



HAL
open science

Weighted residual solution methods

Mathias Legrand, Stéphane Junca

► **To cite this version:**

| Mathias Legrand, Stéphane Junca. Weighted residual solution methods. 2025. <hal-05155148>

HAL Id: hal-05155148

<https://hal.science/hal-05155148v1>

Preprint submitted on 9 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Weighted residual solution methods

Mathias Legrand¹ and Stéphane Junca²

¹Structural Dynamics and Vibration Laboratory, Department of Mechanical Engineering,
McGill University, Montréal, Canada

²Department of Mathematics, LJAD, Inria & CNRS, Université Côte d'Azur, Nice, France

Abstract Nonlinear dynamic systems are rarely solvable in closed form, requiring robust numerical methods. This chapter focuses on the Method of Weighted Residuals (MWR), a general approach for approximating solutions to differential equations. Through simple examples, the chapter shows how selecting appropriate trial and test functions leads to schemes capable of handling a variety of challenging formulations. It also explains why weak solutions should be considered with care. The chapter highlights modern trends, such as integrating MWR with machine learning which blend data-driven modeling with physical constraints.

1	History	2
1.1	Variational principles	2
1.2	Theory of distributions	5
1.3	Modern developments	5
2	General formulation	6
2.1	Duality	6
2.2	Weak formulation	9
2.3	Discretization: test and trial functions	9
3	Strengths and weaknesses	11
3.1	Flexibility	11
3.2	Loss of uniqueness	12
3.3	Weak limits, nonlinearity & numerical schemes	13
4	Convergence analysis of weighted residual schemes	17
4.1	Discretization of derivatives	17
4.1.1	Finite differences scheme	17
4.1.2	Weighted residual scheme	18
4.1.3	Correspondence between schemes	19
4.2	General discussion	19
5	Artificial Intelligence: future of weighted residuals?	20
5.1	AI-powered extensions	20
5.2	Advantages and open questions	22
5.3	AI-integrated WRMs for Nonlinear Dynamics	22
6	Conclusion	22
	References	23

Notations

Partial derivatives are denoted using subscripts with the name, as opposed to the position, of the argument versus which differentiation is conducted. Given a function $w(x, y)$, we have $\frac{\partial w}{\partial y} \equiv \partial_y w \equiv \partial_2 w$. Ordinary derivatives are denoted with $'$ unless stated otherwise.

Acronyms

AI Artificial Intelligence	HBM Harmonic Balance Method
BC Boundary Conditions	MWR Method of Weighted Residuals
BV Bounded Variation	NN Neural Network
CG Continuous Galerkin	ODE Ordinary Differential Equation
DAE Differential Algebraic Equation	PDAE Partial Differential Algebraic Equation
DE Differential Equation	PDE Partial Differential Equation
DG Discontinuous Galerkin	PINN Physics-Informed Neural Network
FEM Finite Element Method	SFEM Spectral Finite Element Method
FVM Finite Volume Method	WAN Weak Adversarial Network

1 History

1.1 Variational principles

The exact history of weighted residual methods is challenging to track back, and while not explicitly formulated at that time, weighted residual techniques were probably glimpsed in the work of Euler (1707–1783) on variational principles in the 18th century. Along with Joseph-Louis Lagrange (1736–1813), he developed the calculus of variations that can be stated as follows: Given a functional, mapping functions to reals, find the function that makes the functional take an extremum value. Such a function is said to be a *stationary point* and stationarity does not determine whether the function is a maximum, a minimum or a saddle point. More specifically, given the function $L(t, y, y')$ ¹, the problem is to find a function $y(t)$ that makes stationary the functional (Bérest 1997)

$$I[y] = \int_a^b L(t, y, y') dt \quad (1)$$

where L depends on t , the independent variable, along with $y(t)$, and its derivative with respect to $y'(t)$. Well-posedness is guaranteed once appropriate boundary conditions acting on $y(a)$ and $y(b)$ are satisfied. From variational principles, the celebrated Euler-Lagrange equation, which provides a necessary condition for y to be a stationary point reads

$$\partial_y L - \frac{d}{dt}(\partial_{y'} L) = 0. \quad (2)$$

Previous derivations are limited to a one-dimensional framework. However, variational principles are known to extend to a vast variety of multi-dimensional problems and provide simple, elegant and powerful methodologies for finding solutions to formulations enjoying the existence of a functional. In the context of optimization, they are known to provide deep conceptual insights into the nature of physical (conservative) laws, suggesting that the behavior of systems can be understood in terms of optimization principles, rather than solely through cause-and-effect dynamics (Finlayson 1972).

Equation (2) commonly takes the form of a nonlinear [Ordinary Differential Equation \(ODE\)](#) in the unknown function $y(t)$ complemented with appropriate boundary conditions. In most cases, this [ODE](#) cannot be solved exactly and we should resort to (numerical) approximations. It is also important to note that the derivation of the Euler–Lagrange equation involves essential integrations by parts.

¹The quantity L should be seen as a classical function in three arguments, that is $L(t, y, z)$.

Example 1 (Variational calculus: Shorted path in the plane (Bérest 1997)). *What is the curve joining two points A and B in the plane whose length is the minimum? If the curve is represented by its Cartesian equation, the problem reduces to finding the minimum of the function*

$$I[y] = \int_a^b \sqrt{1 + y'^2(t)} dt \quad (3)$$

such as $y(a) = \alpha$ and $y(b) = \beta$. In this example, it can be shown that the stationary point is a minimizer of the functional. From $L(t, y, y') = \sqrt{1 + y'^2}$, the corresponding Euler-Lagrange equation reads

$$\frac{d}{dt} \left(\frac{y'}{\sqrt{1 + y'^2}} \right) = 0 \quad \Rightarrow \quad y'' = 0 \quad (4)$$

along with $y(a) = \alpha$ and $y(b) = \beta$, and the segment joining points A and B .

The Method of Weighted Residuals, which is the focus of the present chapter, proceeds in the opposite direction by starting with a differential equation, complemented by boundary and/or initial conditions, collectively transformed into a formulation involving integrals, although not necessarily with the aim of finding stationary points (Finlayson 1972; Fletcher 1998). These methods have expanded over time to a point where it is fair to state that variational principles form a subclass of the more general weighted residual methods. We also note that the terms “weak formulation” and “variational formulation” are often used interchangeably in the literature. Both refer to rewriting a [Differential Equation \(DE\)](#) in a form that weakens the derivative requirements and introduces test functions. However, there is a subtle distinction: the weak formulation focuses on satisfying the [DE](#) in an integral sense, even for less regular solutions. The variational formulation, on the other hand, specifically arises from minimizing or finding stationary points of a functional. While all variational formulations are weak, not all weak formulations are variational.

Starting from [Equation \(2\)](#) in the case of [ODE](#), the procedure involves premultiplying the differential equation by a *test function* denoted $w(t)$, followed by applying integration by parts (potentially utilizing Green’s identities in a multidimensional context):

$$\forall w, \quad \int_a^b w (\partial_y L - \frac{d}{dt}(\partial_{y'} L)) dt = \int_a^b (w \partial_y L + w' \partial_{y'} L) dt + w \partial_{y'} L \Big|_a^b = 0. \quad (5)$$

This process ultimately ensures that the original equation holds in an integral sense. Depending on the mathematical meaning assigned to the function w , sometimes denoted δy , with appropriate [Boundary Conditions \(BC\)](#), [Equation \(5\)](#) can be read as the variation of the Lagrangian L , that is

$$\int_a^b (\partial_y L \delta y + \partial_{y'} L \delta y') dt = \int_a^b \delta L(x, y, y') dt = \delta \left(\int_a^b L(x, y, y') dt \right) \quad (6)$$

where δ is understood as the *variation* operator in the sense of Calculus of Variations defined by Euler. When the unknown y of the formulation is a displacement, then δy is commonly called a *virtual* displacement. In the context of Weighted Residuals, the test function is rather called a *weighting function* and is just an arbitrary function which should satisfy some appropriate functional features discussed later.

The [Method of Weighted Residuals \(MWR\)](#) follows a similar approach to the procedure outlined in [Equations \(5\) and \(6\)](#), aiming to transform a [DE](#) into an integral formulation. It serves as the abstract foundation for various numerical techniques, including the finite difference method, the [Finite Element Method \(FEM\)](#), or spectral methods (Chen and Ney 2009); even [Finite Volume Method \(FVM\)](#) and [Discontinuous Galerkin \(DG\)](#) methods can be recovered using arguments from the Weighted Residual formalism. The key differences between the above numerical strategies lie in: (i) the application of integration by parts, (ii) the approach used to approximate the weighting (or test) function, and (iii) the method chosen to approximate the unknown function, also referred to as the *trial* function. In residual-based methods, the test function serves as a tool to assess how well an approximate solution satisfies a [DE](#). The fundamental idea behind using test functions is to “weight” the residual—the discrepancy introduced by the approximation—and making it vanish in an integral (or average) sense across the domain.

The trial function (or *approximate solution*) is a function used to approximate the exact yet often inaccessible solution of a [DE](#). In approaches such as the [MWR](#), the [FEM](#), and the [Continuous Galerkin \(CG\)](#) Method, trial functions are typically expressed as a finite sum of basis functions, with their contributions to the unknown solution determined by parameters that need to be computed.

