Methods for Learning Mechanics: Inverse Problems, Constitutive Modeling, and Design

Thesis by
Aakila Rajan

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Mechanical Engineering

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 14, 2025

Aakila Rajan
ORCID: [0000-0001-9218-5062]

ACKNOWLEDGEMENTS

This thesis is a reflection of the support and love of many people. I could not have completed it without the encouragement of my professors, coworkers, and family. I sincerely thank all of them.

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Kaushik Bhattacharya. His continued guidance, expertise, and mentorship have been a source of immense motivation and growth throughout my journey. I am forever grateful to him for giving me the opportunity to pursue research in his group. My journey at Caltech could not have even started five years ago without his support, especially when I was stuck in India for months during the COVID-related travel ban. Kaushik has not only taught me academics but has also profoundly impacted my personal growth. He has always listened to my concerns and supported me through difficult times during my PhD. I am deeply thankful for his mentorship, which has shaped me into a better researcher and person.

I am also grateful to my thesis committee members, Dr. Philip Buskohl, Prof. Guruswami Ravichandran, and Prof. Ruby Fu, for their time, valuable feedback, and academic advice. Phil invited me to the Air Force Research Laboratory (AFRL) to collaborate, giving me the opportunity to visit a top-class research facility and work on an exciting project. His family also welcomed me warmly into their home. Ravi has been a mentor since my first year, and our interactions during group meetings were always insightful and helpful. Ruby has been in contact with me since the day I received my Caltech acceptance letter and later served as the chair of my candidacy committee. Both Ravi and Ruby taught me fundamental mechanics during my first-year coursework, for which I am very grateful.

My research would not have been possible without my collaborators: Andrew Stuart, Andrew Akerson, Nikola Kovachki, Margaret Trautner, Jonathan Chen, and Nate Hertlein. Andrew Stuart collaborated with and guided me on multiple projects; I owe much of my knowledge in deep learning and algebra to him. I will always cherish his courses in linear analysis and multiscale modeling. Andrew Akerson taught me much of the computational knowledge I have regarding solid mechanics, and I had an amazing time collaborating with him

on three papers. Nikola and Margaret worked with me on machine learning and homogenization, while Jonathan and Nate collaborated with me on topology optimization and deep learning. I would especially like to thank Nate for hosting me at AFRL.

My office at Caltech fostered a nurturing and fun environment for me to work. I cherish every second I spent hanging out, eating, and playing games with my friends from the "boba crew group," namely — Vatsa Gandhi, Barry Lawlor, Adeline Wihardja, and Shengduo Liu. I am also grateful to all my current and former officemates — Eric Ocegueda, Mary Agajanian, Carter Cocke, Mayank Raj, Theresa Tsaggaris, Maximo Cravero Baraja, and Taeho Kim — for making the office a productive and enjoyable environment.

I am deeply thankful to my parents, Suganthi and Rajan, and to my sister's family — Helan, Johnson, and little Joanna — for their unwavering support and love. This thesis would not have been possible without the steadfast encouragement of my closest circle: Sathvik, Vivian, and Tita Mary Ann. Finally, I would like to express my heartfelt gratitude to my girlfriend, Gyllie, for her love, support, and encouragement through all the peaks and valleys of my life.

ABSTRACT

Predicting the behavior of physical systems under complex loading conditions remains a central challenge in engineering and the applied sciences. In solid mechanics, accurate prediction requires not only knowledge of the governing physical laws but also appropriate constitutive models that describe material behavior. These constitutive models are parameterized functions typically obtained through material characterization techniques. However, traditional techniques have limited applicability due to their reliance on challenging experiments designed to produce homogeneous fields, especially for complex materials.

First, we propose a method to accurately and efficiently identify the constitutive behavior of complex materials from full-field observations. We formulate the problem of inferring constitutive relations as an indirect inverse problem constrained by the balance laws. Specifically, we seek a constitutive relation that minimizes the difference between experimental observations and model predictions while strictly enforcing physical laws. The forward problem is posed as a boundary value problem corresponding to the experiment, and sensitivities are computed using the adjoint method. The resulting framework is robust and applicable to constitutive models of arbitrary complexity. We focus on elasto-viscoplasticity and demonstrate the method using synthetic data on two problems, one quasistatic and the other dynamic.

We then extend this methodology to infer constitutive parameters from experimental dynamic contact data, where challenges such as noise, incomplete information, and model-form uncertainties complicate inversion. Using force–depth measurements from split Hopkinson pressure bar experiments on steel and aluminum samples, and incorporating regularization techniques to mitigate ill-posedness, the framework robustly recovers material properties. Results show the successful extraction of meaningful parameters, highlighting the method's effectiveness over traditional direct inversion approaches, particularly for nonlinear models.

A key limitation of the proposed method is its reliance on a pre-specified form of the constitutive model. The choice of such forms becomes harder with the evergrowing catalog of materials. To address this, we develop a neural network-

based framework where the constitutive relation emerges directly from data. Using a recurrent neural operator to model history-dependent behavior with internal variables, we demonstrate on synthetic high strain rate compression experiments that this approach significantly outperforms traditional models in capturing path-dependent material responses. Therefore, this proves that neural networks can be powerful tools for constitutive modeling.

Building on the ability of neural networks to approximate complex functions, we further explore their use in approximating solutions to partial differential equations (PDEs). In following study, we investigate neural operators for multiscale modeling of elliptic PDEs, focusing on learning effective macroscopic behavior from complex microstructures without repeatedly solving fine-scale problems. Targeting the classical homogenization setting of linear elliptic PDEs with discontinuous coefficients, we analyze the challenges posed by sharp interfaces and degraded solution regularity. We provide theoretical guarantees on approximation capabilities and demonstrate the method across various representative microstructures, showing that neural operator-based homogenization offers a scalable and non-intrusive approach to multiscale modeling.

Finally, in the last study, we extend these data-driven tools to topology optimization, where the need for numerous PDE solves presents a major computational bottleneck. By integrating a reduced-order PCA-based neural network into the optimization loop, we represent complex structures in a low-dimensional latent space and achieve efficient, high-quality updates. This approach significantly accelerates optimization while preserving design flexibility, highlighting the potential of machine learning to enhance and transform classical design methodologies.

Together, these contributions present a unified strategy that integrates physics-based modeling, machine learning, and optimization to advance material characterization, multiscale modeling, and design in solid mechanics.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] A. Akerson, A. Rajan, and K. Bhattacharya. "Learning constitutive relations from experiments: 1. PDE constrained optimization." In: *Journal of the Mechanics and Physics of Solids* (2025), p. 106128. DOI: https://doi.org/10.1016/j.jmps.2025.106128.
Contributions: A.R. and A.A. developed the computational code and performed simulations. A.A. and K.B. conceptualized the methodology. A.R., A.A., and K.B. wrote the manuscript.

[2] K. Bhattacharya, N. B. Kovachki, A. Rajan, A. M. Stuart, and M. Trautner. "Learning homogenization for elliptic operators." In: *SIAM Journal on Numerical Analysis* 62.4 (2024), pp. 1844–1873. DOI: 10.1137/23M1544954.
Contributions: A.R. developed the code for a portion of the numerical results. N.B.K. and M.T. took the lead in the mathematical proofs, developed the code, performed simuations. A.M.S. and K.B. supervised the mathematical proofs. N.B.K., M.T., A.M.S., and K.B. wrote the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Modeling the behavior of physical systems under diverse and often complex conditions remains a foundational challenge in engineering and the applied sciences. Whether dealing with physical systems at the microscale or large-scale engineered systems, the ability to make accurate predictions is critical for understanding performance, ensuring safety, and enabling innovation. These systems are influenced by a broad range of factors—such as external stimuli, environmental variability, evolving conditions, and interactions across multiple scales—all of which contribute to their complexity. Over the years, increasingly sophisticated techniques have been developed to predict the behavior of engineering systems while accounting for this growing spectrum of factors. While these methods have achieved remarkable success, the complexity of real-world systems continues to grow, driven by emerging technologies, novel materials, and increasingly ambitious design goals. As a result, there is a continual and evolving need for more flexible, accurate, and robust predictive techniques—tools that can adapt to new contexts and push the boundaries of what engineering models can achieve.

This thesis addresses solid mechanics, the study of the mechanical behavior of solids and structures. The approach involves deriving a set of governing equations based on Newton's laws, which describes how the system responds to the external conditions. However; these equations are not closed and an extra set of relations, called the constitutive laws, are necessary to solve the system [31, 8]. These constitutive laws are independent of the governing equations, depend only on the material under consideration, and incorporate the complexity that is not explicitly resolved. They are typically obtained from controlled experiments that measure how a material deforms under different loading conditions. Therefore, the problem of predicting the behavior of a structure requires two things: a proper constitutive law describing the material and a set of governing equations describing the physics of the problem.

In the recent decade, neural networks have emerged as a powerful tool for modeling complex systems, especially when traditional approaches fall short.

Their ability to approximate functions and learn from data makes them useful for capturing complicated behaviors in the physical sciences. Recently, they have been used to speed up computations and offer flexible, data-driven models.

In this thesis, we address two related challenges: the extraction of constitutive behavior from experimental observations, and the use of neural networks in mechanics. We show examples of the latter for constitutive modeling, multiscale modeling and topology optimization.

1.1 Novel method for material characterization

Constitutive laws describe the relationship between stress and strain in a material, capturing its mechanical response under different loading conditions. Traditionally, these laws are postulated from phenomenological observations or micromechanical principles and expressed through mathematical equations, such as elasticity, viscoelasticity or plasticity models. However, they require material-specific parameters—such as elastic moduli, yield strength, or hardening coefficients—that are often unknown and must be inferred from experiments. This method of obtaining the necessary material parameters is known as material characterization [9]. Conventional characterization techniques are limited in accuracy and applicability, especially when the experimental data are noisy, sparse, or when the material exhibits non-linear, non-local behavior. Therefore, improved methods are needed to characterize material behavior under different stress and strain conditions. Furthermore, traditional constitutive models, while effective in many scenarios, often fail to capture the full complexity of real-world material responses. These limitations arise from the fundamental assumptions inherent in phenomenological models, which impose pre-defined functional forms on the stress-strain relationship. Such models may be adequate for simple materials under controlled conditions, but they often break down when dealing with complex materials, multi-axial loading, or history-dependent behaviors [16, 38, 20]. This motivates the need for a more flexible, data-driven approach that can adapt to a wide range of mechanical phenomena [23].

A common approach to material characterization is to perform experiments and use the data collected to recover the parameters of a chosen constitutive model by solving an inverse problem. Conventionally, this method typically

consists of three steps: (i) measuring deformation fields, surface tractions, or other macroscopic observables, (ii) converting these measurements into stress-strain data, and (iii) extracting constitutive parameters from the data [8, 22]. While this technique is conceptually straightforward, it is often inefficient and impractical for complex materials. It requires carefully controlled experiments to ensure a reliable conversion of measurable data into stress-strain pairs. Such conversion is only possible if the experiment is conducted under sterile conditions to obtain homogeneous fields. As material complexity increases, these experiments become expensive and cumbersome. Additionally, each experiment typically provides only a single data point for inversion, necessitating a large number of experiments for constitutive models with numerous parameters. Another key limitation is that material parameters are learned from experiments conducted in simple configurations but are later applied to scenarios with highly complex loading conditions, creating a gap in applicability. These limitations motivate the exploration of alternative approaches that leverage data-driven techniques and machine learning to improve robustness and generalizability.

To overcome the above limitations, we employ an alternative framework. We formulate a boundary value problem corresponding to the experiment, and compute the measurable quantities using finite element analysis (FEA). The material parameters are inferred by minimizing the discrepancy between experimental observations and model predictions. This minimization is an optimization problem with a partial differential equation (PDE) as a constraint [13, 4, 37]. This approach offers several advantages: it does not require sterile experiments as we do not need to convert measurable data into stress-strain pairs, it is more robust as it leverages the governing differential equations to improve inversion accuracy; and it can handle complex, nonlinear constitutive models more effectively. Furthermore, we aim to design this framework to scale effectively with the complexity of the constitutive law. Additionally, this framework enables learning material behavior directly from complex loading conditions, enhancing its predictive capabilities for future applications.

We demonstrate the capability of this framework with synthetically generated quasistatic and dynamic data. The data is obtained by solving finite element simulations of a known constitutive model under different loading conditions, providing a controlled setting for evaluating the methodology. The material

parameters are recovered by minimizing an objective which measures the difference between the computed and synthetic reference data. The governing equations serve as a constraint to this minimization problem, which is solved using a gradient based method. An adjoint-based optimization approach is employed to efficiently compute gradients, enabling rapid convergence of the material parameters. The results demonstrate that the framework can accurately recover the underlying constitutive parameters highlighting its robustness and effectiveness for material characterization in both static and dynamic regimes.

This methodology is extended to infer constitutive parameters from experimental data, specifically focusing on dynamic contact problems. Unlike synthetic data, experimental measurements introduce additional complexities such as noise, incomplete information, and model-form uncertainties, which make direct inversion approaches unreliable. Our method considers dynamic contact scenarios, where both inertia effects and contact interactions must be accurately captured within the inversion process [24, 2]. Experimental data is obtained using split Hopkinson pressure bars to indent steel and aluminum samples [12, 27, 11]. Indentation force versus depth data is used to recover material properties. Regularization techniques are incorporated to improve robustness against measurement noise and ill-posedness, enabling stable and reliable parameter recovery. The results demonstrate that this framework can successfully extract meaningful material parameters from experimental data in dynamic contact settings, making it a powerful alternative to traditional direct inversion methods, particularly for nonlinear constitutive models.

The problems outlined above rely on a postulated form of the constitutive law, with the goal of inferring the parameters involved. However, conventional constitutive models often lack physical reasoning, require *a priori* knowledge, and are based on phenomenological laws or analytical approximations. Moreover, as the catalog of newly discovered alloys and materials continues to expand, existing constitutive laws may struggle to generalize across a wide range of materials. The intersection of machine learning (ML) and mechanics has opened new avenues for modeling complex material behaviors beyond the capabilities of traditional constitutive models [17, 14, 30, 21]. By leveraging data-driven techniques, ML offers a powerful alternative for constructing constitutive relations directly from experimental or simulated data, reducing the need for

explicit assumptions about material behavior [17, 5, 33]. In this context, the framework can be extended by replacing the explicit constitutive model with a recurrent neural network (RNN), allowing for a fully data-driven approach to characterize materials [7, 29]. This is particularly useful in cases where the underlying constitutive law is unknown or too complex to be expressed in a closed-form equation. Rather than extracting traditional material parameters, the methodology optimizes the weights and biases of the neural network using dynamic data within a PDE-constrained optimization framework. The neural network serves as a black-box constitutive model, learning the material response directly from experimental or simulated data while ensuring consistency with the governing equations of solid mechanics. Since the adjoint-based optimization strategy scales efficiently with the number of parameters, the size of the neural network does not significantly affect computational time. This approach provides a powerful alternative to traditional constitutive modeling, enabling the characterization of materials without predefined assumptions and allowing for adaptive learning of complex material behaviors directly from data.

1.2 Multiscale modeling

Going a step further, constitutive laws on the continuum scale truly arise from mechanisms occurring at a lower scales [32]. It is quite cumbersome to capture these mechanisms and bridge the gap between the layers, and therefore constitutive laws are assumed to be of some form dependent on *a priori* knowledge. Multiscale modeling is an effective method in dealing with this problem as it derives closure relations from lower scale mechanisms. A set of theories have been developed to model these mechanisms at their individual scales, for instance - Density Functional Theory at the electron scale [18], Molecular Dynamics at the atomistic scale [32], Crystal Plasticity at the polycrystal scale [3], and Continuum Mechanics at the structural scale [19]. Multiscale modeling approaches these complex theories by solving relevant mechanisms at each scale, and stitching the hierarchy back by passing information between scales [35]. This hierarchy is shown in Figure 1.1. The interaction between different scales occurs pairwise with the larger scale filtering the equations and regulating the lower scale. Conventionally, there are two methods of implementing multiscale models. The first approach is known as the sequential multiscale/parameter

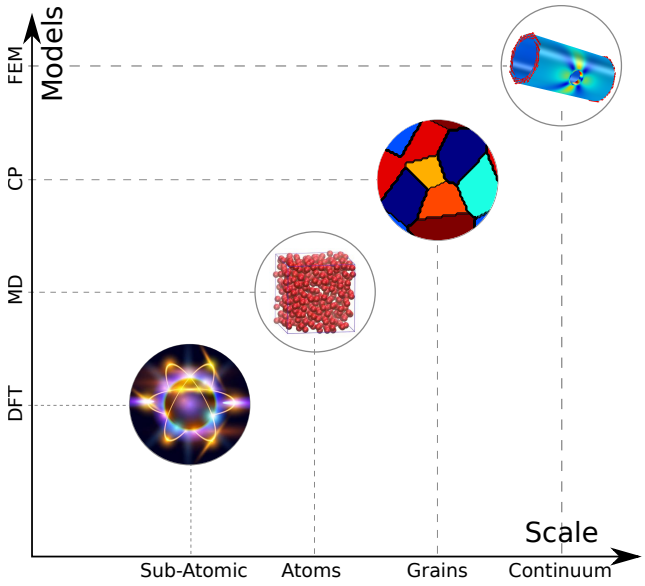


Figure 1.1: The approach of multiscale modeling in mechanics at various levels using different theories.

passing method, which involves solving for model parameters at any given scale using the information from lower scale mechanisms [36]. This approach is motivated by empirical models. The second approach involves evaluating the behavior at every scale simultaneously [10, 34, 15]. This results in higher fidelity compared to the sequential method. However, it is computationally expensive and can require high amounts of processing power. Furthermore, this approach also requires empirical information (e.g., order parameters) to connect models at two scales [6]. Both the above methods also use a very small portion of information for the increased amount of complexity, which does not justify their usage. Therefore, the current methods in multiscale modeling suffer from three challenges — (1) need of empirical knowledge on the interaction between models at different scales, (2) complex equations need to be solved repeatedly only to use trivial amounts of the solution, and (3) expensive computational costs.

Previous research has addressed the complexity of multiscale partial differential equations (PDEs) by employing deep neural networks to learn surrogate models that approximate fine-scale behavior based on offline simulations [28, 29, 26]. These learned models can be seamlessly integrated into coarse-scale computations, bypassing the need for repeated fine-scale evaluations or predefined state variables. Such approaches capture essential physical characteristics, while enabling high-fidelity simulations of macroscopic phenomena, such

as impact problems, at significantly reduced computational cost. By eliminating the need to repeatedly solve the same fine-scale equations, these methods focus only on the necessary macroscopic information. Additionally, because they are based on machine learning, they can be efficiently parallelized on GPUs, further accelerating computations. Central to this paradigm is the use of neural operators, which learn mappings between function spaces and provide a flexible, data-driven framework for approximating solutions to elliptic PDEs [25]. These methods stem from the theory of homogenization, replacing explicit upscaling with learned solution maps that generalize across microstructures and enable fast, coarse-scale predictions without repeated micro-scale solves.

We examine this learning framework, targeting elliptic operators in multiscale problems. Homogenization seeks to derive effective macroscopic equations that describe the behavior of heterogeneous media characterized by multiple spatial scales. We focus on scenarios with complex microstructures—such as discontinuities and sharp interfaces—that pose challenges for traditional learning approaches due to their impact on solution regularity. Our method leverages specialized neural operators to learn coarse-scale solutions directly from fine-scale data, without relying on empirical closure models or handcrafted feature selection. We provide theoretical insights into the learning process and validate our approach through a series of numerical experiments, demonstrating its effectiveness in capturing key macroscopic behaviors while remaining computationally efficient. This work offers a scalable, non-intrusive path toward learning homogenized models in complex multiscale systems.

The development of a neural operator framework capable of efficiently approximating solutions to multiscale elliptic PDEs, we recognize the broader potential of this approach for solving other challenging problems in mechanics. They provide a powerful tool for learning solution maps directly from data, bypassing the need for repeated fine-scale computations once trained. This capability makes them well-suited for computational tasks that involve iterative PDE solves, such as topology optimization, uncertainity quantification and evolutionary problems. We demonstrate the case of topology optimization, where sensitivities must be computed repeatedly over many design iterations.

1.3 Topology optimization

Topology optimization is a computational design method that optimally distributes material within a given domain to achieve a desired performance objective while satisfying constraints such as weight, stress limits, or manufacturability. It has been widely used in structural mechanics, aerospace engineering, and materials design to develop lightweight and high-performance structures [1]. Classical topology optimization methods rely on gradient-based or heuristic approaches to iteratively refine the material distribution. These techniques have enabled significant advancements in automated design but also suffer from several limitations [4]. The iterative nature of these methods requires repeatedly solving finite element analyses (FEA), making the process computationally intensive. Additionally, conventional approaches struggle with capturing complex microstructural patterns, often requiring ad hoc regularization techniques or additional post-processing steps to ensure manufacturability. Another key limitation is the dependency on predefined interpolation schemes for material properties, which can constrain the diversity of achievable designs and limit the method's applicability to emerging material systems.

