Introduction to Deep Ritz Method

Yaohua Zang

June 23, 2025

0.1 Background: The Variational Method

The DeepRitz method for solving PDEs is built upon variational principles from classical analysis. In this section, we review the key mathematical foundations—especially how solving certain PDEs can be recast as an optimization problem over a function space.

0.1.1 From Equations to Minimization Problems

Suppose we want to solve a system of equations:

$$F(u) = 0, \quad F : \mathbb{R}^n \to \mathbb{R}^n,$$

where $u \in \mathbb{R}^n$ is the unknown. If we can find a scalar-valued function $J : \mathbb{R}^n \to \mathbb{R}$ such that:

$$F = \nabla J,$$

and if J has a minimizer u, i.e.,

$$J(u) = \min_{v \in \mathbb{R}^n} J(v),$$

then this minimizer satisfies $F(u) = \nabla J(u) = 0$.

Conversely, if we aim to minimize a functional J, then any point u satisfying $\nabla J(u) = 0$ is a critical point. If J is convex, such a point is guaranteed to be a global minimizer. Otherwise, further analysis is needed (e.g., checking second-order conditions).

This optimization viewpoint naturally extends to PDEs through the **calculus of variations**, where we seek to minimize functionals defined over infinite-dimensional function spaces.

0.1.2 Dirichlet's Principle and the Poisson Equation

Consider the classical Dirichlet problem for the Poisson equation:

$$\begin{cases} -\Delta u = f, & x \in \Omega \\ u = g, & x \in \partial\Omega \end{cases} \tag{1.1}$$

We define an energy functional ($|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^n$):

$$J(v) = \int_\Omega \left(\frac{1}{2} |\nabla v(x)|^2 - f(x) v(x) \right) dx.$$

We also define a set of admissible functions:

$$K = \{ v \in C^2(\bar{\Omega}) \mid v(x) = g(x), \forall x \in \partial\Omega \}.$$

Dirichlet's principle states: If $u \in K$ minimizes $J(v)$ over all $v \in K$, then u is the solution to the Poisson equation. The reverse is also true: if u solves the Poisson equation, then it minimizes J.

To see this, consider a small perturbation of u in the direction of a smooth function $\varphi \in C_0^\infty(\Omega)$, which vanishes on the boundary. Define:

$$\tilde{J}(\lambda) = J(u + \lambda\varphi).$$

A Taylor expansion yields:

$$\begin{aligned}\tilde{J}(\lambda) &= \int_\Omega \left(\frac{1}{2}|\nabla(u+\lambda\varphi)(x)|^2 - f(x)(u+\lambda\varphi)(x)\right) dx \\ &= J(u) + \lambda \int_\Omega \langle \nabla u(x), \nabla\varphi(x)\rangle - f(x)\varphi(x)\, dx + \lambda^2 \int_\Omega \frac{1}{2}|\nabla\varphi(x)|^2\, dx \qquad (1) \\ &= J(u) + \lambda \int_\Omega \langle \nabla u, \nabla\varphi\rangle - f\varphi\, dx + \mathcal{O}(\lambda^2).\end{aligned}$$

Since u minimizes J, the derivative at $\lambda = 0$ must vanish:

$$0 = \tilde{J}'(0) = \int_\Omega \left(\langle \nabla u, \nabla\varphi\rangle - f\varphi\right)\, dx.$$

Using integration by parts and the fact that $\varphi = 0$ on $\partial\Omega$, we get:

$$\int_\Omega \left(-\Delta u - f\right)\varphi\, dx = 0,$$

for all $\varphi \in C_0^\infty(\Omega)$. This implies (in the weak sense):

$$-\Delta u = f \quad \text{in } \Omega.$$

Conversely, suppose u is a classical solution of the Poisson equation. For any $v \in K$, we can show through energy comparisons that:

$$J(u) \leq J(v),$$

so u minimizes J. This leads to the following equivalence, known as Dirichlet's principle:

$$u \text{ solves the Poisson equation} \iff u \text{ minimizes } J \text{ on } K.$$

Here, the PDE:

$$-\Delta u = f$$

is the **Euler-Lagrange equation** associated with the variational problem.

