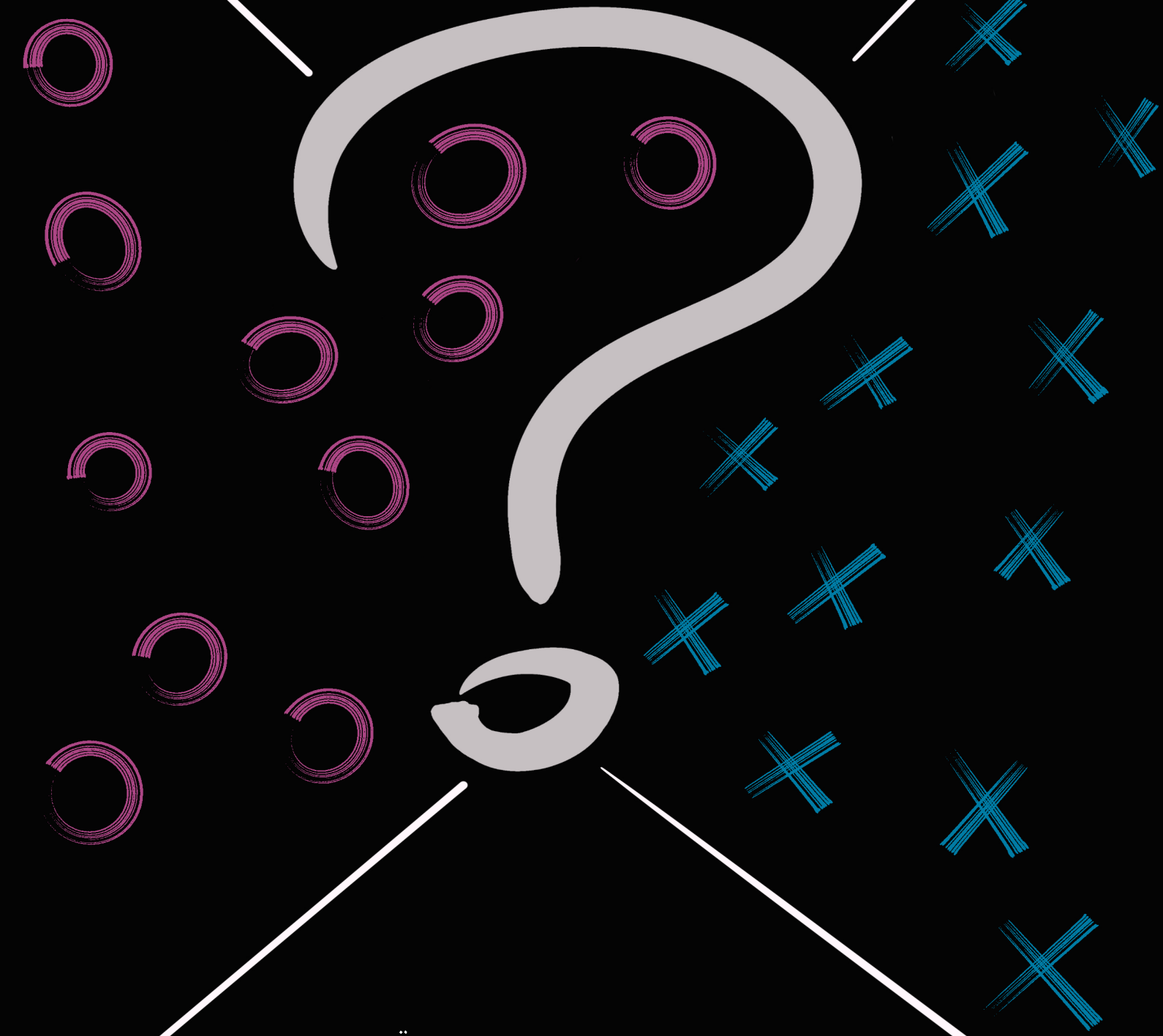
New Machine Learning Strategies in Data-scarce Material Science Problems

Go/No-Go Report



Özgür Taylan Turan

New Machine Learning Strategies in Data-scarce Material Science Problems

by

O.T. Turan

Pattern Recognition & Bioinformatics Group
Intelligent Systems Department
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Employee number:	919405	
Project duration:	November 1, 2020 – November 1, 2025	
Supervisory Members:	Prof. dr. ir. M. J. T. Reinders,	PRB/INSY/EEMCS/TU Delft, Promoter
	Dr. M. Loog,	PRB/INSY/EEMCS/TU Delft, Daily Supervisor
	Dr. D. M. J. Tax,	PRB/INSY/EEMCS/TU Delft , Daily Supervisor
	Dr. M. Bessa,	MSE/3ME/TU Delft , Daily Supervisor

TU Delft

Abstract

Data scarcity is one of the main issues while creating surrogate constitutive manifolds for complex solid mechanics applications. This makes existing data immensely important or requires the new data collection scheme to be as efficient and smart as possible. In the current state of the solid mechanics field, conventional machine learning strategies are being utilized in a wide range of problems, by either generating too much data or using models with physical information embedded in them. However, deeper machine learning paradigms are not being focused on as much. In this context, this research aims to investigate and come up with different possible methods to utilize the existing data and to efficiently sample data. This aim is specifically constructed to accelerate a multi-scale numerical analysis method called FE^2. This document presents FE^2 in a context-free manner with the downsides of the method. Then, the current strategies to solve these problems are presented with a literature review. Afterward, the problem is formulated in the most general way and the research questions arising from this formulation are presented. Then, the most relevant and critical research path is identified and its relation to the machine learning field is established with a preliminary literature review. Finally, the tentative path to be followed in this PhD project is presented.

Contents

1 | Introduction

The increasing complexity of engineering problems creates the need to design new materials with better properties that are suitable for the problem at hand (*e.g.* durable, strong, sustainable materials). Composite materials, an assembly of two or more materials with different properties that together have superior properties than its separate constituents [44], is one of the prime examples of this endeavor. Composites can derive their macroscopic behavior from a specific microstructure, in other words, interactions at lower scales between constituents govern the overall behavior of the material [87]. Due to difficulties encountered in the experimentation of these types of materials in small scales, high-fidelity numerical analysis is seen as the way for the composite structure design [77]. Concurrent finite element method (FE^2) is a high-fidelity numerical solution technique that is commonly utilized for modelling multi-scale heterogeneous continua [25] (See Figure 1.1).

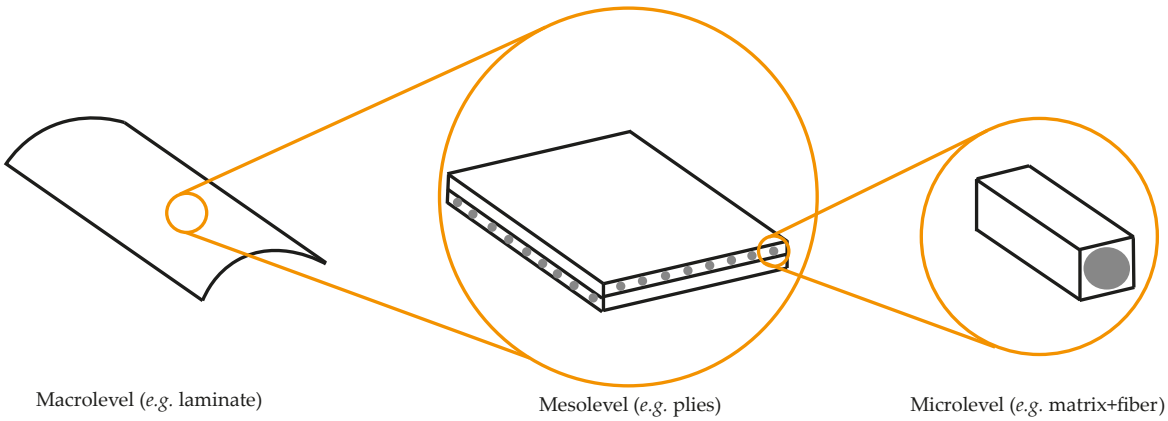


Figure 1.1: Engineered composite make use of the strong properties of its constituents in different scales to create superior materials to be used in desired engineering problems.

Doubling of the computational power according to Moore's Law[1] [63] allowed great advances in numerical modeling endeavors like the finite element method in the last 5 decades. However, although the engineering problems get more complex, there are certain limits of computing power. Moreover, it is evident that compute power increase rate might decrease or even hinder due to physical limitations of the standard silicone technology [4]. Non-classical computing strategies like quantum computing offers enormous potential, but the integration of promising quantum technologies in science and engineering remains hindered due to both hardware and theoretical limitations [35, 50]. Thus, there exists a need to put an effort towards acceleration strategies of the current numerical modeling techniques to be able to efficiently solve problems

[1]Number of transistors on a chip doubles about every 2 years.

with growing complexity.

FE^2 method is computationally expensive owing to its multi-scale nature. This computational burden increases more depending on the complexity of the problem at hand. Serious effort regarding the acceleration strategies of FE^2. Model Order Reduction (MOR) is one of the acceleration strategies that reduce the complexity of the problem while preserving the fidelity of the solution [81]. Applications of finite element method based model order reduction techniques [12, 34, 37, 49, 77, 88] can help ease the computational burden of FE^2.

Machine Learning (ML)/Pattern Recognition (PR) is another field that benefited from increased computational power. To the authors knowledge the mechanics applications of Machine Learning (ML) can be traced back to the second-boom[2] of ML in the fields of civil and mechanical engineering (*e.g* [1, 7, 66, 74, 75]). Then, with the gold-rush[3] of ML, its prevalence in various fields and availability of open-source toolboxes ML techniques are starting to get embedded engineering ventures more frequently.

One of those ventures is the aforementioned case of acceleration of high-fidelity numerical methods. This is achieved either by developing new numerical methods with the help of machine learning, which are referred as *physical* methods (*e.g.* [9, 53, 55, 72]), or creating surrogate models with the help of experimental data or direct numerical solutions (DNS), which are referred as *non-physical* methods (*e.g.* [6, 45, 64, 77, 78]). Since physical information is involved with *physical* methods, their extrapolation capabilities to unseen scenarios are surprisingly well. However, the *non-physical* methods suffer deeply when the surrogate models face scenarios in the extrapolation regime. One way to circumvent this is to use both *physical* and *non-physical* methods at the same time. By generating an abundance of data with an *physical* method and utilizing an *non-physical* method to fit the generated data, surrogate models are aimed to be situated in an interpolation regime at all times.

Although the intention of incorporation of machine learning strategies for the betterment of the computational field. This intersection of field is still struggling with its growing pains which contribute to the pool of reasons why machine learning is not well integrated into the mechanics community. For instance, depending on the problem at hand small tweaks might be needed for *physical* methods to incorporate more physical information that is essential to the problem or some parts of the model are still lacking physical interpretability. Similarly, *non-physical* methods might need abundant data depending on the problem at hand or the predictions made by the machine learning models are oblivious to the underlying physical phenomenon.

The main aim of this work is to bring the mechanics and machine learning communities together with the intention of tackling the problems that arise from the utilization of non-physics-based tools in the fields of mechanics. As the title of the PhD position suggests particular focus will be on the problems where there is data scarcity that hinders the usage of machine learning tools in a physically driven environment where every input has an output that is reasoned by the physical laws. To achieve the over-arching goal of this PhD the ways of exploiting the similarities between problems will be of particular interest. In other words, by using the information from

[2] From 1990's till 2000's [31]
[3] Starting from the 2000s till now [31].

previous problems that had abundant data, the need for data that is coming from DNS will be tried to be mitigated.

This document is intended for formulating a research path for data-scarce material science problems, for the PhD position opened in collaboration with the Mechanical, Materials, and Maritime Department of the Delft University of Technology. Chapter 2 is devoted to the physical problem definition in order to explain from which physical assumptions and methods data for this PhD will mostly come. Chapters 3 a preliminary literature review for the mechanics community is provided. After this review, the learning problem is reformulated and some research questions, that naturally arise from this formulation, are provided in Chapter 4. Given the reformulated learning problem the literature review of the meta-learning paradigm is presented in Chapter 5. Finally, Chapter 6 presents the progress and the future planning for the given PhD project.

2 | Numerical Composite Modeling

The idea of dividing the problem into smaller pieces dates back to Archimedes's idea of dividing the complex shapes into many smaller but easily computable constituents to obtain the area of that shape. The Finite Element Method (FEM) takes this concept and applies it in a similar fashion to the world of physical laws. The concept of FEM dates back to the 1960 and before that, there was unpublished research for technological advancements during the World War II. [84]. Although the capabilities of the method the computational limitations during the times were problematic for the application of this method for multi-scale problems. However, this limitation did not prevent the research regarding multi-scale investigations dating back to the 1970's [76]. However, the computational availability nowadays is still not enough to tackle the multi-scale problem to the full extent. Thus, multi-scale numerical methodologies related to research preserve its activity. These methods provide high-fidelity solutions to micro-heterogeneous material modeling ventures.

One of the most popular approaches is the FE^2 method which relies on computational homogenization to average the microscopic behavior with the aim to get macroscopic material behavior. In this PhD the aimed application data is expected to come from the inner loop (Micro-scale problem) of the FE^2 method. Which constitutes the main bottleneck of the method. Surrogate models that are aimed to be created involve the inner loop only with the intention to accelerate the overall procedure. To achieve the desired acceleration the computational burden originating from the need to obtain Direct Numerical Solutions (DNS) is tried to be kept at a minimum level. This intent of trying to keep the DNS at a minimum level can be considered as the data-scarcity. However, it should be noted that in the given setting data scarcity does not mean that the DNS cannot be realized. In most cases data is obtainable, but the cost of getting it is extremely high.

This chapter is dedicated to the introduction of the numerical composite modeling approach FE^2 followed by a brief explanation of how machine learning fits into the given framework and some preliminary literature review on applied machine learning in the given context of trying to accelerate the conventional numerical modeling techniques. (*e.g.* surrogate modeling, physics-informed machine learning strategies, reduced-order modeling, etc.)

2.1 FE^2

This section is dedicated to concurrent multi-scale analysis overview. Interested readers can find more detailed information in [25, 40, 60, 61, 77].