To overcome the computational and representational limitations of traditional topology optimization methods, we integrate a Principal Component Analysis (PCA)-based neural network within the optimization loop. Instead of directly optimizing the full high-dimensional design space, a reduced-order representation is learned using PCA, capturing the most significant design variations while drastically reducing computational costs . At each iteration of the topology optimization process, the design is represented in this low-dimensional latent space, and updates are performed efficiently before reconstructing the full structure. This approach not only accelerates convergence by reducing the number of required finite element analyses but also enhances the diversity of achievable designs by enabling smooth and data-driven design updates.

1.4 Outline of the thesis

Here we present five studies which address the above problems of constitutive modeling, material characterization and accelerating computations in the following five chapters. Each chapter is self-contained and contains a detailed literature survey related to that study.

In the first study included included in Chapter 2, we provide a detailed overview of the existing material characterization techniques and highlight their limitations. We then introduce a novel approach for characterizing material behavior by recovering material properties from partial observations. The formulation, solution strategy, and associated numerical techniques are presented in detail. To demonstrate the efficacy of the method, we recover viscoplastic material parameters from synthetically generated quasistatic and dynamic datasets.

In Chapter 3, we extend our framework to dynamic contact problems and demonstrate its applicability using experimental data. Specifically, we infer material parameters for RHA (Rolled Homogeneous Armor) steel and polycrystalline aluminum alloy obtained using indentation tests. Two key things highlighted in the chapter are the formulation of the inverse problem in the presence of contact constraints and the regularization used to stabilize noisy data. Our results show that the proposed method can accurately recover material properties for real-world applications.

A key limitation of the previous framework is its reliance on a pre-specified form of the constitutive law. Furthermore, traditional constitutive models suffer from predicting the response of the evergrowing catalog materials. We move on to chapter 4 where we depart from traditional constitutive models and use a neural network constitutive law, allowing the constitutive relation to emerge from data without prescribing a form. A recurrent neural operator is used to describe a history dependent material response using internal variables. We demonstrate our framework using synthetic high strain rate compression experiments, where the RNO-based model significantly outperforms classical models in capturing path-dependent responses.

In chapter 5, we explore neural networks' ability to approximate solutions to partial differential equations. We shift our focus to the problem of multiscale modeling of elliptic PDEs and use neural operators to learn effective macroscopic behavior from complex microstructured media, eliminating the need to repeatedly solve fine-scale problems. We target the classical homogenization setting of linear elliptic PDEs with discontinuous coefficients, which is representative of many applications in solid mechanics. A key challenge in this setting is the presence of sharp interfaces and corners in the microstructure, which degrade solution regularity and make learning difficult. We analyze the

approximation capabilities of neural operators in this regime, provide theoretical guarantees, and demonstrate the approach on several representative microstructure classes. This work shows that data-driven homogenization using neural operators provides a scalable and non-intrusive path toward multiscale modeling, and sets the stage for future applications in nonlinear and history-dependent materials.

Finally, we apply these data-driven tools to topology optimization in chapter 6, where design tasks typically require many expensive PDE solves. By integrating a reduced-order PCA-based neural network into the optimization loop, we represent complex structures in a low-dimensional latent space and perform efficient updates. This significantly reduces computational cost while maintaining design quality and flexibility. The results illustrate how machine learning can accelerate and enhance classical optimization pipelines, paving the way for next-generation design frameworks.

We conclude in chapter 7 with a summary of results and contributions toward constitutive modeling and usage of neural networks in solid mechanics. Lastly, we provide some suggestions for future work on the discussed approaches.

References

[1] A. Akerson. *Optimal Design of Soft Responsive Actuators and Impact Resistant Structures.* California Institute of Technology, 2023.

[2] R. J. Anton and G. Subhash. "Dynamic Vickers indentation of brittle materials." In: *Wear* 239.1 (2000), pp. 27–35.

[3] R. J. Asaro. "Crystal Plasticity." In: *Journal of Applied Mechanics* 50.4b (1983), pp. 921–934.

[4] M. P. Bendsoe and O. Sigmund. *Topology optimization: theory, methods, and applications.* Springer Science & Business Media, 2013.

[5] M. A. Bessa, R. Bostanabad, Z. Liu, A. Hu, D. W. Apley, C. Brinson, W. Chen, and W. K. Liu. "A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality." In: *Computer Methods in Applied Mechanics and Engineering* 320 (2017), pp. 633–667.

[6] K. Bhattacharya. "Phase boundary propagation in a heterogeneous body." In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 455.1982 (1999), pp. 757–766.

[7] K. Bhattacharya, B. Liu, A. Stuart, and M. Trautner. "Learning Markovian homogenized models in viscoelasticity." In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 641–679.

[8] E. Billington and A. Tate. *Physics of Deformation and Flow: An Introduction.* McGraw Hill, 1981.

[9] W. D. Callister Jr and D. G. Rethwisch. *Materials science and engineering: An introduction.* John Wiley & Sons, 2020.

[10] R. Car and M. Parrinello. "Unified approach for molecular dynamics and density-functional theory." In: *Physical Review Letters* 55.22 (1985), p. 2471.

[11] D. Casem and E. Retzlaff. "A kolsky bar for high-rate indentation." In: *Journal of Dynamic Behavior of Materials* 9.3 (2023), pp. 300–314.

[12] J. Clayton, J. Lloyd, and D. Casem. "Simulation and dimensional analysis of instrumented dynamic spherical indentation of ductile metals." In: *International Journal of Mechanical Sciences* 251 (2023), p. 108333.

[13] A. Constantinescu and N. Tardieu. "On the identification of elastoviscoplastic constitutive laws from indentation tests." In: *Inverse Problems in Engineering* 9.1 (2001), pp. 19–44.

[14] J. Dornheim, L. Morand, H. J. Nallani, and D. Helm. "Neural networks for constitutive modeling: From universal function approximators to advanced models and the integration of physics." In: *Archives of computational methods in engineering* 31.2 (2024), pp. 1097–1127.

[15] F. Feyel and J.-L. Chaboche. "Multi-scale non-linear FE2 analysis of composite structures: damage and fiber size effects." In: *Revue européenne des éléments finis* 10.2-4 (2001), pp. 449–472.

[16] N. Fleck and J. W. Hutchinson. "Strain gradient plasticity." In: *Advances in Applied Mechanics* 33 (1997), pp. 295–361.

[17] J. N. Fuhg, G. Anantha Padmanabha, N. Bouklas, B. Bahmani, W. Sun, N. N. Vlassis, M. Flaschel, P. Carrara, and L. De Lorenzis. "A review on data-driven constitutive laws for solids." In: *Archives of Computational Methods in Engineering* (2024), pp. 1–43.

[18] F. Giustino. *Materials modelling using density functional theory: Properties and predictions.* Oxford University Press, 2014.

[19] M. E. Gurtin, E. Fried, and L. Anand. *The mechanics and thermodynamics of continua.* Cambridge University Press, 2010.

[20] G. A. Holzapfel. *Nonlinear solid mechanics: a continuum approach for engineering science.* Kluwer Academic Publishers Dordrecht, 2002.

[21] A. Hussain, A. H. Sakhaei, and M. Shafiee. "Machine learning-based constitutive modelling for material non-linearity: A review." In: *Mechanics of Advanced Materials and Structures* (2024), pp. 1–19.

[22] K. T. Kavanagh and R. W. Clough. "Finite element applications in the characterization of elastic solids." In: *International Journal of Solids and Structures* 7.1 (1971), pp. 11–23.

[23] T. Kirchdoerfer and M. Ortiz. "Data-driven computational mechanics." In: *Computer Methods in Applied Mechanics and Engineering* 304 (2016), pp. 81–101.

[24] B. Koeppel and G. Subhash. "An experimental technique to investigate the dynamic indentation hardness of materials." In: *Experimental Techniques* 21.3 (1997), pp. 16–18.

[25] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. "Neural operator: Learning maps between function spaces with applications to pdes." In: *Journal of Machine Learning Research* 24.89 (2023), pp. 1–97.

[26] N. Kovachki, B. Liu, X. Sun, H. Zhou, K. Bhattacharya, M. Ortiz, and A. Stuart. "Multiscale modeling of materials: Computing, data science, uncertainty and goal-oriented optimization." In: *Mechanics of Materials* 165 (2022), p. 104156.

[27] A. Lee and K. Komvopoulos. "Dynamic spherical indentation of elastic-plastic solids." In: *International Journal of Solids and Structures* 146 (2018), pp. 180–191.

[28] B. Liu, N. Kovachki, Z. Li, K. Azizzadenesheli, A. Anandkumar, A. M. Stuart, and K. Bhattacharya. "A learning-based multiscale method and its application to inelastic impact problems." In: *Journal of the Mechanics and Physics of Solids* 158 (2022), p. 104668.

[29] B. Liu, E. Ocegueda, M. Trautner, A. M. Stuart, and K. Bhattacharya. "Learning macroscopic internal variables and history dependence from microscopic models." In: *Journal of the Mechanics and Physics of Solids* 178 (2023), p. 105329.

[30] X. Liu, S. Tian, F. Tao, and W. Yu. "A review of artificial neural networks in the constitutive modeling of composite materials." In: *Composites Part B: Engineering* 224 (2021), p. 109152.

[31] L. E. Malvern. *Introduction to the Mechanics of a Continuous Medium.* Monograph. 1969.

[32] R. Phillips. *Crystals, defects and microstructures: Modeling across scales.* Cambridge University Press, 2001.

[33] 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." In: *Journal of Computational Physics* 378 (2019), pp. 686–707.

[34] E. B. Tadmor, M. Ortiz, and R. Phillips. "Quasicontinuum analysis of defects in solids." In: *Philosophical magazine A* 73.6 (1996), pp. 1529–1563.

[35] E. Van Der Giessen, P. A. Schultz, N. Bertin, V. V. Bulatov, W. Cai, G. Csányi, S. M. Foiles, M. G. Geers, C. González, M. Hütter, et al. "Roadmap on multiscale materials modeling." In: *Modelling and Simulation in Materials Science and Engineering* 28.4 (2020), p. 043001.

[36] E. Weinan. *Principles of multiscale modeling.* Cambridge University Press, 2011.

[37] A. Wihardja and K. Bhattacharya. "Constitutive relations from images." In: *arXiv preprint arXiv:2504.00898* (2025).

[38] A. Wineman. "Nonlinear viscoelastic solids—A review." In: *Mathematics and Mechanics of Solids* 14.3 (2009), pp. 300–366.

Chapter 2

LEARNING CONSTITUTIVE RELATIONS FROM EXPERIMENTS: I. PDE CONSTRAINED OPTIMIZATION

Akerson, A., Rajan, A. and Bhattacharya, K., 2025. "Learning constitutive relations from experiments: 1. PDE constrained optimization." *Journal of the Mechanics and Physics of Solids*, p.106128.

Contributions: A.R. and A.A. developed the computational code and performed simulations. A.A. and K.B. conceptualized the methodology. A.R., A.A. and K.B. wrote the manuscript.

2.1 Introduction

Continuum mechanics is an approach to solving problems involving complex phenomena directly at the scale of applications (e.g., [18, 8]). It exploits universal laws of physics: the balance of mass, momenta, energy as well as the second law of thermodynamics. However these equations are not closed and a constitutive relation that characterizes the material is necessary to do so. This constitutive relation is typically obtained empirically by conducting experiments. However, we cannot measure the constitutive relation directly and it has to be obtained through the solution of an inverse problem. In fact, we cannot even directly measure the quantities like stress, strain, strain rate and energy density that comprise the constitutive relation. Instead, they have to be inferred from quantities like displacements and total forces that can be measured in the laboratory. Thus, the problem of inferring constitutive relations from experiments, and thereby completing the continuum mechanical formulation, requires the solution of an indirect inverse problem.

The classical approach is to design an experimental setup that is consistent with either uniform states of stress and strain/strain-rate („ uniaxial tension in solids or shear rheometers in complex fluids), or a universal solution (e.g., torsion in solids or viscometric flow in complex fluids); these enable a semi-

analytic solution to the indirect inverse problem (e.g., [8]). There are a number of difficulties with this approach. First, we obtain information about the constitutive relation only in simple, idealized states of stress (history) while we seek to deploy them in more complex situations. For example, we may obtain only uniaxial tensile data while we may want to use the model in multiaxial, non-proportional loading scenarios. Second, one has to conduct a large number of tests since each test can only provide a limited amount of information on the constitutive behavior. For example, we may have to repeat a test at various strain rates, or with specimens with different orientation. Third, one may need sterile conditions with precise alignment (e.g., plate impact) to obtain the desired state. This limits the number of tests that one can conduct. Together, these and other shortcomings contribute significantly to the uncertainty, and limit the fidelity of the constitutive relations that can be obtained from experiments.

It is also classical to postulate *a priori* an explicit formula for the constitutive relation with a few constants, and then use above-described experiments to fit the constants (e.g., [8]). Since it is common practice to use a very few constants, the limited experimental information is sufficient to fit them. However, the ability of explicit formulas with a few constant to represent complex behavior is limited. Even in the setting of isotropic, incompressible hyperelasticity, we have a still-evolving menagerie of models (e.g., [33, 26]). It is therefore natural to try to go beyond hypo-parametrized explicit formulas to hyper-parametrized neural networks and other approximations that have proven to be enormously successful in a variety of fields. However, this requires significant amounts of data, much more than the traditional approach can deliver.

There is a closely related issue. In history dependent and structured continua, one also has to postulate *a priori* the internal or state variables that describe the history dependence/internal structure [8]. For example, we introduce plastic strain, accumulated plastic strain and a kinematic hardening variable as internal variables in plasticity. However, the choice of internal variables is unclear in complex anisotropic situations like in composite media or shape memory alloys [23]. Instead, one would like to infer these directly from experimental measurements, further increasing the need for empirical data.

The recent decade has seen a revolution in experimental methods primarily

driven by the growth of full-field observations techniques. In particular, digital image correlation (DIC) [12, 46, 21], where one infers deformation by comparing images of a complex (speckle) pattern imprinted on a surface taken before and after deformation, is widely used. The method has been extended to stereo DIC to obtain out-of-plane deformation [45], the high resolution setting using the scanning electron microscope [24], to dynamics [42] and to digital volume correlation [10, 53]. DIC enables the measurement of the entire deformation and strain field on a surface. This opens the intriguing possibility of breaking away from homogeneous (or idealized) states: instead, one can probe many different strain paths in a single test by working with specimens with complex geometries. However, there are challenges. While one can measure the strain field, it is limited to a single surface. More importantly, it is not possible to measure the stress field, and instead, we typically can only measure the total reaction force on a surface. So we need to address the difficult indirect inverse problem discussed earlier.

Another rapidly evolving technique that provides full-field information is high energy diffraction microscopy (HEDM) that uses high energy synchrotron radiation and diffraction to obtain the crystal structure on a three-dimensional voxelated volume [7]. Comparing the obtained crystal structure with the stress-free structure gives us the lattice strain on a voxelated volume. One can convert this to stress by using the elastic modulus of the material. However, there are two issues. First, one only obtains the lattice strain, and not the total strain. Therefore, one still has to solve an indirect inverse problem in inelastic phenomena. Second, this is averaged over the volume of the voxel with some unknown kernel, and therefore subject to errors. For example, the resulting stress fields obtained by this approach are not equilibrated [54]. While we do not address HEDM in this paper, the ideas presented may be extended to this technique.

A variety of approaches have been proposed to solve this indirect inverse problem. The key idea is to use the balance laws, and specifically equilibrium in some form. We review a few contributions to provide context to our work, without attempting a comprehensive review of this rapidly growing literature. One approach is to use a combination of model reduction and Gaussian process regression to obtain the parameters associated with a particular constitutive relation (see [49, 50] for the use of this approach in viscoelastic materials). It is

possible to combine Bayesian uncertainty quantification with such approaches (see [52] on toughness). While such approaches are effective in identifying a small number of parameters, they scale exponentially with the number of parameters and are prohibitively expensive for complex materials or neural network representations.

Miller, Dawson and collaborators pioneered the concurrent use of finite element analysis with given constitutive relations and experimental observations using high energy x-ray diffraction with the goal of understanding the relationship between single crystal elastic modulus and hardening laws and heterogeneity in stress distribution in polycrystals.[15, 36, 51, 11]. However, they do not seek to obtain the full constitutive relation.

An emerging approach is to use a physics informed neural network (PINN) where the constitutive relation is represented as a neural network and the (failure of the fields to satisfy) balance laws are used as a part of the objective (loss function) to be minimized as a part of the fitting [19, 16, 48]. The balance laws are not enforced exactly in this approach.

In the virtual field method (VFM) (for example, [17, 40, 29, 31, 27]), the balance laws are integrated against a set of "virtual fields" to obtain a system of equations for the unknown constants in a postulated constitutive relation. The system is linear when the constitutive relation is linear in the constants as in the case of hyperelasticity, but is typically solved by linearization otherwise. The application of the method requires one to specify kinematically admissible virtual fields, and one may take them to be the measured strain field. However, the strain field is typically measured on a part of the boundary of the specimen, and has to be extrapolated to obtain it over the entire domain. Further, one needs as many independent full field measurements as there are unknown constitutive constants even for complex domains. This limits the complexity of the constitutive model. Finally, only a finite dimensional approximation of the balance laws is enforced.

A widely used approach is the finite element updating (FEMU) (for example, [25, 38, 32, 34, 20, 44, 28, 43, 13]). The idea is to formulate the problem as an inverse method which is solved iteratively by using a finite element method to solve the forward problem, and then to update the model parameters using a numerically computed sensitivity. This has been applied to a variety of history dependent phenomena and can be integrated with the inverse problem

of digital image correlation [32]. An important consideration here is that the sensitivity, or the gradient of the objective with respect to the objective, is computed numerically. While a number of ideas can be used to speed this up, this approach scales poorly with the number of material parameters. A closely related idea is to formulate the problem as a PDE-constrained optimization problem where the balance laws act as the PDE constraint, and then use the adjoint method to compute the sensitivity. This has been applied to elasticity [35], viscoelasticity [14] and Norton-Hoff viscoplasticity [9].

In this work, we build on the last two lines of work, and propose a method to accurately and efficiently identify the constitutive behavior through experiments. We formulate this as an indirect inverse problem that is constrained by the balance laws (PDE constraint). The objective that we seek to minimize is the difference between the experimental observation and the corresponding quantities computed using the constitutive model. We then formulate the forward problem as a boundary value problem corresponding to the experiment, and the problem of computing sensitivity of the solution to the parameters as an adjoint problem. This is a partial differential equation that is linear in space and quasilinear in time. The adjoint equation has widely been used in optimal design in mechanics (e.g., [6, 2]). In this work, we implement both the forward and the adjoint problem using finite element approximation. The cost of solving the adjoint problem is that of a single iteration of the forward problem, and the parameters can be updated at little cost from the solution of the adjoint problem. Thus, the core computation of the method is independent of the number of parameters, and thus suited for complex models with numerous parameters (including neural networks).

We formulate and demonstrate the method for both quasistatic and dynamic experiments with synthetic data in the current Part 1. We extend the formulations to include contact in Part 2 and demonstrate it using experimental data from a dynamic indentation test. We demonstrate the method for models formulated as neural networks in Part 3. Together, these show that the method is robust and can be adapted for various experimental situations. While we focus on elasto-viscoplasticity, the ideas are valid for any local constitutive relation.

The paper is organized as follows. We present the method in Section 2.2. We do so for an arbitrary inelastic material described by a rate-dependent internal

variable theory in Section 2.2.1, and for a J2 elasto-viscoplastic model with yield in Section 2.2.2. We overview the numerical method in Section 2.2.3 with details provided in the appendix. We demonstrate and validate the method using synthetic data on two examples in Section 2.3: a thick plate with a hole in quasistatic compression in Section 2.3.1 and an extended dynamic impact test in Section 2.3.2. We conclude with some comments in Section 2.4.

2.2 Formulation and method

We first present our method for a general history-dependent constitutive relation, and then specialize to elasto-viscoplasticity.