Remark: Let $v \in K$ be arbitrary. Then

$$\int_\Omega \left(-\Delta u(x) - f(x)\right)\left(u(x) - v(x)\right) dx. \qquad (2)$$

Since $u - v = 0$ on $\partial\Omega$, partial integration gives

$$\int_\Omega \langle \nabla u(x), \nabla u(x) - \nabla v(x)\rangle\, dx - \int_\Omega f(x)(u(x) - v(x))\, dx. \qquad (3)$$

It follows that (we drop the argument x of the functions)

$$\int_\Omega |\nabla u|^2 - fu\,dx = \int_\Omega \langle \nabla u, \nabla v\rangle - fv\,dx \le \int_\Omega \frac{1}{2}(|\nabla u|^2 + |\nabla v|^2) - fv\,dx, \tag{4}$$

since for arbitrary vectors $y, z \in \mathbb{R}^n$, we have

$$|\langle y, z\rangle| \le |y| \cdot |z| \le \frac{1}{2}(|y|^2 + |z|^2). \tag{5}$$

Subtracting $\frac{1}{2}\int_\Omega |\nabla u|^2\,dx$ on both sides, we have

$$J(u) \le J(v), \tag{6}$$

so u is a minimizer of J on K.

0.1.3 The Variational Problem and Energy Minimization

To summarize, solving the Poisson equation:

$$\begin{cases} -\Delta u = f, & x \in \Omega \\ u = g, & x \in \partial\Omega \end{cases} \tag{1.2}$$

is equivalent to solving the **variational problem**:

$$\min_{u\in\mathcal{U}} I(u) = \int_\Omega \left(\frac{1}{2}|\nabla u(x)|^2 - f(x)u(x)\right) dx, \tag{1.3}$$

where $\mathcal{U}$ denotes a set of admissible functions satisfying the boundary condition $u = g$ on $\partial\Omega$. This energy functional $I(u)$ is often called the **energy or loss functional**. Its minimization forms the core idea of the DeepRitz method.

Important note: Dirichlet's principle provides equivalence but not necessarily the existence of a solution. Additional functional analysis tools (e.g., Sobolev spaces, coercivity, and compactness) are needed to establish existence and uniqueness in general.

0.2 The Deep Ritz Method (DeepRitz)

The DeepRitz Method (DRM) is a deep learning-based approach for solving partial differential equations (PDEs) by leveraging **variational formulations** rather than the strong (differential) form of the equations, as used in Physics-Informed Neural Networks (PINNs). The DeepRitz method frames the solution of a PDE as an energy minimization problem and uses deep neural networks (DNNs) to approximate the solution.

0.2.1 The Basic Idea

Recall the variational formulation of the Poisson equation from Section 0.1:

$$\min_{u\in\mathcal{U}} I(u) = \int_\Omega \left(\frac{1}{2}|\nabla u(x)|^2 - f(x)u(x)\right) dx. \tag{2.1}$$

Instead of solving the differential equation directly, the DeepRitz method minimizes the energy functional $I(u)$ over a class of neural network functions u_θ, where θ represents the network parameters. Thus, solving the PDE becomes an optimization problem:

$$\min_\theta I(u_\theta),$$

subject to the boundary condition $u_\theta(x) = g(x)$ for $x \in \partial\Omega$.

0.2.2 The Procedure of DeepRitz

The DeepRitz method can be implemented in three main steps:

Step 1: Approximating the Solution with a Neural Network

We approximate the unknown solution $u(x)$ by a feedforward neural network $u_\theta(x)$. The network is composed of multiple layers:

$$u_\theta(x) = (\mathcal{N}^L \circ \mathcal{N}^{L-1} \circ \cdots \circ \mathcal{N}^0)(x),$$

where $\mathcal{N}^l$ denotes the operation of the l-th layer, and θ includes all the weights and biases of the network.

Step 2: Formulating the Loss Function

The total loss function $L(\theta)$ in the DeepRitz method consists of two components:

- **Energy Loss (PDE Term)**: Encodes the physical laws via the energy functional.
- **Boundary Loss**: Enforces the boundary condition $u = g$ on $\partial\Omega$.

(a) Boundary Condition Loss Let $\mathcal{D}_{bd} = \{x_i, g(x_i)\}_{i=1}^{N_{bd}}$ be a dataset of sampled points $x_i \in \partial\Omega$, where $g(x_i)$ gives the prescribed boundary values.