The 20th century saw the emergence of more refined definitions of [MWR](#). The Rayleigh-Ritz method was developed in the 1910’s to numerically approximate the eigenvalues in the context of finite-dimensional Boundary Value Problems via the minimization of a quadratic function representing the system energy. The Rayleigh-Ritz method, developed separately by Rayleigh (Rayleigh 1877) and Ritz (Ritz 1909) can be seen as an early form of the [MWR](#). Similarly, one of the most well-known approximate methods was introduced by the Russian engineer B. Galerkin (Galerkin 1915). In this approach, both the test and trial functions are constructed using the same functional basis, which, given a sufficient number of terms, is assumed to provide an accurate approximation of the exact solution. Unlike the Rayleigh-Ritz method, which is rooted in modern optimization and seeks to minimize the residual in an integral sense, the [CG](#) method instead enforces orthogonality of the residual with respect to each function in a complete set. This characteristic places it within the broader class of *projection methods*. It was used for solution to problems arising in the stability of beams, rods, and plates but the term “Method of Weighted Residuals” emerged later and is attributed to S. Crandall (Crandall 1956). The methodology became a cornerstone for the development of [FEM](#) pioneered by Courant, Argyris and Hrennikoff in the 1960s’. [FEM](#) provided a systematic approach to discretizing complex geometries supporting [Partial Differential Equation \(PDE\)](#) stemming from engineering and physics. During the 1970s, [MWR](#) were extensively developed and applied to a wide range of problems in fluid dynamics, structural analysis, heat transfer, and more. The [CG](#) method and [FEM](#) are now standard tools in engineering analysis.

In the field of Mechanics and Elasticity, more advanced formulations of the [MWR](#) have been developed, including the three-field Hu-Washizu principle, which incorporates displacement, strain, and stress, and the two-field Reissner principle (Zienkiewicz et al. 2005), which considers displacement and stress. The main idea is to consider the [MWR](#) on independent governing equations, classically, (i) equilibrium condition in the form of Newton’s second law, (ii) the strain-displacement relationship and (iii) the constitutive equation where stress is expressed as a function of strain, rather than attempting to substitute (ii) and (iii) into (i). The Hu-Washizu principle was instrumental in the development of mixed finite element methods, as it provides greater flexibility in constructing approximate models

for beams and plates. Notably, it allows the displacement and its associated velocity to vary differently across the thickness of a structure, leading to improved accuracy (Barr 1966; Sun 2022).

1.2 Theory of distributions

The modern mathematical framework for formulating and analyzing the MWR is based on the Theory of Distributions, introduced by L. Schwartz (Schwartz 1997). This framework extends classical calculus by allowing the treatment of equations involving generalized functions, which go beyond traditional smooth functions. It provides a rigorous foundation for handling discontinuities, Dirac delta functions, measures, concentrated forces on small subsets, and even fractal structures.

By leveraging this theoretical foundation, the MWR can be effectively applied to a broad range of engineering and physics problems, including those involving singularities, shocks, boundary layers, and discontinuous solutions. The ability to approximate distributions with structured function sets also enhances the stability and accuracy of numerical methods, making this framework an indispensable tool in modern computational mechanics. Distributions are not functions in the traditional sense but are useful in many applications involving DE. A distribution is a linear functional that acts on a space of test functions. The most common space of test functions is the set of infinitely differentiable functions with compact support, denoted by $\mathcal{D}(\Omega)$: if T is a distribution and φ is a test function, the action of T on φ is denoted by the *duality* pairing $\langle T, \varphi \rangle$ which generalizes the classical inner product between functions.

Example 2 (Dirac’s distribution). *For any function $\varphi \in \mathcal{D}(\Omega)$, the Dirac delta δ is defined by $\langle \delta, \varphi \rangle = \varphi(0)$. This means that when δ acts on φ , it simply picks its value at $x = 0$. In other words, the Dirac distribution is concentrated at $x = 0$.*

The theory extends the concept of differentiation to objects that may not be smooth or even continuous. Since distributions are defined in terms of their action on test functions and their derivatives follow the Integration by Parts identity, as already used in Equation (5), which is fundamental in functional analysis: the distributional (or weak) derivative T' of a distribution T is defined by $\langle T', \varphi \rangle = -\langle T, \varphi' \rangle$ which should hold true for all smooth test functions φ . This definition ensures that differentiation of distributions coincides with classical differentiation for smooth functions.

The MWR is closely related to the Theory of Distributions, and associated weak formulations, because it does not require the differential equation of interest to be satisfied exactly at every point. Instead, it ensures that the weighted equation holds in an integral sense where derivatives are possibly considered in the weak sense which reduces the differentiability requirements of the solution by transferring derivatives to the test function.

1.3 Modern developments

The 1980s and 1990s marked significant progress in the development of adaptive finite element methods, where both the mesh and the basis functions are dynamically refined based on error estimates to enhance solution accuracy. These advances were made possible by increasing computational power and the emergence of robust finite element software, which enabled the practical implementation of mesh adaptivity and error control strategies.

In recent years, the most advanced evolution of the **MWR** can be found in **DG** methods, which incorporate adaptivity, local remeshing, and high-order accuracy. **DG** methods are a class of finite element methods derived from a weak (or variational) formulation of the governing equations, but unlike traditional **CG** methods, they allow for discontinuities across element interfaces. This non-conformity makes **DG** methods particularly well-suited for handling complex geometries, sharp gradients, and conservation laws. In contrast, **CG** methods are conforming in the sense that the discrete solution space where the approximation is built is a subspace of the continuous function space in which the exact solution resides.

Recently, **MWR** have been extended to handle multiphysics problems, where different physical phenomena are coupled together. Techniques such as isogeometric analysis, spectral methods, and meshless methods have also been developed to address specific challenges in numerical approximation (Cottrell et al. 2009; Canuto et al. 2006; Liu and Gu 2005). If instead of making the size of the mesh, denoted h , smaller, one increases the degree p of the polynomials used in the basis function, one has a p -method. If one combines these two refinement types, one obtains an hp -**FEM**, where the polynomial degree can even vary from element to element. High-order methods with large uniform p are called **Spectral Finite Element Method (SFEM)**, not to be confused with spectral methods.

2 General formulation

The **MWR** is a technique for obtaining approximate solutions to differential equations when exact solutions are intractable. The core idea is that the residual, which is the function obtained by substituting the approximate solution into the original **DE**, should be orthogonal to a chosen set of test functions. In order to keep the derivations simple, the domains considered in the remainder are finite intervals $\Omega = [a, b] \subset \mathbb{R}$ and the independent variable is time t because dynamics is the focus of the present chapter. The unknown of the problem is assumed to be a scalar function $u(t)$.

2.1 Duality

Is considered an operator \mathcal{N} acting on the unknown function u :

$$\mathcal{N}[u(t)] = f(t) \quad \text{on } t \in]a, b[\tag{7}$$

where f is a specified function. The order of the differential operator plays a crucial role in determining the required regularity of the corresponding test functions. Typically, operators involving higher-order derivatives require test functions with correspondingly higher regularity.

Example 3 (Duffing oscillator). *We could have $\mathcal{N}[u] \equiv u'' + u + u^3$ and $f(t) = \cos t$.*

Equation (7) is commonly supplemented with additional conditions restricting the space of solutions and possibly guaranteeing uniqueness of solutions. The **MWR** relies on the notion of duality through the following classical result, adapted to our needs.

Lemma 1 (Fundamental Lemma). *Consider an open domain $\Omega \subset \mathbb{R}$ and suppose $g \in \mathcal{C}^0(\Omega)$. Then,*

$$\forall w \in \mathcal{C}_c^\infty(\Omega), \quad \int_{\Omega} w(t)g(t) dt = 0 \quad \implies \quad g = 0 \text{ on } \Omega \tag{8}$$

where $C_c^\infty(\Omega)$ is the set of infinitely differentiable functions with compact support.

The lemma essentially states that if the integral of the product of a function g and any arbitrary test function w is zero, then g itself must be zero on the interval. Another interpretation is to consider an approximation of the unity function given by $w_\alpha(t) = w_0(t - \alpha)$ where $w_0(t) = \frac{1}{2\Delta t}\chi_{[-\Delta t, \Delta t]}(t)$, with $\chi_{[-\Delta t, \Delta t]}(t)$ denoting the characteristic function of the interval $[-\Delta t, \Delta t]$, and $\Delta t > 0$. For this *weight* function w_α , the integral in Equation (8) represents the mean value of the function g around α . As $\Delta t \rightarrow 0$, this integral converges to $g(\alpha)$, recovering the pointwise value of the function. However, relying solely on the pointwise value $g(t)$ for each t is not always the most meaningful perspective. Mean values over a region are often more relevant than isolated point evaluations. Moreover, if g is discontinuous at a given point, its value at that point may be ill-defined or meaningless. For this reason, the integral formulation of Equation (8) provides a more general and robust way to express that $g = 0$ on Ω compared to the pointwise condition $\forall t \in \Omega, g(t) = 0$.

This integral perspective extends naturally to various function spaces and interpretations: the L^2 integral (Hilbert space interpretation), the L^1 integral (Lebesgue integration), the L^1_{loc} integral, meaning local integrability on any bounded subset (though possibly not over the entire real line), or cases where the integral is not a classical integral but rather a measure-theoretic interpretation. For instance, if g is a measure, we can still interpret it as an integral related to a measure, denoted as $g = \mu$, meaning that instead of a function in the classical sense, g is understood in terms of measure theory and integration, that is

$$\int_{\Omega} wg \, dt = \int_{\Omega} w\mu \, dt = \int_{\Omega} w \, d\mu. \quad (9)$$

This notation may seem somewhat informal from a strict mathematical standpoint (except for the last integral), but it is well understood within the scientific and engineering communities. Additionally, we can mention oscillatory integrals (such as Airy integrals) and divergent integrals, which fall under alternative integration theories beyond the classical Riemann or Lebesgue frameworks. More significantly, what appears to be an integral may, in some cases, not be an integral at all. For instance, if g represents the derivative of a continuous but nowhere differentiable function, the integral interpretation may break down. Such cases arise in nonsmooth dynamics, particularly when a system makes contact with a rough surface. In practical applications, these pseudo-integral expressions can often be transformed into true integrals after some mathematical operations, such as integration by parts. Additionally, the function w can be restricted to a discrete space, allowing numerical evaluation of the integrals.