2.2.1 Internal variable theory

Governing equations

We consider an open, bounded domain $\Omega \subset \mathbb{R}^n$ occupied by a solid body undergoing a deformation with displacement $u(x,t)$ over time interval $(0,T)$. We assume the body is composed of a material of density ρ, and one whose constitutive behavior is described by a set of internal parameters $\xi := \{\xi_i\}$, $i = 1,\dots,N_\xi$ that evolve according to a Perzyna-type kinetic law:

$$\begin{aligned} \sigma &= S(F,\xi;P), \\ \dot{\xi} &= R(F,\xi;P), \end{aligned} \tag{2.1}$$

where σ is the first Piola-Kirchoff stress, F is the deformation gradient, and S and R are constitutive functions parametrized by material parameters $P := \{P_i\}$, $i = 1,\dots,N_p$. The deformation of the body is then governed by the equations

$$\begin{aligned} &\rho\ddot{u} - \nabla\cdot\sigma = b && \text{on } \Omega, \\ &\sigma = S(\nabla u,\xi;P) && \text{on } \Omega, \\ &\dot{\xi} = R(\nabla u,\xi;P) && \text{on } \Omega, \\ &\sigma n = f && \text{on } \partial_f\Omega, \\ &u = u_0 && \text{on } \partial_u\Omega, \\ &u|_{t=0} = \dot{u}|_{t=0} = \xi|_{t=0} = 0, && \text{on } \Omega, \end{aligned} \tag{2.2}$$

where b is the applied body force, f is the applied (dead) traction on part $\partial_f\Omega$ of the boundary while u_0 is the applied displacement on the complement $\partial_d\Omega$ of

the boundary. Note that $\partial_f\Omega$ typically includes regions that are traction-free. We assume quiescent initial conditions.

Indirect inverse problem of parameter identification

We are given experimental data, usually displacements at certain instances of time on just a portion of the boundary of the domain (some subset of the traction-free portion of $\partial_f\Omega$) and some overall force or moment acting on some part of the boundary (some subset of $\partial_u\Omega$). We label the experimental data $D^{\text{exp}} := \{u^{\text{exp}}, M^{\text{exp}}\}$. Here, u^{exp} is the set of partial displacement measurements, and M^{exp} is a collection of macroscopic measurable quantities which could consist of net loads or averaged strains. The goal is to find the parameter set P such that the modeled trajectory and computed macroscopic quantities from solving (2.2) match the experimental data D^{exp}. We write this as an optimization problem

$$\inf_{P\in\mathcal{P}} \mathcal{O}(P, u, \xi, D^{\text{exp}}) \qquad \text{subject to } \{u, \xi\} \text{ satisfying (2.2)}, \tag{2.3}$$

where $\mathcal{P}$ defines a physical range of parameters and $\mathcal{O}(P, u, \xi.D)$ is the objective or loss function.

We need to choose an objective function $\mathcal{O}$ that is both computationally efficient and one that attains a minimum when the displacement and forces computed using the solutions $\{u, \xi\}$ to (2.2) are equal to the measured values D^{exp}. A somewhat subtle point in internal variable theories like (2.1) is that the internal variable is only defined pointwise, and thus does not have a mathematical meaning (specifically a trace) on the boundary. However, the traction is mathematically defined through the interior stress distribution. Thus, we have to interpret the measurement of boundary forces accordingly. In this work we use a finite element discretization, and the governing equations are used in their weak form. Specifically, the displacements are imposed on the nodes, but the strains, stress and internal variables are defined on the quadrature points and inherit their meaning by integration. Therefore, it is natural to write the objective as a volume integral even though the force measurements are made on the boundary. So, we take

$$\mathcal{O}(P, u, \xi, D^{\text{exp}}) := \int_0^T \int_\Omega o(P, u, \xi, D^{\text{exp}})\, d\Omega\, dt. \tag{2.4}$$

for some o that is concentrated near the boundary. We discuss specific choices of o in Section 2.3.

Adjoint method for sensitivity

A gradient based optimization approach requires that we compute the sensitivities, that is, the total derivative of the objective with respect to the parameter set P while enforcing the PDE constraint of the governing equations. We may compute this by applying standard chain rule. However, this would require expressions for $\frac{\mathrm{d}u}{\mathrm{d}P}$ and $\frac{\mathrm{d}\xi}{\mathrm{d}P}$, which are the sensitivity of the solutions of (2.2) to changes in parameters. These are difficult to compute. Therefore, we use the adjoint method [22, 41] to circumvent the difficulty. We rewrite the objective by adding a term that is zero according to the weak form of the evolution equation,

$$\begin{aligned}\mathcal{O} = &\int_0^T \int_\Omega o(P, u, \xi, D^{\exp})\ d\Omega\ dt \\ &+ \int_0^T \left\{ \int_\Omega \left[-\rho \ddot{u} \cdot v - \sigma \cdot \nabla v + b \cdot v - \phi(\dot{\xi} - R) \right]\ d\Omega \right. \\ &\left. + \int_{\partial_f \Omega} f \cdot v\ dS \right\} dt,\end{aligned} \tag{2.5}$$

where v and ϕ are test functions associated with momentum balance and internal variable evolution. The idea of the adjoint method is to choose the test functions v and ϕ such that there is no need to explicitly compute the sensitivity of the solutions of (2.2) to changes in parameters ($\frac{\mathrm{d}u}{\mathrm{d}P}$ and $\frac{\mathrm{d}\xi}{\mathrm{d}P}$). Differentiating this rewritten objective with respect to P gives

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg[&\frac{\partial o}{\partial P} + \frac{\partial o}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}P} + \frac{\partial o}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} - \rho \frac{\mathrm{d}\ddot{u}}{\mathrm{d}P} \cdot v \\ &- \left(\frac{\partial \sigma}{\partial P} + \frac{\partial \sigma}{\partial \nabla u} \cdot \nabla \frac{\mathrm{d}u}{\mathrm{d}P} + \frac{\partial \sigma}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} \right) \cdot \nabla v \\ &- \phi \left(\frac{\mathrm{d}\dot{\xi}}{\mathrm{d}P} - \frac{\partial R}{\partial P} - \frac{\partial R}{\partial \nabla u} \cdot \nabla \frac{\mathrm{d}u}{\mathrm{d}P} - \frac{\partial R}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} \right) \Bigg]\ d\Omega\ dt.\end{aligned} \tag{2.6}$$

We group terms to obtain

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg[& \left(\frac{\partial o}{\partial P} - \frac{\partial \sigma}{\partial P}\cdot\nabla v + \phi\frac{\partial R}{\partial P}\right) - \rho v \cdot \frac{\mathrm{d}\ddot{u}}{\mathrm{d}P} \\ & - \left(\nabla v \cdot \frac{\partial \sigma}{\partial \nabla u} - \phi\frac{\partial R}{\partial \nabla u}\right)\cdot\left(\nabla\left(\frac{\mathrm{d}u}{\mathrm{d}P}\right)\right) \\ & + \frac{\partial o}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}P} - \phi\frac{\mathrm{d}\dot{\xi}}{\mathrm{d}P} + \left(\frac{\partial o}{\partial \xi} - \nabla v\cdot\frac{\partial \sigma}{\partial \xi} + \frac{\partial R}{\partial \xi}\phi\right)\frac{\mathrm{d}\xi}{\mathrm{d}P}\Bigg]\, d\Omega\, dt.\end{aligned} \tag{2.7}$$

The terms that include $\frac{\mathrm{d}\ddot{u}}{\mathrm{d}P}$ and $\frac{\mathrm{d}\dot{\xi}}{\mathrm{d}P}$ are integrated by parts temporally, while the term that includes $\nabla\left(\frac{\mathrm{d}u}{\mathrm{d}P}\right)$ is treated with divergence theorem in space. This yields

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg[& \left(\frac{\partial o}{\partial P} - \frac{\partial \sigma}{\partial P}\cdot\nabla v + \phi\frac{\partial R}{\partial P}\right) \\ & + \left[-\rho\ddot{v} + \frac{\partial o}{\partial u} + \nabla\cdot\left(\nabla v\cdot\frac{\partial \sigma}{\partial \nabla u} - \phi\frac{\partial R}{\partial \nabla u}\right)\right]\cdot\frac{\mathrm{d}u}{\mathrm{d}P} \\ & + \left(\dot{\phi} + \frac{\partial o}{\partial \xi} - \nabla v\cdot\frac{\partial \sigma}{\partial \xi} + \frac{\partial R}{\partial \xi}\phi\right)\frac{\mathrm{d}\xi}{\mathrm{d}P}\Bigg]\, d\Omega\, dt \\ - \int_0^T \int_{\partial\Omega} & \left(\nabla v\cdot\frac{\partial \sigma}{\partial \nabla u} - \phi\frac{\partial R}{\partial \nabla u}\right) n\cdot\frac{\mathrm{d}u}{\mathrm{d}P}\, dS\, dt \\ + \Bigg[\int_\Omega & \left(-\rho v\cdot\frac{\mathrm{d}\dot{u}}{\mathrm{d}P} + \rho\dot{v}\cdot\frac{\mathrm{d}u}{\mathrm{d}P} - \phi\frac{\mathrm{d}\xi}{\mathrm{d}P}\right)\, d\Omega\Bigg]_0^T.\end{aligned} \tag{2.8}$$

We eliminate the dependence on $\frac{\mathrm{d}u}{\mathrm{d}P}$ and $\frac{\mathrm{d}\xi}{\mathrm{d}P}$ by choosing a particular v and ϕ such that the respective terms multiplying them become zero. As $\frac{\mathrm{d}\dot{u}}{\mathrm{d}P}\Big|_{t=0} = \frac{\mathrm{d}u}{\mathrm{d}P}\Big|_{t=0} = 0$, and $\frac{\mathrm{d}\xi}{\mathrm{d}P}\Big|_{t=0} = 0$ (since $u, \dot{u}$ and ξ are prescribed at $t = 0$ as initial conditions), we eliminate the contributions from boundary terms by choosing $v|_{t=T} = \dot{v}_{t=T} = 0$ and $\phi|_{t=T} = 0$.

We conclude that the expression for the sensitivity is

$$\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \left[\frac{\partial o}{\partial P} - \frac{\partial \sigma}{\partial P}\cdot\nabla v + \phi\frac{\partial R}{\partial P}\right]\, d\Omega\, dt \tag{2.9}$$

where the adjoint variables v and ϕ satisfy the evolution

$$\begin{aligned}
&\rho\ddot{v} - \nabla\cdot\left(\nabla v\cdot\frac{\partial\sigma}{\partial\nabla u} - \phi\frac{\partial R}{\partial\nabla u}\right) = \frac{\partial o}{\partial u} && \text{on } \Omega,\\
&\dot{\phi} = \nabla v\cdot\frac{\partial\sigma}{\partial\xi} - \frac{\partial R}{\partial\xi}\phi - \frac{\partial o}{\partial\xi} && \text{on } \Omega,\\
&\left(\nabla v\cdot\frac{\partial\sigma}{\partial\nabla u} - \phi\frac{\partial R}{\partial\nabla u}\right)n = 0 && \text{on } \partial_f\Omega,\\
&v = 0 && \text{on } \partial_d\Omega,\\
&v|_{t=T} = \dot{v}|_{t=T} = 0,\ \phi|_{t=T} = 0.
\end{aligned} \tag{2.10}$$

We note that the conditions for the adjoint variables are given at the end time $t = T$. Thus, the adjoint system should be solved backwards in time starting from $t = T$ and moving to $t = 0$.

Putting all of this together, we use the following procedure to solve the optimization problem (2.3). We start with an initial value for the parameter set P. We then solve the forward problem (2.2) for $u(t),\ \xi(t)$ followed by the adjoint problem (2.10) for $v(t),\ \phi(t)$. These are used in (2.9) to compute the sensitivities. The sensitivities are used to update the parameters and the process is repeated until convergence. The schematic for the algorithm is shown in Figure 2.1.

Scaling

We now discuss how the numerical cost of the algorithm shown in Figure 2.1 scales respect to the number of parameters. The forward problem ((2.2) labeled (F) in the figure) is largely independent of the number of parameters, as is the adjoint problem ((2.10) labeled (A)). The calculation of the sensitivity ((2.9) labeled (S)) and the parameter update (labeled (U)) scale linearly, but do not involve the solution of any equations. Thus the overall cost of each iteration is

$$\text{Cost}(N_P) = A + BN_P, \qquad \text{where } A >> B \tag{2.11}$$

where N_P is the number of parameters. In other words, the cost is linear with a small coefficient. This is in contrast with gradient-free approaches that scale as $A\alpha^{N_P}$ for some α since N_P is the dimension of the search space.

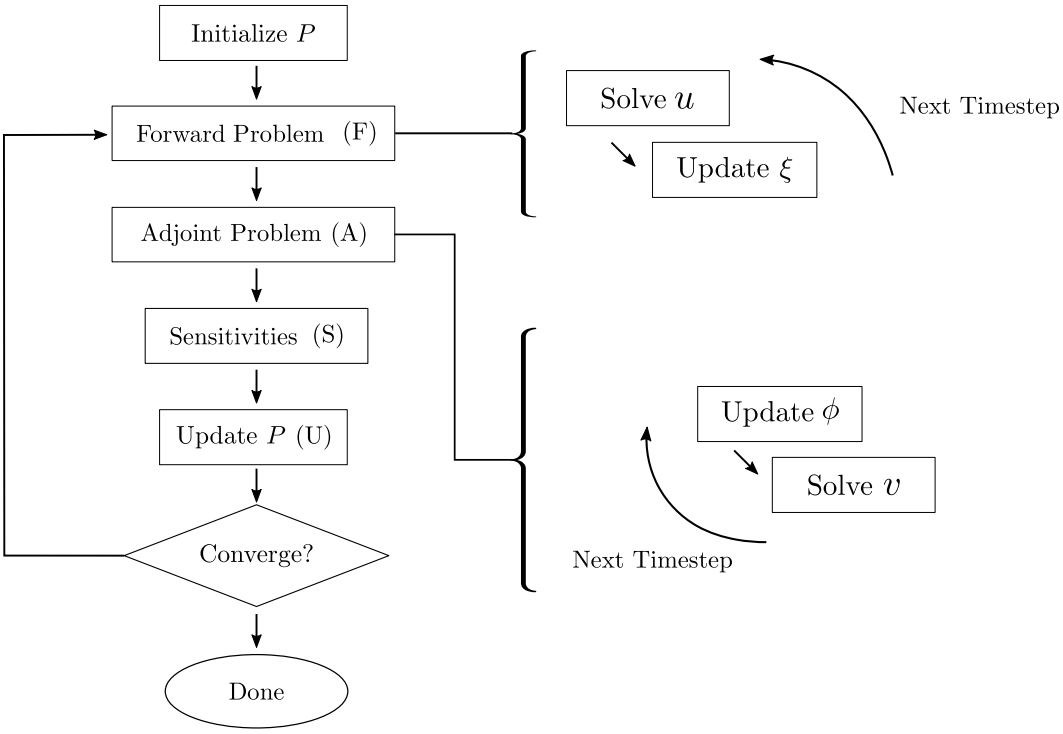


Figure 2.1: Schematic representation of the iterative algorithm to obtain material parameters. The forward and adjoint problem are given by (2.2) and (2.10), respectively. The sensitivity is given by (2.9).

2.2.2 J2 Elasto-viscoplasticity

Governing equations

We now specialize to a specific example of a J2 plastic material with isotropic power-law hardening and rate dependence in the small strain setting [37, 30]. We denote the displacement as u and the strain as $\varepsilon = (\nabla u + \nabla u^T)/2$. The plastic strain is ε^p and the accumulated plastic strain is

$$\dot{q} = \sqrt{\frac{2}{3}\dot{\varepsilon^p} \cdot \dot{\varepsilon^p}}. \tag{2.12}$$

The body is linearly elastic before yield, with the stress and stored elastic energy density as

$$\sigma = \mathbb{C}\varepsilon^e \quad \text{and} \quad W^e = \frac{1}{2}\varepsilon^e \cdot \mathbb{C}\varepsilon^e, \tag{2.13}$$

respectively, where $\mathbb{C}$ is the elastic modulus and $\varepsilon^e = \varepsilon - \varepsilon^p$ is the elastic strain. We assume a von Mises yield criterion with yield strength σ_y, power

law isotropic hardening with potential

$$W^p(q) = \sigma_y \left[q + \frac{n\varepsilon_0^p}{n+1} \left(\frac{q}{\varepsilon_0^p} \right)^{(n+1)/n} \right], \tag{2.14}$$

with ε_0^p the reference plastic strain and n the hardening exponent, and isotropic power law rate hardening with dissipation potential

$$\psi(\dot{q}) = \begin{cases} g^*(\dot{q}) := \frac{m\sigma_y \dot{\varepsilon}_0^p}{m+1} \left(\frac{\dot{q}}{\dot{\varepsilon}_0^p} \right)^{(m+1)/m} & \dot{q} \geq 0, \\ \infty & \dot{q} < 0 \end{cases}, \tag{2.15}$$

where $\dot{\varepsilon}_0^p$ is the reference plastic strain rate, and m is the rate-sensitivity exponent.

The governing equations for the fields $\{u, q, \varepsilon^p\}$ are

$$\begin{aligned} \rho\ddot{u} &= \nabla \cdot \mathbb{C}\varepsilon^e && \text{on } \Omega, \\ 0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot{q}) && \text{on } \Omega, \\ \dot{\varepsilon}^p &= \dot{q}M && \text{on } \Omega, \\ u &= u_0 && \text{on } \partial_u\Omega, \\ f &= \mathbb{C}\varepsilon^e \cdot \hat{n} && \text{on } \partial_f\Omega, \\ u|_{t=0} &= \dot{u}|_{t=0} = 0,\ q|_{t=0} = 0,\ \varepsilon^p|_{t=0} = 0, \end{aligned} \tag{2.16}$$

where u_0 is the applied displacement on $\partial_u\Omega$, f is the applied traction on $\partial_u\Omega$, $\sigma_M = \sqrt{\frac{2}{3}s \cdot s}$ is the Mises stress, $s := \mathbb{C}\varepsilon^e - (1/N)\,\mathrm{tr}(\mathbb{C}\varepsilon^e)$ is the stress deviator, and $M := (1/\sigma_M)s$ is the flow direction. We have assumed quiescent initial conditions.

The governing equations for quasistatic evolution can be obtained by ignoring the inertial term in the momentum balance, and the initial conditions for the displacement.

Adjoint method for sensitivities

We seek to find the material parameters by solving the optimization problem (2.3) with (2.16) replacing (2.2) as the constraint for a suitable objective $\mathcal{O}$. We follow Section 2.2.1 to find the sensitivity using the adjoint method; the details are provided in Appendix A.1. The sensitivity is

$$\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \left[\gamma\dot{q} \left(-\frac{\partial^2 W^p}{\partial q \partial P} - \frac{\partial^2 g^*}{\partial \dot{q} \partial P} \right) \right] d\Omega\, dt, \tag{2.17}$$

where the adjoint variables v, γ, and ζ, being dual to the forward variables u, q, and ε^p, satisfy the evolution relations

$$\begin{aligned}
&\rho\ddot{v} - \nabla\cdot\left(\mathbb{C}\nabla v + \gamma\dot{q}\frac{\partial\sigma_M}{\partial\varepsilon} - \dot{q}\zeta\cdot\frac{\partial M}{\partial\varepsilon}\right) = -\frac{\partial o}{\partial u} && \text{on } \Omega,\\
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\gamma\left(\sigma_M - \frac{\partial W^p}{\partial q} - \frac{\partial g^*}{\partial\dot{q}}\right) - \gamma\dot{q}\frac{\partial^2 g^*}{\partial\dot{q}^2} - \zeta\cdot M\right] = -\gamma\dot{q}\frac{\partial^2 W^p}{\partial q^2} && \text{on } \Omega,\\
&\frac{\mathrm{d}\zeta}{\mathrm{d}t} = \nabla\xi\cdot\frac{\partial^2 W^e}{\partial\varepsilon\partial\varepsilon^p} + \gamma\dot{q}\frac{\partial\sigma_M}{\partial\varepsilon^p} - \dot{q}\zeta\cdot\frac{\partial M}{\partial\varepsilon^p} && \text{on } \Omega,\\
&v|_{t=T} = \dot{v}|_{t=T} = 0,\ \gamma|_{t=T} = 0, \quad \zeta|_{t=T} = 0.
\end{aligned} \tag{2.18}$$

Note that these equations are solved backward in time as before.

We obtain the equation in the quasistatic setting by ignoring the inertial ($\ddot{v}$) term in the first equation, and the final conditions on $\dot{v}$ in the last.

2.2.3 Numerical implementation

We discretize the equations in space using standard $P = 1$ Largrange polynomial finite elements for the displacement u and the corresponding adjoint variable v. The plastic quantities q and ε^p are spatially discretized at quadrature points. In dynamics, we adopt a mixed scheme for the temporal discretization. For the governing equations (2.16), we use an explicit central difference update for the displacement, but an implicit backwards Euler scheme for the plastic updates. Similarly, for the adjoint equations (2.18) we use an explicit central difference method for the adjoint displacement v, and implicit backwards Euler scheme for the plastic adjoints γ, ζ. In quasistatic evolution, we use an implicit backward Euler scheme for temporal discretization of all variables. Further details are provided in Appendix A.2.