2.1.1 Macro-scale Problem

Let Ω be the macroscopic domain, the displacement field $\mathbf{u}^{\Omega}$, as a product of applied Neumann and Dirichlet boundary conditions at the boundary of the macroscopic domain Γ. It should be noted that each point at undeformed configuration $\mathbf{X}^{\Omega}$ is mapped on to $\mathbf{x}^{\Omega}$ in the deformed configuration continuously during deformation. Then deformation gradient can be represented by,

$$\mathbf{F}^{\Omega} = \frac{\partial \mathbf{x}^{\Omega}}{\partial \mathbf{X}^{\Omega}} = \mathbf{I} + \frac{\partial \mathbf{u}^{\Omega}}{\partial \mathbf{X}^{\Omega}} = \mathbf{I} + \nabla_{\mathbf{X}}\mathbf{u}, \tag{2.1}$$

where $\mathbf{I}$ is the identity tensor. Moreover, stress equilibrium in the reference configuration is expressed as,

$$\begin{aligned} \nabla \cdot \mathbf{P}^{\Omega} + \mathbf{b}^{\Omega} &= \mathbf{0} \quad \text{on} \quad \Omega, \\ \mathbf{P}^{\Omega} \cdot \mathbf{N}^{\Omega} &= \mathbf{t}^{\Omega} \quad \text{on} \quad \Gamma^{N}, \quad \text{and} \\ \mathbf{u}^{\Omega} &= \mathbf{u}_{D}^{\Omega} \quad \text{on} \quad \Gamma^{D}, \end{aligned} \tag{2.2}$$

where $\mathbf{u}_{D}^{\Omega}$ is the prescribed displacement, $\mathbf{t}^{\Omega}$ is the traction, $\mathbf{N}^{\Omega}$ is the surface normal, $\mathbf{b}^{\Omega}$ is the body forces and $\mathbf{P}^{\Omega}$ is the first Piola-Kirchhoff stress tensor. In order to solve Equation 2.2 a constitutive relation ($\mathcal{C}$) between stresses ($\mathbf{P}^{\Omega}$) and deformation gradient $\mathbf{F}^{\Omega}$ is needed,

$$\mathbf{P}^{\Omega} = \mathcal{C}^{\Omega}(\mathbf{F}^{\Omega}). \tag{2.3}$$

It should be noted that there are no assumptions made until now regarding the constitutive relation. In the given setting $\mathcal{C}^{\Omega}$ will represent the information coming from the smaller scales.

After, obtaining the weak form and discretization of the domain Ω the equilibrium problem can be solved by minimization of the residual force ($\mathbf{r}^{\Omega} = \mathbf{f}^{\Omega} - \mathbf{f}^{\Gamma} = 0$). It should be noted that,

$$\mathbf{f}^{\Omega} = \int_{\Omega} \mathbf{B} \cdot \mathbf{P}^{\Omega} d\Omega \quad \text{and} \quad \mathbf{f}^{\Gamma} = \int_{\Gamma} \mathbf{N} \cdot \mathbf{t}^{\Omega} d\Gamma, \tag{2.4}$$

where $\mathbf{N}$ are the shape functions and $\mathbf{B}$ their spatial derivatives.

Thus, the finite element problem definition of the macro-scale completes with the definition of the tangent stiffness matrix $\mathbf{K}^{\Omega}$, that is needed to compute displacement update ($\Delta\mathbf{u}^{\Omega} = -(\mathbf{K}^{\Omega})^{-1} \cdot \mathbf{r}^{\Omega}$),

$$\mathbf{K}^{\Omega} = \int_{\Omega} \mathbf{B} : \mathbf{D} : \mathbf{B} d\Omega \tag{2.5}$$

where $\mathbf{D}$ is the stiffness tensor.

2.1.2 Micro-scale Problem

Let ω represent the microscopic domain of a Representative Volume Element (RVE), then similar to the macroscopic problem with the aim to find the microscopic displacement field $\mathbf{u}^{\omega}$ the microscopic problem can be represented similarly assuming periodic boundary conditions on the boundary of the RVE (γ),

$$\begin{aligned} \nabla \cdot \mathbf{P}^{\omega} + \mathbf{b}^{\omega} &= \mathbf{0} \quad \text{on} \quad \omega, \\ \mathbf{u}^{\omega}_{\text{master}} &= \mathbf{u}^{\omega}_{\text{slave}} \quad \text{on} \quad \gamma. \end{aligned} \tag{2.6}$$

Under the first-order computational homogenization assumption, the microscopic deformed configuration of the RVE is coupled with macro-scale and decomposed as

$$\mathbf{x}^{\omega} = \underbrace{\mathbf{F}^{\Omega}\mathbf{X}^{\omega}}_{\text{homogenous}} + \underbrace{\tilde{\mathbf{u}}^{\omega}}_{\text{fluctuation}} . \tag{2.7}$$

Then, microscopic deformation gradient can be expressed as

$$\mathbf{F}^{\omega} = \mathbf{F}^{\Omega} + \tilde{\mathbf{F}}^{\omega} \tag{2.8}$$

where $\tilde{\mathbf{F}}^{\omega} = \nabla_{\mathbf{X}^{\omega}} \tilde{\mathbf{u}}^{\omega}$. Unlike, macroscopic problem here the constitutive relation $\mathbf{P}^{\omega} = \mathcal{C}^{\omega}(\mathbf{F}^{\omega})$ has to be explicitly defined.

2.1.3 Coupling of Scales

Noting that according the Hill-Mandel Lemma deformation gradients ($\mathbf{F}^{\Omega}$ and $\mathbf{F}^{\omega}$) and their work-conjugate stress measures ($\mathbf{P}^{\Omega}$ and $\mathbf{P}^{\omega}$) can be related between macro-scale and micro-scale with their volume averages as,

$$\begin{aligned} \mathbf{P}^{\Omega} &= \frac{1}{\mathrm{V}} \int_{\omega} \mathbf{P}^{\omega} d\Omega \quad \text{and} \\ \mathbf{F}^{\Omega} &= \frac{1}{\mathrm{V}} \int_{\omega} \mathbf{F}^{\omega} d\Omega \end{aligned} \tag{2.9}$$

where V is the volume of the RVE.

Thus, after solving the microscopic boundary value problem, the stress field can be recovered for the RVE and averaged over the volume to obtain the macroscopic stresses. To complete the whole scheme of the FE^2 method calculation of the tangent stiffness is needed, but for the sake of brevity it is left out, the reader is referred to [83] for detailed information.
The overall scheme of the whole procedure is summarized in Figure 2.1.

2.2 FE^2 via Machine Learning

In a simplified manner the above FE^2 procedure can be explained as follows:

Remark 1. There exists a domain Ω for which we are interested in solving Equation 2.2. This equation requires the mapping $\mathcal{C}^{\Omega}$ to be defined ($\mathcal{C}^{\Omega} := \mathbf{F}^{\Omega} \rightarrow \mathbf{P}^{\Omega}$). For composite materials, this mapping is determined by the properties of its constituents in microscopic/mesoscopic levels. Thus, to obtain a complex macroscopic mapping $\mathcal{C}^{\Omega}$ Equation 2.6 has to be solved and averaged over the domain ω. At the end of this procedure a fairly good approximation for the mapping $\mathcal{C}^{\Omega}$.

In cases, the constitutive relation $\mathcal{C}$ (general mapping in any part of the problem) is nonlinear the linear system of equations has to be solved iteratively at every so-called time step[1]. Furthermore, FE^2 method tries to solve a discretized macro-scale problem, by solving a discretized

[1]Here the mentioned time step does not correspond to time as one might perceive. It is the load application step in a vague time sense.

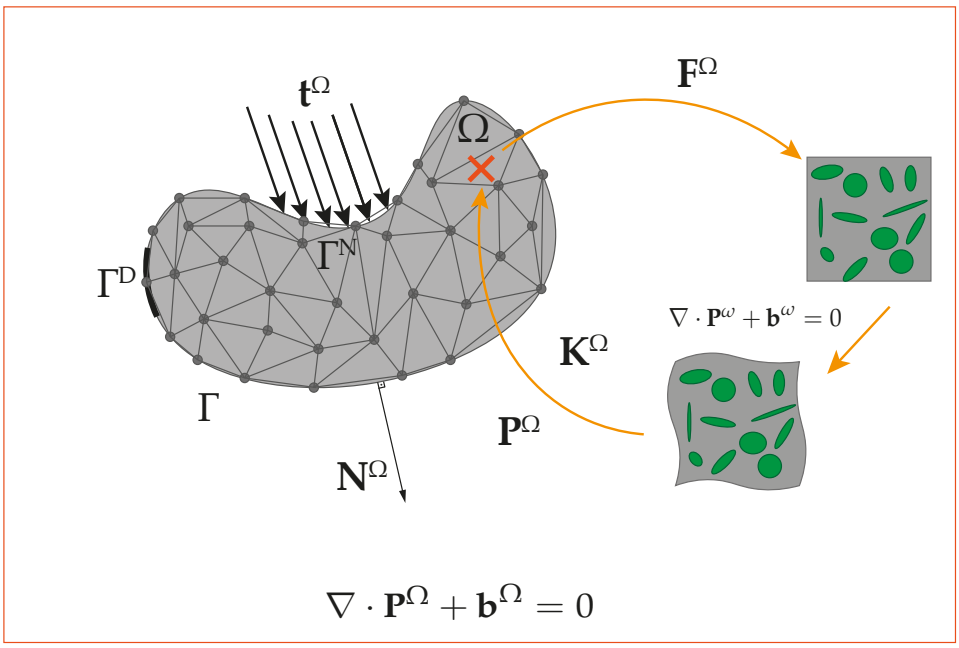


Figure 2.1: Schematic representation of the FE^2 method in reference configuration.

micro-scale problem at every integration point[2]. This introduces a huge bottleneck especially for the non-linear material behavior where the minimization of the force residual takes several iterations even for one incremental step. Thus, it is not hard to imagine that with an increasing number of degrees of freedom affiliated with the fine discretization needed for arbitrary geometries, the computational burden of the method increases tremendously. This hinders the applicability of the method when considering real-world prediction tasks or when searching the design space for an optimal combination of parameters in consideration.

Looking at the macro-scale problem it is explicitly mentioned that there exists no assumption regarding the constitutive relation. In other words, there is no information in the formulation of the problem regarding the underlying materials. Thus, in essence, the computational homogenization scheme introduced can be perceived as querying for macroscopic stress $\mathbf{P}^{\Omega}$ for a given $\mathbf{F}^{\Omega}$ from a micro-scale problem. This nature of the method makes it indifferent to the underlying micro-scale model. To rephrase it, as long as the micro-scale model gives the necessary information to the macro-scale problem the method is still capable of generating solutions. Instead of solving a micro-scale problem, one might decide to switch to a supervised learner $\mathcal{M}$. As long as the learner can approximate the mapping $\mathcal{C}^{\Omega}$ in a fairly accurate manner the overall procedure will be indifferent to the change of the micro-scale problem to the learner $\mathcal{M}$. (See Figure 2.2 for the FE^2 method with a supervised learner instead of solving a macroscopic problem.)

The machine learning applications are not limited to only in the explained fashion. In fact, with the deep understanding of the physics involved and the machine learning methods the field witnessed quite impactful work that might replace the conventional approaches in the near future. The next section is devoted to the preliminary literature review of the machine learning applications in the field of mechanics.

[2]Points within the elements caused by the discretization scheme where all the integrals are evaluated numerically

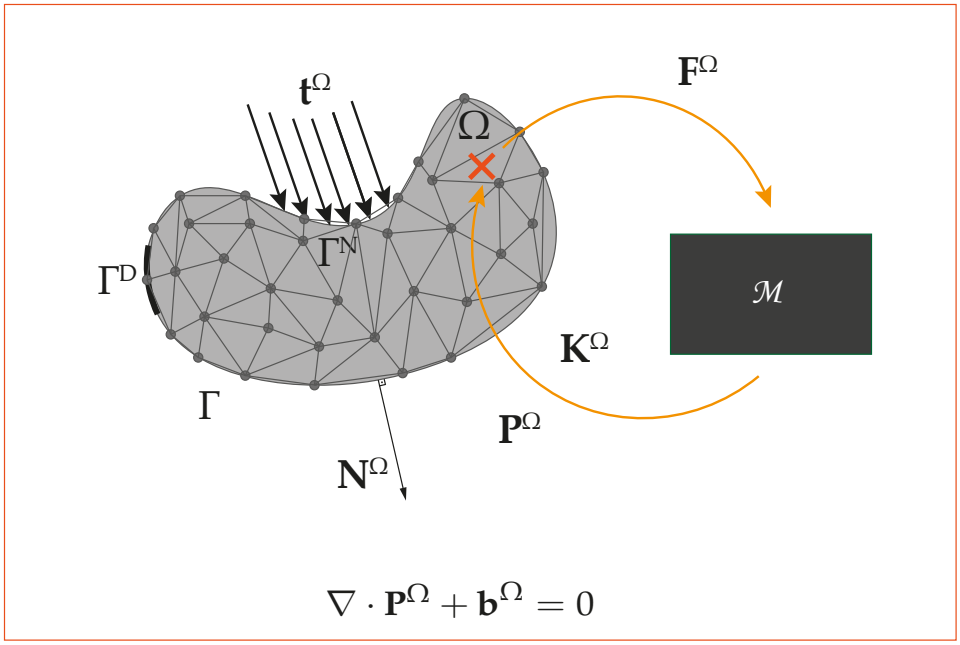


Figure 2.2: Schematic representation of the FE^2 method in reference configuration via a learner $\mathcal{M}$.

3 | Preliminary Literature Review of Machine Learning in Mechanics

As mentioned in Chapter 1, the application of ML methods in the field of mechanics dates way back. However, the increased convenience of the machine learning toolboxes allowed mechanics and material science research endeavors to incorporate it in a mainstream manner. Although the applications of machine learning in material science are relatively diverse and plentiful, the main aim of this chapter is to establish machine learning applications in the mechanics field. However, the interested reader can find a detailed description of the material science field in [82].

3.1 Literature Review

The mechanics community has adopted a wide variety of machine learning applications. The aim of this section is to depict the current landscape of the field regarding the machine learning applications roughly to identify the knowledge gaps, which can guide the direction of the given PhD position. It should be noted that due to the vast application landscape the intricacy of the detailed methods is not of primary importance. Thus, readers are encouraged to divert to the referred work given that unknown terminology is present.