As already noted, a powerful and flexible mathematical framework that encapsulates the theory of weighted residuals is the theory of distributions which provides rigorous mathematical foundations while remaining highly adaptable for engineering applications. It also offers a mathematically sound yet practical approach for validating computations in applied sciences and engineering. Schwartz's perspective is to interpret a function (or more generally, a measure) as a linear functional acting on test functions, which correspond to weight functions in the MWR. Specifically, let T_g be the linear functional defined by

$$T_g(w) = \int_{\Omega} wg \, dt. \quad (10)$$

By definition, $T_g = 0$ means that the functional is identically zero if, for all weight functions w , we have $T_g(w) = 0$. If the function g is continuous (or even just locally integrable), this

condition implies that $g = 0$, through Lemma 1. For a more general functional T , which does not necessarily correspond to a well-behaved function g (e.g., when g is not in L^1_{loc}), Schwartz instead introduced the duality bracket notation $T(w) = \langle T, w \rangle$ since it does not always represent a classical integral.

In the theory of distributions, functionals are continuous, meaning they remain well-defined under the limit of any convergent sequence of weight functions. This property is particularly important as it allows us to restrict the set of weight functions to a countable subset without losing generality. In practical applications, we do not verify the equality in Equation (8) for all weight functions w , but rather for a carefully chosen subset that provides sufficient information. Furthermore, the choice of weight functions must align with the regularity and properties of the distribution. The more singular the distribution, the smoother the test functions need to be. This interplay between the nature of the distribution and the required regularity of the test functions is a crucial aspect of the formulation.

Definition 1 (Distribution: Fundamental Lemma of Variational Calculus). Consider an open domain $\Omega \subset \mathbb{R}$ and suppose $f \in \mathcal{D}'(\Omega)$. Then,

$$\forall w \in \mathcal{D}(\Omega), \quad \langle w, g \rangle = 0 \quad \implies \quad g = 0 \quad \text{in } \mathcal{D}'(\Omega) \quad (11)$$

where $\langle \cdot, \cdot \rangle$ indicates the duality pairing product on $\mathcal{D}(\Omega) \times \mathcal{D}'(\Omega)$.

When g is more regular, the weights can be less regular, by the duality principle. For instance, if g is at least continuous, the weights can be Dirac measures, leading to collocation methods, briefly described below.

Definition 1 is then used on the quantity $\mathcal{R}[u] \equiv \mathcal{N}[u] - f$ called the residual (of the governing equation). An obvious aim when solving Equation (7) is to enforce $\mathcal{R}[u] = 0$ identically on Ω . However, this is impossible unless the exact solution to the equation is known. Instead, the aim is to approximately solve the formulation through the following statement: Find the function u such that

$$\langle w, \mathcal{R}[u] \rangle = 0, \quad \forall w \in \mathcal{D}(\Omega) \quad (12)$$

It is clear that exactly solving Equation (12) would imply that Equation (7) is also exactly solved. However, this is generally impossible but Equation (12) is more conducive to constructing approximate solutions, primarily because the solution space is broader.

The requirement that the residual be orthogonal to the weight functions can be interpreted through duality where the space of residuals and the space of weight functions are dual to each other. This duality ensures that the residual vanishes, distributing the error according to the chosen weight functions.

The main attractive aspect of duality is the flexibility given in the definition of \mathcal{D} , which essentially reduces to the differentiability of its elements, and \mathcal{D}' . It is immediately clear that in the MWR, careful attention must be given to the selection of the elements of the two spaces in duality. For instance, if a Dirac δ arises in the residual (through f in Equation (7) for instance), then it would not be permitted to select δ distributions in the test space. Moreover, the notion of duality here does not reduce to the classical spaces \mathcal{D} and \mathcal{D}' from distribution theory. Instead, the dual spaces—denoted \mathcal{H} and \mathcal{H}' in what follows—are properly defined in accordance with the DE under consideration, typically with respect to its highest-order derivative.

2.2 Weak formulation

Integration by parts is a fundamental tool for reformulating a DE as a real integral equation, such as in Equation (8). Problems with boundary conditions can be incorporated into this formalism through a suitable choice of weights, specifically by exploiting the properties of the weights on the boundary.

Integration by parts is often used in the MWR with the aim to transform the residual to make it more amenable for approximation and easier to evaluate, often reducing the order of derivatives and allowing the application of boundary conditions more naturally. The strong form $R[u] = 0$ is converted into a weak form which involves lower derivatives of the unknown function. This process is not systematically implemented in weighted residual methods but is critical in FEM and other variational methods.

Example 4 (Reduction of derivatives). *We consider the differential operator $\mathcal{N}[u] \equiv -u''$ and the corresponding residual $R[u] \equiv -u'' - f$. In the sense of distributions, the weighted residual reads $\langle w, R[u] \rangle = -\langle w, u'' + f \rangle$ or $\langle w', u' \rangle + \langle w, f \rangle = 0$, by definition of the distributional derivative $\langle w, u'' \rangle = -\langle w', u' \rangle$.*

The distributional derivative imitates the classical integration by parts, and transforms the original differential equation into a form that (1) lowers the order of the differential operator, (2) facilitates the application of boundary conditions when appropriate, and (3) provides a potentially symmetric weak form that is more amenable to approximation techniques. The weak formulation of the problem thus places fewer restrictions on the smoothness of the trial functions. This approach is especially advantageous when dealing with Neumann and Robin conditions, which are automatically incorporated in the residual formulation without direct specification.

2.3 Discretization: test and trial functions

Once \mathcal{H} and attendant \mathcal{H}' are selected, it is required to find a finite family of functions to first discretize the solution and then to compute the duality pairings, or inner products. It is common to assume the solution² as a linear combination of basis functions, that is

$$u(t) \approx u_h[\phi](t) = \sum_{i=1}^n u_i \phi_i(t), \quad (13)$$

where the $\phi_i(t)$ are the chosen trial functions and the u_i are the unknown coefficients to be determined. In Equation (13), the notation $u_h[\phi](t)$ highlights the dependence of u_h to the basis trial functions ϕ_i , when appropriate. The expression of the *discrete* residual readily follows as

$$R[u] \approx R[u_h] = \mathcal{N}[u_h] - f \neq 0, \quad \forall t \in \Omega. \quad (14)$$

The residual $R[u_h]$ represents the error introduced by substituting the approximate solution into the original differential equation. This residual is then annihilated *in the sense of the dual pairing* condition

$$\langle w_h, \mathcal{R}[u_h] \rangle = 0, \quad \forall w_h \in \mathcal{H}(\Omega). \quad (15)$$

²However, this is not a mandatory choice, see Equation (50) and Section 5.1 on Neural Network (NN).

with $R[u_h] \in \mathcal{H}'(\Omega)$. A possible choice to describe $w_h(t)$ is

$$w_h[\psi](t) = \sum_{i=1}^n w_i \psi_i(t), \quad (16)$$

which means that w_h lies in the subspace $\mathcal{W}_h \subset \mathcal{H}$ where $\mathcal{W}_h \equiv \{\psi_1, \psi_2, \dots, \psi_n\}$. These equations ensure that the residual is orthogonal to the weight functions, in the sense of the duality bracket, which can be seen as a generalized yet non symmetric inner product. The choice of weight functions $w_j(t)$ determines the specific type of weighted residual method (Zienkiewicz et al. 2005, p. 61).

Galerkin Method In this incarnation of the MWR, the basis functions are the same for both the trial and test functions, that is $\psi_j = \phi_j$, such that Equation (15) becomes

$$\int_{\Omega} \phi_i(t) R[u_h[\phi]](t) dt = 0, \quad \forall i \quad (17)$$

where a reduction of derivatives is classically implemented resulting in a weak form. The Galerkin method effectively minimizes the residual in an integral sense for self-adjoint and linear differential equations, particularly those arising from variational problems. The residual being orthogonal to the chosen test function space leads to an optimal approximation in the energy norm (Ern and Guermond 2004).

Petrov-Galerkin Method Unlike the Galerkin method, the basis functions for the trial and test spaces are chosen independently, meaning that $\psi_j \neq \phi_j$. This results in the formulation

$$\int_{\Omega} \psi_i(t) R[u_h[\phi]](t) dt = 0, \quad \forall i. \quad (18)$$

By allowing different spaces for the trial and test functions, the Petrov-Galerkin method provides greater flexibility, especially useful in problems involving non-self-adjoint operators, stabilization techniques, and variational formulations where weighted test spaces improve approximation properties. However, Petrov-Galerkin often leads to non-symmetric and sometimes ill-conditioned systems, and solving the resulting equations may require sophisticated numerical solvers compared to the symmetric and well-conditioned systems typical of the Galerkin method.

Least Squares Method In this version, the residual itself is minimized in a least-squares sense:

$$\min_{u_i, \forall i} \int_{\Omega} (R[u_h](t))^2 dt \quad \rightarrow \quad \frac{\partial}{\partial u_i} \int_{\Omega} (R[u_h](t))^2 dt = 0. \quad (19)$$

The least squares method minimizes the total squared error and does not require orthogonality to the test functions. No integration by parts is conducted on the residual term which is less convenient for the selection of the trial functions since a certain degree of differentiability will have to be satisfied. Upon enforcement of the minimization, the test functions can be understood as implicitly defined as the variations of the residual with respect to the coefficients $u_i, \forall i$. However, this assumes that the integral in Equation (19) and sum in Equation (13) can commute at the limit, which is not always clear. By optimally minimizing the squared residual, the method leads to a stable and well-posed numerical scheme, however, not weak in the distributional sense.