We solve the optimization problem iteratively starting from an arbitrary initial guess. We update the parameter set using the gradient-based method of moving asymptotes optimization scheme [47]. This technique for solving constrained optimization problems has been used extensively in the optimal design community to solve PDE-constrained design problems over both small [3] and large parameter sets [39, 1]. We compare this with other optimization approaches in Appendix A.3. We allow the optimization to run for a set number of iterations, with this number chosen large enough such that the objective plateaus. We implement the numerical method in the deal.ii Finite Element Library [5, 4].

Synthetic and Converged Viscoplastic Material Parameters							
	σ_y/μ	ε_0^p	n	$\dot{\varepsilon}_0^p$ $(/s)$	m	$\tilde{\mathcal{O}}$	$\mathcal{O}_{\text{ind}}$
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00	–	–
P^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	–	–
P^{QS}	0.002127	0.03932	3.05	4.75×10^5	1.88	5.59×10^{-9}	0.0069
						(2.74×10^{-4})	(1.452)
P^{DC}	0.001734	0.02859	2.87	4.05×10^5	4.50	1.44×10^{-4}	0.021
						(1.81×10^{-1})	(1.452)

Table 2.1: Summary of results. The parameters used to generate the synthetic data, the initial guess and ones obtained through the proposed method in the quasistatic tension (QS) and dynamic compression (DC) experiments. The final objectives are reported in the table with the initial objectives in the parenthesis.

2.3 Demonstration and validation

We now demonstrate and validate the method described above using two examples with synthetic data. The first example is a quasistatic compression test on a plane-strain specimen while the second is an extended dynamic impact test. In both cases, we generate synthetic data using numerical simulations of a material model with properties similar to that of copper. Specifically, we take the shear modulus to be $\mu = 46.7$ GPa, a Poisson ratio of $\nu = 0.3656$, and the density to be $\rho = 8.93 \times 10^3$ kg/m^3, with the other parameters $P = \{\sigma_y,\ \varepsilon_0^p,\ n,\ \dot{\varepsilon}_0^p,\ m\}$ as shown in the the first row (P^{gen}) of Table 2.1[1]. We use these parameters, and the forward problem described in Section 2.2.2 above, to simulate the respective experiments to generate the data D^{exp}. We then use this data to obtain the material parameters using the indirect inversion method described in Section 2.2.2. We initialize these calculations with an initial guess P^{init} that are significantly different from the ground truth P^{gen}. The $P^{\text{gen}}, P^{\text{init}}$ and recovered parameters for both experiments are summarized in Table 2.1.

2.3.1 Quasistatic compression test of a plane strain specimen

The first example is a thick plate with an off-center hole subjected to quasistatic compression shown in Figure 2.2(a). We consider a geometry with

[1]The stress scales as the shear modulus μ and therefore we use it to normalize all our calculations.

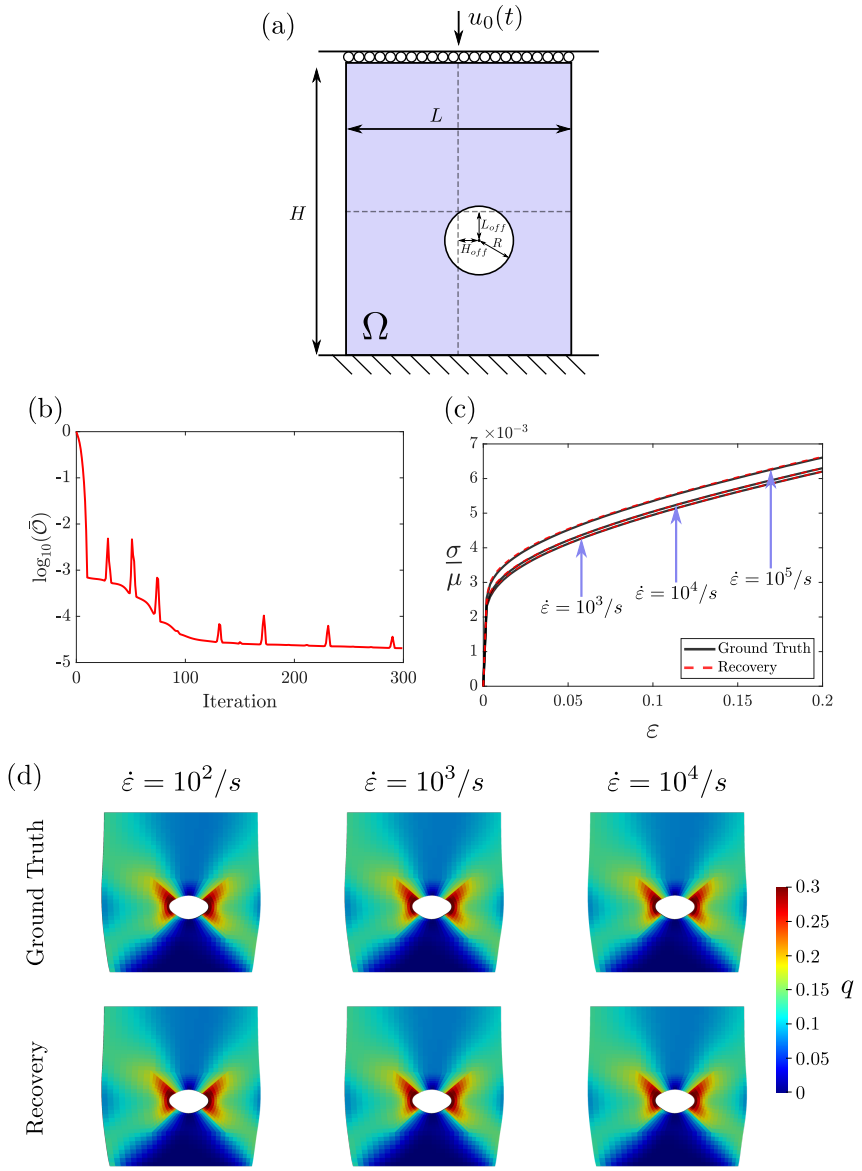


Figure 2.2: Results for the quasistatic compression of a thick plate with offset hole. (a) Geometry and boundary conditions. (b) Normalized objective from 2.20 vs iteration number. (c) Results of an independent uniaxial stress-strain test. (d) Deformed configurations at the final time-step for the ground truth data with the accumulated plasticity q.

$L/H = 0.8$, $R/H = 0.1$, $L_{\text{off}}/H = 0.1$, and $H_{\text{off}}/H = 0.04$. We assume plane strain conditions. We impose uniform vertical displacement

$$u_0(t) = \dot{\bar{\varepsilon}} H t, \tag{2.19}$$

with a nominal or macroscopic strain rate $\dot{\bar{\varepsilon}}$ on the top boundary, $\mathcal{T}$, while the bottom surface is held fixed. The experimental data D^{exp} consists of the full-

field displacement $u^{\text{exp}}(x,T) = u(x,T)$ at the final time T, and the reaction load history $f_R^{\text{exp}}(t) := \int_{\mathcal{T}} \mathbb{C}\varepsilon^e n \; d\Gamma$. We emphasize that we only use the full-field displacement at the final time step[2]. We further assume that we have data from n tests with varying rates. We define our objective to be

$$\begin{aligned}\mathcal{O}(P) := \sum_{i=1}^{n} \Bigg(& \frac{\alpha_u}{L^4} \int_\Omega |u^i(x,T) - u^{\text{exp},i}(x,T)|^2 d\Omega \\ & + \frac{\alpha_f}{T_i \mu^2 L^2} \int_0^{T_i} \left\| f_u^i(t) - f_R^{\text{exp},i} \right\|^2 dt \Bigg) \\ & \text{where } \{u^i, q^i, \varepsilon^{p,i}\} \text{ satisfy 2.16 for each } i = 1, \ldots, n\end{aligned} \tag{2.20}$$

where the superscript indexes each of the tests, and α_u and α_f are scaling factors that balance the weights of the two objective terms.

We consider a data-set from three tests ($n = 3$) on the same geometry and material, but differing in the loading rate. The loading rates span two orders of magnitude, with macroscopic strain rates of 10^2, 10^3, and 10^4/s, reaching a final macroscopic strain of $\bar{\varepsilon} = 0.1$ over 100 time-steps. The data-set is generated synthetically from forward simulations using the parameters P^{gen} in Table 2.1. We then use the proposed approach with the objective (2.20). Full numerical details of the forward and adjoint problem can be found in Appendix A.2.1 and Appendix A.2.3

The results are shown in Figure 2.2(b-d) and the recovered parameters are listed in Table 2.1. Figure 2.2(b) shows the change in the objective as we iterate. It drops rapidly but then gradually stabilizes after a little over 100 iterations decreasing by a roughly a factor 10^5. The values of the recovered parameters after 300 iterations is listed as P^{QS} in the Table 2.1. We recover the normalized yield strength σ_y/μ to within 10%, but the strain and rate hardening parameters differ significantly. This is true despite the fact that our objective is extremely small, with an objective value on the order of 10^{-9}. This means that the experiment with three tests at three strain rates are unable to distinguish between the two sets of parameters, the parameters P^{gen} used to generate the data and the recovered parameters P^{rec}. This is further

[2]It is typical in such tests to use a high speed camera to capture multiple snapshots and thus have the displacement field over many snapshots. We could extend our method to include multiple snapshots, but we chose not to do so to understand how much information we can obtain from a minimal amount of data since the inversion from DIC to displacement fields can also add computational cost.

demonstrated in Figure 2.2(d) which compares the accumulated plastic strain computed using the two sets of parameters, P^{gen} and P^{QS}.

Independent objective We found that the parameters we recover P^{QS} is different from those used to generate the data P^{gen} even though the difference in the measured quantities (objective) are very small. Therefore, we consider a *zero-shot test* where we evaluate our results against an independent objective that is not used in the inverse problem. We simulate the material response with the two sets of parameters with an independent uniaxial tension test. The resulting stress-strain curves are shown in Figure 2.2(c), and they agree well. Thus, even the independent uniaxial stress tension test is also unable to distinguish between the two sets of parameters.

To assess this quantitatively, we define a *independent objective* to be the average relative root mean square error of the material response for tests performed at n different strain rates:

$$\mathcal{O}_{\text{ind}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\int\left(\sigma_{\text{rec}}(\varepsilon)-\sigma_{\text{gen}}(\varepsilon)\right)^2 d\varepsilon}{\int \sigma_{\text{gen}}^2\, d\varepsilon}\right)^{1/2}, \tag{2.21}$$

where σ_{gen} is the uniaxial stress computed using P^{gen} while σ_{rec} is the uniaxial stress-strain curve generated using the recovered parameters. The value of $\mathcal{O}_{\text{ind}}$ in Table 2.1 confirms the results in Figure 2.2(c) that the two sets of parameters can not be distinguished in independent uniaxial stress-strain curves.

We conclude this section by studying the robustness of the method to the initial guess, objective and mesh size.

Sensitivity to initial guess We consider three tests with different and quite distinct initialization sets $\{P_1^{\text{init}}, P_2^{\text{init}}, P_3^{\text{init}}\}$. These are shown, along with their corresponding recovered parameter sets in Table 2.2. We observe that the method is insensitive to the initial parameter guess, yielding only slight variations in the recovered parameters with the objective remaining on the order of 10^{-9} for all of the recovered parameter values.

Sensitivity to objective The parameters α_u and α_f in the objective function represent the weights for the displacement data and the force data, re-

	σ_y/μ	ε_0^p	n	$\dot{\varepsilon}_0^p$ $(/s)$	m	$\mathcal{O}$	$\mathcal{O}_{\text{ind}}$
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00		
Sensitivity to initial guess							
P_1^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	2.74×10^{-4}	1.452
P_1^{rec}	0.002127	0.03932	3.05	4.75×10^5	1.88	5.59×10^{-9}	0.0069
P_2^{init}	0.01	0.1	2.0	2.0×10^5	6.0	6.43×10^{-4}	3.791
P_2^{rec}	0.002064	0.03352	3.09	8.16×10^5	2.23	2.81×10^{-9}	0.0060
P_3^{init}	0.0075	0.075	4.0	4.0×10^5	1.0	3.49×10^{-4}	1.91
P_3^{rec}	0.002142	0.04039	3.05	3.62×10^5	1.71	7.37×10^{-9}	0.0128
Sensitivity to objective							
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00	–	–
P_1^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	–	1.452
$\alpha_f/\alpha_u = 0.1$	0.002081	0.03456	3.08	5.34×10^5	1.96	4.44×10^{-9}	0.0049
$\alpha_f/\alpha_u = 1$	0.002127	0.03932	3.05	4.75×10^5	1.88	5.59×10^{-9}	0.0069
$\alpha_f/\alpha_u = 10$	0.002132	0.04009	3.04	4.98×10^5	1.87	2.44×10^{-9}	0.0048
Sensitivity to meshsize							
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00	–	–
P^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	–	1.452
$h = h_0$	0.002036	0.03233	3.08	6.81×10^5	2.26	9.94×10^{-10}	0.0031
$h = h_0/2$	0.002127	0.03932	3.05	4.75×10^5	1.88	5.59×10^{-9}	0.0060
$h = h_0/4$	0.002150	0.04410	3.00	4.56×10^5	1.95	7.50×10^{-9}	0.0089
$h = h_0/8$	0.002129	0.04410	3.00	4.47×10^5	2.18	8.44×10^{-9}	0.0090

Table 2.2: Robustness of the proposed method for the quasistatic experiment.

spectively. Table 2.2 shows the recovered set of parameters for three different orders of α_f/α_u, with all of them showing similar performance.

Sensitivity to mesh size To confirm the convergence of solution with respect to the mesh size, we perform the optimization on different mesh refinement levels. Table 2.2 shows the recovered parameters for four mesh resolutions ranging from 384 to 24,576 elements. As the recovered parameters only differ slightly between the four meshes, we conclude that the solution does not depend on mesh resolution.

We conclude that the method accurately recovers elasto-viscoplastic material parameters from data obtained from the final snapshot and force-history of three quasistatic tests. While the values of the recovered and generate parameters differ slightly, these parameters are unable to distinguish either observations of the test, or the zero-shot independent test, and possibly reflects the degeneracy of the constitutive model.

2.3.2 Dynamic compression of thin specimens

The second example we consider is that of dynamic impact of a cylinder. We consider a thin cylindrical specimen that $\Omega := \overline{\Omega} \times (0, h)$ with cross-section $\overline{\Omega} \in \mathbb{R}^2$ and thickness of $h \in \mathbb{R}$. Figure 2.3(a) shows the case where $\overline{\Omega}$ is a circular annulus. The cylinder is placed on an anvil and impacted on the top with a striker whose cross section is larger than that of the cylinder. Ignoring an irrelevant rigid translation, this corresponds to the following imposed boundary conditions:

$$u_3(X_1, X_2, 0, t) = 0, \quad u_3(X_1, X_2, h, t) = (\lambda(t) - 1)h \tag{2.22}$$

where $\lambda(t)$ is the imposed nominal axial stretch. We assume that the contact with the anvil and the striker is friction-free, and that the specimen is sufficiently small so that we may assume uniform axial strain,

$$u(X_1, X_2, X_3, t) = \bar{u}(X_1, X_2, t) + (\lambda(t) - 1)X_3 \; e_3, \tag{2.23}$$

where $\bar{u} := \bar{\Omega} \mapsto \mathbb{R}^2$ is the in-plane displacement. This enables us to reduce this to two space dimensions, see Appendix A.4 for details.

Our experimental data D^{exp} consists of the final in plane displacement $\bar{u}^{\text{exp}}(X_1, X_2, T)$ and the net axial force history $f_R^{\text{exp}}(t)$. Notice that this is more data than is typically collected in a classical dynamic impact experiment, where only f_R^{exp} is measured. However, we may measure $\bar{u}^{\text{exp}}(X_1, X_2, T)$ by placing a speckle pattern on the face of the specimen, imaging it before and after the impact, and using digital image correlation. We repeat the test n times, and consider the following objective

$$\begin{aligned} \mathcal{O}(P) := \sum_{i=1}^{n} \Bigg(& \frac{\alpha_u}{L^4} \int_{\bar{\Omega}} |\bar{u}^i(X_1, X_2, T) - \bar{u}^{exp,i}(X_1, X_2, T)|^2 \; d\bar{\Omega} \\ & + \frac{\alpha_f}{T_i \mu^2 L^4} \int_0^T |f_u^i - f_R^{exp,i}|^2 \; dT \Bigg) \\ & \text{where } \{\bar{u}^i, q^i, \varepsilon^{p,i}\} \text{ satisfy (A.27) for } i = 1, \ldots, n, \end{aligned} \tag{2.24}$$

and $f_u^{exp,i} : \int_{\bar{\Omega}} (\mathbb{C}\bar{\varepsilon}^{e,i})_{33} \; dS$ is the net vertical force. α_u and α_f are weights associated with the displacement and force components of the objective. L, μ and T_i are characteristic length, shear modulus, and characteristic time, respectively.

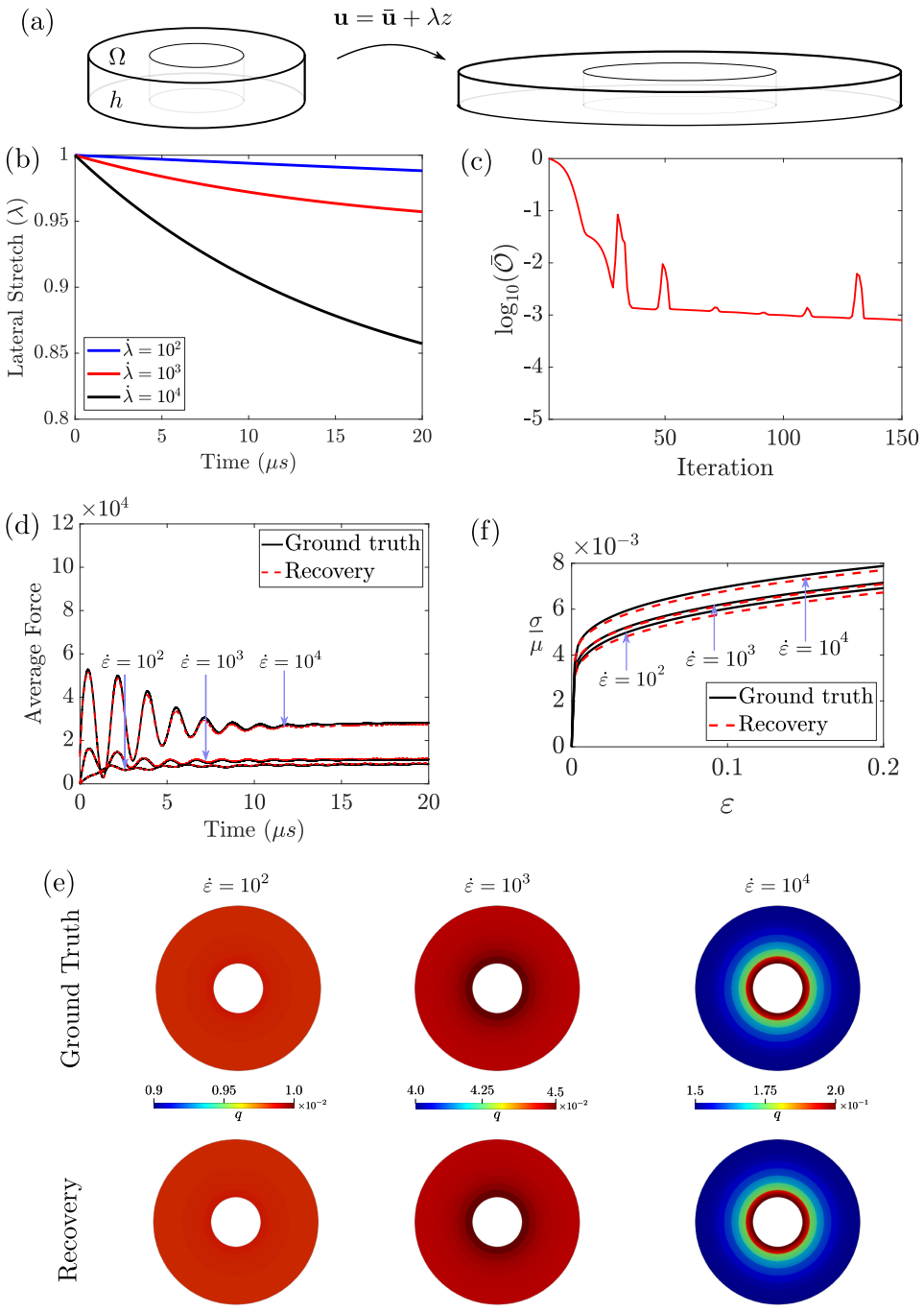


Figure 2.3: Results for the dynamic compression of circular annulus. (a) Geometry and deformation. (b) Applied stretch along the annulus thickness (c) Normalized objective from (2.24) vs iteration number. (d) The net vertical force applied to the annulus over time. (e) Comparison of accumulated plasticity q. (f) Results of an independent uniaxial stress-strain test.