The current state of the field can be roughly divided to mainly two branches, namely *physical* and *non-physical* methods. As the name suggests the *non-physical* methods rely on generating data and using that data on a machine learning pipeline to be used in later applications of mechanics. And, the *physical* methods rely on little to no offline data and try to approximate theory in one way or another. Furthermore, these methods can be further divided into two branches named *theory-informed* and *theory-driven*. For *theory-driven* methods there is no need to supply any additional data to the methods, whereas *theory-informed* methods rely heavily on a small and informative amount of data for the initial push of the model. It should be noted that all of these methods are aimed to be developed in order to circumvent the computational bottleneck of the Direct Numerical Solutions (DNS). Thus, the overall aim of all of these methods is to get good approximations for the well-established methods results in a short period of time which is the main reason for creating surrogate models via machine learning strategies.

3.1.1 Non-physical Methods

In [6] one of the first multi-scale mechanics applications can be found where an offline dataset is generated for the micro-scale problem. The specific applications of machine learning algo-

rithms therein are Artificial Neural Networks (ANN) and Gaussian Process Regression (GPR) on hyper-elastic and plastic material models on an RVE where the overall material behavior depends on multiple descriptors[1]. Due to scalability problems of the GPR half of the application focuses purely on ANN's. Moreover, the computational cost of obtaining the data is decreased by generating the data with the help of a *physical* method. In the given paper a vast dataset is created with all the descriptors that have major contributions to the material behavior and trying to get surrogate replacements with the help of supervised learning methods.

Similarly, in [64] the same procedure is used to create surrogate models. But, the main difference with [6] is the effort that is put on, to model the history-dependency of the plasticity problem. Again a dataset is generated and a surrogate model is created. This is achieved by employing a Recurrent Neural Network (RNN) architecture.

A similar type of application for the plasticity problem can be found in [77]. Again, without considering all the descriptors that lead to different macro-scale behavior a surrogate model is created with the help of ANN's. One of the main contributions of this paper is the implementation of the whole surrogate to the FE^2 framework and the use of the surrogate as a material model (Similar application can be seen in [92] for finite strain elasticity problem). One of the conclusions of the paper is the limited extrapolation capabilities of the *non-physical* method, utilized. Without enough data, the material model collapses, which hinders the utilization of the surrogate in a data-scarce application. Moreover, from the same author, a novel framework is proposed in [78] where the surrogate is created actively. By utilizing a GPR whenever the uncertainty of the prediction exceeds a certain value new data is added to the surrogate model. This work aims to tackle the problem of creating a good enough surrogate in a smart way. This way the surrogate is kept at an interpolation regime at all times, by triggering data collection as the uncertainty exceeds a certain value.

Noting that the FEM is a general method that is utilized for solving PDEs in a structured manner, it can be applied to any other type of problem. For instance, the same FE^2 method is used for multi-scale a nonlinear conduction problem in [58]. In the given work considering volume fraction as the source of uncertainty measure a Monte Carlo simulation is conducted with the help of the created surrogate models to quantify the uncertainty. The only reason for the Monte Carlo simulations to be possible in the given setting is the ability of the surrogate models to make accurate predictions in a short amount of time. Moreover, similar to [77, 78] the surrogate models are plugged into the FE^2 framework to get the macro-scale response of arbitrary domains too. Although ANN's are used as surrogates, this work acknowledges the data dependence of the data-driven approaches. To circumvent this only the input feature that has prime importance for the problem is taken as input and other descriptors are disregarded. An ensemble of networks is used for multiple descriptors which arrive at a final prediction value by a Lagrangian interpolation scheme.

Similar types of *non-physical* methodologies are being published with small tweaks. An example of this can be seen in [40] where supervised learning techniques are used after applying Proper Orthogonal Decomposition on a collection of micro-scale simulation data, for a fast and accurate reduced-order model, to be later utilized in the FE^2 framework. Furthermore, applica-

[1]Descriptors mean the parameters of the model.

tion areas of faster FE^2 with the help of the surrogates created with machine learning strategies are applied to specific problems of their interest.

It should be noted that machine learning applications are not limited to the FE^2 framework. Many problems that have similar structures can be tackled via various ML approaches (*e.g.* [45, 69]), but for the sake of brevity, these applications are left out.

3.1.2 Physical Methods

It should be noted that, unlike the literature, this document has a different nomenclature for the existing *physical* methods than the rest of the mechanics community. Since all the *physical* methods are referred to as physics-informed and the clear distinction between them is often disregarded. However, this literature review takes it into account, although the names of the method might get confusing due to conflicting names that are used in the literature. However, when there exists a certain clash due to nomenclature it is highlighted.

This line of work stems from the downsides of the *non-physical* methods where the ML methods are used as black-box predictors. In other words, the predictions are only based on the information included in the data. This limits the usage of the surrogate models that are created in the extrapolation regime, which exacerbates especially if the models are needed to be used in different settings. Moreover, it is argued in [48] that the physical-information inclusion in the machine learning models decreases the need for abundant data.

Theory-informed Methods
One of the first machine learning-based reduced-order models aimed to tackle heterogeneous multi-scale material models is Self-consistent Clustering Analysis (SCA) presented in [53]. This method uses clustering algorithms to combine similarly behaving regions in a given RVE later to be related using Lippmann-Schwinger [2] equation to obtain a homogenized behavior. Initial dependence of the given method to elastic response and physical information makes this method a *theory-informed* method. This method can be accurately utilized to compensate for the abundant data need of the *non-physical* methods as it is shown in [6]. Similarly, in [91] an equivalent formulation is presented. Recently, this method is improved further in [24] with the addition of the adaptive strategy to eliminate the unnecessary discretization of the RVE in places where it is not necessary. Thus, focusing only on the localized regions more to increase the efficiency and the accuracy of the given clustering-based reduced-order models.

Another important step in the given direction of the theory-informed methods originated from the main author of [53] named Deep Material Network (DMN) [56]. With the hopes of tackling the same multi-scale homogenization problem. This method is a bit more theory-based but still relies on offline data obtained from the elastic material response. Overall, the aim is to obtain a homogenized compliance matrix with the help of a deep neural network like architecture where each component of the compliance matrix is assumed to be a neuron. This method is shown to capture complex history-dependent plasticity and hyper-elasticity material behaviors

[2] A special version of Schrödinger equation which is a linear PDE that governs the wave function of a quantum mechanical system.

and provide a good basis for the acceleration of the method due to the little data that is needed to obtain a surrogate. Although its effectiveness authors acknowledge the improvement paths to be in the direction of understanding the training procedure and investigating different data-sampling strategies for *non-physical* stage.

Theory-driven Methods

One of the most significant works in this line of research is presented in [71] under the name Physics-Informed Neural Network (PINN). This type of methodology is developed with the intention of general usage of solving PDEs in a more general way unlike the methods presented up to now, so the application range of this method is incredibly wide. As mentioned before the nomenclature that is used in this work conflicts with the overall literature. This line of work is coined as Physics-Informed (PI), although the process is not just informed, but completely driven by it. That is the reason here this method is categorized under the *theory-driven* method. The main reason for the development of the PINNs is the unreliability of the modern machine learning techniques under data-scarce settings. Acknowledging the accumulated prior physics knowledge and incorporating it in the process of learning is the main solution for the unreliability problem. Incorporating the desired differential equations to be satisfied together with the boundary conditions in the loss function the solution of the minimization problem with the spatial coordinates as inputs and the desired field as the outputs PDEs can be solved in a time-efficient manner. It is worth mentioning that the automatic-differentiation tools come in handy in the given formulation of the problem, which makes the whole process easy to handle. In the subsequent works, many different problems are tackled with the proposed approach in [79, 86, 90, 94, 96]. The authors of [71] do not propose this general framework to replace the conventional methods, but they claim this method will favor rapid development and easy testing for various problems in a reliable fashion. The most obvious downside of this formulation is the inability to have a specific application. In other words, every problem that is newly encountered has to be reformulated and a separate model has to be created, but this can be perceived as an advantage as well.

Upon introduction of this general framework more specific applications are derived from the same principles. For instance, in [32] the strain energy density function is tried to be approximated in terms of invariances with GPR in a completely physical formulation thanks to easy analytical expressions of GPR. Moreover, in [59] a thermodynamics-based ANN approach is proposed to cope with more general material states without additional tweaking of the model. With the help of automatic differentiation similar to PINNs first and second-order relationships are established overall.

A recent and novel approach that is similar to [59] is presented in [52] where the strain energy density function is tried to be approximated with a physically-sound NN architecture. The most important contribution of [52] is the separation of essential inputs from the additional information that can come in and for a computational solid mechanics problem that is trying to determine the relationship between stress and strain the only essential component is these two components. The additional information is added to the approximation of strain energy density function at a later stage if there is any. This division between the essential and the non-essential information regarding the representation of material behavior in a computational sense can be seen as a big step in the general applicability of the theory-driven methods in a specific applica-

tion field.

3.2 Possible Problems with the Current Approaches

This section is dedicated to identifying the possible problems with the current ML applications in the context of solid mechanics described in Chapter 2. Moreover, a personal reflection on those problems is presented.

Risky Extrapolation As mentioned before, one of the main downsides of the *non-physical* methods is the inability to have extrapolation capabilities. The input space of the deformation gradient can be thoroughly collected with the hopes of always staying in the interpolation while querying the surrogate model. The feasibility of data collection via DNS might be unfeasible. This necessitates the need for data collection employing surrogate models too. An example can be seen in [6] where a surrogate replacement is utilized, named Self-consistent Clustering Analysis (SCA) [53], to replace DNS solutions. It should be noted that any surrogate model that is capable of generating accurate enough data can be utilized can be a deep material network [55]. However, if no surrogate modeling technique is available and there is too little data the problem is data-scarce.

Data scarcity This problem is the main reason why the *non-physical* methods are not useful. Moreover, as more information is involved in the micro-scale constitutive relation like damage modeling or multi-physics applications the data scarcity problem exacerbates. The most suitable direction, in this case, is to utilize theory-driven methods where without additional data need the underlying phenomenon is tried to be captured.

Single Parameter Configuration Most of the proposed methods focus on a particular type of problem with a limited parameter configuration of the selected material model. This is a problem, partially addressed in [6] where the most relevant parameters are selected via sensitivity analysis then added to the ANN as a feature while creating the surrogate model. This allows relevant parameter space to be sampled for just a combination of two materials. But, a combination of other material models is left out, which indicates that although there exists an effort for general applicability, there is still room for improvement as the finite amount of input data can be provided a more general framework need is evident. Moreover, this problem is the main limiting factor for the general applicability of *non-physical* methods.

General Applicability The most general surrogate modeling technique in the literature is PINN presented in [71]. In the spirit of [71] various more specific applications flourished. However, highly specific applications can be a limiting factor as well. Thus, a sweet spot should be found to ensure the general applicability of the sets of problems that might be faced. Considering the problem related to constitutive relation determination, the separation of the essential inputs from the non-essential ones presented in [52] serves well for general applicability.

Personal Perspective Most of the mechanics groups are not as open with their data as the computer science field as often one may fail to find a place for the data and the source of the data.

However, with the years of experience, these groups are assumed to have collected/generated data that is relevant to their problems of interest over the years. For instance, some groups might focus on polymers and some can focus on crystalline materials. Although the introduction of the ML methods in these mechanics communities is shifting in terms of sharing data and methods, there is still problem-specific applications as a general downside of most publications in terms of the relation between ML and mechanics. This problem combined with the semi-structured and volatile data [20] in the mechanics community makes the *non-physical* models not generally applicable. On the other hand, due to the generality of the *physical* methods (at least to some extend), they are getting more traction, with a big impact on the mechanics community. Looking at both *physical* and *non-physical* research endeavours. It can be observed that due to immense amount of physics knowledge accumulated over the years has serious advantages. However, getting the most out of *non-physical* methodologies is being overlooked while trying to incorporate strong prior information.

Hypothesis 1. After considering and formulating the problem in a better way most of the problems that stem from the current applications can be tackled more effectively with more general use cases.

Due to this hypothesis, this work intends to have a different perspective for ML techniques applied in the context of solid mechanics and more specifically multi-scale composite modeling, which leads to the following hypotheses.

Hypothesis 2. Although, many people work on different problems with different parametrizations there is something intrinsically similar due to the above-described nature of the problem at hand, which is the determination of the constitutive relation. In other words, even though modeling assumptions or validity regions of these models change, the core of the problem pertains, finding a good mapping between a deformation measure (*e.g.* deformation gradient, strain measures) and a force measure (*e.g.* stress measures).

Hypothesis 3. Similarity between different problems is not exploited fully in the current literature with the current available *non-physical* methodologies, due to the current formulation of the problem in terms of ML.

Upon this discussion, the next chapter is dedicated to the reformulation of the current problem at hand given in Equation 4.1 and possible research questions that might arise, given the problem definition.

4 | Learning Problem

This chapter aims to reformulate the learning problem in the context of the computational homogenization scheme presented in Chapter 2. As mentioned before the learning problem involves the substitution of micro-scale BVP solution with a learning algorithm denoted by $\mathcal{M}$ as can be seen in Figure 2.2.

A material in a computational sense is nothing but this constitutive relation, where for every deformation-induced there exists a certain force generation within the continuum. In other words, this constitutive relation is the most essential constituent to represent a material computationally. Again, considering the querying of macroscopic stress $\mathbf{P}^{\Omega}$, the input of macroscopic deformation gradient $\mathbf{F}^{\Omega}$ is a must and this process is indifferent to all the underlying assumptions regarding modeling. Thus, all the other variables (*e.g.* material model selection, parameters of the given model, geometrical parameters of the RVE, etc.) can be decoupled from the basic mapping problem and can be viewed as different learning tasks. Then, a set of M material models can be represented as stress-strain pairs with different assumptions as,

$$\left(\mathbf{P}^{\Omega}{}_i = \mathcal{C}_i^{\Omega}(\mathbf{F}^{\Omega})\right)_{i=1}^{M} \tag{4.1}$$

where, subscript i intrinsically represents a set of possible parameters that can come from modeling assumptions, thus affecting the overall constitutive relation. This equation means that for every assumption involved there exists a constitutive manifold. And by not fixing the input of the function $\mathcal{C}^{\Omega}$ to a certain number of parameters, the problem definition is kept on a general basis. However, this generality comes with a cost, that is sampling of this infinite-dimensional assumption space is a near-impossible job. The best one can do is to learn continuously from the space of assumptions as new assumptions are unraveled from that space. This unraveling might happen in bunches or one at a time depending on the origin of the data (stress-strain pairs), but this is not the prime importance at this stage.

The next section is devoted to establishing the nomenclature of the learning problem and research questions that naturally arise from the formulation of the problem.