Collocation Method In this version, the basis functions used in the discretization of the test function are Dirac delta distributions centered at specific collocation points t_i , that is $\psi_j(t) = \delta_{t_i}$ and the residual then satisfies

$$\langle \delta_{t_i}, R[u_h[\phi]](t) \rangle = R[u_h[\phi]](t_i) = 0, \quad \forall i. \quad (20)$$

The selection of collocation points can be tailored to enhance accuracy in critical regions, such as those with steep solution gradients. A common misconception is that the error *disappears* at the collocation points. It is more accurate to state that the residual—highly dependent on the quality of the trial function—vanishes at these points. This does not necessarily ensure that the approximate solution precisely matches the exact solution at those locations.

In Equation (20), the distributional properties are on the test functions (via Dirac delta functions) rather than on the residual, which is quite unique. This approach thus assumes that the solution to the problem is sufficiently smooth to ensure the proper definition of the dual pairing. In other words, the collocation method is not well-suited for problems involving impulsive forces, for example. Also, no derivative reduction is commonly introduced in this approach.

Subdomain Method The subdomain method involves dividing the problem domain into smaller subdomains and ensuring that the average of the residual vanishes over each subdomain. The test functions w_i are chosen to be characteristic functions of subdomains of $\Omega_i \subset \Omega$, commonly with $\cup_i \Omega_i = \Omega$ and $\cap_i \Omega_i = \emptyset$, yielding

$$\int_{\Omega_i} R[u_h](t) dt = 0, \quad \forall i. \quad (21)$$

which is equivalent to choosing ψ_i as a constant equal to one on each Ω_i , and zero otherwise. This approach is closely related to the **FVM**, which discretizes the governing equations by integrating them over small control volumes and converting volume integrals to surface integrals using the divergence theorem.

Mathematically, true orthogonality arises only when both the test and trial function spaces are Hilbert spaces. The above list highlights that this *orthogonality*, as used in the context of **MWR**, becomes somewhat questionable when distributions are introduced in the formulation. Since distributions do not belong to Hilbert spaces, only the dual pairing remains applicable. Nevertheless, in many contexts, two functions are still referred to as orthogonal when their interaction through the dual pairing vanishes.

3 Strengths and weaknesses

3.1 Flexibility

Weighted residual techniques are powerful methods for solving differential equations, especially in the context of engineering and applied mathematics. Sharing similarities with weak formulations, they are far more flexible than strong formulations because they essentially allow solutions to exist in broader function spaces where the level of differentiability might be reduced. While Dirichlet boundary conditions are commonly strictly enforced in the definition of the considered function spaces, Neumann boundary

conditions are only weakly enforced through the integration by parts mechanism when used.

They also allow for flexible approximations by choosing various trial and test functions, making them adaptable to various problem types outside the classical world of ODE or PDE. For instance, as shown in [Example 13](#), possibly non-smooth [Differential Algebraic Equation \(DAE\)](#) can also tackled by these methods.

Finally, error control in the weighted residual formulation is achieved through a posteriori error estimation and adaptive refinement. Residual-based estimates assess error by evaluating the residuals of the governing equations and boundary conditions. In Galerkin and Petrov-Galerkin methods, error control is enhanced by adjusting weight functions and selecting adaptive basis functions. Least-squares formulations minimize squared residuals for stable solutions. Convergence is monitored using norm-based criteria, ensuring the estimated error remains below a predefined tolerance.

3.2 Loss of uniqueness

The primary limitation of the MWR is the potential loss of solution uniqueness, as shown in [Examples 5](#) and [6](#). Extending differential calculus to non-differentiable functions through distribution theory may excessively broaden the solution space, introducing ambiguity and further challenging the uniqueness of solutions. Additionally, convergence achieved in a mean weighed sense can complicate the interpretation of results, as subtle features such as high-frequency oscillations might be neglected or smoothed out.

Example 5 (Burger’s equation). *In fluid mechanics, it is well known that the initial value problem for the inviscid Burgers’ equation on $u(x, t)$ admits infinitely many weak solutions for the simple Riemann problem*

$$\partial_t u + \partial_x(u^2/2) = 0 \tag{22}$$

$$u_0(x) = \text{sign}(x), \tag{23}$$

where $\text{sign}(x) = \pm 1, \pm x > 0$. *To recover uniqueness, a single entropy inequality is sufficient in some cases (Panov 1994), while for general fluxes, infinitely many entropy inequalities are required (Kruzhkov 1970). Selecting non-classical weak solutions is more involved (LeFloch 2002), and typically necessitates the use of appropriate numerical schemes (Boutin et al. 2008).*

Example 6 (Solution to $tu(t) = 0$ in the sense of distributions). *The problem asks to determine a function u such that $tu(t) = 0$ for all t . A straightforward continuous solution on \mathbb{R} leads to the trivial solution $u(t) = 0$. However, what happens if we relax the condition and consider less regular solutions, such as impulsive or measure-valued solutions? Using the formalism of distributions, we note that*

$$\langle t\delta(t), \varphi(t) \rangle = \langle \delta(t), t\varphi(t) \rangle = (t\varphi(t))_{t=0} = 0 \times \varphi(0) = 0. \tag{24}$$

and the corresponding solutions are $u(t) = 0 + C\delta(t)$ where C is any real number and δ is the Dirac delta distribution. There are thus infinitely many impulsive solutions to the equation, as opposed to only one in the classical sense.

Uniqueness can be recovered by enforcing additional conditions such as “ u is a continuous function” or “ $\int u(t)dt$ is prescribed”, and alike. [Example 6](#) can then be solved

numerically through a simple yet not optimal method enforcing a prescribed integral C on the approximation. The unknown function u can be approximated using Dirac-like functions, namely bump functions ϕ_i with unit integral, centered at points t_j and having disjoint supports. These functions satisfy $\phi_i(t_j) = \alpha_i \delta_{ij} = 0$ except when $i = j$, where δ_{ij} denotes the Kronecker delta. Here, α_i is a positive constant such that $\int \phi_i = 1$. Assume that one of the points t_j is at 0, and denote the corresponding index by k . This assumption is essential to apply the collocation method at the points t_j . Seeking a solution of the form $u = \sum u_i \phi_i$ leads to all $u_i = 0$ except for u_k , which is determined by the prescribed integral, i.e., $u_k = C$. Hence, the function $C\phi_k$ provides a good approximation of the unique solution, provided the discretization is sufficiently fine. This example can be generalized to $t^n u(t) = 0$ where $\delta, \delta', \dots, \delta^{(n-1)}$ are solutions (Golse 2020).

3.3 Weak limits, nonlinearity & numerical schemes

The concept of a *weak limit* emerges naturally within MWR, extending beyond the classical notion of convergence, referred to as *strong limit*. Specifically, when considering formulations dependent on parameters or numerical schemes parameterized by a cell size or a number of basis functions, solutions may converge weakly towards unexpected limits. This phenomenon poses significant challenges in the context of convergence analysis, particularly in the presence of nonlinearities.

Definition 2 (Weak limit). A sequence of functions (or distributions) (u_n) converges weakly towards u , if for all smooth weights w (or test functions), the numeric sequence $\langle u_n, w \rangle$ converges to the number $\langle u, w \rangle$. This weak convergence is denoted $u_n \rightharpoonup u$.

Usually, the strong convergence implies the weak convergence but the converse is wrong. The examples shown below are relevant to a chapter on MWR because of the way weak limits are defined via the use of a weight.

Example 7 (Weak convergence of a highly oscillating periodic function). *Let u be a periodic function, integrable on the period, with mean \bar{u} , then $u(nt) \rightharpoonup \bar{u}$. Equivalently, for all test functions w with compact support,*

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} u(nt)w(t) dt = \bar{u} \int_{\mathbb{R}} w(t) dt. \quad (25)$$

The weak limit thus corresponds to a convergence for all weighted averages. Highly oscillating terms in the Fourier expansion converge in the mean towards 0, but the constant term remains. For instance, $\sin(nt) \rightharpoonup 0$, but $\sin^2(nt) \rightharpoonup 1/2 \neq 0^2$. The first limit is recovered through an integration by parts. For the second, notice that

$$\frac{1}{\pi} \int_0^\pi \sin^2 \theta d\theta = \frac{1}{2}. \quad (26)$$

Thus, the weak limit of the square may not be the square of the weak limit.

In practice, when a nonlinearity is involved in the formulation, careful attention should be paid to the weak limit in the nonlinear terms. However, for linear problems, the concept of weak limit is a powerful tool.

Example 8 (Oscillator with weakly vanishing forcing). Consider the linear second-order ODE $u_n'' + u_n = f_n$ indexed by n , with initial conditions $u_n(0) = 0$ and $u_n'(0) = 0$. Using the Green kernel expression

$$u_n(t) = \int_0^t K(t-s)f_n(s) ds, \quad K(t) = \sin t \quad (27)$$

it can be easily checked that if $f_n \rightharpoonup f$ then $u_n \rightharpoonup u$ where u satisfies the limit equation $u'' + u = f$, with $u(0) = 0$ and $u'(0) = 0$, that is $u = 0$. The kernel K can be seen as a smooth test function.