	σ_y/μ	ε_0^p	n	$\dot{\varepsilon}_0^p$ $(/s)$	m	$\mathcal{O}$	$\mathcal{O}_{\text{ind}}$
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00		
Sensitivity to initial guess							
P_1^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	1.81×10^{-1}	1.452
P_1^{rec}	0.001734	0.02859	2.87	4.05×10^5	4.50	1.44×10^{-4}	0.021
P_2^{init}	0.01	0.1	2.0	2.0×10^5	6.0	5.55×10^{-1}	3.791
P_2^{rec}	0.001793	0.06623	2.11	2.40×10^5	4.50	8.57×10^{-4}	0.079
P_3^{init}	0.0075	0.075	4.0	4.0×10^5	1.0	2.12×10^{-1}	1.91
P_3^{rec}	0.0022	0.07812	2.70	6.02×10^5	1.09	2.93×10^{-4}	0.090
Sensitivity to objective							
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00	–	–
P_1^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	1.81×10^{-1}	1.452
$\alpha_f/\alpha_u = 0.1$	0.001726	0.03583	2.71	3.46×10^5	4.58	1.47×10^{-4}	0.050
$\alpha_f/\alpha_u = 1$	0.001734	0.02859	2.87	4.05×10^5	4.50	1.44×10^{-4}	0.021
$\alpha_f/\alpha_u = 10$	0.001778	0.03929	2.69	3.51×10^5	4.81	0.87×10^{-4}	0.043
Sensitivity to geometry							
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00	–	–
P_1^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00	1.81×10^{-1}	1.452
Circular	0.001734	0.02859	2.87	4.05×10^5	4.50	1.44×10^{-4}	0.021
Ellipse	0.001921	0.04590	2.55	3.10×10^5	3.14	1.81×10^{-5}	0.020
Flower hole	0.001735	0.03892	2.63	3.51×10^5	4.43	2.92×10^{-4}	0.060
Flower	0.001800	0.02768	2.93	3.93×10^5	3.04	8.49×10^{-5}	0.023

Table 2.3: Robustness of the proposed method for the dynamic compression experiment.

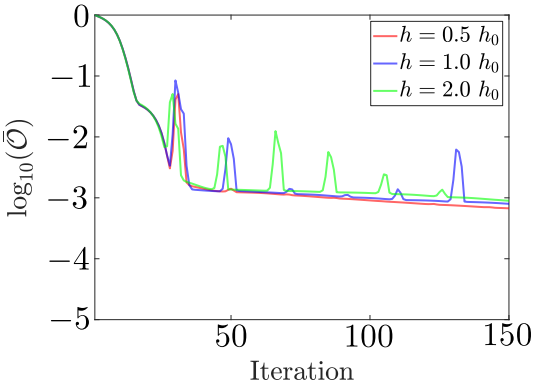


Figure 2.4: Objective curves for the dynamic compression test with respect to iterations for different mesh sizes.

We apply our method to a specimen whose cross-section is a circular annulus shown in Figure 2.3(a). We generate synthetic data for $n = 3$ tests with imposed thickness strains shown in Figure 2.3(b) that correspond to initial strain rate in the range $\dot{\varepsilon} = 10^2 - 10^4$ s^{-1} with the parameters P^{gen}. We initialize our iterative optimization approach with the parameters P^{init}. α_u and α_f are chosen such that the contribution of both parts of the objective are approximately equal, $\alpha_f/\alpha_u = 1$. Figure 2.3(c) shows the evolution of the objective with it-

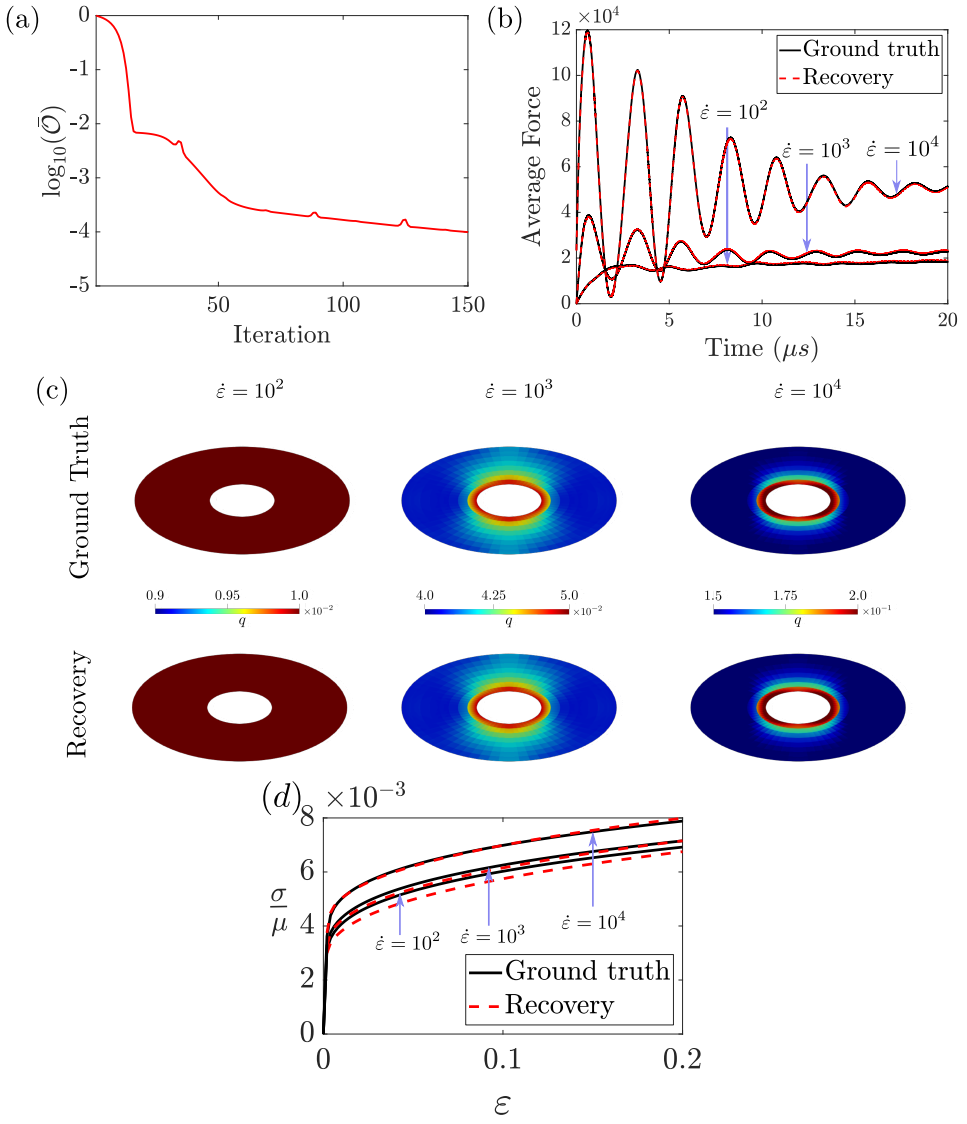


Figure 2.5: Results for the dynamic compression of elliptic annulus . (a) Normalized objective from 2.24 vs iteration number. (b) The net vertical force applied to the annulus over time. (c) Comparison of accumulated plasticity q. (d) Results of an independent uniaxial stress-strain test.

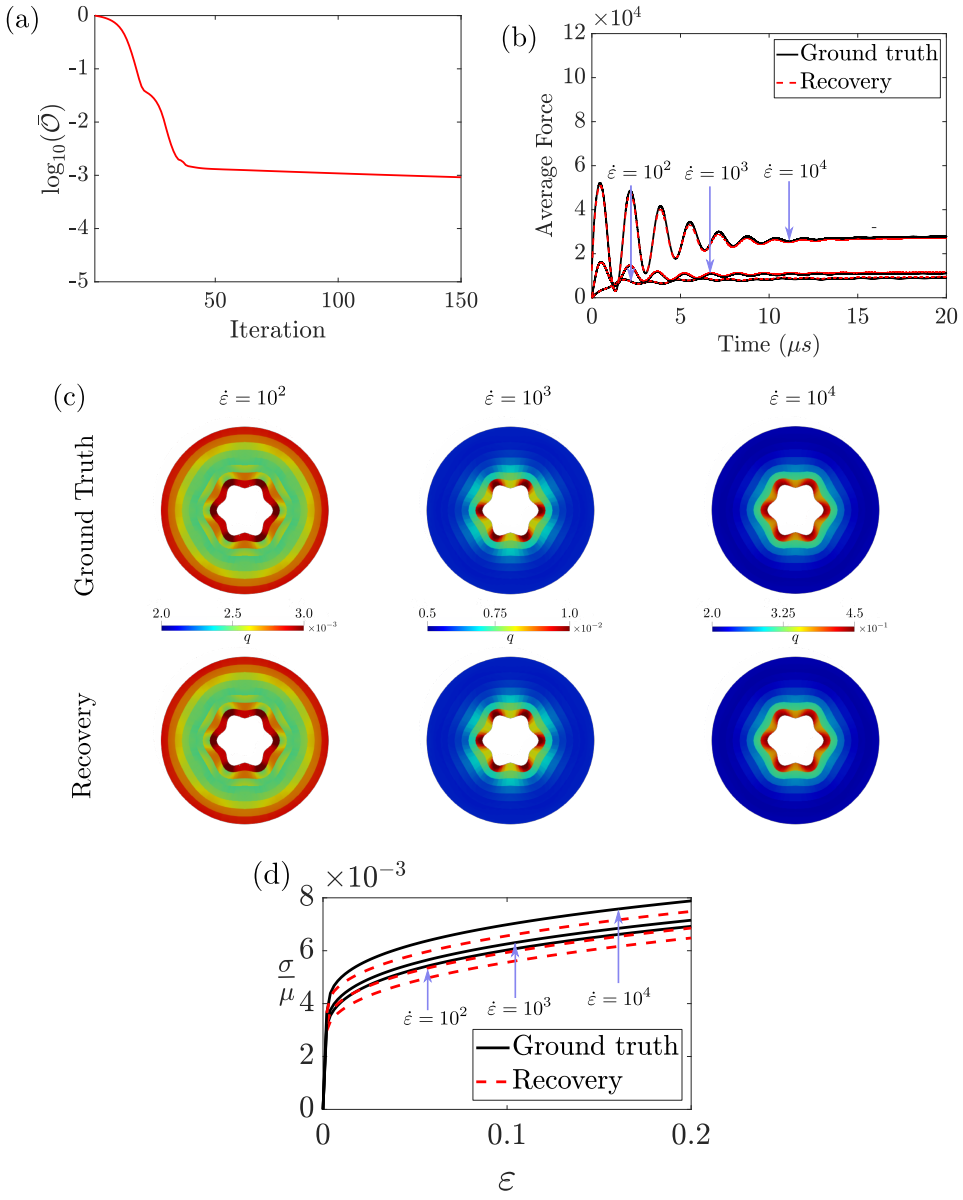


Figure 2.6: Results for the dynamic compression of flower-shaped hole . (a) Normalized objective from 2.24 vs iteration number. (b) The net vertical force applied to the annulus over time. (c) Comparison of accumulated plasticity q. (d) Results of an independent uniaxial stress-strain test.

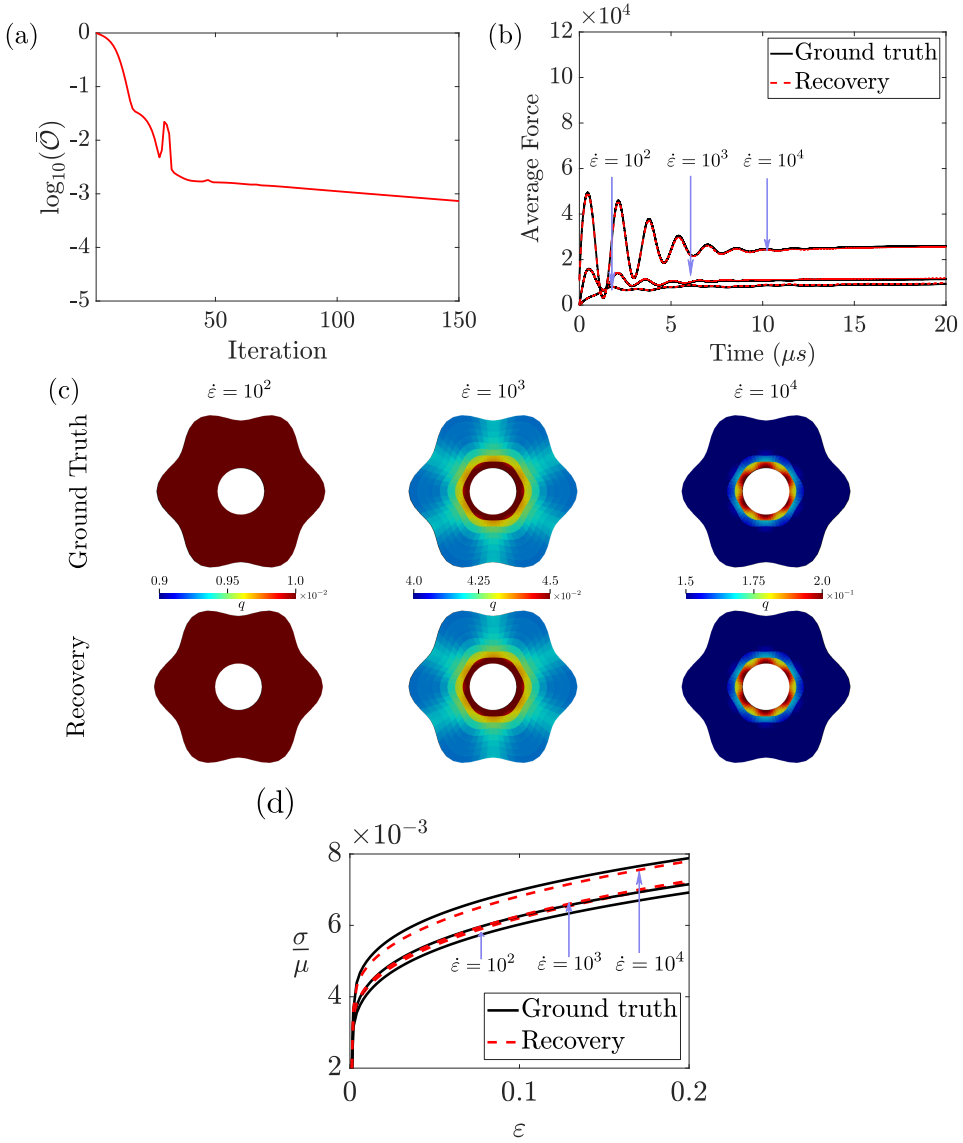


Figure 2.7: Results for the dynamic compression of disc with flower-shaped specimen . (a) Normalized objective from 2.24 vs iteration number. (b) The net vertical force applied to the annulus over time. (c) Comparison of accumulated plasticity q. (d) Results of an independent uniaxial stress-strain test.

eration. The objective decreases rapidly at first and then stabilizes at a value with a factor greater than 10^3 than the initial value. The resulting parameters are shown in Table 2.1 as P^{DC}. As in the previous quasistatic compression test, there is good agreement with the yield strength, but not for the hardening parameters. Despite the fact that the objective is very small, the experimental observations agree well, as shown in Figures 2.3(d,e). Figure 2.3(d) compares the experimental reaction force with the one simulated with the recovered parameters, and we observe excellent agreement. Figure 2.3(e) compares the accumulated plastic strain computed with the original parameters P^{gen} with those computed with the recovered parameters P^{DC}. Again, we see good agreement. Finally, we conduct the zero-shot test of comparing the response in uniaxial tensile tests. Figure 2.3(f) shows the results of the independent stress-strain test, and we again see very good agreement.

Finally, we demonstrate the robustness of our method.

Sensitivity to initial guess We consider three tests with different and quite distinct initialization sets $\{P_1^{\mathrm{init}}, P_2^{\mathrm{init}}, P_3^{\mathrm{init}}\}$. These are shown, along with their corresponding recovered parameter sets in Table 2.3. We observe that the method is insensitive to the initial parameter guess, yielding only slight variations in the recovered parameters with the objective remaining on the order of 10^{-4} for all of the recovered parameter values.

Sensitivity to objective The parameters α_u and α_f in the objective function represent the weights for the displacement data and the force data, respectively. Table 2.3 shows the recovered set of parameters for three different orders of α_f/α_u, with all of them showing similar performance.

Sensitivity to mesh size To confirm the convergence of solution with respect to the mesh size, we perform the optimization for three different mesh sizes. Figure 2.4 shows the objective versus optimization iterations for different mesh sizes. Since the objective remains the same for all mesh sizes, we conclude that it is safe to assume the solution does not depend on the size of the mesh.

Sensitivity to configuration The geometry of the specimen can significantly impact the inversion methodology. Certain geometries can induce a wide range of strains during compression, providing datasets with richer information about the material properties. To investigate this, we performed the inversion on different cross-sectional geometries, specifically circular, elliptical, and flower-shaped configurations. The results for the elliptic annulus, disc with flower-shaped hole and flower-shaped specimen are shown in Figures 2.5, 2.6 and 2.7, respectively. The recovered parameters for these configurations are compared in Table 2.3. Due to the geometry of the ellipse configuration, the average forces experienced is much higher than the other two configurations, leading to more efficient recovery of parameters and a lower overall objective value. Furthermore, the lack of angular symmetry of the strain field in the ellipse provides a richer dataset, resulting in a more efficient inversion.

2.4 Conclusion

Material characterization through a constitutive model is necessary to close the balance laws and allow for continuum modeling. Traditional approaches rely on experimental configurations that yield uniform states of stress and strain (rate), and can therefore be easily inverted to characterize the constitutive law. However, these approaches only probe highly selected and idealized states and trajectories, thereby leading to large uncertainties in the complex states and trajectories that arise in the applications of the resulting models. Further, some of these configurations require careful alignment to achieve the idealized states, further limiting the amount of data that can be collected. In this work, we have introduced an alternate approach that uses a combination of PDE-constrained optimization and full-field observation techniques to obtain material models from complex experiments.

We formulate the problem of learning the constitutive model as an indirect inverse problem. We assume that the model is given in a parametrized form: this can be either classical where one has a few parameters or a hyperparametrized formulation like a neural network. We recast this inverse problem as a PDE-constrained optimization where we seek to find the parameters that minimize the difference between experimentally measured quantities, and the corresponding quantities computed using the parameters subject to the balance laws (PDE-constraint). We follow an iterative gradient-based approach where

we use the adjoint method to compute the sensitivities. This results in an adjoint partial differential equation (linear in space and quasilinear in time) that is no more expensive to solve than the forward problem. The resulting algorithm is shown in Figure 2.1. As noted in the introduction, our work builds on prior work in the literature on the finite-element updating and PDE-constrained optimization.

In this paper, we describe the formulation, first for a general history dependent material and then for the specific model of a J2 elasto-viscoplasticity. We then demonstrate it with synthetic data in two experimental configurations: quasistatic compression of a thick plate with a hole, and extended dynamic compression. In forthcoming work, we demonstrate the method against experimental data (part 2).

We point out a number of salient features of this approach. First, the method scales linearly (see Section 2.2.1) with the number of parameters in a model with a small pre-factor to the linear term. Therefore, this method is ideally suited for complex models with a large number of parameters including a hyperparametrized formulation like a neural network. We demonstrate this in forthcoming work (part 3).

Second, the method is able to recover complex behavior with relatively little data. In the quasistatic test described in Section 2.3.1, we only used the strain field at the final time instead of a series of snapshots. This is because this test used a complex geometry and full-field information, and thus probed multiple trajectories through the adjoint method. At the same time, the computational cost of the method is relatively independent of the amount of data: so we could have easily incorporated multiple snapshots.

Third, the approach is versatile. We have demonstrated this by applying it to two different configurations. Further, the approach is capable of addressing non-smooth problems. We applied the method here to elasto-viscoplasticity including a yield criterion. This is difficult to do with a method that uses a numerical derivative. In forthcoming work (part 2), we extend the approach to indentation including contact.

Finally, the formulation does not make any assumptions about isotropy or homogeneity. Our current work addresses learning single crystal behavior from observations on polycrystals.