4.1 Notation

As mentioned before a constitutive relation is determined by a series of assumptions and those hierarchical assumptions lead to a parameterized model (*e.g.* deformation assumptions, behavior assumptions, material model assumptions, geometrical assumptions, etc.). For a specific material,

this collection of assumptions can considered be arranged hierarchically as,

$$\mathcal{A} := \{a_n \subset a_{n-1} \subset \cdots \subset a_1\}, \tag{4.2}$$

where a represents a specific assumption in the given hierarchical assumption space $\boldsymbol{\mathcal{A}}$. Then, under a series of assumptions $\mathcal{A}$ learning task can be defined as $\mathcal{T}_{\mathcal{A}} : \mathbf{F}^{\Omega} \mapsto \mathbf{P}^{\Omega}$, then a material response manifold with series of assumptions can be represented as

$$\mathcal{C}^{\Omega} := \{\mathcal{T}_{\mathcal{A}} | \mathcal{A} \in \boldsymbol{\mathcal{A}}\}. \tag{4.3}$$

Remark 2. A draw of a series of assumptions lead to a task generation. In other words, the task space is created by assumptions.

Moreover, due to the hierarchical structure and the ability to be parametrized of after the last assumption, the subset of assumptions can give rise to a family of material response manifolds that can be expressed as

$$\boldsymbol{\mathcal{C}}^{\Omega} := \{\mathcal{T}_{\mathcal{A}_i} | \mathcal{A}_i \in \boldsymbol{\mathcal{A}}, \quad i \in \mathbb{Z}^+\}. \tag{4.4}$$

4.2 Preliminary Research Questions

Although, for now only a vague explanation is given for the assumption space, the next subsections are dedicated to the overarching questions that arise from the given problem definitions.

4.2.1 Prediction of a new mapping given another mapping

Research Question 1. Given an assumption $\mathcal{A} \in \boldsymbol{\mathcal{A}}$ which results in the mapping $\mathcal{T}_{\mathcal{A}}$ how and to what extent can the mapping $\mathcal{T}_{\mathcal{A}'}$ be predicted where $\mathcal{A}' \in \boldsymbol{\mathcal{A}}$ is a new and unseen assumption.

Remark 3. This research question aims to tackle a similar problem to a supervised learning setting where the learning task involves the prediction of a label. However, instead of labels Research Question 1 aims to predict a whole mapping.

In the light of the Research Question 1 several interesting research questions naturally arise. These research questions can be summed up in 3 branches, namely sampling, data, and model-related questions.

Sampling Related Question

Remark 4. It should be noted that these two sampling Research Questions are ill-defined since there is no underlying $\mathbf{F}^{\Omega}$ distribution. Thus, referred sampling strategies are non-probabilistic (*e.g.*pseudo-random, pseudo-uniform, active learning or deterministic similar to collecting major directions (See Figure 4.1) as proposed in Research Question 2). Denoting the $\mathbf{F}^{\Omega}$ sampling rule as $q_{\mathbf{F}^{\Omega}}$ and given a conditional constraint c, conditional sampling rule can be represented as $q_{\mathbf{F}^{\Omega}|c}$.

Research Question 2. To what extent sampling only the major directions (see Figure 4.1) in $\mathbf{F}^{\Omega}$ can allow the prediction of a new and unseen mapping $\mathcal{T}_{\mathcal{A}'}$?

Research Question 3. If the major direction sampling is not enough, can a good sampling rule $q_{\mathbf{F}^{\Omega}|c}$ be found to improve the prediction for unseen mapping $\mathcal{T}_{\mathcal{A}'}$?

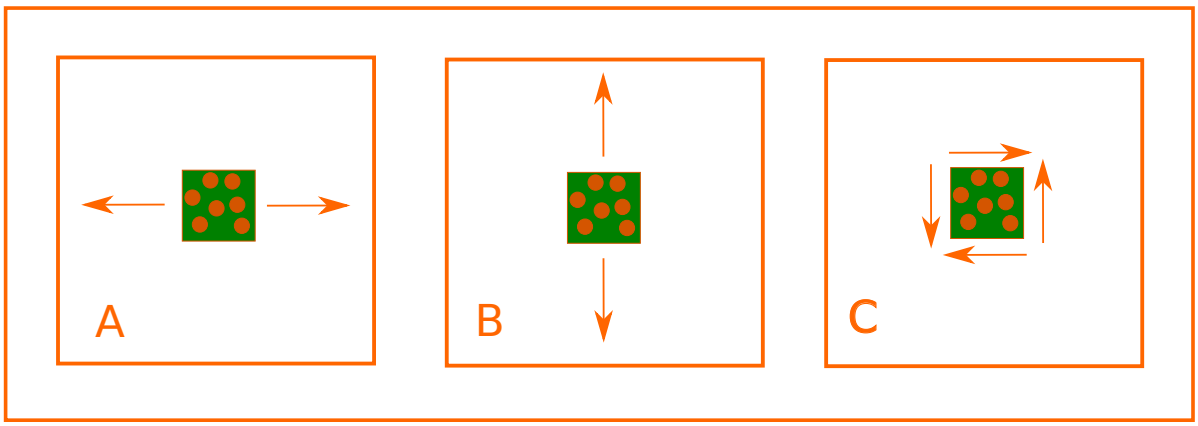


Figure 4.1: Major loading directions for $\Omega \in \mathbb{R}^2$, which implies $\mathbf{X}^{\Omega} \in \mathbb{R}^2$ and $\mathbf{x}^{\Omega} \in \mathbb{R}^2$. Then, deformation gradient $\mathbf{F}^{\Omega} \in \mathbb{R}^{2x2}$ components corresponding to the major loading directions are A:f_{11}, B:f_{22} and C:f_{12} and f_{21}

Data Related Question

Research Question 4. To what extend information coming from the $(\mathbf{F}^{\Omega}, \mathbf{P}^{\Omega})$ pairs in the $\mathcal{T}_{\mathcal{A}}$ can improve the prediction of the new mapping $\mathcal{T}_{\mathcal{A}'}$?

Remark 5. Since the data coming from the previous task can be utilized in various ways (*e.g.*reweighing the instances, finding a feature space that the predictions in both tasks are knowledge transfer, etc.) all of these questions can be considered under the Research Question 4.

Model Related Question

Research Question 5. Given a model $\mathcal{M}$ parametrized by $\boldsymbol{\theta}$ trained for $\mathcal{T}_{\mathcal{A}}$, to what extent does the $\boldsymbol{\theta}_{\mathcal{A}}$ help to the new and unseen prediction of mapping $\mathcal{T}'_{\mathcal{A}}$?

Remark 6. Research Questions 5 only considers the model parameters as the medium and nothing more. It can be assumed that suddenly all access to the data coming from $\mathcal{T}_{\mathcal{A}}$ is magically lost.

4.2.2 Prediction of a mapping given a batch of mappings

Most of the time the data at hand might come from different places and from different applications, which makes it interesting to investigate to what extent the information obtained from multiple tasks can improve the prediction of a single mapping. This scenario is expressed as can be seen in Equation 4.1, which describes M, different material models. Thus, in a general sense, the learning problem at hand given M material simulations can be represented as

$$\left(\mathcal{T}_{\mathcal{A}_i}\right)_{i=1}^{M}. \tag{4.5}$$

Remark 7. This problem assumes that there exists a commonality/similarity exists between the tasks.

If such similarities exist between the tasks several other research questions naturally arise.

Research Question 6. Given a batch of assumptions $\{\mathcal{A}_i \in \boldsymbol{\mathcal{A}}, \quad i \in \mathbb{Z}^+\}$ which results in a batch of mappings $\{\mathcal{T}_{\mathcal{A}_i} \quad i \in \mathbb{Z}^+\}$ how and to what extent can the mapping $\mathcal{T}_{\mathcal{A}'}$ be predicted where $\mathcal{A}' \in \boldsymbol{\mathcal{A}}$ is a new and unseen assumption.

Remark 8. It should be noted that Research Question 6 can be structured similarly to the problem given in Subsection 4.2.1. In other words, data, sampling, and model-related research questions are all valid research questions in the given problem setting too.

Given that such similarities exist between the tasks, another question arises naturally from the formulation of the problem, where the parameters defining the material models are not taken as input for the learning problem.

Research Question 7. Given a batch of assumptions that lead to different mappings, can we find a latent space where the similarities between different tasks can be exploited to improve the prediction of $\mathcal{T}_{\mathcal{A}'}$ compared to individual learning?

An interesting research endeavor that is not mentioned in previous subsections can be related to the training of a model $\mathcal{M}$.

Research Question 8. Given a batch of tasks $\{\mathcal{T}_{\mathcal{A}_i}, \quad i \in \mathbb{Z}^+\}$ can a model $\mathcal{M}$ parametrized by $\boldsymbol{\theta}$ be trained to allow fast adaptation to predict new and unseen task $\mathcal{T}_{\mathcal{A}'}$?

Remark 9. It should be noted that the Research Question 8 is related to a base model training, from which adaptation is fast and accurate. In other words, this research question mainly focuses on finding the best $\boldsymbol{\theta}$ to start the training for task $\mathcal{T}_{\mathcal{A}'}$.

Another interesting research path is related to sampling on a higher level. Due to the unique nature of the problem (*e.g.*each task representing a mapping), if there exists a possibility to sample tasks non-probabilistically from the assumption space $\boldsymbol{\mathcal{A}}' \subseteq \boldsymbol{\mathcal{A}}$.

Research Question 9. Given $\boldsymbol{\mathcal{A}}' \subseteq \boldsymbol{\mathcal{A}}$ how can the assumption space $\boldsymbol{\mathcal{A}}'$ be sampled to improve the prediction of a new and unseen task $\mathcal{T}_{\mathcal{A}'}$ where $\mathcal{A}' \in \boldsymbol{\mathcal{A}}$ or $\mathcal{A}' \in \boldsymbol{\mathcal{A}}'$.

Remark 10. The question regarding the assumption space can be posed through task space as well as one is the cause of the other.

Remark 11. In the grand scheme of predicting a mapping given a batch of different mappings, there can be a scenario where after the prediction of the single mapping it is added to the batch as well in a continual manner. This opens up a whole new research path for the given problem.

Research Question 10. To what extend $\mathcal{T}_{\mathcal{A}'}$ can be predicted given $\{\mathcal{T}_{\mathcal{A}}, \quad i \in \mathbb{Z}^+\}$ and after each prediction, $\mathcal{T}_{\mathcal{A}'}$ is added to the batch to be utilized in the prediction of $\mathcal{T}_{\mathcal{A}''}$ where $\mathcal{A}'' \in \boldsymbol{\mathcal{A}}'$.

4.2.3 Prediction of a batch of mappings

In certain settings, the aim can be to find a predictive model for the bunch of mappings resulting from different assumptions. As mentioned before due to the hierarchical response of the assumption space, subsets of this space can give rise to material response manifolds $\boldsymbol{\mathcal{C}}^{\Omega}$. In the cases where the interest is not only in the application of a certain material with certain parameter values, but the whole physical response of the given material independent from the material model parameters. Moreover, similar to Subsection 4.2.2 this batch of tasks can be combined with predicting a single mapping task $\mathcal{T}_{\mathcal{A}'}$ too. However, for the sake of simplicity these two research paths can be separated. Then the broad research question can be expressed as,

Research Question 11. Given a subset of assumptions $\boldsymbol{\mathcal{A}}' \subseteq \boldsymbol{\mathcal{A}}$, to what extent $\boldsymbol{\mathcal{C}}^{\Omega}$ can be predicted?

Remark 12. Research Question 11 is concerned with only the construction of the whole $\mathcal{C}^{\Omega}$ construction rather than focusing on a single task. This type of construction can be utilized to store general material behaviors in an endeavor to predict unseen material behaviors $\mathcal{C}^{\Omega'}$. Moreover, this research question is especially important as the more general the assumption space is the mappings tasks associated inside the given subset of assumption space will be related.

Another important research question can be posed assuming that for any given assumption $\mathcal{A} \in \boldsymbol{\mathcal{A}}$ the mapping $\mathcal{T}_{\mathcal{A}}$ is available from an oracle. Denoting the sampling rule for a given task $\mathcal{T}_{\mathcal{A}}$ as $q_{\mathcal{T}_{\mathcal{A}}}$ sampling related question can be posed as,

Research Question 12. How to design a good sampling rule $q_{\mathcal{T}_{\mathcal{A}}}$ to come to a good estimate of $\mathcal{C}^{\Omega}$?

Remark 13. It should be noted that Research Questions 2 and 12 can be combined to find the optimal design of experiments in the given setting. Otherwise just considering a certain sampling strategy for $\mathbf{F}^{\Omega}$ and looking for a good sampling strategy for tasks is an interesting investigation by itself.

4.2.4 Discussion of the Proposed Research Questions

Although there are similarities between the research questions, Section 4.2.1 questions can be related to the transfer learning, Section 4.2.2 questions can be related to meta-learning (learning-to-learn) and Section 4.2.3 can be related to multi-task learning paradigms in a vague sense. One can argue that all of these paradigms are related in a general sense, where information transfer is the aim from different aspects, however, there are some fundamental differences between these paradigms.

Transfer Learning Considering there is a source and a target domain, it tries to exploit the similarities between two different tasks to improve the prediction of the model created with the source domain [93]. In the explained setting one can match the source domain to an observed task and the target domain to the targeted task explained in Section 4.2.1.

Meta Learning(Learning-to-learn) Given multiple learning tasks this paradigm aims to find ways to search for the best strategy of learning to have a good prediction in the possible upcoming task [89]. As mentioned before this paradigm is related to investigating the posed research questions given in Section 4.2.2.

Remark 14. It should be noted that the Research Question 10 can be related to the continual learning paradigm, but it can be considered under the incremental meta-learning paradigm too since both of them try to tackle a similar type of problem where the batch of tasks is increasing and information obtained from the previous batch of tasks are used to improve the possible upcoming task prediction.

Multi-task Learning Given a batch of tasks this learning paradigm tries to learn all of the tasks at the same time. In other words, it aims to find a learner that is capable of performing well overall the tasks in a given batch [10].

As can be inferred from the above descriptions the all of the paradigms try to exploit latent information under certain scenarios to improve the prediction capabilities of a learner. Thus, one might conclude that there are no essential differences, but subtle ones that originate from

the problem at hand and the aimed result.