Example 9 (Steady-state solutions with weakly vanishing oscillatory forcing). Consider a linear oscillator subject to a weakly vanishing forcing

$$u_n'' + 2\xi\omega u_n' + \omega^2 u_n = \sin(n\Omega t) \quad (28)$$

where $\xi\omega > 0$. The limit equation when $n \rightarrow \infty$ with $u_n \rightarrow u$ reads

$$u'' + 2\xi\omega u' + \omega^2 u = 0 = \lim_{n \rightarrow \infty} \sin(n\Omega t) \quad (29)$$

and the steady-state solution to Equation (28) is

$$u_n(t) = \frac{(\omega^2 - n^2\Omega^2) \sin n\Omega t - 2n\xi\omega\Omega \cos n\Omega t}{n^4\Omega^4 + 2n^2(2\xi^2 - 1)\omega^2\Omega^2 + \omega^4} \quad (30)$$

from which it is clear that $u_n \rightarrow u = 0$. Moreover, u_n' strongly converges to its limit, but the convergence of the acceleration u_n'' is only weak. In the sense of distributions, the limit and derivative commute in a weak sense, hence

$$u_n''(t) = \frac{-n^2\Omega^2(\omega^2 - n^2\Omega^2) \sin n\Omega t + 2n^3\xi\omega\Omega^3 \cos n\Omega t}{n^4\Omega^4 + 2n^2(2\xi^2 - 1)\omega^2\Omega^2 + \omega^4} \rightharpoonup 0. \quad (31)$$

In Example 9, the strong convergence of $u_n(t)$ behaves as expected because the second derivative of the sequence is uniformly bounded. Consequently, by Ascoli's theorem and up to a subsequence, the sequence converges strongly in C^0 and also in C^1 . However, the second derivative converges only weakly to zero, as the source term does. This nice property is a consequence of the continuity of the derivation for weak convergence sequences, that is $u_n \rightharpoonup u$ implies $u_n' \rightharpoonup u'$. However, for nonlinear problems it can be very different. Consider the problem

$$u_n'' + f_n u_n' + u_n = 0, \quad (32)$$

with fixed initial data $u_n(0) = 1$ and $u_n'(0) = 0$. The nonlinear term causing difficulties is the product $f_n u_n'$ of two weakly converging sequences. Assume that $f_n \rightharpoonup f$. It is not always true that $u_n \rightharpoonup u$ where u is the solution of the naive limit equation $u'' + fu' + u = 0$ with the same initial data: the limit of the equation is not necessarily the equation of the limit, see also Example 7.

However, it can be true if u_n' converges strongly, since if $u_n' \rightarrow u'$ and $f_n \rightharpoonup f$ then $f_n u_n' \rightharpoonup f u'$ (Brezis 2011). How can we ensure the strong convergence of the sequence u_n' ? In practice, the existence of a uniform bound on the second derivative is a sufficient condition. Indeed, in simple terms, if the second derivative remains uniformly bounded, high-frequency oscillations in u_n' are prevented, thereby guaranteeing strong convergence. Such bounds can be obtained using an energy method if the problem exhibits dissipation, for instance.

Let us continue the investigation of the linear ODE (32). Although the equation itself is linear, the presence of the product term is not compatible with weak limits.

Example 10 (Dissipative second-order linear ODE with weakly converging non-constant coefficients). In [Equation \(32\)](#), consider the function $f_n(t) = 2 + \sin nt$, which satisfies $f_n(t) > 0$ and $f_n \rightharpoonup 2$. Through the positivity of f_n , the problem exhibits dissipation, ensuring that u'_n converges strongly to u' , where u is the solution of the differential equation $u'' + 2u' + u = 0$, with $u(0) = 1$, $u'(0) = 0$.

Example 11 (Non-dissipative second-order linear ODE with weakly converging non-constant coefficients). Starting from [Example 10](#), a counterexample is given by $f_n(t) = \cos nt$. In this case, the positivity of f_n is lost and the problem is no longer dissipative. To illustrate the attendant difficulty, consider the toy model

$$u''_n + f_n u'_n = 0, \quad u'_n(0) = 1. \quad (33)$$

A naive limit approach suggests that $u'' = 0$ with $u'(0) = 1$, implying that the expected limit function is $u' = 1$. However, this is incorrect. The solution u'_n and its weak limit can be explicitly computed as:

$$u'_n(t) = \exp(\sin nt) \rightharpoonup \frac{1}{2\pi} \int_0^{2\pi} \exp(\sin \theta) d\theta > 1. \quad (34)$$

Accordingly, the weak limit differs from the naive expectation of 1.

Example 12 (Duffing oscillator with very singular forcing). We consider the nonlinear ODE with a distributional right hand side term

$$u'' + \omega^2 u + \epsilon u^3 = \delta'' \quad (35)$$

where δ'' is the second distributional derivative of the classical Dirac δ .

The problem in [Example 12](#) is nonlinear and can be ill-posed. It is expected that the non-smoothness of δ'' propagates to u'' since the latter is the highest derivative in the equation and hence, the least smooth. By integrating twice in the sense of distributions, the unknown displacement u will feature a contribution of the δ distribution and the term u^3 becomes ill-defined.

Example 13 (Coulomb's dry friction with periodic Dirac delta forcing). Consider a one degree-of-freedom oscillator $u(t)$ subject a dry friction force $r(t)$, satisfying Coulomb's model, and externally forced by a T -periodic Dirac comb $\text{III}_T(t)$. The equations governing the dynamics take the form of the non-smooth vector DAE

$$u'' + \omega^2 u = r + \text{III}_T(t) \quad (36a)$$

$$u + \min(\rho(r + \mu N) - u, 0) + \max(\rho(r - \mu N) - u, 0) = 0 \quad (36b)$$

in the unknowns (u, r) , where μ is the friction coefficient and N , the prescribed normal force applied on the oscillator. [Equation \(36b\)](#) is an equality-based version of Coulomb's friction model where $\rho > 0$ ([Legrand and Pierre 2024](#)).

Dissipation is introduced through dry friction, and as a result, u'' is expected to be a measure. Consequently, u' belongs to the space of functions with [Bounded Variation \(BV\)](#). Such functions are continuous except at a countable set of discontinuity points. As a result, the sequence u'_n converges strongly in L^1 , but not pointwise at the discontinuities of the expected [BV](#) limit. The Coulomb friction term is more singular than a smooth nonlinearity and requires pointwise convergence, so that issues may arise in [Equation \(36b\)](#). This aspect warrants further investigation.

Example 14 (Sawtooth forcing and nonlinear damping). *Consider the periodically forced nonlinear oscillator*

$$u'' + \xi|u|u' + \omega^2 u = \text{Sqr}_T(t), \quad 0 < \xi \ll 1. \quad (37)$$

where $\text{Sqr}_T(t)$ is a periodic sawtooth function of period T .

The ODE is both dissipative and nonlinear. The **Harmonic Balance Method (HBM)** is generally effective for approximating the expected T -periodic solution u . However, concerns remain regarding its well-posedness. To begin with, the expected solution—assuming it exists—is approximated using a finite number of modes, and the formal computations proceed term by term (only cos terms are considered for simplicity):

$$u_N = \sum_{k=0}^N a_k \cos k\omega t, \quad u'_N = -\sum_{k=0}^N k\omega a_k \sin k\omega t, \quad u''_N = -\sum_{k=0}^N k^2 \omega^2 a_k \cos k\omega t. \quad (38)$$

Substituting the above finite Fourier expansions into [Equation \(37\)](#) and seeking orthogonality via a Galerkin procedure leads to a nonlinear problem in which the coefficients $(a_k)_{k=0}^N$ become the unknowns which can be computed using a nonlinear solver. Due to the dissipative nature of the ODE, an L^∞ bound is expected for u''_N . This bound implies weak convergence of u''_N and strong convergence (up to a subsequence) of u'_N and u_N . Consequently, one can pass to the weak limit in the second derivative term and to the strong limit in the nonlinear term. In other words, the **HBM** is expected to yield a C^1 solution whose second derivative is bounded in the sense of distributions, even for large values of ξ .

Example 15 (Space and time coupling in friction system). *We consider a simple one-dimensional bar system of length $L = 1$ clamped at $x = 0$ and subject to a normalized sinusoidal unit force at $x = L$. It is also subject to dry friction along its length. The unknowns are the displacement $u(x, t)$ and the friction force $r(x, t)$. The governing equations read*

$$u_{tt} - u_{xx} = r \quad (39a)$$

$$u_t + \min(\rho(r + \mu N) - u_t, 0) + \max(\rho(r - \mu N) - u_t, 0) = 0 \quad (39b)$$

$$u(0, t) = 0, \quad u_x(1, t) = \sin t \quad (39c)$$

where μ is the friction coefficient and N , the specified normal force.

[Equation \(39\)](#) is a (non-smooth) **Partial Differential Algebraic Equation (PDAE)** (Simeon and Arnold 2000) or a Partial Differential Inclusion. Implementing a space-time weighted residual formulation leads to consider the trial approximation

$$u(x, t) \approx u_h(x, t) = \sum_{i=1}^N \sum_{j=1}^M \phi_i(x) \theta_j(t) u_{ij} \quad (40)$$

and similarly for $r_h(x, t) \approx r(x, t)$, where $\phi_i(x)$ are the spatial shape functions, $\theta_j(t)$, the temporal shape functions, and u_{ij} , the coefficients to be found. The functions $\phi_i(x)$ are chosen to satisfy the Dirichlet BC in [Equation \(39c\)](#). The Neumann BC is satisfied in the weak form of [Equation \(39a\)](#). The approximate test function q_h is built in a similar fashion and the weighted residual form of [Equation \(39b\)](#) reads

$$\iint_{\Omega} q_h [u_{ht} + \min(\rho(r_h + \mu N) - u_{ht}, 0) + \max(\rho(r_h - \mu N) - u_{ht}, 0)] dx dt = 0 \quad (41)$$

where Ω is an appropriate space-time domain. The unknowns functions u_h and r_h being encapsulated in the min and max operators, integral and sum permutation is impossible, and the integral has to be computed at each iteration of a nonlinear solver working on the coefficients u_{ij} . The questions on the type of convergence expected, as well as on existence and uniqueness remain open. Nonetheless, the PDAE exhibits favorable properties rooted in the underlying physics. For example, the friction term r is expected to be a bounded function, which implies that u_t may possess sufficient regularity to satisfy the algebraic Coulomb condition pointwise and to allow passage to the limit within a MWR framework.