References

[1] A. Akerson, B. Bourdin, and K. Bhattacharya. "Optimal design of responsive structures." In: *Structural and Multidisciplinary Optimization* 65 (4 2022). ISSN: 16151488. DOI: 10.1007/s00158-022-03200-5.

[2] A. Akerson. "Optimal structures for failure resistance under impact." In: *Journal of the Mechanics and Physics of Solids* 172 (Mar. 2023), p. 105172. ISSN: 00225096. DOI: 10.1016/j.jmps.2022.105172. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022509622003489.

[3] A. Akerson and T. Liu. "Mechanics, modeling, and shape optimization of electrostatic zipper actuators." In: *Journal of the Mechanics and Physics of Solids* 181 (Oct. 2023), p. 105446. DOI: 10.1016/j.jmps.2023.105446.

[4] D. Arndt, W. Bangerth, D. Davydov, T. Heister, L. Heltai, M. Kronbichler, M. Maier, J.-P. Pelteret, B. Turcksin, and D. Wells. "The deal.II finite element library: Design, features, and insights." In: *Computers & Mathematics with Applications* 81 (2021), pp. 407–422. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2020.02.022. URL: https://arxiv.org/abs/1910.13247.

[5] W. Bangerth, R. Hartmann, and G. Kanschat. "Deal.II - -A general-purpose object-oriented finite element library." In: *ACM Transactions on Mathematical Software* 33 (2007), pp. 1–27. ISSN: 00983500. DOI: 10.1145/1268776.1268779.

[6] M. Bendsøe and O. Sigmund. *Topology optimization. Theory, methods, and applications. 2nd ed., corrected printing.* Springer, Berlin, Heidelberg, 2004. DOI: 10.1007/978-3-662-05086-6.

[7] J. V. Bernier, R. M. Suter, A. D. Rollett, and J. D. Almer. "High-energy X-ray diffraction microscopy in materials science." In: *Annual Review of Materials Research* 50 (2020), pp. 395–436. DOI: https://doi.org/10.1146/annurev-matsci-070616-124125.

[8] E. Billington and A. Tate. *Physics of Deformation and Flow: An Introduction.* McGraw Hill, 1981.

[9] M. Bonnet and A. Constantinescu. "Inverse problems in elasticity." en. In: *Inverse Problems* 21 (2005), R1–R50. ISSN: 0266-5611, 1361-6420. DOI: 10.1088/0266-5611/21/2/R01. URL: https://iopscience.iop.org/article/10.1088/0266-5611/21/2/R01 (visited on 12/12/2023).

[10] A. Buljac, C. Jailin, A. Mendoza, J. Neggers, T. Taillandier-Thomas, A. Bouterf, B. Smaniotto, F. Hild, and S. Roux. "Digital volume correlation: Review of progress and challenges." In: *Experimental Mechanics* 58 (2018), pp. 661–708.

[11] R. Carson, M. Obstalecki, M. Miller, and P. Dawson. "Characterizing heterogeneous intragranular deformations in polycrystalline solids using diffraction-based and mechanics-based metrics." en. In: *Modelling and Simulation in Materials Science and Engineering* 25.5 (July 2017), p. 055008. ISSN: 0965-0393, 1361-651X. DOI: 10.1088/1361-651X/aa6dc5. URL: https://iopscience.iop.org/article/10.1088/1361-651X/aa6dc5 (visited on 11/17/2023).

[12] T. Chu, W. Ranson, and M. A. Sutton. "Applications of digital-image-correlation techniques to experimental mechanics." In: *Experimental Mechanics* 25 (1985), pp. 232–244.

[13] M. Conde, Y. Zhang, J. Henriques, S. Coppieters, and A. Andrade-Campos. "Design and validation of a heterogeneous interior notched specimen for inverse material parameter identification." In: *Finite Elements in Analysis and Design* 214 (2023), p. 103866.

[14] A. Constantinescu and N. Tardieu. "On the identification of elastoviscoplastic constitutive laws from indentation tests." en. In: *Inverse Problems in Engineering* 9 (2001), pp. 19–44. ISSN: 1068-2767, 1029-0281. DOI: 10.1080/174159701088027751. URL: http://www.tandfonline.com/doi/abs/10.1080/174159701088027751 (visited on 10/07/2022).

[15] C. Efstathiou, D. E. Boyce, J. .-. Park, U. Lienert, P. R. Dawson, and M. P. Miller. "A method for measuring single-crystal elastic moduli using high-energy X-ray diffraction and a crystal-based finite element model." In: *Acta Materialia* 58 (2010), pp. 5806–5819.

[16] A. Eghtesad, J. N. Fuhg, and N. Bouklas. *NN-EVP: A physics informed neural network-based elasto-viscoplastic framework for predictions of grain size-aware flow response under large deformations.* en. arXiv:2307.04301. 2023. URL: http://arxiv.org/abs/2307.04301 (visited on 12/06/2023).

[17] M. Grediac, F. Pierron, S. Avril, and E. Toussaint. "The virtual fields method for extracting constitutive parameters from full-field measurements: A review." In: *Strain* 42 (2006), pp. 233–253.

[18] M. Gurtin. *Introduction to Continuum Mechanics.* Academic Press, 1970.

[19] C. M. Hamel, K. N. Long, and S. L. B. Kramer. *Calibrating constitutive models with full-field data via physics informed neural networks.* en. arXiv:2203.16577. 2022. URL: http://arxiv.org/abs/2203.16577 (visited on 12/06/2023).

[20] J. Henriques, J. Xavier, and A. Andrade-Campos. "On the inverse identification of wood elastic properties using a DIC-based FEMU approach." In: *International Journal of Structural Integrity* 14 (2022), pp. 74–90.

[21] F. Hild and S. Roux. "Digital image correlation: From displacement measurement to identification of elastic properties ? A review." In: *Strain* 42 (2006), pp. 69–80.

[22] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints.* Springer, 2009.

[23] M. F. Horstemeyer and D. J. Bammann. "Historical review of internal state variable theory for inelasticity." In: *International Journal of Plasticity* 26 (2010), pp. 1310–1334.

[24] A. D. Kammers and S. Daly. "Digital image correlation under scanning electron microscopy: Methodology and validation." In: *Experimental Mechanics* 53 (2013), pp. 1743–1761.

[25] K. T. Kavanagh and R. W. Clough. "Finite element applications in the characterization of elastic solids." In: *International Journal of Solids and Structures* 7 (1971), pp. 11–23.

[26] H. B. Khaniki, M. H. Ghayesh, R. Chin, and M. Amabili. "Hyperelastic structures: A review on the mechanics and biomechanics." In: *International Journal of Non-Linear Mechanics* 148 (2023), p. 104275.

[27] C. Kim and M.-G. Lee. "Finite element-based virtual fields method with pseudo-real deformation fields for identifying constitutive parameters." en. In: *International Journal of Solids and Structures* 233 (2021), p. 111204. ISSN: 00207683. DOI: 10.1016/j.ijsolstr.2021.111204. URL: https://linkinghub.elsevier.com/retrieve/pii/S0020768321002948 (visited on 12/11/2023).

[28] V. Kosin, A. Fau, C. Jailin, F. Hild, and T. Wick. "Parameter identification of a phase-field fracture model using integrated digital image correlation." In: *Computer Methods in Applied Mechanics and Engineering* 420 (2024), p. 116689.

[29] S. L. B. Kramer Scherzinger W.M. *Implementation and Evaluation of the Virtual Fields Method: Determining Constitutive Model Parameters From Full-Field Deformation Data, Sandia Report, SAND2014- 17871.* 2014.

[30] J. Lubliner. *Plasticity Theory.* Dover books on engineering. Dover Publications, 2008. ISBN: 9780486462905. URL: https://books.google.com/books?id=MkK-BLbHtcAC.

[31] A. Marek, F. M. Davis, and F. Pierron. "Sensitivity-based virtual fields for the non-linear virtual fields method." en. In: *Computational Mechanics* 60 (Sept. 2017), pp. 409–431. ISSN: 0178-7675, 1432-0924. DOI: 10.1007/s00466-017-1411-6. URL: http://link.springer.com/10.1007/s00466-017-1411-6 (visited on 12/11/2023).

[32] F. Mathieu, H. Leclerc, F. Hild, and S. Roux. "Estimation of elastoplastic parameters via weighted FEMU and integrated-DIC." In: *Experimental Mechanics* 55 (2015), pp. 105–119.

[33] S. K. Melly, L. Liu, Y. Liu, and J. Leng. "A review on material models for isotropic hyperelasticity." In: *International Journal of Mechanical System Dynamics* 1 (2021), pp. 71–88.

[34] J. Neggers, F. Mathieu, F. Hild, and S. Roux. "Simultaneous full-field multi-experiment identification." In: *Mechanics of Materials* 133 (2019), pp. 71–84.

[35] A. A. Oberai, N. H. Gokhale, and G. R. FeijÂ o. "Solution of inverse problems in elasticity imaging using the adjoint method." en. In: *Inverse Problems* 19 (Apr. 2003), pp. 297–313. ISSN: 0266-5611. DOI: 10.1088/0266-5611/19/2/304. URL: https://iopscience.iop.org/article/10.1088/0266-5611/19/2/304 (visited on 12/14/2023).

[36] M. Obstalecki, S. L. Wong, P. R. Dawson, and M. P. Miller. "Quantitative analysis of crystal scale deformation heterogeneity during cyclic plasticity using high-energy X-ray diffraction and finite-element simulation." In: *Acta Materialia* 75 (2014), pp. 259–272.

[37] M. Ortiz and L. Stainier. "The variational formulation of viscoplastic constitutive updates." In: *Computer Methods in Applied Mechanics and Engineering* 171 (1999), pp. 419–444. ISSN: 00457825. DOI: 10.1016/S0045-7825(98)00219-9.

[38] E. Pagnacco, A.-S. Caro-Bretelle, and P. Ienny. "Parameter identification from mechanical field measurements using finite element model updating strategies." In: *Full-Field Measurements and Identification in Solid Mechanics*. John Wiley & Sons, Ltd, 2013, pp. 247–274.

[39] C. Pedersen. "Topology optimization design of crushed 2D-frames for desired energy absorption history." In: *Structural and Multidisciplinary Optimization* 25 (Jan. 2003), pp. 368–382. DOI: 10.1007/s00158-003-0282-y.

[40] F. Pierron and M. Grédiac. *The virtual fields method: extracting constitutive mechanical parameters from full-field deformation measurements.* Springer Science & Business Media, 2012.

[41] R. E. Plessix. "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications." In: *Geophysical Journal International* 167.2 (2006), pp. 495–503. ISSN: 0956540X. DOI: 10.1111/j.1365-246X.2006.02978.x.

[42] S. Ravindran, V. Gandhi, A. Joshi, and G. Ravichandran. "Three-dimensional full-field velocity measurements in shock compression experiments using stereo digital image correlation." In: *Review of Scientific Instruments* 94.2 (2023).

[43] J. Rojicek, Z. Paska, M. Fusek, F. Fojtik, and D. Lickova. "A study on FEMU: Influence of selection of experiments on results for ABS-M30 material." In: *MM Science Journal* Dec (2021), pp. 5426–5434.

[44] E. Sakaridis, C. C. Roth, B. Jordan, and D. Mohr. "Post-Necking full-field FEMU identification of anisotropic plasticity from flat notched tension experiments." In: *International Journal of Solids and Structures* 305 (2024), p. 113076.

[45] M. Sutton, J. Yan, V. Tiwari, H. Schreier, and J. Orteu. "The effect of out-of-plane motion on 2D and 3D digital image correlation measurements." In: *Optics and Lasers in Engineering* 46 (2008), pp. 746–757.

[46] M. A. Sutton, J. J. Orteu, and H. Schreier. *Image correlation for shape, motion and deformation measurements: Basic concepts, theory and applications.* Springer Science & Business Media, 2009.

[47] K. Svanberg. "The method of moving asymptotes - A new method for structural optimization." In: *International Journal for Numerical Methods in Engineering.* 24 (1987), pp. 359–373.

[48] K. Upadhyay, J. N. Fuhg, N. Bouklas, and K. T. Ramesh. *Physics-informed Data-driven Discovery of Constitutive Models with Application to Strain-Rate-sensitive Soft Materials.* en. arXiv:2304.13897. 2023. URL: http://arxiv.org/abs/2304.13897 (visited on 12/06/2023).

[49] J. Wang, T. Li, F. Cui, C.-Y. Hui, J. Yeo, and A. T. Zehnder. "Metamodeling of constitutive model using Gaussian process machine learning." en. In: *Journal of the Mechanics and Physics of Solids* 154 (2021), p. 104532. ISSN: 00225096. DOI: 10.1016/j.jmps.2021.104532. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022509621001939 (visited on 12/06/2023).

[50] J. Wang, T. Li, C.-Y. Hui, J. Yeo, and A. Zehnder. "Gaussian process to identify hydrogel constitutive model." In: *Challenges in Mechanics of Time Dependent Materials, Mechanics of Biological Systems and Materials & Micro-and Nanomechanics, Volume 2: Proceedings of the 2021 Annual Conference & Exposition on Experimental and Applied Mechanics*. Springer. 2022, pp. 79–83.

[51] E. Wielewski, D. E. Boyce, J.-S. Park, M. P. Miller, and P. R. Dawson. "A methodology to determine the elastic moduli of crystals by matching experimental and simulated lattice strain pole figures using discrete harmonics." en. In: *Acta Materialia* 126 (2017), pp. 469–480. ISSN: 13596454. DOI: 10.1016/j.actamat.2016.12.026. URL: https://linkinghub.elsevier.com/retrieve/pii/S1359645416309624 (visited on 11/17/2023).

[52] T. Wu, B. Rosić, L. De Lorenzis, and H. G. Matthies. "Parameter identification for phase-field modeling of fracture: A Bayesian approach with sampling-free update." en. In: *Computational Mechanics* 67 (2021), pp. 435–453. ISSN: 0178-7675, 1432-0924. DOI: 10.1007/s00466-020-01942-x.

URL: http://link.springer.com/10.1007/s00466-020-01942-x (visited on 11/17/2023).

[53] J. Yang, L. Hazlett, A. Landauer, and C. Franck. "Augmented Lagrangian Digital Volume Correlation (ALDVC)." In: *Experimental Mechanics* 60 (2020), pp. 1205–1223.

[54] H. Zhou, R. A. Lebensohn, P. Reischig, W. Ludwig, and K. Bhattacharya. "Imposing equilibrium on experimental 3-D stress fields using Hodge decomposition and FFT-based optimization." In: *Mechanics of Materials* 164 (2022), p. 104109.

Chapter 3

LEARNING CONSTITUTIVE RELATIONS FROM EXPERIMENTS: II. DYNAMIC INDENTATION

Akerson, A., Rajan, A., D. Casem and Bhattacharya, K., 2025. "Learning viscoplastic constitutive behavior from experiments: II. Dynamic indentation." In preparation.

Contributions: A.R. and A.A. developed the computational code and performed simulations for the synthetic and experimental section, respectively. D.C. performed the experiments required for material characterization. K.B. supervised the project and conceptualized the methodology. A.R., A.A. and K.B. wrote the manuscript.

3.1 Introduction

The engineering of solutions involving complex materials and phenomena requires a constitutive relation that describes the properties of the material. This constitutive relation is typically obtained empirically by conducting experiments. However, we cannot measure the constitutive relation directly and it has to be obtained through the solution of an inverse problem. In fact, we cannot even directly measure the quantities like stress, strain, strain rate and energy density that comprise the constitutive relation. Instead, they have to be inferred from quantities like displacements and total forces that can be measured in the laboratory. Thus, the problem of inferring constitutive relations from experiments, and thereby completing the continuum mechanical formulation, requires the solution of an indirect inverse problem.

In part I of this work [2], we proposed an approach where this indirect inverse problem is formulated as a partial differential equation constrained optimization problem. The constitutive relation is postulated as a parametrized relation; this could either be a classical constitutive relation or a hyper-parametrized neural network. The problem of identifying the constitutive

relation, then, is to find the parameters that minimize the different between the experimental observation and the corresponding quantities computed by solving the governing equations (PDE constraint). We adopt a gradient-based optimization approach, where we compute the sensitivity using the adjoint method. We survey the literature, describe the method broadly and demonstrate it against quastistatic and dynamic experiments using synthetic data.

In this part II, we extend the work to dynamic indentation, and demonstrate the method with both synthetic and experimental data. Since their introduction by J.A. Brinnell over a century ago, indentation tests have been widely used in the static setting to study the hardness of materials. Koeppel and Subhash [11] introduced dynamic indentation by adapting a split Hopkinson pressure bar to study the dynamic hardness of materials. Since then, this methods and adaptations based on it have been used by various researchers due to the relative ease of use [4, 18, 21, 8]. However, obtaining quantitative information about the constitutive behavior has remained a challenge for a number of reasons (e.g., [13]). Dynamic indentation leads to a complex time-dependent state of stress, and elastic-plastic deformation. Addressing these is the purpose of the approach that we are developing. Further, we have contact between indenter and the specimen. Contact, is a one-sided constraint, and this requires additional theoretical development to that presented in part I. This is one of the main objectives of this paper. The second objective is to test the method against experimental data. We do so against experiments in both rolled hardened armor (RHA) steels, and aluminum. In each case, we demonstrate the ability of our method to recover elasto-plastic constitutive behavior with a very few tests.

This paper is divided into 5 sections. We begin in section 3.2.1 by presenting the details for the constitutive law and the evolution laws. Section 3.2.2 describes the method for performing inversion. Following which, we show results of the inversion algorithm in section 3.3 and 3.4. Section 3.5 concludes the study.

3.2 Formulation and method

We build on the method developed in [2] by generalizing to include contact.

3.2.1 Governing equations

We begin by recalling the constitutive relations, balance laws and contact conditions. We formulate the latter two in weak form.

Constitutive Model We consider an elastic-plastic body of mass density ρ occupying the domain $\Omega \subset \mathbb{R}^N$ over time $(0,T)$, governed by small strain, J2 plasticity with power-law isotropic hardening and power law rate-sensitivity [19, 15]. We assume the displacement field $u : (0,T) \times \Omega \mapsto \mathbb{R}^N$ is fixed to $u = u_0$ on a portion of the boundary $\partial_u\Omega$. Additionally, we consider a surface traction f applied to a portion of the boundary $\partial_f\Omega$ and an applied body force b. The elasto-plastic potential of the body has contributions from the elastic energy, the plastic potential, and the plastic dissipation,

$$\mathcal{U}(u,q,\varepsilon^p) = \int_\Omega \left[\frac{1}{2}\varepsilon^e \cdot \mathbb{C}\varepsilon^e + W^p(q) + \int_0^t \psi(\dot{q})\ dt\ \right] d\Omega, \tag{3.1}$$

where $\varepsilon^p : (0,T) \times \Omega \mapsto \mathbb{R}^{N\times N}$ is the plastic strain and $q : (0,T) \times \Omega \mapsto \mathbb{R}$ is the accumulated plastic strain

$$\dot{q} = \sqrt{\frac{2}{3}\dot{\varepsilon^p} \cdot \dot{\varepsilon^p}}. \tag{3.2}$$

$\mathbb{C}$ is the stiffness tensor, $\varepsilon^e := \varepsilon - \varepsilon^p$ is the elastic strain, and $\varepsilon := (\nabla u + \nabla u^T)/2$ is the symmetric gradient of the displacement field. W^p is the power-law plastic potential

$$W^p(q) = \sigma_y \left[q + \frac{n\varepsilon_0^p}{n+1}\left(\frac{q}{\varepsilon_0^p}\right)^{(n+1)/n}\right], \tag{3.3}$$

where σ_y is the yield stress, ε_0^p is the reference plastic strain, and n is the hardening exponent. We consider a power-law dissipation potential which enforces the monotonicity of the accumulated plasticity,

$$\psi(\dot{q}) = \begin{cases} g^*(\dot{q}) := \frac{m\sigma_y\dot{\varepsilon}_0^p}{m+1}\left(\frac{\dot{q}}{\dot{\varepsilon}_0^p}\right)^{(m+1)/m} & \dot{q} \geq 0, \\ \infty & \dot{q} < 0 \end{cases}, \tag{3.4}$$

where $\dot{\varepsilon}_0^p$ is the reference plastic strain rate, and m is the rate-sensitivity exponent.