As mentioned before the mechanics problem can often be regarded as learning a mapping $\mathcal{T}$ for a single parameter configuration of the given material and geometrical model. Since the material models are generally parameterized models, the relationship given in Equation 4.1 is often the case. In other words, given a material model and a geometry, although the interest lies with the specific parameters of the model at prediction time, there is access to a batch of material models (in the given context a task $\mathcal{T}$). This makes the investigation of research questions presented in Section 4.2.2 relevant. Although the other research questions cannot be disregarded completely due to interwoven relationship between all of the presented questions the weight of this research endeavor will be given to the Section 4.2.2. Moreover, the fact that the solid mechanics community has not yet shifted attention to learning-to-learn material models from the data with the help of *non-physical* methods, this research path seems to have a lot of potential for the investigation of the learning-to-learn paradigm in a setting, which machine learning community has not access too. Because of this reason the next section is dedicated to the preliminary literature review of the meta-learning (learning-to-learn) paradigm.

5 | Preliminary Literature Review of Meta Learning (Learning-to-Learn)

This chapter is dedicated to the literature review of the meta-learning paradigm. However, before diving into this paradigm some preliminary definition of what is meant by learning in this document is needed.

Definition 1. Given a task, and training experience with a performance measure, a learner assumed to be learning if the performance increased with the experience [85].

Following the Definition 1 the learning-to-learn (meta learning) concept can be defined as,

Definition 2. Given a set of tasks and set of training experiences with a collection of performance measures, the learner is learning-to-learn as the training experience improves with the increase of experience and with task set size. [85].

Noting the fact that all learning systems try to reduce the bias of a hypothesis space $\mathcal{H}$ during the adaptation/training [89], learning-to-learn can be viewed as a system that learns its bias dynamically while distilling the information from the tasks that the learner is exposed to. With an analogy to the learning of humans, learning-to-learn can be explained as the ability of a human to adapt and use the information obtained from previously experienced tasks to a new and unseen task which later on is added to the experience of tasks to make another new task learning easier. For example, knowing how to ride a bike and a motorcycle share a common structure, so upon learning how to bike, a human will have the ability to detect the common structure between biking to using a motorcycle to make its learning of motorcycle riding task easier, compared to the case of not knowing how to ride a bike. In a similar context learning from multiple tasks at the same time is known to improve the predictive performance of the learner, given the tasks share a common underlying structure [21]. One can make the same analogy to human learning by, learning to ride a motorbike becomes an easier task if a person tries to master riding a bike.

Abstractly the learning-to-learn paradigm can be depicted by considering Figure 5.1. Considering the region D can be learned the best by learner $\mathcal{M}_D$, similarly, for the region, A it can be learned the best by $\mathcal{M}_A$. The main aim of the meta-learning is to learn intrinsic properties of the region A and D that make the learners $\mathcal{M}_A$ and $\mathcal{M}_D$ and both learners intrinsic properties that make them favorable in their corresponding regions. It should be noted that the application area of the meta-learning paradigm is not limited to the observed task regions. Its generalization capabilities can be utilized for tasks that are outside the scope of the learner. For instance

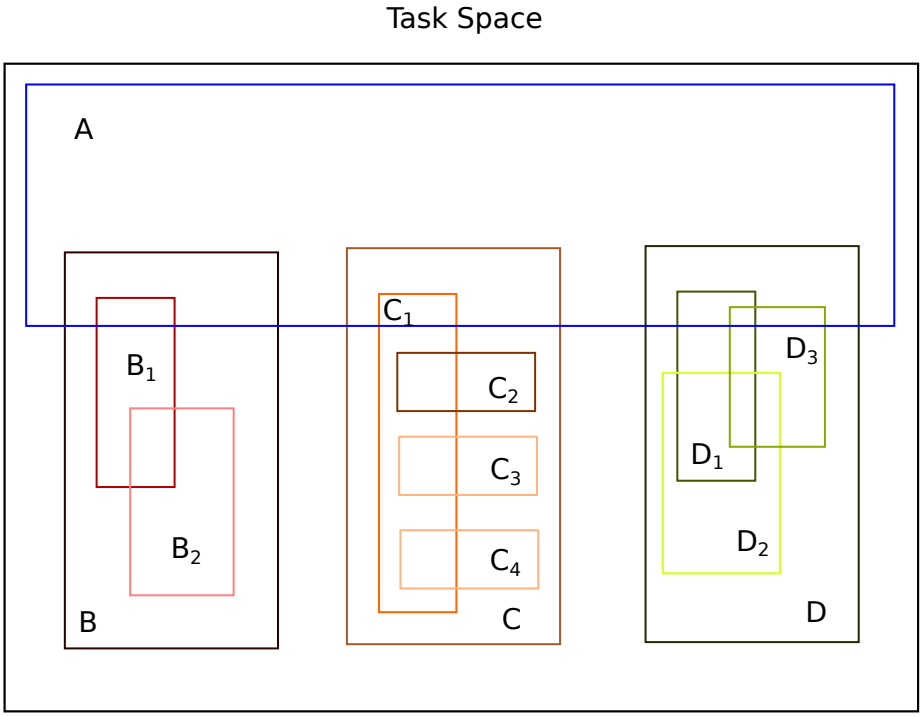


Figure 5.1: Illustration for Task space which is an arbitrary space. For solid mechanics problem this space can be considered as the space of assumptions regarding the material modelling.

a meta-learner $\mathcal{M}_{meta}$ can be utilized to combine base predictions of $\mathcal{M}_A$ and $\mathcal{M}_D$ to make predictions about region C.

Considering the upon descriptions, a preliminary literature review can be found in the next sections. Looking at the survey of meta-learning paradigm given in [89] the meta-learning endeavor can be traced back to the 1980s, but for the sake of depicting the current state of the field related to this paradigm the big part of this literature review is dedicated to the recent meta-learning approaches. Moreover, it should be noted that [70] provides a comprehensive overview of the meta-learning paradigm, in which the paradigm is summed up under 4 main branches. However, due to fact that the line between the methods that are put there is blurred, this document does not categorize this paradigm even further.

5.1 Preliminary Literature Review

Given that there are multiple learning tasks memory can be considered an important aspect for the learner. The memory in this context can be understood as the ability of the learner to detect the similarities between tasks that it encounters. In [43] it is used to train recurrent neural networks that are used as meta learners. Moreover, [3] utilizes a similar idea to train Long Short-Term Memory (LSTM) architectures, [80] utilizes Neural Turing Machines, [65] uses a Meta Network that has internal memory, for meta-learning problems. Moreover, [62] uses attention

blocks to get a similarity measure between tasks to get the most relevant ones from the memory.

Most of the above-mentioned methods are introduced around the same time, but one of the most traction is gained in gradient descent based meta-learning can be traced to Model-Agnostic Meta-Learning (MAML) [26]. This method utilizes the model parameters as a memorizing agent in a sense compared to the above-mentioned methods built-in memory architectures. Thus, MAML is easily applicable to any differentiable model and can be utilized in supervised and reinforcement learning problems. This allowed the MAML to be applied on a wide basis. The main of the MAML approach is to search for a warm initialization point in the parameter space of a model (can be any parametrized model), given the distribution of tasks. This warm initial point is found by minimizing the future empirical risk. One of the main disadvantages of MAML is the inclusion of the Hessian products and this problem is tackled in the same paper by creating a first-order approximation, although later on in [22] it is shown that there are no convergence guarantees of the first-order approximation, thus another version of MAML which is Hessian free is introduced. The dependence of the model on the learning rate and adaptation rate of MAML is acknowledged in [51] through another method called Meta-SGD. Another MAML improvement is presented in [30], which tries to learn to create a loss landscape where the minimization problem is easier to solve for the task distribution. In the given set of MAML, all the tasks have equally weighted contributions to the minimization problem, but in [8] a MAML variant for the case where the tasks are not equally important is presented. In a similar fashion in [13], the worst possible future empirical risk is minimized to find an initialization in the MAML framework.

In [39], an empirical study is presented for the generalization of gradient-based meta learners, specifically the MAML method. The paper concludes that the loss while adapting to new tasks becomes flatter and lower in the loss. Moreover, it is shown that, even when the generalization degrades this is the case. This is provided as an indication of the generalization capabilities not originating from the flatter loss landscape. Furthermore, recently in [23], MAML is theoretically investigated in a convex setting, and its ability to generalize well to a recurring task is acknowledged. However, the ability of MAML to a new task is shown to depend on the variance of task distribution. In other words, MAML's generalization ability decreases as the variance of the task increases.

For the settings where the prediction confidence is important (*e.g.* fingerprint detection), it is important to have a certainty measure along with the inference as well. In [27] a method named PLATIPUS, which is an extension to deterministic MAML method. By introducing noise to the gradient descent during the meta training stage and optimizing concerning a variational lower bound on the likelihood, an approximate posterior is obtained conditioned on the training set. Moreover, another probabilistic treatment of MAML named VAMPIRE can be found in [67], where a prior is assumed over the parameters of the model and the posterior is obtained via variational inference. Another recasting of MAML as a hierarchical Bayesian Model can be found in [38] where a Laplace approximation technique is used for likelihood calculation, which performs badly in a multi-modal distribution setting.

As mentioned the early meta-learning approaches are focused on memory-based architectures and the last 5-6 years are dominated by MAML and its variants, but there are several other

works in other directions as well. ALPaCA [41] is one of these methods, where all the prior information (kernel selection and hyper-parameters) for a Gaussian Process Regression is tried to be obtained by employing a meta-learning algorithm. Another interesting endeavor is presented in [5] where one of the most important parts of a machine learning pipeline, the loss function, is aimed to be meta learned from the given tasks. In [15, 16] it is shown that meta-learning problems can be tackled by more fundamental approaches in convex problem settings. In this method, Ridge Regression is employed, and by minimizing the transfer risk which measures the true expected risk of a learner on the task distribution meta-learning is performed. It should be noted that the study of the convex problem allows the bounds for the presented methods to be presented as well.

The strong assumptions of many meta-learning methods regarding the task distribution is a known issue in the field not only for MAML and its variants but also for the other type of methods. One such assumption is that there are enough tasks available in the task space. In case that is the case [95] proposes a framework that interpolates between the given tasks to populate the task space and make the meta-learning easier. Another assumption is that meta training and meta test tasks are originating from the same distribution, to solve that issue domain adaptation techniques are employed in the task space in [54] (See [47] for theoretical investigation of this domain shift problem.). The heterogeneity of the task distribution presents a multitude of challenges for the meta-learning methods. Especially for the MAML like methods where the available task information is tried to be compressed only to deterministic parameters of the model. There are several MAML variants where this problem is tried to be tackled. In [29] by minimizing the loss trajectory for each task drawn from the task distribution while trying to minimize the future empirical risk. TreeMAML [33] is proposed where a tree-like adaptation step is utilized to avoid the performance degradation caused by the task distribution involving unrelated tasks. Methods other than MAML variants try to tackle this problem in various ways as well. In [18] this problem is tackled by learning a conditional function mapping on the meta-level, for the problems where multiple clusters of tasks exist in the task space.

Although the above-mentioned meta-learning methods are quite useful in batched learning settings, their applicability decorates for the unknown, ever-changing task distribution, which relates the meta-learning paradigm to the continual learning paradigm. In [46] a Dirichlet process mixture of hierarchical Bayesian models over the parameters is employed to tackle this problem caused by squeezing to a set of hyper-parameters. In [17, 19, 28, 42, 73] the catastrophic forgetting is prevented in a sequential task learning setting for unknown ever-changing task distribution.

5.2 Learning-to-learn Material Models

As mentioned before traditional machine learning applications often try to tackle the solid mechanics problems separately, however, this PhD aims to investigate the effect of learning multiple problems at the same time in increments or batches compared to individual task learning. To the author's knowledge, there is no current work that is published in the mechanics community that relates the meta-learning paradigm to solid mechanics. However, some of the recently published papers utilize transfer learning approaches to overcome the excessive amount of data needed for some applications. Employed strategies do not go beyond keeping some parameters

from a previously trained model to allow faster convergence or better accuracy of the problem in consideration.

One of the examples of transfer learning can be found in [97]. The main of the paper is to analyze the Kirchhoff plate model with a special auto-encoder architecture that takes into account the total energy inspired by the methodology PINNs that is presented in Chapter 3. The transfer learning comes in to picture, for the case when there exists a trained model that has certain parameters and the problem of interest is on another parametrization. In this case the, current model is used as a starting point for the problem of interest to transfer the inductive bias with the help of model parameters. Another PINN informed method for fracture modeling with phase-field type problem solution technique in [36]. Due to the current formulation, the model has to be trained for every incremental displacement step and transfer learning (through the old displacement step as the starting point for the next incremental step) is employed. It should be noted that for both of these papers the main aim is not to perform well with a small amount of data (since the PINN type of methodologies does not require training data), but to get a faster convergence for a new problem.

A framework not only aimed at acceleration but also decrease the need for the data can be found in [57]. As mentioned in Chapter 3 the DMN method needs linear elastic simulation data. Since it is not feasible to train for every microstructure realization (an RVE with different parameters and geometry) the goal is to train only a few realizations of RVEs with DMN and interpolate the rest of the task by transfer learning. By doing so the whole database can be created more efficiently for a given microstructure assumption. This method is then combined with the FE^2 methodology to decrease the computational burden.

5.2.1 Current Formulation for Material Model Learning

Considering the subspace of tasks which is presented by the region D in Figure 5.1. In the current material model learning setting this space can be considered as a family of material responses. For instance, D_1 can represent a certain parametrization of a material, whereas D_2 and D_3 represent other parametrizations. Moreover, focusing on the Figure 5.2 specifically the region marked by the dashed pink circle represents an intersection where some different parametrizations of models live. Zooming into this region we end up with material points where each label marked by crosses, circles, and triangles represent a certain $\mathcal{C}_i^{\Omega}$ as given in Equation 4.1. It should be noted that these points correspond to whole material response manifolds as can be deducted from the figure. Moreover, the tasks spaces being a result of assumptions as mentioned in Chapter 4 the relation between learning material models and the learning-to-learn problem can be seen much vividly.

In the light of the literature review presented in this chapter and Chapter 3, combined with the proposed research questions given in Chapter 4, the next chapter is dedicated to the PhD project planning.

