}$ is symmetric. All functions in the null space of D are unaffected by this smoothing term or the regularization constraints imposed by $|D\ f|$. This regularization constraint is specific to the problem at hand.

(Refer Slide Time: 25:23)

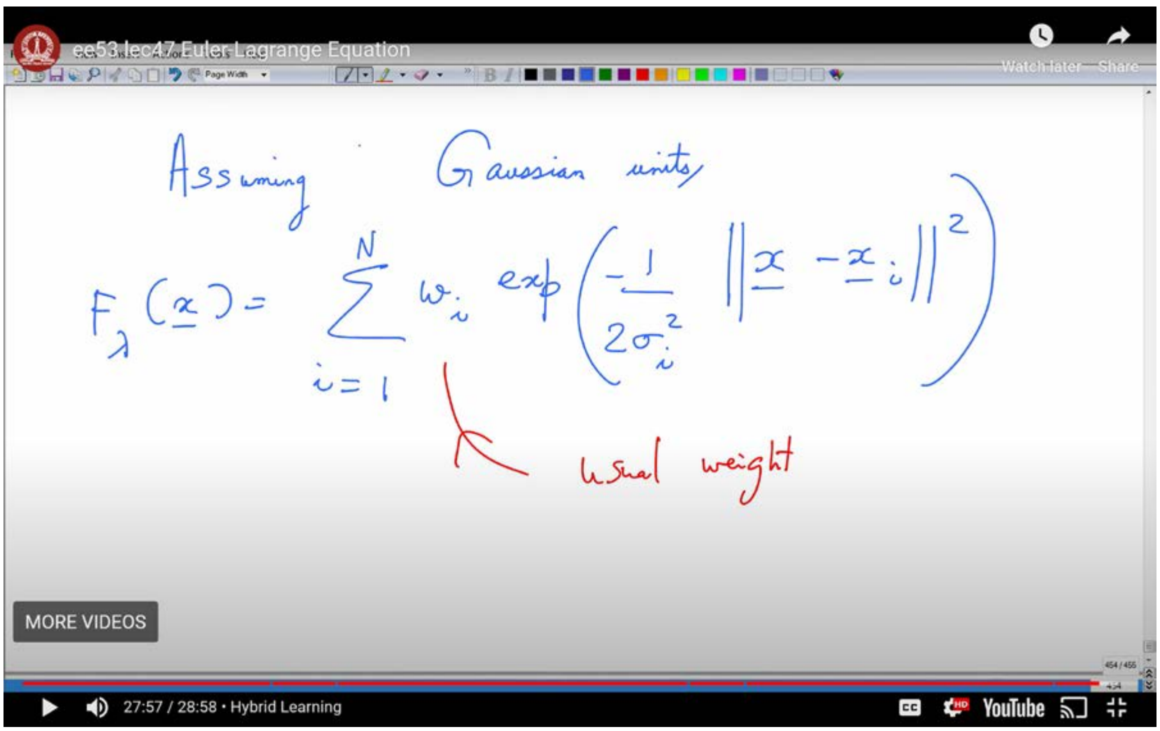



(Refer Slide Time: 26:48)



Now, let us examine what happens in the context of Radial Basis Functions (RBF). Specifically, we may choose Green's functions to be translationally and rotationally invariant. This can be achieved if we set the Green's function g to be a function of the norm $|x - x_i|$ rather than a general $g(x, x_i)$. In this case, $f_\lambda(x)$ can be expressed as a linear combination of these special Green's functions, where g is a function of $|x - x_i|$. This represents a linear function space dependent on the data, which is crucial because the weights $w_i$ depend on $x_i$ and $d_i$.

(Refer Slide Time: 27:57)




In the Radial Basis Function network, we use Gaussian units. Here, the Green's function is defined as:

$$g(x, x_i) = \exp\left(-\frac{1}{2}\frac{|x - x_i|^2}{\sigma_i^2}\right)$$

where $x_i$ is the data point and $\sigma_i$ is a spread parameter. In our hybrid learning algorithm, we partition the space into different clusters and examine the variance within each cluster. To avoid overly peaked or overly flat distributions, we use a uniform $\sigma^2$ for all radial basis function units, replacing $\sigma_i^2$ with $\sigma^2$ in the equation. This provides the form for our RBF response.

This approach leads us to the idea of using Gaussian units and weights, which we have explored in the RBF network. However, if we generalize beyond Gaussian units and treat the Green's function in a broader context, we obtain a general network composed of various Green's functions. This framework forms the basis for Green's function-based regularization networks and learning networks, which we will study further in this module.