4 Convergence analysis of weighted residual schemes

A numerical scheme is said to be convergent if the solution of the discretized problem approaches the solution of the continuous problem as the discretization is refined, by increasing the number of basis functions in the approximation. We reproduce a classical case where the Lax Equivalence Theorem states that, for linear problems, convergence is ensured if the scheme is both consistent and stable. However, in more general nonlinear settings, the situation becomes considerably more intricate, and the Lax Theorem may not apply directly. In such cases, (1) both consistency and stability may require reevaluation, or (2) convergence may need to be addressed directly. Proving convergence is often highly challenging, if not infeasible, yet partial results can still offer meaningful insights.

4.1 Discretization of derivatives

A classical case in one space dimension is presented where the notions of *consistency*, *stability*, *convergence* are stated. The ODE

$$u''(t) = f(t); \quad t \in (0, T) \quad (42)$$

is considered along with homogeneous boundary conditions

$$u(0) = 0 = u(T). \quad (43)$$

The function f is fixed and sufficiently smooth on $(0, T)$ and the unknown is u . Problem in Equations (42) and (43) is usually presented in space variable as the Poisson equation with a minus sign.

4.1.1 Finite differences scheme

A finite difference scheme is considered first. This presentation can be found in textbooks on numerical analysis, see for instance (Arnold 2015). The continuous differential operator is $L[u] = u''$ and the time interval is discretized classically as t_i , $i = 0, \dots, N$ with $h = t_{i+1} - t_i = 1/N$. The corresponding approximation reads $u_h = (u_0, u_1, \dots, u_N)$ with $u_i \approx u(t_i)$. Moreover, $t_0 = 0$, $t_N = T$, and $u_N = u_0 = 0$. The restriction operator r_h acts on the continuous function u and outputs its values at the discrete time steps, that is $r_h[u] = (u(t_0), \dots, u(t_N))$. The discretization of the derivative terms using central difference is expressed as

$$v_h = L_h[u_h], \quad v_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2}, \quad \forall i \quad (44)$$

where L_h is the finite difference discrete operator. Let $f_h = r_h[f]$, the discretized version of [Equations \(42\) and \(43\)](#) is thus

$$L_h[u_h] = f_h. \quad (45)$$

The corresponding consistency error is $L_h[u_h] - L_h[r_h[u]]$ and is known to be $O(h^2)$ for a suitable norm $\|\cdot\|_h$ (such as the supremum norm or the Euclidean norm ([Ciarlet 1978](#))) when $u \in C^4([0, T], \mathbb{R})$ as easily verified using a Taylor expansion. This error quantifies how well the true solution satisfies the discrete equation. Consistency is a necessary condition that any numerical scheme must satisfy.

Stability of the scheme implies that errors in data yield errors in solutions of the same order. The amplification coefficient of the error must be controlled independently of the parameters of the discretization. For this purpose, a norm $\|\cdot\|_h$ is chosen depending on the size of the linear system in [Equation \(45\)](#). This is usually the same norm as the one above used for consistency. It is typically a Euclidean norm, normalized to be of the order as the max norm. Stability is achieved, for instance for a linear operator, when, there exists a constant C independent of h , such that if $L_h[u_h] = f_h$ and $L_h[\tilde{u}_h] = \tilde{f}_h$ then $\|u_h - \tilde{u}_h\|_h \leq C\|f_h - \tilde{f}_h\|_h$, where the norms of the source term and the norm of the solution could be different. Stability is more challenging to verify than consistency, as it requires a uniform bound on $\|L_h^{-1}\|_h$ that is independent of h . This property is inherently dependent on the choice of the norm.

In the case of linear problems, stability can be studied using techniques such as von Neumann stability analysis ([Euvrard 1988](#)), while more sophisticated methods, such as energy estimates, are typically employed for nonlinear problems ([Eymard et al. 2000](#)). To verify that consistency and stability imply convergence, we evaluate how the consistency error $L_h[u_h] - L_h[r_h[u]]$ is transferred to $u_h - r_h[u]$, through the stability of the scheme.

4.1.2 Weighted residual scheme

The MWR is presented for the same boundary value problem discussed in [Equations \(42\) and \(43\)](#). However, the discretization procedure differs, and the notion of a time step is not applicable in this context. We consider the following function spaces: $\mathcal{U} = H_0^1([0, T])$, and a finite-dimensional subspace $\mathcal{U}_h \subset \mathcal{U}$ spanned by the basis functions $\{\phi_i\}_{i=1}^N$ to define the approximate solution u_h . The test functions are drawn from a subspace $\mathcal{W}_h \subset \mathcal{W}$, spanned by $\{\psi_i\}_{i=1}^N$. The discretized solution takes the form $u(t) \approx u_h(t) = \sum_{i=1}^N u_i^\phi \phi_i(t)$, where the coefficients u_i^ϕ are the unknowns. These are collected into the vector $u_h^\phi = (u_1^\phi, \dots, u_N^\phi)$, and similarly for $w(t) \approx w_h(t)$. The notation is introduced to avoid confusion with quantities used in [Section 4.1.1](#).

A mandatory requirement is that \mathcal{U}_h must approximate \mathcal{U} when $h \rightarrow 0$, that is

$$\forall u \in \mathcal{U}, \quad \lim_{h \rightarrow 0} d(u, \mathcal{U}_h) = 0 \quad \text{with} \quad d(u, \mathcal{U}_h) = \inf_{v_h \in \mathcal{U}_h} \|u - v_h\|. \quad (46)$$

Then, the $N \times N$ linear system of equations in u_h^ϕ corresponding to [Equations \(42\) and \(43\)](#) reads

$$\langle w_h, L[u_h] - f \rangle = \int_0^T u_h'' w_h \, dt - \int_0^T f w_h \, dt = 0, \quad \forall w_h \in \mathcal{W}_h. \quad (47)$$

[Equation \(47\)](#) takes the matrix form $M_h^{\psi\phi} u_h^\phi = f_h^\psi$ which is the weighted residual version of [Equation \(45\)](#). Convergence analyses does not necessarily follow the consistency and

stability route, see (Ciarlet and Lunéville 2023a, p. 52) or (Ciarlet and Lunéville 2023b). Equation (46) is sometimes called consistency error for the MWR, but this is actually misleading. The convergence proof for MWR relies on specific analytical tools: namely, the quality of the approximate solution space discussed in Equation (46), and the coercivity of the bilinear operator associated with the considered elliptic problem (see (Demkowicz 2020, p. 41)). As will be explained later, different types of problems—such as certain hyperbolic PDEs—necessitate distinct analytical approaches.

4.1.3 Correspondence between schemes

In one dimension (*ie* time t , here), it is possible to recover various classical time-marching procedures using appropriate test and trial basis functions in the MWR, among which the finite difference method, see (Zienkiewicz 1977). For our need, derivations are conducted on two consecutive time intervals centered on t_i using nodes t_{i-1} , t_i , t_{i+1} with $t_{i+1} - t_i = h$. However, the MWR offers a versatile framework that enables the derivation of a wide variety of approximation schemes and unifies the development of both classical and novel time integration methods, including those for first-order systems.

Without integration by parts on the second derivative, the test function is chosen to be $w_h(t) = w_i \delta_{t_i}(t)$ while the trial function $u_h(t) = u_{i-1} \phi_{i-1}(t) + u_i \phi_i(t) + u_{i+1} \phi_{i+1}(t)$ involves the quadratic polynomials

$$\begin{aligned}\phi_{i-1}(t) &= \frac{(t - t_i)(t - t_{i+1})}{2h^2}; \\ \phi_i(t) &= -\frac{(t - t_{i-1})(t - t_{i+1})}{h^2}; \\ \phi_{i+1}(t) &= \frac{(t - t_i)(t - t_{i-1})}{2h^2}\end{aligned}\tag{48}$$

defined for $t \in (t_i - h, t_i + h)$, which also means constant acceleration on each time subinterval, leading to $h^2 u_h'' \equiv u_{i-1} - 2u_i + u_{i+1}$. From Equation (47), the weighted residual formulation leads to

$$\langle \delta_{t_i}, L[u_h] - f \rangle = \langle \delta_{t_i}, u_h''(t) - f \rangle = \frac{1}{h^2}(u_{i+1} - 2u_i + u_{i-1}) - f_i = 0\tag{49}$$

and the classical central finite difference scheme is recovered. Using $\int w_h u_h'' dt = -\int w_h u_h dt$ along with the classical hat shape functions for both the test and trial functions leads to the same scheme. It is then clear that errors in the scheme stem from the crucial choice of *both* the trial and test function spaces.

4.2 General discussion

The MWR serves as a powerful framework for developing effective numerical schemes across a broad range of problems. However, the question of convergence remains intricate and far from trivial. In particular, the applicability of weighted residual methods to nonlinear hyperbolic conservation laws poses significant challenges, even in the one-dimensional case. The convergence analysis of schemes based on this approach does not always adhere to the conventional framework of consistency and stability. In particular, establishing stability is often the most challenging aspect and is not always assured, especially when the underlying problem lacks uniqueness.

Consistency Consistency serves as a bridge between the exact solution and the approximate scheme—or alternatively, between the approximate solution and the exact equation. In other words, one can assess consistency either by evaluating how well the exact solution satisfies the approximate scheme, or by examining how closely the approximate solution satisfies the exact equation.

The first perspective is typically employed in the context of numerical methods such as Euler and Runge–Kutta methods for solving ODE or initial boundary value problems. In contrast, for hyperbolic conservation laws, one constructs an approximate solution and then verifies that it satisfies the exact weighted residual formulation up to an error term. One finally pass to the limit, when possible. In the case of the Wave Front Tracking algorithm developed in (Dafermos 1972), the consistency error is *effectively* reduced to zero by approximating the nonlinear flux with a piecewise linear flux, thereby modifying the exact equation itself. This modified equation, which preserves the structure of the initial equation, simplifies the analysis of the convergence of the solution scheme based on the MWR. Similarly, for certain special systems, the initial data can be approximated by piecewise constant functions for which the initial (exact) problem can be solved exactly, thus eliminating the consistency error (Bourdarias et al. 2022). For general systems, however, such consistency errors cannot be avoided and must be carefully estimated (Bressan 2000).