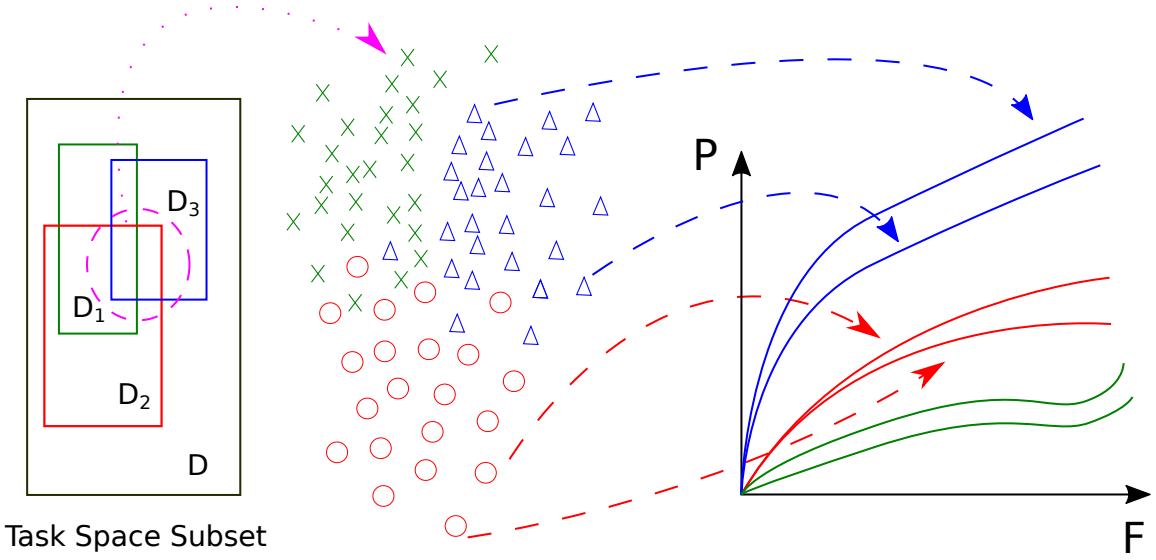


Figure 5.2: Visual for the current formulation of learning material models.

6 | Aim, Progress and Planning

As mentioned in the previous chapters the main aim of this work is to accelerate a multi-scale numerical analysis technique called FE^2. This will be achieved by replacing the micro-scale boundary value problem with a learner $\mathcal{M}$ as can be seen in Figure 2.2. The learner is aimed to have the following properties:

- Capability of utilizing the past information either by means of data or any other memorization technique of the past knowledge gained to be utilized in prediction of an unseen task.
- Ability to extract information from a batch of tasks, to be utilized in the prediction of an unseen task.
- Limited data demand from the unseen task.
- Generalize well for the unseen task.

These above-mentioned properties can be identified with the objective of the learning-to-learn paradigm. Looking back at the literature review it can be observed that generalization problems are originating from the assumptions of the task distribution as it is mentioned in the previous chapter. There is a whole line of work that tries to tackle this complex task distribution problem. However, for the sake of completeness, the start of this work will focus on simpler task distributions, then more complex task distributions will be tackled. Moreover, this type of approach is especially convenient in terms of data generation. Since the data to be utilized does not exist yet, the effort to get the data can be distributed evenly, and machine learning-related aspects of the project can be focused on a more equal basis.

6.1 1^{st} Year Progress

The first year of this PhD consists of doing the literature review and laying the foundation for the numerical analysis tools to be utilized for data generation.

6.1.1 Data Generation Framework

Data generation for the problem at hand will be made by utilizing open-source software for every possible step.

RVE Generation

For the RVE generation, the GMSH-Model from the Technical University of Dresden will be utilized with some minor modifications. Some example RVE's can be found in Figure 6.1.

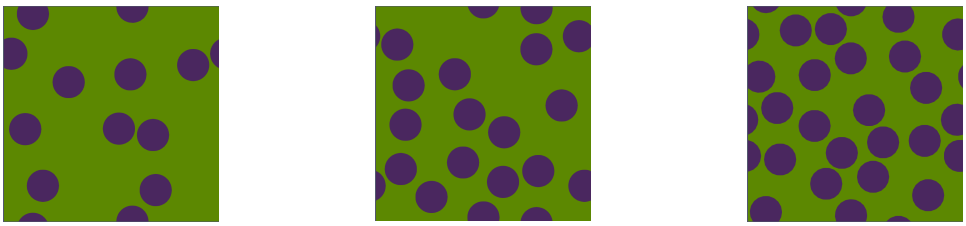


Figure 6.1: Representative volume elements for V_f(Volume fraction): $\{19\%, 30\%, 40\%\}$ **(material-1 → green & material-2 → purple)**

Constitutive Relations

The foundations of the computational homogenization scheme are developed with the help of an open-source software FEniCS [2] for solving Partial Differential Equations (PDE). For the time being, two materials are implemented namely Neo-Hookean and Saint Venant-Kirchhoff materials models to represent. Their energy density functions per unit reference volume is given in Equations 6.1 [11] and 6.2 [68] respectively.

$$\psi_{\text{neo}}(\mathbf{C}) = \frac{\lambda}{2}(\ln(J))^2 - \mu\ln(J) + \frac{\mu}{2}(\text{tr}(\mathbf{C}) - 3) \tag{6.1}$$

$$\psi_{\text{kirchhoff}}(\mathbf{E}) = \frac{\lambda}{2}(\text{tr}(\mathbf{E}))^2 + \mu\mathbf{E} : \mathbf{E} \tag{6.2}$$

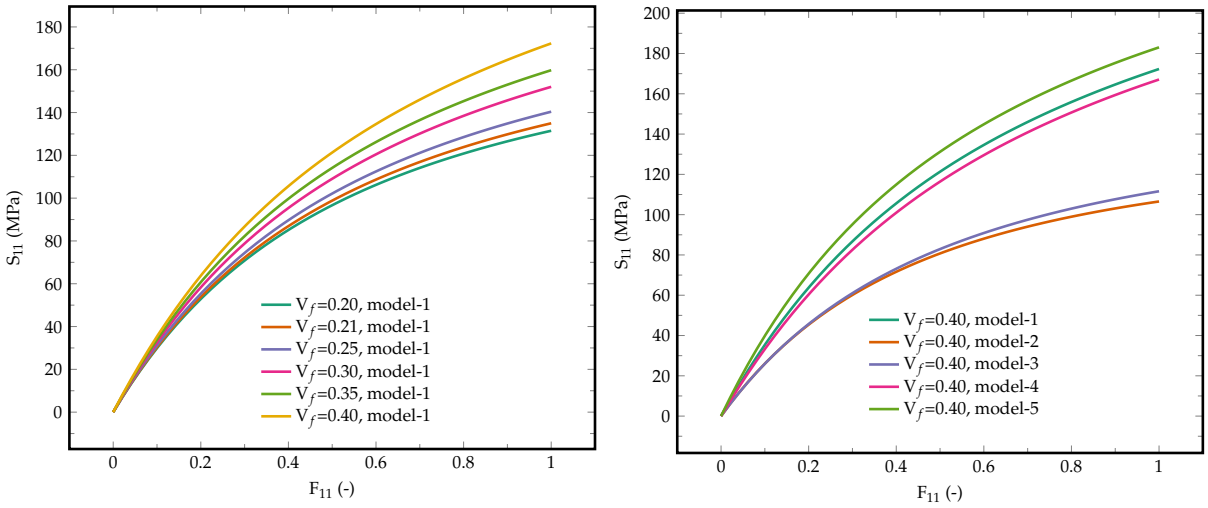
It should be noted that both of these models are parametrized by Lame's constants ($\mu = E/(2(1+\nu))$ and $\lambda = E\nu/((1+\nu)(1-2\nu))$), which are calculated from E (Young's Modulus) and ν (Poisson's ratio). Moreover, $\mathbf{C} = \mathbf{F}^T \cdot \mathbf{F}$ is the right Cauchy-Green strain tensor, $\mathbf{E} = 1/2(\mathbf{C} - \mathbf{I})$ is the Green strain tensor and J is the Jacobian of the deformation gradient $\mathbf{F}$. Moreover, in order to formulate the energy densities with Equation 2.6 the first Piola-Kirchhoff stress tensor is needed. This relation can be established by,

$$\mathbf{P} = \frac{\partial\psi}{\partial\mathbf{F}}. \tag{6.3}$$

Moreover, after the micro-scale problem is solved any other measure of stress can be used to construct the constitutive manifold we can use the first Piola-Kirchhoff stress tensor or use another stress measure of interest like similar to [6], where the second Piola-Kirchhoff stress tensor ($\mathbf{S}$). Some material responses resulting from this constitutive relation can be seen in Figure 6.2.

6.1.2 Learning Problem

Regarding the learning problem, a small investigation of the popular algorithm MAMLs [26] generalization capabilities are investigated in a small study on a linear regression problem is investigated for a convex setting. This method is investigated due to the effect of this publication, a better understanding of its capabilities in simpler settings where the problem is convex is an important step. The setting of the problem is presented below.



$$\text{model1}: \overbrace{\text{NeoHookean(E:300 [MPa], } \nu\text{:0.1 [-])}}^{\text{material-1}} + \overbrace{\text{SVenant(E:500 [MPa], } \nu\text{:0.3 [-])}}^{\text{material-2}}$$

$$\text{model2}: \text{NeoHookean(E:300 [MPa], } \nu\text{:0.1 [-])} + \text{NeoHookean(E:300 [MPa], } \nu\text{:0.1 [-])}$$

$$\text{model3}: \text{NeoHookean(E:300 [MPa], } \nu\text{:0 [-])} + \text{NeoHookean(E:100 [MPa], } \nu\text{:0 [-])}$$

$$\text{model4}: \text{NeoHookean(E:300 [MPa], } \nu\text{:0.1 [-])} + \text{SVenant(E:500 [MPa], } \nu\text{:0.1 [-])}$$

$$\text{model5}: \text{NeoHookean(E:300 [MPa], } \nu\text{:0 [-])} + \text{SVenant(E:800 [MPa], } \nu\text{:0.3 [-])}$$

Figure 6.2: Visualization of different tasks for following a displacement path defined by the deformation gradient tensor entries intervals as F_{11}: [0,1], F_{12}: [0,0.1] & F_{22}: [0,-0.1].Selected example tasks for visual purposes. See Figure 6.1 for material-1 and material-2 definitions. Moreover, the relation between to the task space can be associated with Figure 5.2.

Problem

- We are after linear regression problem, with tasks representing the slope of the below Equation 6.4 (*a*). Task distribution is $p_A \sim \mathcal{N}(m\boldsymbol{I}, c\boldsymbol{I})$
- Samples drawn are represented by Z and distribution is p_Z

$$y = a^{\mathrm{T}}x + \underbrace{\varepsilon}_{\mathcal{N}(0,\sigma=1)} \tag{6.4}$$

$$Z := ((x_i, y_i))_{i=1}^{N} \tag{6.5}$$

- For an estimator $\hat{a}_N$ trained with N samples from Z we are after the expected error over the whole task space.
- Input distribution is $p_x \sim \mathcal{U}(\mathbf{0}, b\mathbf{1})$

$$\int\int\int (\hat{a}_N(Z)^{\mathrm{T}}x - y)^2 p(x,y)dxdyp_Z dZ p_A da \tag{6.6}$$

Results This small investigation of a future empirical risk-minimizing algorithm MAML will be summed up in the upcoming month of the Go-NoGo meeting.

6.2 Planning for the Upcoming Years

The tentative planning for the upcoming years is presented in Figure 6.3. It should be noted that the orange bars represent the duration of the activity with the red diamonds as milestones that are expected to be achieved. As can be seen from this tentative plan the main task ahead for the next year is the developmental stage where the research problems are given in Section 4.2.2. The first stage for the research questions presented there will be investigated for the simple batch of tasks where the relationship is guaranteed. For instance, learning a family of hyperelastic materials can be considered as a simple task distribution. Then, plastic materials can be added to the distribution of the tasks involving hyperelastic material models which makes the detection of the relationship between the tasks is a bit more obscured. This difficulty comes from the fact that the different parameters are needed for the $\mathcal{C}^{\Omega}$. This makes the learning problem more difficult as the similarities between the tasks are not clear anymore. The development stage for this problem is aimed to be finished at the end of the 4th year. The data generation problem is a continuous process where effort is needed for problem data to be generated. Open-source software and the utilization of existing open-source tools can accelerate this process. For instance, it is known that the code of [24] will be available for utilization in the coming years decrease the burden for data generation processes so that the focus can be directed primarily on the learning problem. Moreover, collaborations (in consultation with every member of the supervisory team) with other groups that rely on open-source software and in-house data generation tools will also serve the same purpose and decrease the effort that is needed for data generation tool development. Moreover, such utilization of various tools in the given PhD will make the effectiveness of the proposed method, since the aim of this PhD is to provide a general enough machine learning solution for the acceleration of current numerical analyses methods.

The methods presented in [14–16] provide a really good starting point for the problems posed in Section 4.2.2. MAML [26] and its variants are promising tools too, however, the non-convex setting that they operate in makes it hard to investigate this learning algorithm. Moreover, the performance difference of a future empirical risk-minimizing algorithm MAML and the transfer risk-minimizing algorithm presented in [14] should be investigated.

On another line, a new method development can be done via latent variable modeling for the same purpose of meta-learning. This type of method might help us find a latent space where the correlation between different models is more evident and prove to be more useful compared to other methods that are being utilized in the current machine learning literature. Moreover, this new method development might be tailored specifically to the problem at hand too. Moreover, investigation of domain adaptation and generalization paradigms can open up interesting research paths in the given context of tasks to relate the similar tasks together.

The research questions posed in Sections 4.2.1 and 4.2.3 cannot be completely separated from the research questions presented in Section 4.2.2. However, these questions will be formulated in a structured manner for a few MSc thesis work to be possible. This process will start at the end of the second year of the PhD as most of the course work will be eased, with more time to

help MSc students.

Currently, an MSc student is being supervised by the Civil Engineering and Geosciences Department, for concluding my work on Multi-fidelity Gaussian Process Regression for the prediction of constitutive relations. This work is aimed to be concluded in the first half of 2022. With this work, my MSc thesis-related work will be wrapped up overall.

Doctoral Education Planning showing the lectures to be taken during the PhD are can be found in 6.1.