Dynamic Rigid Contact We now discuss rigid contact in the context of dynamic indentation. Here, we consider the simplified case of frictionless contact. However, the presented formulation may be readily extended to include

friction. We consider a convex, rigid indentor occupying $\Omega_I \subset \mathbb{R}^N$ which may come into contact with the original elastic, J2 plastic body occupying $\Omega \subset \mathbb{R}^N$. Here, we consider $N = 3$. The centroid position and orientation of the indentor is described by $q_I := \{x_I,\ o_I\}$, $x_I \in \mathbb{R}^N$, with $o_I \in \mathbb{R}^{N+1}$ describing the quaternion orientation. We consider a continuous contact function $\mathcal{C}_I : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^{N+1} \mapsto \mathbb{R}$ such that $\mathcal{C}_I(x, q_I) < 0$ if $x \in \Omega_I$ and $\mathcal{C}_I(x, q_I) \geq 0$ if $x \notin \Omega_I$. In the case of a spherical indentor of radius R, the contact function may take the form

$$\mathcal{C}_I^{sphere}(x, q_I) := |x - x_I| - R. \tag{3.5}$$

Then, we consider the contact condition on boundary $\partial\Omega$. We enforce $\mathcal{C}_I(x, q_I) \geq 0$ through the Lagrange multiplier and slack variables $\lambda, \ell : \partial\Omega \mapsto \mathbb{R}$.

Balance Laws We derive the governing equations by writing the action integral based on the elasto-plastic potential, with a Lagrange multiplier for contact:

$$\begin{aligned}\widetilde{\mathcal{A}}(u, q, \varepsilon^p, \lambda) := \int_{t_1}^{t_2} \Bigg[\int_\Omega \frac{\rho}{2}|\dot{u}|^2 \, d\Omega - \mathcal{U}(u, q, \varepsilon) \\ + \int_\Omega b \cdot u \, d\Omega + \int_{\partial_f\Omega} f \cdot u \, dS \\ - \quad \int_{\partial\Omega} \lambda(\mathcal{C}_I(X + u, q_I) - \ell^2) \, dS \Bigg] \, dt.\end{aligned} \tag{3.6}$$

Stationarity of this action integral over the variable set $\{u, q, \varepsilon^p, \lambda\}$ gives the evolution relations

$$\begin{aligned}0 &= \int_\Omega \left[\rho\ddot{u} \cdot \delta u + \mathbb{C}\varepsilon^e \cdot \nabla\delta u - b \cdot \delta u\right] \, d\Omega \\ &\quad - \int_{\partial_f\Omega} f \cdot \delta u \, dS - \int_{\partial\Omega} \lambda \frac{\partial \mathcal{C}_I}{\partial u} \cdot \delta u \, dS \qquad \forall \delta u \in \mathcal{K}_0, \\ 0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot{q}) \qquad \text{on } \Omega, \\ \dot{\varepsilon}^p &= \dot{q}M \qquad \text{on } \Omega, \\ 0 &= \mathcal{C}_I(X + u, q_I) - \ell^2 \qquad \text{on } \partial\Omega, \\ u|_{t=0} &= \dot{u}|_{t=0} = 0,\ q|_{t=0} = 0,\ \varepsilon^p|_{t=0} = 0,\end{aligned} \tag{3.7}$$

where we now have the additional contact condition and associated contact force.

Numerics We discuss the discretizations and numerical schemes we use to solve for the deformation and plasticity evolution for each of the previously presented scenarios. The full details of this is presented in B.1. In all of these, the spatial discretization for the displacement field u is through standard $P = 1$ Largrange polynomial finite elements. The plastic quantities q and ε^p are spatially discretized at quadrature points. To handle the contact condition of (3.7), we adopt a staggered mixed update through a predictor-corrector algorithm [7, 17]. The displacements and velocities are first updated explicitly through a central difference scheme assuming no contact. Then, if penetration is detected, a correction force is computed and applied which maintains the contact condition. Then, the plastic quantities are updated with the usual implicit backwards Euler scheme.

3.2.2 Indirect inverse problem of parameter identification

We now turn to the problem of obtaining the elasto-viscoplastic material parameters, specifically $P := \{\sigma_y,\ \varepsilon_0^p,\ n,\ \dot{\varepsilon}_0^p,\ m\}$, from dynamic indentation tests.

Optimization Problem We formulate the inverse problem for finding the elasto-viscoplastic material parameters as an optimization problem. Following Section 3.2.1, we assume experimental data $D^{exp} := \{u^{exp},\ M^{exp}\}$, where u^{exp} is partial displacement data and M^{exp} is a collection of macroscopic measurable quantities. Then, the optimization problem may be written as

$$\begin{aligned} &\inf_{P\in\mathcal{P}}\ \mathcal{O}(P, u, q, \varepsilon^p, \lambda, D^{exp}) \\ &\text{Subject to : Evolution for } \{u, q, \varepsilon^p, \lambda\} \text{ from (3.7).} \end{aligned} \tag{3.8}$$

Here, the objective function is left general, however, it is expected that it attains its minimum value when the associated values computed from $\{u, q, \varepsilon^p, \lambda\}$ match the data D^{exp}.

Adjoint Method for Sensitivities We apply the adjoint method to obtain the sensitivities following [1] and the general approach of 3.2.1. We assume an objective of integral form,

$$\mathcal{O}(P, u, q, \varepsilon^p, \lambda, D^{exp}) = \int_0^T \int_\Omega o(P, u, q, \varepsilon^p, \lambda, D^{exp})\ d\Omega\ dt. \tag{3.9}$$

We introduce the adjoint variable set $\{v, \gamma, \mu, b\}$ associated, respectively, with $\{u, q, \varepsilon^p, \lambda\}$. Performing the adjoint method gives the sensitivities as

$$\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \left[\gamma \dot{q} \left(-\frac{\partial W^p}{\partial q \partial P} - \frac{\partial^2 g^*}{\partial \dot{q} \partial P} \right) \right] d\Omega \, dt, \tag{3.10}$$

where the adjoint variables ξ, γ, and μ satisfy the evolution relations

$$\begin{aligned}
0 = &\int_\Omega \left[\rho \ddot{v} \cdot \delta u + \left(\mathbb{C} \nabla v + \gamma \dot{q} \frac{\partial \sigma_M}{\partial \varepsilon} - \dot{q} \mu \cdot \frac{\partial M}{\partial \varepsilon} \right) \cdot \nabla \delta u \right. \\
&\left. + \frac{\partial o}{\partial u} \cdot \delta u \right] d\Omega + \int_{\partial\Omega} \left(l\lambda \frac{\partial \mathcal{C}_I}{\partial u} - \lambda v \cdot \frac{\partial^2 \mathcal{C}_I}{\partial u \partial u} \right) \cdot \delta u \; dS \quad \forall \delta u \in \mathcal{K}_0 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \left[\gamma \left(\sigma_M - \sigma_0 - \frac{\partial g^*}{\partial \dot{q}} \right) - \gamma \dot{q} \frac{\partial^2 g^*}{\partial \dot{q}^2} - \mu \cdot M \right] = -\gamma \dot{q} \frac{\partial \sigma_0}{\partial q} \qquad \text{on } \Omega \\
&\frac{\mathrm{d}\mu}{\mathrm{d}t} = \nabla \xi \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p} + \gamma \dot{q} \frac{\partial \sigma_M}{\partial \varepsilon^p} - \dot{q} \mu \cdot \frac{\partial M}{\partial \varepsilon^p} \qquad \text{on } \Omega \\
&0 = \frac{\partial C_I}{\partial u} \cdot v - \frac{\partial o}{\partial \lambda} \qquad \text{on } \partial\Omega_{\lambda \neq 0} \\
&v|_{t=T} = \dot{v}|_{t=T} = 0, \; \gamma|_{t=T} = 0, \quad \mu|_{t=T} = 0.
\end{aligned}$$

Here, we may interpret λl as a Lagrange multiplier enforcing the constraint that $\frac{\partial C_I}{\partial u} \cdot v = \frac{\partial o}{\partial \lambda}$ on regions of the boundary with nonzero λ.

Numerics The spatial discretization for each of the adjoint variables are equivalent to that of their associated forward variables. We use temporal discretizations which are analogous to the forward problem which accommodate the backwards in time nature of the adjoint problem. In the case of dynamic evolution of the displacement field u, the adjoint displacements are approximated with a explicit central difference method. Then, the adjoint Lagrange multiplier which satisfies the adjoint constraint is solved through a predictor-corrector scheme similar to the contact condition for the forward problem. Finally, the adjoint plastic variables are solved implicitly through a forward Euler scheme. The full details of this can be found in B.3.

Following the solution of the forward and adjoint problem, the objective and sensitivities are computed. Both are approximated with a simple Riemann sum in time and integrated with Gauss-quadrature in space.

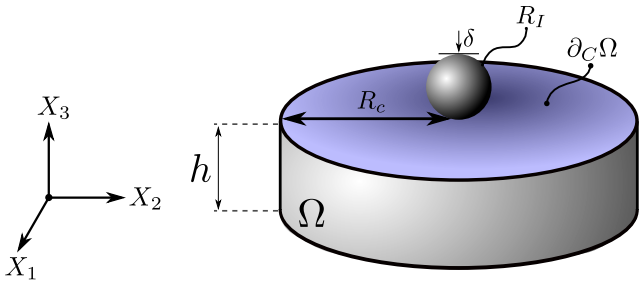


Figure 3.1: Diagram for the rigid indentation test. We consider a cylindrical domain Ω of radius R_C and height h indented on its top surface $\partial_C\Omega$ by a rigid sphere of radius R_I to a prescribed depth δ.

3.3 Demonstration using synthetic data

We consider a rigid sphere of radius R_I described by position $x_I(t)$ indenting a body which occupies a circular cylindrical domain Ω, which has radius R_C and height h in the reference configuration. This is shown in Figure. Then, following the formulation of Section 3.2.1, we look to match the net vertical load applied to the indentor $F_{net}^{sim}(t) := \int_{\partial_C\Omega} -\lambda N \cdot e_3 dS$ to the corresponding experimental data $F_{net}^{exp}(t)$. However, the explicit contact model introduces slight oscillations in the net load. Additionally, experimental data for indentation forces tends to be noisy. Thus, simply matching the L^2 norm in time of these two quantities, as is done in the previous examples, is inappropriate. Thus, we consider a mollification where a convolution filter in time is applied to both the modeled and experimental loads

$$\widetilde{F}_{net}^{sim,exp}(t) := \frac{1}{\int_0^T K(t-\tau)d\tau} \int_0^T K(t-\tau) F_{net}^{sim,exp}(t)\, d\tau, \tag{3.11}$$

where $K \in H^1((-T,T))$ is a positive kernel. Here, a re-normalization is applied to ensure unit weighting. Figure 3.2 shows a typical net force vs indentation depth plot for a dynamic indentation simulation with and without the filter.

Then, we consider the optimization problem

$$\begin{aligned} &\inf_{P\in\mathcal{P}} \ \mathcal{O} = \sum_{i=1}^{r} \left\| \widetilde{F}_{net}^{sim,i} - \widetilde{F}_{net}^{exp,i} \right\|_{L^2((0,T))}, \\ &\text{Subject to : Evolution for } \{u, q, \varepsilon^p, \lambda\} \text{ from (3.7) for } i = 1, \ldots, r. \end{aligned} \tag{3.12}$$

We again consider r trajectories at different indentation velocities.

Again, we generate synthetic data by running a forward set of simulations for a set parameters P^{gen}. Then, starting from P^{init}, we apply the gradient-based

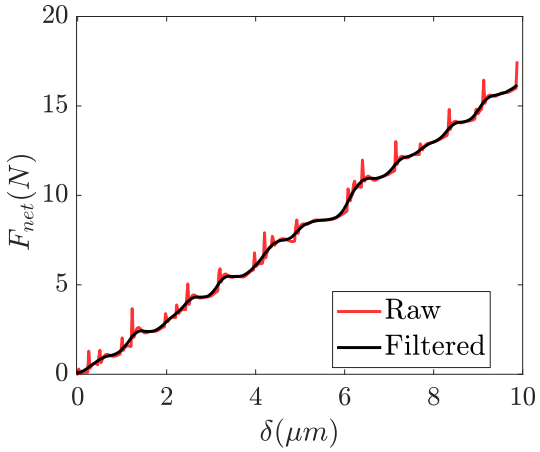


Figure 3.2: Raw and filtered net indentation force F_{net} vs indentation depth δ for a typical dynamic indentation simulation.

Synthetic and converged viscoplastic material parameters					
	σ_y/μ	ε_0^p	n	$\dot{\varepsilon}_0^p\ (/s)$	m
P^{gen}	0.001935	0.02245	3.23	5.00×10^5	2.00
P^{init}	0.005000	0.05000	5.00	3.00×10^5	5.00
P^{conv}	0.001961	0.03018	2.88	4.98×10^5	2.43

Table 3.1: Parameters for synthetic data generation and ones obtained through the adjoint method for the quasi-static tension, dynamic compression and dynamic indentation test.

optimization scheme. We consider $r = 4$ trajectories differing only by their indentation velocities of $V = 1.56$, 6.25, 25, and $100m/s$.

Table 3.1 shows the parameter values for those used to generate the synthetic data (P^{gen}), as well as the final converged parameter values (P^{conv}) and the initial parameter values (P^{init}). Figure 3.5a shows the relative objective function through optimization iterations. We see that the objective function decreases by 4.8 orders of magnitude over 500 iterations. This is more iterations than was required by both the quasi-static (300) and the dynamic annular compression case (150). Figure 3.4 shows the final deformed configurations corresponding to the synthetic data and the converged parameter sets for each of the tested impact velocities, while Figure 3.3 shows the load displacement curves. We see excellent agreement for both the deformed configurations and net load. Additionally, the converged set of parameters is close to the values used to generate the synthetic data. Figure 3.5b shows the quasi-static, uniaxial stress-strain behavior computed for a single element in tension for both the synthetic pa-

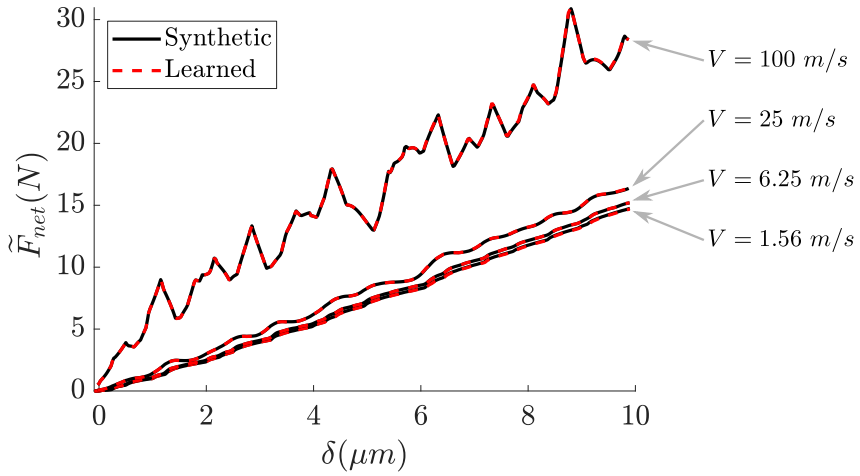


Figure 3.3: Filtered force vs indentation depth curves for the synthetic and the converged paraemter sets for the four indenter velocities tested.

rameters and learned parameters at different strain rates. While the values of the optimized parameters differ somewhat from those used to generate the synthetic data, their behavior in tension is quite similar. However, it does begin to differ at the larger strain rate (10^6).

Overall, the parameter optimization for the dynamic indentation test behaves worse than the quasi-static and annular compression example. This maybe due to a few factors. Firstly, we have only considered the force values in the objective function. This is the most easily accessible data from dynamic indentation testing. Secondly, the filtering applied to the loads in the objective function may affect the optimization. While this is necessary to clean up oscillations in the computations and those expected in real data, it may also expand the solution space of the optimization problem. Finally, the computational cost associated with the 3-dimensional nature of this example requires coarser-grained simulations. This may also contribute to the larger error seen in this example when compared to the 2D ones previously studied.

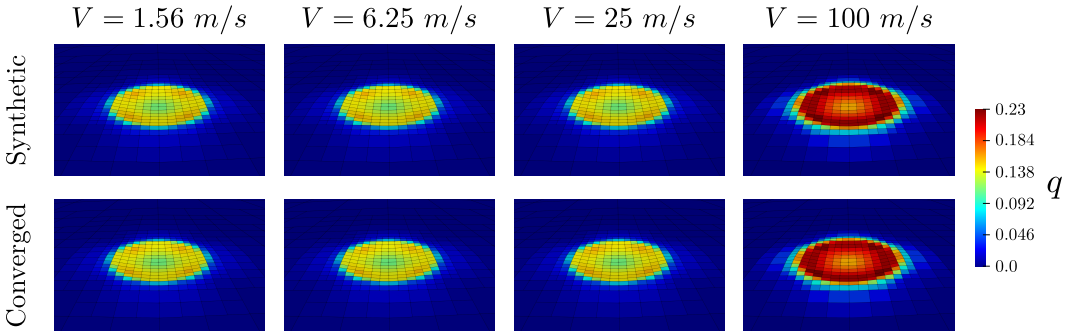
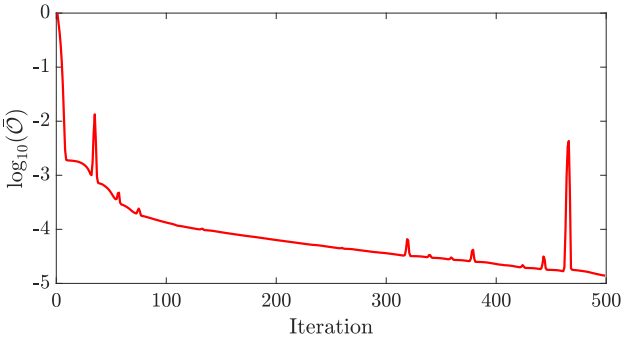
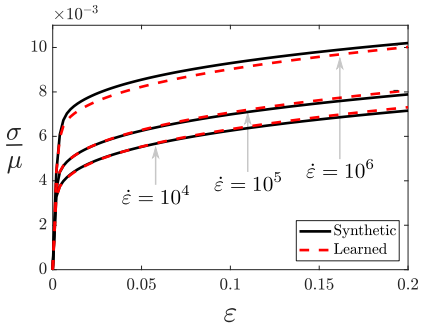


Figure 3.4: Deformed configurations at the final timestep for the indentation tests at the impact site for the synthetic and converged parameter sets at each of the impacting velocities of $V = 1.56$, 6.25, 25, and $100m/s$.



(a) Normalized objective vs iteration number for the indentation test parameter optimization.



(b) Uniaxial stress-strain response of a single element in quasi-static tension for the synthetic and learned parameters from the indentation test at different strain rates.

Figure 3.5: Optimization results and validation for synthetic data.

3.4 Demonstration using experimental data

3.4.1 Experimental setup

In this section, we attempt to invert experimental data obtained from dynamic indentation tests. Here, a specimen is impacted by a comparatively rigid indenter moving at an initial velocity. The indentation is performed using a Kolsky bar impact testing system [23, 8]. The setup consists of an input bar, an indenter bar, a striker and the specimen, which is placed between the input and output bars. For the test, the striker impacts the opposite end of the input bar, causing the sample to get compressed between the input and indenter bars. A combination of interferometers and strain-gages are used to measure indentation forces and velocity simultaneously. Thus, the set of

measurables include the indentation force and velocity profile of the indenter as a function of time. The spherical indenter was made of tungsten-carbide and mounted on a 38.6mm output bar. For simplicity of computations, the indenter is assumed to be rigid. In our tests, the peak indentation force was O(100N), total penetration depth O(10μm) and deformation rate of O(1 m/s). The optimization problem we are trying to solve is

$$\inf_{\{E,\sigma_y,n,\varepsilon_0^p\}} \mathcal{O} = \alpha_{IN} \sum_{i=1}^{r} \left\| \widetilde{F}_{net}^{sim,i} - \widetilde{F}_{net}^{exp,i} \right\|_{L^2((0,T))}, \tag{3.13}$$

Subject to : Evolution for $\{u, q, \varepsilon^p, \lambda\}$ from (3.7) for $i = 1, \ldots, r$.

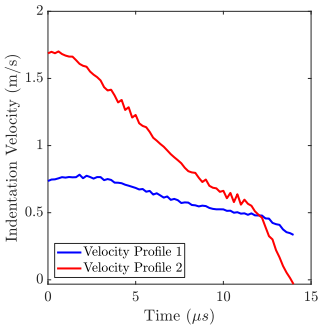
We seek to obtain the elastic modulus, yield stress and hardening parameters. Indentation data was performed at the same order of velocities, and therefore does not posses rich information about the rate effects. The inversion results for rolled homogeneous armous steel and polycrystalline aluminum alloy Al 6061-T6 are presented below.