The loss that enforces the boundary condition is:

$$L_{bd}(\theta) = \frac{1}{N_{bd}} \sum_{i=1}^{N_{bd}} |u_\theta(x_i) - g(x_i)|^2.$$

(b) Energy Loss (Variational/PDE Loss) We approximate the energy functional $I[u_\theta]$ using numerical integration. Suppose we sample $\{x_i, w_i\}_{i=1}^{N_{\text{int}}}$ from Ω, where w_i are integration weights (e.g., from Monte Carlo or quadrature methods). Then:

$$L_{PDE}(\theta) = I[u_\theta] \approx \sum_{i=1}^{N_{\text{int}}} w_i \left(\frac{1}{2}|\nabla u_\theta(x_i)|^2 - f(x_i)u_\theta(x_i)\right).$$

(c) Total Loss Function We combine the two losses using weighting parameters ω_{PDE} and ω_{bd} to balance their contributions:

$$L(\theta) = \omega_{\text{PDE}} \cdot L_{PDE}(\theta) + \omega_{\text{bd}} \cdot L_{bd}(\theta).$$

These weights can be tuned depending on the problem setup, especially to ensure that the boundary condition is sufficiently enforced.

Step 3: Training the Neural Network

Once the loss function is defined, we train the neural network using standard optimization algorithms such as stochastic gradient descent (SGD) or Adam. The update rule for parameter θ at iteration t is:

$$\theta_{t+1} = \theta_t - l_r \nabla_\theta L(\theta_t),$$

where l_r is the learning rate.

Remark:

- In practice, training is done over multiple epochs, where each epoch uses batches of sample points for integration and boundary evaluation.
- In the original DeepRitz paper, the authors propose using a special activation function:

  $$\sigma(x) = \max\{x^3, 0\},$$

  This cubic ReLU-like activation increases the smoothness of the approximated solution and better supports the variational framework, where higher-order derivatives may appear in the Euler-Lagrange equations.

0.3 The Pros and Cons of the DeepRitz Method

0.3.1 Advantages

1. Efficient Training and Faster Convergence Since the DeepRitz method minimizes a scalar-valued energy functional, it avoids the need to enforce PDE residuals at every point in the domain (as in PINNs). This typically results in a **simpler optimization landscape** and can lead to **faster convergence** during training. Moreover, the energy functional often provides a more global measure of error compared to pointwise residual loss, making the optimization more stable.

2. Supports Weak Solutions One of the most significant advantages of the DeepRitz method is that it naturally accommodates **weak solutions**. In many physical problems, the true solution of a PDE may not be differentiable in the classical sense (especially near boundaries or singularities). Since the variational form only requires the solution to be in a **Sobolev space** (i.e., the function and its weak derivatives are square-integrable), DeepRitz is capable of approximating these weak solutions, which the strong-form-based methods may fail to capture accurately.

0.3.2 Disadvantages

1. Requires a Large Number of Integration Points To accurately evaluate the energy functional $I[u_\theta]$, especially in high-dimensional spaces, **a large number of integration points** is typically needed. This can be computationally expensive. While Monte Carlo integration helps mitigate the curse of dimensionality to some extent, the variance in estimation may slow down convergence, especially for complex geometries or sharp solution features.

2. Only Applicable to PDEs with Variational (Energy) Formulations Perhaps the biggest limitation is that DeepRitz can **only be applied to PDEs that admit a variational (energy) formulation**. While many classical PDEs in physics do arise from energy minimization principles (e.g., Poisson, elasticity, and some forms of the Schrödinger equation), this does **not** include all PDEs. For example, general **time-dependent** PDEs or **hyperbolic equations** like the wave equation do not typically come with an associated energy functional suitable for this approach. This restricts the scope of applicability of the DeepRitz method compared to PINNs, which can handle a broader class of equations via their strong form.

0.3.3 When to Use DeepRitz?

The DeepRitz method is especially suitable when:

- The PDE has a known variational (energy) form.
- The solution may be non-smooth, requiring weak solution concepts.
- You want faster and more stable convergence during training.

However, it is less suitable for:

- PDEs that lack an energy functional.
- High-dimensional problems with limited computational resources (due to expensive numerical integration).