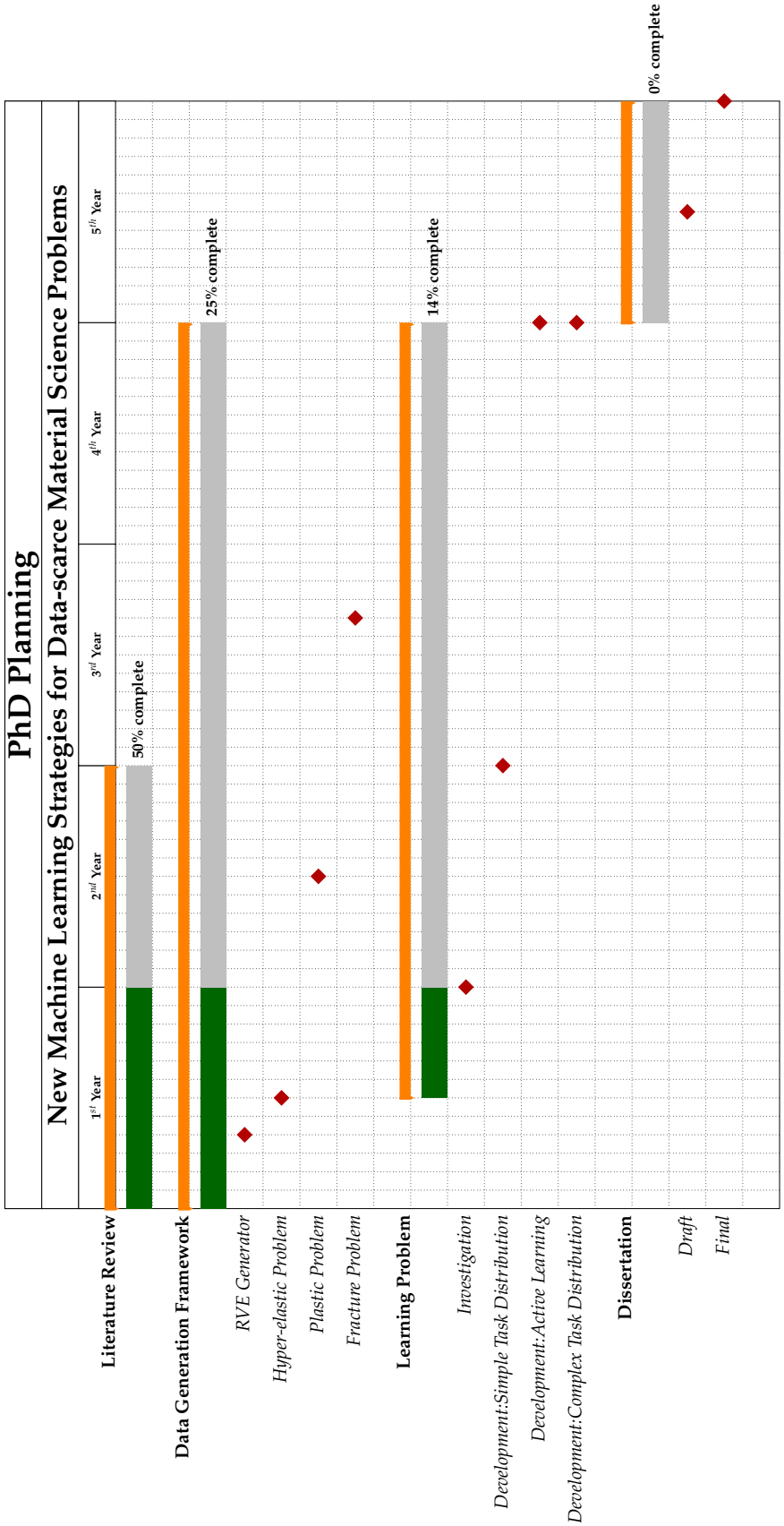


Figure 6.3: PhD Planning Gantt Chart

Table 6.1: Doctoral Education Planing (Maximum of 15 credits are needed for each skill.)

Doctoral Education Planning				
Skills-Credit Balance	Course	Year	Credit	Status
Discipline Related-(10/16)	Machine Learning-1	1	5	✓
	Deep Learning	1	5	✓
	Linear Algebra and Optimisation for Machine Learning	2	6	✗
Research Related-(4/24)	Speaker at a national (or minor international) conference	1-5	1	✗
	Poster presentation conference/workshop	1-5	1	✗
	Internship of at least 1 month with another institute	1-5	2	✗
	Writing a research proposal	1-5	2	✗
	Writing a conference paper	1-5	1	✗
	Writing a journal article	1-5	3	✗
	Supervising a MSc. Student	1-2	4	**in-progress**
	Supervising a BSc. Student	1	4	✓
	Teaching assistance:laboratory course	2	2	**in-progress**
	Teaching assistance:providing material, correcting exams	2	2	**planned**
Research Related-(6/17)	PhD Startup Module (A-B-C)	1	2	✓
	Mental fitness Intervention Program	1	1	✓
	Standing up for yourself while keeping good relation	1	1	✓
	Analytical Storytelling	1	2	✓
	Work Smarter, stress less	2	3	**planned**
	Speedreading and Mindmapping	2	1.5	**planned**
	Time Management-Individual Crash Course	2	0.5	**planned**
	Research Design	2	3	**planned**
	Dutch for Foreigners	2	3	✗

Bibliography

[1] H. Adeli. Knowledge engineering. *Archives of Computational Methods in Engineering*, 2(4): 51–68, 1995. ISSN 11343060. doi: 10.1007/BF02736196.

[2] Martin S Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, E Rognes, and Garth N Wells. The FEniCS Project Version 1 . 5. 3(100):9–23, 2015.

[3] Marcin Andrychowicz, Misha Denil, Sergio Gómez Colmenarejo, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in Neural Information Processing Systems*, (Nips):3988–3996, 2016. ISSN 10495258.

[4] Claudio Arenas, Guillermo Herrera, Enrique Muñoz, and Raul C. Munoz. The breakdown of Moore's law induced by weak Anderson localization and by size effects in nanoscale metallic connectors. *Materials Research Express*, 8(1), 2021. ISSN 20531591. doi: 10.1088/2053-1591/abd422.

[5] Sarah Bechtle. Meta Learning via Learned Loss. 2021.

[6] M. A. Bessa, R. Bostanabad, Z. Liu, A. Hu, Daniel W. Apley, C. Brinson, W. Chen, and Wing Kam Liu. A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. *Computer Methods in Applied Mechanics and Engineering*, 320:633–667, 2017. ISSN 00457825. doi: 10.1016/j.cma.2017.03.037.

[7] C. M. Bishop and G. D. James. Analysis of multiphase flows using dual-energy gamma densitometry and neural networks. *Nuclear Inst. and Methods in Physics Research, A*, 327(2-3): 580–593, 1993. ISSN 01689002. doi: 10.1016/0168-9002(93)90728-Z.

[8] Diana Cai, Rishit Sheth, Lester Mackey, and Nicolo Fusi. Weighted Meta-Learning. 2020. URL `http://arxiv.org/abs/2003.09465`.

[9] Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143 (6):1–15, 2021. ISSN 15288943. doi: 10.1115/1.4050542.

[10] Rich Caruana. Learning Many Related Tasks at the Same Time With Backpropagation. *NIPS'94: Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 657–664, 1994.

[11] Eduardo W. V. Chaves. *Notes on Continuum Mechanics*. 2013. ISBN 978-94-007-5985-5. URL `http://link.springer.com/10.1007/978-94-007-5986-2`.

[12] Mathilde Chevreuil, Nouy, and Anthony. Model order reduction based on proper generalized decomposition for the propagation of uncertainties in structural dynamics. *INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING*, 89(July): 241–268, 2012. doi: 10.1002/nme.3249. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3249/full.

[13] Liam Collins, Aryan Mokhtari, and Sanjay Shakkottai. Task-robust model-agnostic meta-learning. *Advances in Neural Information Processing Systems*, 2020-December:1–30, 2020. ISSN 10495258.

[14] Giulia Denevi, Carlo Ciliberto, Dimitris Stamos, and Massimiliano Pontil. Incremental learning-to-learn with statistical guarantees. *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018*, 1:457–466, 2018.

[15] Giulia Denevi, Carlo Ciliberto, Dimitris Stamos, and Massimiliano Pontil. Learning to learn around a common mean. *Advances in Neural Information Processing Systems*, 2018-Decem (NeurIPS):10169–10179, 2018. ISSN 10495258.

[16] Giulia Denevi, Carlo Ciliberto, Riccardo Grazzi, and Massimiliano Pontil. Learning-to-learn stochastic gradient descent with biased regularization. *36th International Conference on Machine Learning, ICML 2019*, 2019-June:2814–2843, 2019.

[17] Giulia Denevi, Dimitris Stamos, Carlo Ciliberto, and Massimiliano Pontil. Online-Within-Online Meta-Learning. (NeurIPS):1–11, 2019.

[18] Giulia Denevi, Massimiliano Pontil, and Carlo Ciliberto. The advantage of conditional meta-learning for biased regularization and fine tuning. *Advances in Neural Information Processing Systems*, 2020-Decem, 2020. ISSN 10495258.

[19] Giulia Denevi, Dimitris Stamos, and Massimiliano Pontil. Online parameter-free learning of multiple low variance tasks. *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence, UAI 2020*, 124:909–918, 2020.

[20] Dennis M. Dimiduk, Elizabeth A. Holm, and Stephen R. Niezgoda. Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering. *Integrating Materials and Manufacturing Innovation*, 7 (3):157–172, 2018. ISSN 21939772. doi: 10.1007/s40192-018-0117-8.

[21] Theodoras Evgeniou and Massimiliano Pontil. Regularized multi-task learning. *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 109–117, 2004. doi: 10.1145/1014052.1014067.

[22] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. On the Convergence Theory of Gradient-Based Model-Agnostic Meta-Learning Algorithms. 2019. URL http://arxiv.org/abs/1908.10400.

[23] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Generalization of Model-Agnostic Meta-Learning Algorithms: Recurring and Unseen Tasks. pages 1–23, 2021. URL http://arxiv.org/abs/2102.03832.

[24] Bernardo P. Ferreira, F. M. Andrade Pires, and Miguel A. Bessa. Adaptive Clustering-based Reduced-Order Modeling Framework: Fast and accurate modeling of localized history-dependent phenomena. 2021. URL http://arxiv.org/abs/2109.11897.

[25] Frédéric Feyel. A multilevel finite element method (FE2) to describe the response of highly non-linear structures using generalized continua. *Computer Methods in Applied Mechanics and Engineering*, 192(28-30):3233–3244, 2003. ISSN 00457825. doi: 10.1016/S0045-7825(03)00348-7.

[26] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *34th International Conference on Machine Learning, ICML 2017*, 3:1856–1868, 2017.

[27] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning. *Advances in Neural Information Processing Systems*, 2018-Decem(NeurIPS):9516–9527, 2018. ISSN 10495258.

[28] Chelsea Finn, Aravind Rajeswaran, Sham Kakade, and Sergey Levine. Online Meta-Learning. 2019.

[29] Sebastian Flennerhag, Pablo G. Moreno, Neil D. Lawrence, and Andreas Damianou. Transferring knowledge across learning processes. *arXiv*, pages 1–23, 2018. ISSN 23318422.

[30] Sebastian Flennerhag, Andrei A. Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, and Raia Hadsell. Meta-Learning with Warped Gradient Descent. pages 1–28, 2019. URL http://arxiv.org/abs/1909.00025.

[31] Alexander L. Fradkov. Early history of machine learning. *IFAC-PapersOnLine*, 53(2):1385–1390, 2020. ISSN 24058963. doi: 10.1016/j.ifacol.2020.12.1888. URL https://doi.org/10.1016/j.ifacol.2020.12.1888.

[32] Ari L. Frankel, Reese E. Jones, and Laura P. Swiler. Tensor Basis Gaussian Process Models of Hyperelastic Materials. *Journal of Machine Learning for Modeling and Computing*, 1(1):1–17, 2020. ISSN 2689-3967. doi: 10.1615/jmachlearnmodelcomput.2020033325.

[33] Jezabel R Garcia, Federica Freddi, Feng-Ting Liao, Jamie Mcgowan, Tim Nieradzik, Da-Shan Shiu, Ye Tian, and Alberto Bernacchia. Cross-Lingual Transfer with MAML on Trees. pages 72–79, 2021.

[34] F. Ghavamian, P. Tiso, and A. Simone. POD–DEIM model order reduction for strain softening viscoplasticity. *Computer Methods in Applied Mechanics and Engineering*, 317:458–479, 2017. ISSN 00457825. doi: 10.1016/j.cma.2016.11.025. URL http://dx.doi.org/10.1016/j.cma.2016.11.025.

[35] Sukhpal Singh Gill, Adarsh Kumar, Harvinder Singh, Manmeet Singh, Kamalpreet Kaur, Muhammad Usman, and Rajkumar Buyya. Quantum Computing: A Taxonomy, Systematic Review and Future Directions. pages 1–37, 2020. URL http://arxiv.org/abs/2010.15559.

[36] Somdatta Goswami, Cosmin Anitescu, Souvik Chakraborty, and Timon Rabczuk. Transfer learning enhanced physics informed neural network for phase-field modeling of fracture. *Theoretical and Applied Fracture Mechanics*, 106(July 2019):102447, 2020. ISSN 01678442. doi: 10.1016/j.tafmec.2019.102447. URL https://doi.org/10.1016/j.tafmec.2019.102447.

[37] Olivier Goury, David Amsallem, Stéphane Pierre Alain Bordas, Wing Kam Liu, and Pierre Kerfriden. Automatised selection of load paths to construct reduced-order models in computational damage micromechanics: from dissipation-driven random selection to Bayesian optimization. *Computational Mechanics*, 58(2):213–234, 2016. ISSN 01787675. doi: 10.1007/s00466-016-1290-2.

[38] Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griffiths. Recasting gradient-based meta-learning as hierarchical bayes. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pages 1–13, 2018.

[39] Simon Guiroy, Vikas Verma, and Christopher Pal. Towards Understanding Generalization in Gradient-Based Meta-Learning. 2019. URL http://arxiv.org/abs/1907.07287.

[40] Theron Guo, Ondřej Rokoš, and Karen Veroy. Learning constitutive models from microstructural simulations via a non-intrusive reduced basis method. *Computer Methods in Applied Mechanics and Engineering*, 384:113924, 2021. ISSN 00457825. doi: 10.1016/j.cma.2021.113924.

[41] James Harrison, Apoorva Sharma, and Marco Pavone. Meta-learning Priors for Efficient Online Bayesian Regression. *Springer Proceedings in Advanced Robotics*, 14:318–337, 2020. ISSN 25111264. doi: 10.1007/978-3-030-44051-0_19.

[42] Xu He, Jakub Sygnowski, Alexandre Galashov, Andrei A. Rusu, Yee Whye Teh, and Razvan Pascanu. Task Agnostic Continual Learning via Meta Learning. 2019. URL http://arxiv.org/abs/1906.05201.