Stability Ensuring stability is often the most challenging aspect of numerical analysis, sometimes requiring nearly a decade of effort to establish a rigorous mathematical proof (Bressan 2000; Bressan et al. 1999) on hyperbolic systems. Proving that a scheme is stable is highly non-trivial, and in many cases, stability is not even guaranteed. In fact, for problems that do not have a unique solution for specified data or like HBM which can exhibit multiple steady state solutions for a given external force, stability cannot be expected. If some form of stability does appear in such cases, it means that the scheme implicitly selects a particular solution based on a specific criterion to restore uniqueness. This highlights the need to adapt the definition of stability depending on the nature of the problem.

5 Artificial Intelligence: future of weighted residuals?

Weighted residual techniques provide a powerful framework but might encounter hurdles in nonlinear dynamic regimes, such as convergence to the correct solution branch, stability and stiffness difficulties, high dimensionality, and rich solution structures that are hard to approximate with fixed bases, see for instance Equation (13). This has spurred interest in leveraging modern Artificial Intelligence (AI) solutions to enhance or replace parts of the MWR pipeline.

5.1 AI-powered extensions

Recent years have seen an explosion of methods that integrate deep learning into the residual minimization process of DE solvers. These approaches maintain the spirit of the MWR but use NN for greater flexibility and data-driven capability. As a very brief reminder, a NN is a parametric function of composite form

$$f(t) = \sigma_L(W_L \sigma_{L-1}(W_{L-1} \dots \sigma_1(W_1 t + \mathbf{b}_1) \dots + \mathbf{b}_{L-1}) + \mathbf{b}_L), \quad (50)$$

where t is the input (independent variable), W_k , \mathbf{b}_k are layer weights and biases, and σ_k are activation functions, for $k = 1, \dots, L$. They vastly differ from the additive forms classically used to approximate functions, see again Equation (13). Various key techniques and how they build upon classical MWR to tackle nonlinear dynamics are outlined below. While the concept of a *trial* function—or equivalently, an *approximate* solution—is well understood in the NN community, the definition and use of true weak formulations and attendant *test* functions remain less clearly articulated.

Physics-Informed Neural Networks Physics-Informed Neural Network (PINN) embed the DE *strong* residual into the loss function of a NN (Brink et al. 2021). By penalizing the residual at selected collocation points, the network is trained to approximate the solution in a mesh-free and differentiable manner. PINN allow incorporation of observational data and are suitable for problems with unknown physics or partial information. However, training can be slow and sensitive to the balance between residual and boundary/periodicity condition losses. PINN also struggle with stiff systems and multiscale features due to optimization challenges in highly non-convex landscapes (Kharazmi et al. 2021).

Deep Ritz and Variational Neural Methods The Deep Ritz method (E and Yu 2018) targets PDE with an underlying variational structure by minimizing their energy functional using NN, see Section 1.1. This weak-form approach, which assumes the existence of a variational structure of the considered problem, aligns closely with classical CG methods and is often more stable for symmetric problems. Deep Ritz solutions are also mesh-free and show some natural adaptivity (E et al. 2022), though they face difficulties enforcing boundary conditions and navigating non-convex loss surfaces. Despite these limitations, the method has proven effective in high-dimensional and singularity-dominated problems.

Weak Adversarial Networks Weak Adversarial Network (WAN) generalize the CG method by introducing two networks: one representing the trial function and one acting as the test function (Zang et al. 2020). The solution is trained through a minimax optimization process, ensuring the residual remains small across all admissible test functions. This formulation helps overcome issues with residual balancing and offers stability advantages, especially in high dimensions. WAN are well-suited to space-time formulations and can potentially capture full trajectories or periodic solutions in a single training session. Among all the AI-based solution techniques reviewed in this section, this appears to be the only one in which the notion of a test function is clearly defined.

Operator Learning Operator learning shifts focus from solving a single DE instance to learning the mapping from input data (like forcing terms or initial conditions) to solution functions. Architectures such as DeepONets (Lu et al. 2021) and Fourier Neural Operators (Li et al. 2020) are designed for this purpose. These models, once trained, enable fast inference and are particularly useful for applications requiring many simulations. Operator learning can incorporate physics constraints during training, enhancing generalization. However, training such models is data-intensive, and ensuring stability and physical accuracy outside the training set remains a major challenge.

5.2 Advantages and open questions

Integrating AI with MWR brings several benefits. Deep networks can handle high-dimensional problems more effectively than traditional methods (E et al. 2022). AI-based solvers are mesh-free, adaptive, and capable of integrating data and physical models into a unified framework. These models also enable rapid simulations once trained, making them attractive for tasks like parameter studies or real-time prediction (Schultz et al. 2023). The residual itself serves as an error indicator, guiding adaptive sampling strategies during training.

However, theoretical guarantees that underpin classical MWR—such as convergence and stability—are not yet fully developed for NN-based methods, as explained in the above references. Optimization remains a bottleneck, as training NN to minimize DE residuals can be slow and unreliable, especially for complex or multiscale systems. Enforcing boundary/periodicity conditions is less straightforward than in traditional solvers, and current techniques to address this are still evolving. NN also have difficulty resolving widely separated scales, often requiring hybrid or specialized architectures. Computational cost is another concern (Shukla et al. 2021): while trained models offer speed, training itself can be expensive and may not be justified for single-use cases.

5.3 AI-integrated WRMs for Nonlinear Dynamics

Hybrid approaches that combine NN with classical solvers may offer a practical balance—using networks to estimate residual corrections or accelerate convergence (Brink et al. 2021). Adaptive models that grow or shrink based on residuals could improve training efficiency. Uncertainty quantification methods are being developed to provide confidence estimates, particularly useful for sensitive or chaotic systems (Wang et al. 2025). Architectures tailored to specific physical laws, such as conserving energy for instance, show promise in improving accuracy and convergence. Operator learning models are expected to play larger roles in control and design optimization, providing fast, differentiable surrogates for complex systems. Finally, large pretrained models may emerge as reusable tools across a range of physics problems, just as foundation models have in language and vision.

6 Conclusion

This chapter shows that the MWR offers a unified framework for approximating solutions to nonlinear dynamical systems by transforming differential equations into weighted integral forms. This includes variational methods and extends to other numerical schemes like Galerkin, collocation, and finite differences, depending on the choice of test and trial functions. Through examples such as dry-friction oscillators, the chapter highlights MWR’s ability to handle challenging systems. However, convergence for nonlinear problems is not guaranteed by standard stability arguments alone, and demands careful treatment of the approximation space and operator properties.

The chapter also anticipates the evolving role of MWR, especially through its integration with machine learning. AI-based methods like PINN, Deep Ritz methods, and WAN follow the same residual minimization philosophy, using NN as flexible solution approximators. These approaches offer adaptivity, mesh-free formulation, and the ability to blend data with physical laws. Once trained, they can enable fast simulations and support tasks like parametric studies or uncertainty quantification.

Nonetheless, challenges remain: neural models can be hard to train, lack rigorous convergence theory, and may struggle with enforcing additional conditions (boundary, initial, and periodicity conditions) or preserving conservation laws. Hybrid strategies, such as using NN to enhance traditional MWR solvers, are emerging to address these gaps. Looking ahead, physics-informed architectures and operator-learning models are likely to expand the reach of MWR in nonlinear dynamics. Far from being outdated, the MWR framework remains a foundation for numerical methods, and will continue to influence the development of efficient and reliable solvers.

References

- Arnold, D. N. (2015). Stability, Consistency, and Convergence of Numerical Discretizations. *Encyclopedia of Applied and Computational Mathematics*. Ed. by B. Engquist. Springer, pp. 1358–1364. DOI: [10.1007/978-3-540-70529-1_407](https://doi.org/10.1007/978-3-540-70529-1_407).
- Barr, A. D. S. (1966). An Extension of the Hu-Washizu Variational Principle in Linear Elasticity for Dynamic Problems. *Journal of Applied Mechanics* 33.2, pp. 465–465. DOI: [10.1115/1.3625076](https://doi.org/10.1115/1.3625076).
- Bourdarias, C., A. P. Choudhury, B. Guelmame, and S. Junca (2022). Entropy solutions in BV^s for a class of triangular systems involving a transport equation. *SIAM Journal on Mathematical Analysis* 54.1, pp. 791–2022. DOI: [10.1137/20M1351783](https://doi.org/10.1137/20M1351783). HAL: [hal-02895603](https://hal.archives-ouvertes.fr/hal-02895603).
- Boutin, B., C. Chalons, F. Lagoutière, and P. G. LeFloch (2008). Convergent and conservative schemes for nonclassical solutions based on kinetic relations. I. *Interfaces Free Bound* 10.3, pp. 399–421. DOI: [10.4171/IFB/195](https://doi.org/10.4171/IFB/195). ARXIV: [0712.3766](https://arxiv.org/abs/0712.3766).
- Bressan, A. (2000). *Hyperbolic systems of conservation laws*. Oxford University Press. ISBN: 9780198507000.
- Bressan, A., T.-P. Liu, and T. Yang (1999). L^1 stability estimates for $n \times n$ conservation laws. *Archive for Rational Mechanics and Analysis* 149, pp. 1–22. DOI: [10.1007/s002050050165](https://doi.org/10.1007/s002050050165).
- Brezis, H. (2011). *Functional Analysis, Sobolev spaces and partial differential equations*. Springer. DOI: [10.1007/978-0-387-70914-7](https://doi.org/10.1007/978-0-387-70914-7).
- Brink, A. R., D. A. Najera-Flores, and C. Martinez (2021). The neural network collocation method for solving partial differential equations. *Neural Computing and Applications* 33.11, 5591–5608. ISSN: 0941-0643. DOI: [10.1007/s00521-020-05340-5](https://doi.org/10.1007/s00521-020-05340-5).
- Bérest, P. (1997). *Calcul des variations. Application à la Mécanique et à la Physique*. Ellipses. ISBN: 9782729897048.
- Canuto, C., M. Y. Hussaini, A. Quarteroni, and T. A. Zang (2006). *Spectral Methods: Fundamentals in Single Domains*. Springer. DOI: [10.1007/978-3-540-30726-6](https://doi.org/10.1007/978-3-540-30726-6).
- Chen, Z. and M. Ney (2009). The method of weighted residuals: a general approach to deriving time and frequency-domain numerical methods. *IEEE Antennas and Propagation Magazine* 51.1, pp. 51–70. DOI: [10.1109/MAP.2009.4939019](https://doi.org/10.1109/MAP.2009.4939019). HAL: [hal-00460288](https://hal.archives-ouvertes.fr/hal-00460288).
- Ciarlet, P. G. and E. Lunéville (2023a). *La méthode des éléments finis – De la théorie à la pratique. Tome 1: Concepts généraux*. Les Presses de l’ENSTA. ISBN: 9782722509177. HAL: [hal-04039611](https://hal.archives-ouvertes.fr/hal-04039611).
- Ciarlet, P. G. and E. Lunéville (2023b). *The finite element method. From theory to practice*. Wiley. ISBN: 9781394229741. DOI: [10.1002/9781394229758](https://doi.org/10.1002/9781394229758).
- Ciarlet, P. G. (1978). *The finite element method for elliptic problems*. North-Holland Publishing. ISBN: 9780444850287.