[43] S. Hochreiter, A. S. Younger, and P.R. Conwell. Learning to Learn Using Gradient Decent. 2001. URL http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C42CD22C7E7811821D16157FC882CE94?doi=10.1.1.5.323&rep=rep1&type=pdf.

[44] Rachid Hsissou, Rajaa Seghiri, Zakaria Benzekri, Miloudi Hilali, Mohamed Rafik, and Ahmed Elharfi. Polymer composite materials: A comprehensive review. *Composite Structures*, 262(December 2020):0–3, 2021. ISSN 02638223. doi: 10.1016/j.compstruct.2021.113640.

[45] Dong Phill Jang, Piemaan Fazily, and Jeong Whan Yoon. Machine learning-based constitutive model for J2- plasticity. *International Journal of Plasticity*, 138(December 2020):102919, 2021. ISSN 07496419. doi: 10.1016/j.ijplas.2020.102919. URL https://doi.org/10.1016/j.ijplas.2020.102919.

[46] Ghassen Jerfel, Erin Grant, Thomas L. Griffiths, and Katherine Heller. Reconciling meta-learning and continual learning with online mixtures of tasks. *Advances in Neural Information Processing Systems*, 32, 2019. ISSN 10495258.

[47] Sharu Theresa Jose, Osvaldo Simeone, and Giuseppe Durisi. Transfer Meta-Learning: Information-Theoretic Bounds and Information Meta-Risk Minimization. (725731):1–51, 2020. URL http://arxiv.org/abs/2011.02872.

[48] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. ISSN 25225820. doi: 10.1038/s42254-021-00314-5.

[49] P. Kerfriden, P. Gosselet, S. Adhikari, and S. P.A. Bordas. Bridging proper orthogonal decomposition methods and augmented Newton-Krylov algorithms: An adaptive model order reduction for highly nonlinear mechanical problems. *Computer Methods in Applied Mechanics and Engineering*, 200(5-8):850–866, 2011. ISSN 00457825. doi: 10.1016/j.cma.2010.10.009. URL http://dx.doi.org/10.1016/j.cma.2010.10.009.

[50] Gaweł I. Kuś, Sybrand van der Zwaag, and Miguel A. Bessa. Sparse quantum Gaussian processes to counter the curse of dimensionality. *Quantum Machine Intelligence*, 3(1), 2021. ISSN 2524-4906. doi: 10.1007/s42484-020-00032-8.

[51] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-SGD: Learning to learn quickly for few-shot learning. *arXiv*, 2017. ISSN 23318422.

[52] Kevin Linka, Markus Hillgärtner, Kian P. Abdolazizi, Roland C. Aydin, Mikhail Itskov, and Christian J. Cyron. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. *Journal of Computational Physics*, 429:110010, 2021. ISSN 10902716. doi: 10.1016/j.jcp.2020.110010. URL https://doi.org/10.1016/j.jcp.2020.110010.

[53] Zeliang Liu, M. A. Bessa, and Wing Kam Liu. Self-consistent clustering analysis: An efficient multi-scale scheme for inelastic heterogeneous materials. *Computer Methods in Applied Mechanics and Engineering*, 306:319–341, 2016. ISSN 00457825. doi: 10.1016/j.cma.2016.04.004. URL http://dx.doi.org/10.1016/j.cma.2016.04.004.

[54] Zeliang Liu, C. T. Wu, and M. Koishi. Transfer learning of deep material network for seamless structure–property predictions. *Computational Mechanics*, 64(2):451–465, 2019. ISSN 01787675. doi: 10.1007/s00466-019-01704-4.

[55] Zeliang Liu, C. T. Wu, and M. Koishi. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Computer Methods in Applied Mechanics and Engineering*, 345:1138–1168, 2019. ISSN 00457825. doi: 10.1016/j.cma.2018.09.020.

[56] Zeliang Liu, C. T. Wu, and M. Koishi. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Computer Methods in Applied Mechanics and Engineering*, 345:1138–1168, 2019. ISSN 00457825. doi: 10.1016/j.cma.2018.09.020.

[57] Zeliang Liu, Haoyan Wei, Tianyu Huang, and C. T. Wu. Intelligent multiscale simulation based on process-guided composite database. pages 1–15, 2020. URL http://arxiv.org/abs/2003.09491.

[58] Xiaoxin Lu, Julien Yvonnet, Leonidas Papadopoulos, and Ioannis Kalogeris. Materials A Stochastic FE 2 Data-Driven Method for Nonlinear Multiscale Modeling. pages 1–23, 2021.

[59] Filippo Masi, Ioannis Stefanou, Paolo Vannucci, and Victor Maffi-Berthier. Thermodynamics-based Artificial Neural Networks for constitutive modeling. *Journal of the Mechanics and Physics of Solids*, 147, 2021. ISSN 00225096. doi: 10.1016/j.jmps.2020.104277.

[60] C. Miehe, J. Schotte, and M. Lambrecht. Homogenization of inelastic solid materials at finite strains based on incremental minimization principles. Application to the texture analysis of polycrystals. *Journal of the Mechanics and Physics of Solids*, 50(10):2123–2167, 2002. ISSN 00225096. doi: 10.1016/S0022-5096(02)00016-9.

[61] Christian Miehe, Jörg Schröder, and Jan Schotte. Computational homogenization analysis in finite plasticity simulation of texture development in polycrystalline materials. *Computer Methods in Applied Mechanics and Engineering*, 171(3-4):387–418, 1999. ISSN 00457825. doi: 10.1016/S0045-7825(98)00218-7.

[62] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pages 1–17, 2018.

[63] Gordon M Moore. Cramming more components onto integrated circuits With unit cost. *Electronics*, 38(8):114, 1965. URL `https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf`.

[64] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, Jian Cao, and M. A. Bessa. Deep learning predicts path-dependent plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52):26414–26420, 2019. ISSN 10916490. doi: 10.1073/pnas.1911815116.

[65] Tsendsuren Munkhdalai and Hong Yu. Meta networks. *34th International Conference on Machine Learning, ICML 2017*, 5:3933–3943, 2017. ISSN 2640-3498.

[66] Remi Munos. A convergent reinforcement learning algorithm in the continuous case based on a finite difference method. *IJCAI International Joint Conference on Artificial Intelligence*, 2:826–831, 1997. ISSN 10450823.

[67] Cuong Nguyen, Thanh Toan Do, and Gustavo Carneiro. Uncertainty in model-agnostic meta-learning using variational inference. *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, pages 3079–3089, 2020. doi: 10.1109/WACV45572.2020.9093536.

[68] Siamak Niroomandi, Icíar Alfaro, Elías Cueto, Francisco Chinesta, Siamak Niroomandi, Icíar Alfaro, Elías Cueto, and Francisco Chinesta. Model order reduction for hyperelastic materials To cite this version : HAL Id : hal-01007059 Model order reduction for hyperelastic materials. 81(9), 2017.

[69] Anup Pandey and Reeju Pokharel. Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. *Scripta Materialia*, 193:1–5, 2021. ISSN 13596462. doi: 10.1016/j.scriptamat.2020.10.028.

[70] Huimin Peng. A comprehensive overview and survey of recent advances in meta-learning. *arXiv*, 2020. ISSN 23318422.

[71] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 10902716. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

[72] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. (Part I):1–22, 2017. URL http://arxiv.org/abs/1711.10566.

[73] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah. ITAML: An Incremental Task-Agnostic Meta-learning Approach. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 13585–13594, 2020. ISSN 10636919. doi: 10.1109/CVPR42600.2020.01360.

[74] Yoram Reich. Machine learning techniques for civil engineering problems. *Computer-Aided Civil and Infrastructure Engineering*, 12(4):295–310, 1997. ISSN 10939687. doi: 10.1111/0885-9507.00065.

[75] Yoram Reich and Nahum Travitzky. Machine learning of material behaviour knowledge from empirical data. *Materials and Design*, 16(5):251–259, 1995. ISSN 02613069. doi: 10.1016/0261-3069(96)00007-6.

[76] J R Rice. INELASTIC CONSTITUTIVE RELATIONS FOR SOLIDS : AN INTERNAL-VARIABLE THEORY AND ITS APPLICATION TO METAL PLASTICITY. *Nature Biotechnology*, 19:433–455, 1971.

[77] I. B.C.M. Rocha, P. Kerfriden, and F. P. van der Meer. Micromechanics-based surrogate models for the response of composites: A critical comparison between a classical mesoscale constitutive model, hyper-reduction and neural networks. *European Journal of Mechanics, A/Solids*, 82:103995, 2020. ISSN 09977538. doi: 10.1016/j.euromechsol.2020.103995. URL https://doi.org/10.1016/j.euromechsol.2020.103995.

[78] I. B.C.M. Rocha, P. Kerfriden, and F. P. van der Meer. On-the-fly construction of surrogate constitutive models for concurrent multiscale mechanical analysis through probabilistic machine learning. *Journal of Computational Physics: X*, 9:100083, 2021. ISSN 25900552. doi: 10.1016/j.jcpx.2020.100083. URL https://doi.org/10.1016/j.jcpx.2020.100083.

[79] Francisco Sahli Costabal, Yibo Yang, Paris Perdikaris, Daniel E. Hurtado, and Ellen Kuhl. Physics-Informed Neural Networks for Cardiac Activation Mapping. *Frontiers in Physics*, 8 (February):1–12, 2020. ISSN 2296424X. doi: 10.3389/fphy.2020.00042.

[80] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-Learning with Memory-Augmented Neural Networks. *33rd International Conference on Machine Learning, ICML 2016*, 4:2740–2751, 2016.

[81] Wilhelmus H a Schilders, Henk a Van Der Vorst, and Joost Rommes. *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13. 2008. ISBN 978-3-540-78840-9. URL http://books.google.com.sg/books?hl=en{&}lr={&}id=WWi1evmAQPkC{&}oi=fnd{&}pg=PR7{&}dq=model+order+reduction+techniques{&}ots=1mWoMmwb7M{&}sig=NhtNd6iIrIngXCM22wlP0V{_}KWCw{%}5Cnhttp://www.springerlink.com/index/10.1007/978-3-540-78841-6.

[82] Jonathan Schmidt, Mário R.G. Marques, Silvana Botti, and Miguel A.L. Marques. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5(1), 2019. ISSN 20573960. doi: 10.1038/s41524-019-0221-0. URL http://dx.doi.org/10.1038/s41524-019-0221-0.

[83] I. Temizer and P. Wriggers. On the computation of the macroscopic tangent for multiscale volumetric homogenization problems. *Computer Methods in Applied Mechanics and Engineering*, 198(3-4):495–510, 2008. ISSN 00457825. doi: 10.1016/j.cma.2008.08.018. URL http://dx.doi.org/10.1016/j.cma.2008.08.018.

[84] Lazarus Teneketzis Tenek and John Argyris. *Finite Element Analysis for Composite Structures*. ISBN 978-94-015-9044-0. doi: 10.1007/978-94-015-9044-0.

[85] Sebastian Thrun. *Learning to learn*. Kluwer Academic, 1 edition, 1998. ISBN 978-1-4615-5529-2. doi: 10.1007/978-1-4615-5529-2.

[86] Ramakrishna Tipireddy, Paris Perdikaris, Panos Stinis, and Alexandre Tartakovsky. A comparative study of physics-informed neural network models for learning unknown dynamics and constitutive relations. 2019. URL http://arxiv.org/abs/1904.04058.

[87] F.P. van der Meer. *Computational Modeling of Failure in Composite Laminates*, volume PhD. 2010. ISBN 97890902569 79.

[88] Rody A. van Tuijl, Cale Harnish, Karel Matouš, Joris J.C. Remmers, and Marc G.D. Geers. Wavelet based reduced order models for microstructural analyses. *Computational Mechanics*, 63(3):535–554, 2019. ISSN 01787675. doi: 10.1007/s00466-018-1608-3. URL https://doi.org/10.1007/s00466-018-1608-3.

[89] R. Vilalta and D. Youessef. A perspective View and Survey of Meta-Learning. *Annals of Operations Research*, 18(1):77–95, 2002. ISSN 02545330. doi: 10.1007/BF02283734.

[90] Sifan Wang and Paris Perdikaris. Deep learning of free boundary and Stefan problems. *Journal of Computational Physics*, 428:1–27, 2021. ISSN 10902716. doi: 10.1016/j.jcp.2020.109914.

[91] Stephan Wulfinghoff, Fabiola Cavaliere, and Stefanie Reese. Model order reduction of nonlinear homogenization problems using a Hashin–Shtrikman type finite element method. *Computer Methods in Applied Mechanics and Engineering*, 330:149–179, 2018. ISSN 00457825. doi: 10.1016/j.cma.2017.10.019. URL https://doi.org/10.1016/j.cma.2017.10.019.

[92] Tianju Xue, Alex Beatson, Maurizio Chiaramonte, Geoffrey Roeder, Jordan T. Ash, Yigit Menguc, Sigrid Adriaenssens, Ryan P. Adams, and Sheng Mao. A data-driven computational scheme for the nonlinear mechanical properties of cellular mechanical metamaterials under large deformation. *Soft Matter*, 16(32):7524–7534, 2020. ISSN 17446848. doi: 10.1039/dosm00488j.

[93] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer Learning*. 2020. ISBN 9781107016903. doi: 10.1017/9781139061773.

[94] Yibo Yang and Paris Perdikaris. Adversarial uncertainty quantification in physics-informed neural networks. *Journal of Computational Physics*, 394:136–152, 2019. ISSN 10902716. doi: 10.1016/j.jcp.2019.05.027. URL https://doi.org/10.1016/j.jcp.2019.05.027.

[95] Huaxiu Yao, Linjun Zhang, and Chelsea Finn. Meta-Learning with Fewer Tasks through Task Interpolation. 2021. URL http://arxiv.org/abs/2106.02695.

[96] Yinhao Zhu, Nicholas Zabaras, Phaedon Stelios Koutsourelakis, and Paris Perdikaris. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics*, 394:56–81, 2019. ISSN 10902716. doi: 10.1016/j.jcp.2019.05.024. URL https://doi.org/10.1016/j.jcp.2019.05.024.

[97] Xiaoying Zhuang, Hongwei Guo, Naif Alajlan, Hehua Zhu, and Timon Rabczuk. Deep autoencoder based energy method for the bending, vibration, and buckling analysis of Kirchhoff plates with transfer learning. *European Journal of Mechanics, A/Solids*, 87(April 2020):104225, 2021. ISSN 09977538. doi: 10.1016/j.euromechsol.2021.104225. URL https://doi.org/10.1016/j.euromechsol.2021.104225.