- Cottrell, J. A., T. J. R. Hughes, and Y. Bazilevs (2009). *Isogeometric Analysis: Toward Integration of CAD and FEA*. Wiley. ISBN: 9780470749098.
- Crandall, S. (1956). *Engineering Analysis. A Survey of Numerical Procedures*. McGraw-Hill. ISBN: 978-0070134300.
- Dafermos, C.-M. (1972). Polygonal approximations of solutions of the initial value problem for a conservation law. *Journal of Mathematical Analysis and Applications* 38, pp. 33–41. DOI: [10.1016/0022-247X\(72\)90114-X](https://doi.org/10.1016/0022-247X(72)90114-X).
- Demkowicz, L. F. (2020). *Mathematical theory of finite elements*. SIAM. ISBN: 9780444850287. DOI: [10.1137/1.9781611977738](https://doi.org/10.1137/1.9781611977738).
- E, W, J Han, and A Jentzen (2022). Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning. *Nonlinearity* 35.1, pp. 278–310. DOI: [10.1088/1361-6544/ac337f](https://doi.org/10.1088/1361-6544/ac337f). ARXIV: [2008.13333](https://arxiv.org/abs/2008.13333).
- E, W. and B. Yu (2018). The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems. *Communications in Mathematics and Statistics* 6 (1), pp. 1–12. DOI: [10.1007/s40304-018-0127-z](https://doi.org/10.1007/s40304-018-0127-z). ARXIV: [1710.00211](https://arxiv.org/abs/1710.00211).
- Ern, A. and J.-L. Guermond (2004). *Theory and Practice of Finite Elements*. Springer. DOI: [10.1007/978-1-4757-4355-5](https://doi.org/10.1007/978-1-4757-4355-5).
- Euvrard, D. (1988). *Résolution numérique des équations aux dérivées partielles de la physique, de la mécanique et des sciences de l'ingénieur*. Mathematics for Physics. Masson, p. 287. ISBN: 9782225845093.
- Eymard, R., T. Gallouët, and R. Herbin (2000). *Finite volume methods*. Vol. 7. Handbook of Numerical Analysis. North-Holland, p. 1020. DOI: [10.1016/S1570-8659\(00\)07005-8](https://doi.org/10.1016/S1570-8659(00)07005-8). HAL: [hal-02100732](https://hal.archives-ouvertes.fr/hal-02100732).
- Finlayson, B. (1972). *The Method of Weighted Residuals and Variational Principles: With Application in Fluid Mechanics, Heat and Mass Transfer*. Academic Press. ISBN: 9780122570506.
- Fletcher, C. A. J. (1998). Weighted Residual Methods. *Computational Techniques for Fluid Dynamics 1: Fundamental and General Techniques*. Springer Series in Computational Physics. Springer. Chap. 5, pp. 98–162. DOI: [10.1007/978-3-642-58229-5_5](https://doi.org/10.1007/978-3-642-58229-5_5).
- Galerkin, B. (1915). On electrical circuits for the approximate solution of the Laplace equation. Russian. *Vestnik Inzh.* 19, pp. 897–908.
- Golse, F. (2020). *Distributions, analyse de Fourier, équations aux dérivées partielles*. Ecole Polytechnique. ISBN: 9782730216890.
- Kharazmi, E., Z. Zhang, and G. E. Karniadakis (2021). hp-VPINNs: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering* 374, p. 113547. DOI: [10.1016/j.cma.2020.113547](https://doi.org/10.1016/j.cma.2020.113547).
- Kruzhkov, S. N. (1970). First order quasilinear equations with several independent variables. *Mathematics of the USSR-Sbornik* 81.2, pp. 228–255. DOI: [10.1070/SM1970v010n02ABEH002156](https://doi.org/10.1070/SM1970v010n02ABEH002156).
- LeFloch, P. G. (2002). *Hyperbolic systems of conservation laws. The theory of classical and nonclassical shock waves. The Theory of Classical and Nonclassical Shock Waves*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, p. 294. DOI: [10.1007/978-3-0348-8150-0](https://doi.org/10.1007/978-3-0348-8150-0).
- Legrand, M. and C. Pierre (2024). A compact, equality-based weighted residual formulation for periodic solutions of systems undergoing frictional occurrences. *Journal of Structural Dynamics* 2, pp. 144–170. DOI: [10.25518/2684-6500.190](https://doi.org/10.25518/2684-6500.190). HAL: [hal-04189699](https://hal.archives-ouvertes.fr/hal-04189699).

- Li, Z., N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. M. Stuart, and A. Anandkumar (2020). Fourier Neural Operator for Parametric Partial Differential Equations. *CoRR*. ARXIV: [2010.08895](https://arxiv.org/abs/2010.08895).
- Liu, G. and Y. Gu (2005). *An Introduction to Meshfree Methods and Their Programming*. Springer. DOI: [10.1007/1-4020-3468-7](https://doi.org/10.1007/1-4020-3468-7).
- Lu, L., P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence* 3 (3), pp. 218–229. DOI: [10.1038/s42256-021-00302-5](https://doi.org/10.1038/s42256-021-00302-5). ARXIV: [1910.03193](https://arxiv.org/abs/1910.03193).
- Panov, E. (1994). Uniqueness of the solution of the Cauchy problem for a first order quasilinear equation with one admissible strictly convex entropy. *Math Notes* 55, 517–525. DOI: [10.1007/BF02110380](https://doi.org/10.1007/BF02110380).
- Rayleigh, L. (1877). *The Theory of Sound*. Vol. 1. The Macmillan Company.
- Ritz, W. (1909). Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. ger. *Journal für die reine und angewandte Mathematik* 135, pp. 1–61.
- Schultz, J., J. van Delden, C. Blech, S. C. Langer, and T. Lüddecke (2023). Deep learning for frequency response prediction of a multimass oscillator. *Proceedings in Applied Mathematics and Mechanics* 23.3, e202300091. DOI: [10.1002/pamm.202300091](https://doi.org/10.1002/pamm.202300091).
- Schwartz, L. (1997). *Théorie des distributions*. Hermann. ISBN: 9782705655518.
- Shukla, K., A. D. Jagtap, and G. E. Karniadakis (2021). Parallel physics-informed neural networks via domain decomposition. *Journal of Computational Physics* 447, p. 110683. DOI: [10.1016/j.jcp.2021.110683](https://doi.org/10.1016/j.jcp.2021.110683).
- Simeon, B. and M. Arnold (2000). Coupling DAEs and PDEs for Simulating the Interaction of Pantograph and Catenary. *Mathematical and Computer Modelling of Dynamical Systems* 6.2, pp. 129–144. DOI: [10.1076/1387-3954\(200006\)6:2;1-M;FT129](https://doi.org/10.1076/1387-3954(200006)6:2;1-M;FT129).
- Sun, B. (2022). On Chien’s question to the Hu-Washizu three-field functional and variational principle. *Applied Mathematics and Mechanics* 43, pp. 537–546. DOI: [10.1007/s10483-022-2838-5](https://doi.org/10.1007/s10483-022-2838-5).
- Wang, N., Y. Chen, and D. Zhang (2025). A comprehensive review of physics-informed deep learning and its applications in geoenery development. *The Innovation Energy* 2.2, p. 100087. ISSN: 3006-418X. DOI: [10.59717/j.xinn-energy.2025.100087](https://doi.org/10.59717/j.xinn-energy.2025.100087).
- Zang, Y., G. Bao, X. Ye, and H. Zhou (2020). Weak adversarial networks for high-dimensional partial differential equations. *Journal of Computational Physics* 411, p. 109409. DOI: [10.1016/j.jcp.2020.109409](https://doi.org/10.1016/j.jcp.2020.109409).
- Zienkiewicz, O. C. (1977). A new look at the Newmark, Houbolt and other time stepping formulas. A weighted residual approach. *Earthquake Engineering & Structural Dynamics* 5.4, pp. 413–418. DOI: [10.1002/eqe.4290050407](https://doi.org/10.1002/eqe.4290050407).
- Zienkiewicz, O., R. Taylor, and J. Zhu (2005). *The Finite Element Method: Its Basis and Fundamentals*. 6th ed. Butterworth-Heinemann. ISBN: 9780750663200.