3.4.2 Rolled homogeneous armour steel

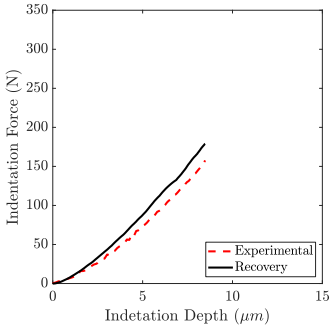
We consider a rectangular RHA steel specimen of dimensions $10mm \times 10mm \times 4mm$. The indentation is directed parallel to the direction of the shortest dimension ($4mm$). The indenter surface is spherical with a radius of $1608\mu m$. Full details of the setup can be found in [8]. Two tests are performed with different velocity profiles of the striker to obtain the experimental data shown in Figure 3.6a. We apply the inversion method, and recover the set of inverted parameters P^{conv} shown in Table 3.2. The normalized objective through optimization iterations are shown in Figure 3.7a. We find the recovered force-indentation depth curves are in close agreement with the experimental data shown in Figure 3.6b and 3.6c.

Converged parameters obtained using experimental indentation data						
	Recovered				Computed	
	E (GPa)	σ_y (MPa)	ε_0^p	n	μ (GPa)	λ (GPa)
P^{conv}	172.919	664.8012	0.0425	10.5537	68.08	79.92

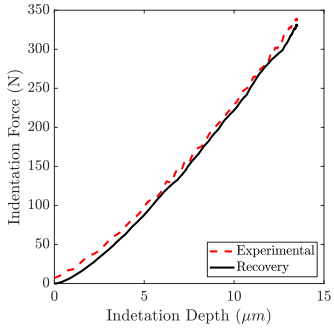
Table 3.2: Recovered parameters for RHA Steel. We perform inversion over elastic modulus (E), yield stress (σ_y) and hardening parameters (ε_0^p, n). We have also presented the computed values of shear modulus (μ) and lame parameter λ.



(a) Velocity profiles for experimental inversion of indetation.

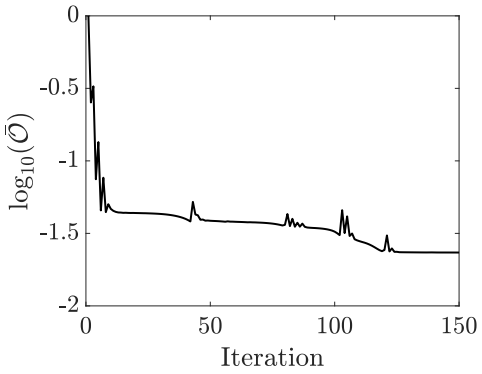


(b) Force versus Indentation curves for velocity profile 1.

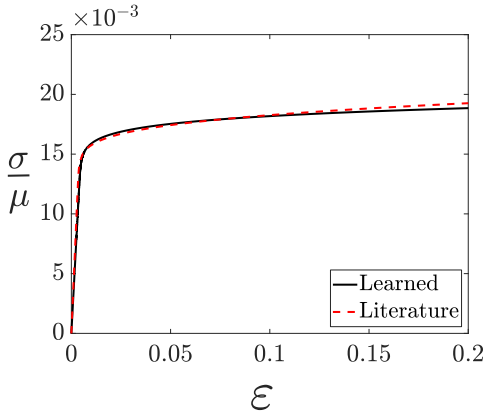


(c) Force versus Indentation curves for velocity profile 2.

Figure 3.6: Indentation inversion results for Rolled Homogeneous Armor Steel



(a) Objective plot for inversion of experimental data.



(b) Material point calculation for uniaxial stress-strain response for tensile loading. For both responses, the stress has been normalized using $\mu = 82.68$ GPa.

Figure 3.7: Optimization results and validation of learnt model from experimental indentation results.

Material properties of RHA steel from literature [16, 5]						
E (Gpa)	μ (GPa)	A(MPa)	B(MPa)	n	C	λ (GPa)
200	82.68	740	780	0.106	0.0891	97.06

Table 3.3: Johnson-Cook parameters for RHA Steel from [16, 5]. We present these parameters to compare the values with our recovered set of parameters. The material point calculation for an uniaxial stress-strain response has been shwon in figure 3.7b.

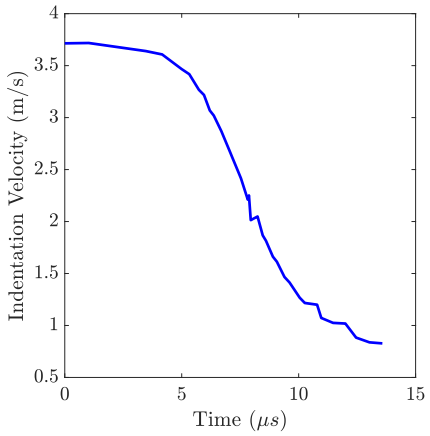
We further compare the efficacy of recovered material properties with existing models as shown in figure 3.7b. Most popularly, RHA Steel is modeled using Johnson-Cook model using the material properties as shown in table 3.3. A material point calculation is performed for uniaxial tensile loading to compare the normalized stress-strain response of the recovered material parameters and Johnson-Cook model obtained from pervious literature. The learnt parameters are able to capture the behaviour with good efficacy.

3.4.3 Polycrystalline aluminum alloy Al 6061-T6

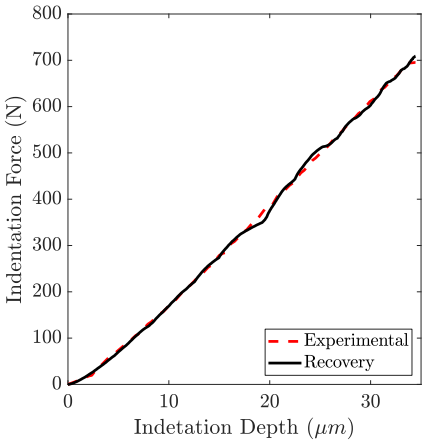
Similarly, we used previously published data on indentation of aluminium [9]. The material properties were determined by performing an inversion analysis using the dynamic spherical indentation data at an average indentation velocity of $\bar{v} = 1.36m/s$. The complete velocity profile of the indenter has been shown in figure 3.8a. The experimental force-depth response was input experimental data to our indirect inversion technique. The recovered response compared with finite element (FE) simulations incorporating the Johnson–Cook plasticity model, which accounts for strain hardening, strain rate sensitivity, and thermal softening effects. The simulation results closely matched the experimental indentation force-depth curves, with minor deviations attributed to boundary condition approximations and potential variations in strain rate sensitivity. The extracted material parameters demonstrated reasonable agreement with previously published values for Al 6061-T6, validating the methodology for characterizing dynamic material behavior using indentation techniques. Additionally, the estimated strain rate sensitivity parameter, obtained through optimization of the force-displacement curves, aligned well with independent high-strain-rate experimental data, further supporting the robustness of the inversion approach.

Converged parameters obtained using experimental indentation data						
	Recovered				Computed	
	E (Gpa)	σ_y (MPa)	ε_0^p	n	μ (GPa)	λ (GPa)
P^{conv}	85.776	386.88	4.94	1.96	32.24	62.65

Table 3.4: Recovered parameters for Al 6061-T6 Steel. We perform inversion over elastic modulus (E), yield stress (σ_y) and hardening parameters (ε_0^p, n). We have also presented the computed values of shear modulus (μ) and lame parameter λ.

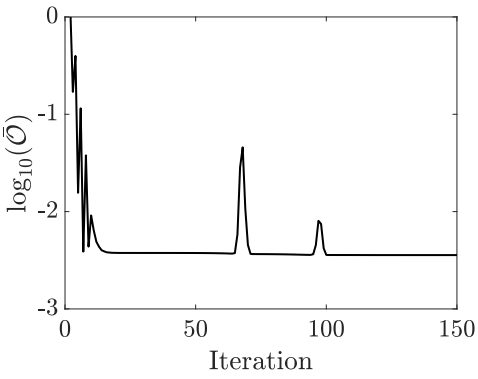


(a) Velocity profile for experimental inversion of indetation.

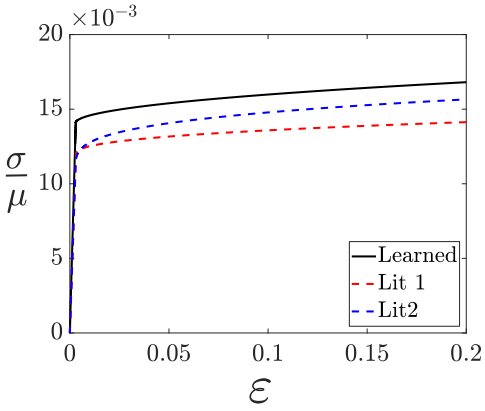


(b) Force versus Indentation curves for velocity profile 1.

Figure 3.8: Indentation results for polycrystalline aluminium allow Al 6061-T6.



(a) Objective plot for inversion of experimental data.



(b) Uniaxial stress-strain response of a single element in quasi-static compression. For both responses, the stress has been normalized using $\mu = 27$ GPa.

Figure 3.9: Optimization results and validation of learnt model from experimental indentation results.

Material properties of Al 6061-T6 from literature [9, 3]						
E (Gpa)	μ (GPa)	A(MPa)	B(Mpa)	n	C	λ (GPa)
70	26	300	200	0.3	0.05	51
69	24	324	114	0.42	0.002	50

Table 3.5: Johnson-Cook parameters for Al6061-T6 from previous literature [9, 3]. We present these parameters to compare the values with our recovered set of parameters. The material point calculation for an uniaxial stress-strain response has been shown in figure 3.9b. Since polycrystalline aluminium alloy is known to have a slight variation in response due to several factors, we present two different set of parameters.

3.5 Conclusions

The persistent challenge in material characterization revolves around the conventional approach of direct inversion, wherein measurable data is initially converted into constitutive quantities before being inverted to deduce material properties. Unfortunately, this method necessitates intricate and expensive experimental setups. Previously, we introduced an alternative approach called the indirect inversion to directly obtain material properties from the measurable data. This methodology offers a more efficient and practical means of material characterization mitigating the shortcomings associated with traditional techniques. We formulate the problem as a PDE-constrained optimization, with an objective function quantifying the disparity between measurable data and predictions and the optimization process is conducted over the material parameter set. Leveraging the adjoint method, we compute the sensitivity of the objective concerning the material parameters. Subsequently, we employ gradient descent to minimize the objective, iterating until the objective falls below a predefined tolerance threshold.

In this study, We present results pertaining to the inversion of synthetically generated data of dynamic rigid indentation data for a J2-isotropic material featuring power-law strain and rate hardening. The material parameters obtained through convergence closely align with the synthetic parameters employed to generate the data. To further validate the accuracy of the converged parameters, we conduct quasi-static uniaxial tensile tests and compare the response with the assumed synthetic parameter set. Additionally, our inversion methodology extends to experimental data, encompassing both indentation data and uniaxial stress-strain data. Notably, we apply this approach to RHA-

steel specimens, successfully obtaining reasonable values for yield and strain hardening parameters. This demonstrates the effectiveness of our technique across synthetically and experimentally acquired datasets.

In future works, we aim to extend our methodology to encompass a generalized constitutive law by integrating a history-dependent neural network as the constitutive model [20, 10, 22]. Recently, recurrent neural operators have been shown to effectively capture the behaviour of plastic materials through an arbitrarily defined set of internal variables [14, 6, 12]. The scalability of our method, operating at order 1 with respect to the number of parameters, ensures that the computation time for inversion remains unaffected even with the incorporation of a recurrent neural operator with large number of parameters. Moreover, there is a need for an automated experimental setup for generating high-throughput data for inversion. Implementing such a system will enable us to acquire data rapidly for robust material characterization. Another additional avenue to be researched is the mathematical formulation of the relationship between prior and posterior distribution of the material parameters. This includes a systematic investigation into efficiently obtaining various parameters from different types of data.

References

[1] A. Akerson. "Optimal structures for failure resistance under impact." In: *Journal of the Mechanics and Physics of Solids* 172 (Mar. 2023), p. 105172. ISSN: 00225096. DOI: 10.1016/j.jmps.2022.105172. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022509622003489.

[2] A. Akerson, A. Rajan, and K. Bhattacharya. "Learning constitutive relations from experiments: 1. PDE constrained optimization." In: *arXiv preprint arXiv:2412.02864* (2024).

[3] S. Akram, S. H. I. Jaffery, M. Khan, M. Fahad, A. Mubashar, and L. Ali. "Numerical and experimental investigation of Johnson–Cook material models for aluminum (Al 6061-T6) alloy using orthogonal machining approach." In: *Advances in Mechanical Engineering* 10.9 (2018), p. 1687814018797794.

[4] R. J. Anton and G. Subhash. "Dynamic Vickers indentation of brittle materials." In: *Wear* 239 (2000), pp. 27–35. ISSN: 0043-1648.

[5] R. Benck. "Quasi-static tensile stress strain curves: II Rolled homogeneous armor." In: *Aberdeen Proving Ground (MD): Ballistic Research Laboratories (US)* (1976).

[6] K. Bhattacharya, B. Liu, A. Stuart, and M. Trautner. "Learning Markovian homogenized models in viscoelasticity." In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 641–679.

[7] G. Camacho and M. Ortiz. "Adaptive Lagrangian modelling of ballistic penetration of metallic targets." In: *Computer Methods in Applied Mechanics and Engineering* 142.3 (1997), pp. 269–301. ISSN: 0045-7825. DOI: https://doi.org/10.1016/S0045-7825(96)01134-6. URL: https://www.sciencedirect.com/science/article/pii/S0045782596011346.

[8] D. Casem and E. Retzlaff. "A Kolsky bar for high-rate indentation." In: *Journal of Dynamic Behavior of Materials* 9.3 (2023), pp. 300–314.

[9] J. Clayton, J. Lloyd, and D. Casem. "Simulation and dimensional analysis of instrumented dynamic spherical indentation of ductile metals." In: *International Journal of Mechanical Sciences* 251 (2023), p. 108333.

[10] Y. Deng, S. Ren, K. Fan, J. M. Malof, and W. J. Padilla. "Neural-adjoint method for the inverse design of all-dielectric metasurfaces." In: *Optics Express* 29.5 (2021), pp. 7526–7534.

[11] B. J. Koeppel and G. Subhash. "An experimental technique to investigate the dynamic indentation hardness of materials." In: *Experimental Techniques* 21 (May 1997), pp. 16–18. ISSN: 1747-1567.

[12] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. "Neural operator: Learning maps between function spaces with applications to pdes." In: *Journal of Machine Learning Research* 24.89 (2023), pp. 1–97.

[13] A. Lee and K. Komvopoulos. "Dynamic spherical indentation of elastic-plastic solids." In: *International Journal of Solids and Structures* 146 (2018), pp. 180–191.

[14] B. Liu, E. Ocegueda, M. Trautner, A. M. Stuart, and K. Bhattacharya. "Learning macroscopic internal variables and history dependence from microscopic models." In: *Journal of the Mechanics and Physics of Solids* 178 (2023), p. 105329.

[15] J. Lubliner. *Plasticity Theory*. Dover books on engineering. Dover Publications, 2008. ISBN: 9780486462905. URL: https://books.google.com/books?id=MkK-BLbHtcAC.

[16] H. Meyer and D. S. Kleponis. "An analysis of parameters for the Johnson-Cook strength model for 2-in-thick Rolled Homogeneous Armor." In: *Army Research Laboratory, Adelphi, MD, Report No. ARL-TR-2528* (2001).

[17] J.-F. Molinari, M. Ortiz, R. Radovitzky, and E. Repetto. "Finite-element modeling of dry sliding wear in metals." In: *Engineering Computations (Swansea, Wales)* 18 (Oct. 2002). DOI: 10.1108/00368790110407257.

[18] M. Nilsson. *Dynamic Hardness Testing Using a Split Hopkinson Pressure Bar Apparatus.* Swedish Defense Research Agency: SE-147 25. 2002.

[19] M. Ortiz and L. Stainier. "The variational formulation of viscoplastic constitutive updates." In: *Computer Methods in Applied Mechanics and Engineering* 171.3-4 (1999), pp. 419–444. ISSN: 00457825. DOI: 10.1016/S0045-7825(98)00219-9.

[20] S. Ren, W. Padilla, and J. Malof. "Benchmarking deep inverse models over time, and the neural-adjoint method." In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 38–48.

[21] B. Si, Z. Li, Y. Yang, L. Qiao, E. Liu, G. Xiao, and X. Shu. "Dynamic indentation testing and characterization of metals based on the split Hopkinson pressure bar (SHPB) device." In: *Mechanics of Materials* 177 (2023), p. 104550.

[22] C. Xiao, Y. Deng, and G. Wang. "Deep-learning-based adjoint state method: Methodology and preliminary application to inverse modeling." In: *Water Resources Research* 57.2 (2021), e2020WR027400.

[23] F. Zhou, J.-F. Molinari, and Y. Li. "Three-dimensional numerical simulations of dynamic fracture in silicon carbide reinforced aluminum." In: *Engineering Fracture Mechanics* 71.9-10 (2004), pp. 1357–1378.

Chapter 4

LEARNING CONSTITUTIVE RELATIONS FROM EXPERIMENTS: III. MACHINE LEARNING

Rajan, A. and Bhattacharya, K., 2025. "Learning viscoplastic constitutive behavior from experiments: III. Machine Learning." In preparation.

Contributions: A.R. developed the computational code and performed simulations. K.B. conceptualized the methodology. A.R. and K.B. wrote the manuscript.

4.1 Introduction

The characterization of constitutive behavior from experimental observations remains a central problem in continuum mechanics. While the governing equations of motion are well-established through the balance laws of mass and momentum, they are not closed without a constitutive relation that captures the material's response to deformation [12, 4]. This relation cannot be measured directly; instead, it must be inferred from observable quantities such as displacements and forces. Recovering constitutive behavior from such data is therefore a challenging task, as it requires reasoning backward from global or partial field measurements to local material response—often under conditions of limited data, measurement noise, and complex loading.

In part I [1], we developed a computational framework for identifying material behavior from experimental data by casting the problem as a partial differential equation (PDE)-constrained optimization. The constitutive relation was represented using a classical isotropic J2 elasto-viscoplasticity model, with a set of parameters governing the stress–strain response. These parameters were inferred by minimizing the discrepancy between measured experimental quantities—such as displacements or reaction forces—and the predictions from solving the governing evolution laws. The optimization was performed using gradient-based methods, with sensitivities efficiently computed via the adjoint

method. This formulation enabled accurate recovery of material parameters in both quasistatic and dynamic settings using synthetic data, while remaining computationally efficient and scalable.

In part II, we built on that foundation and extended the approach to dynamic indentation experiments. These tests, which involve time-dependent effects and contact interactions, present rich yet challenging data for material characterization [18, 17, 6, 24, 27]. We incorporated contact constraints into the variational formulation and tested the method on both synthetic and experimental data. By applying the approach to indentation tests on aluminum and rolled hardened armor steel, we demonstrated that accurate constitutive behavior can be recovered from limited experimental observations—even in complex loading regimes. This extension broadened the method's applicability to more practical and experimentally relevant scenarios.

While classical constitutive models—such as isotropic plasticity or viscoelasticity—have been remarkably successful in many settings, they rely on explicit formulas with a small number of parameters and simplifying assumptions about material behavior. These assumptions often limit their applicability in scenarios involving complex loading histories, anisotropy, or rate-dependent effects. Even for homogeneous and isotropic materials, there remains an ongoing debate about which constitutive models best capture observed behaviors under multiaxial or non-proportional loading conditions [22, 16, 14, 32]. As materials and applications become more complex, so too must the models that describe them. This has led to a growing interest in using hyper-parameterized representations—such as neural networks—to approximate constitutive relations directly from data. These models offer the expressive capacity to capture subtle, nonlinear, and history-dependent behaviors without committing to a specific closed-form structure [20, 9, 15]. However, realizing this potential requires addressing the significant challenges associated with data generation, representation, and inference.

Physics-Informed Neural Networks (PINNs) have emerged as a popular approach for learning constitutive models by embedding physical laws directly into the training process [10, 13, 29, 8]. In this framework, the neural network serves as the constitutive relation—mapping, for instance, strain to stress—and is trained not only to fit experimental observations (such as displacement or force data), but also to minimize residuals of the governing partial

differential equations (PDEs), as well as enforce boundary and initial conditions. However, a central limitation of PINNs is that the PDE is not solved exactly, rather it is only approximately satisfied by penalizing its residuals at a finite number of collocation points.

In the field of plasticity, researchers have started using neural networks for multiscale modeling, where they serve as a surrogate for expensive fine-scale simulations. In traditional multiscale approaches, homogenization techniques are used to relate microscopic structure and behavior to macroscopic response [30]. However, these can be computationally intensive, especially for inelastic or history-dependent materials as equations are required to be solved in multiple length scales [5, 28, 11, 31]. Efforts have been made to computationally acclerate these traditional techniques using GPUs, but they make significant assumptions on the microscopic structure [33, 25]. Learning-based approaches offer an efficient alternative: by training neural networks on the results of high-fidelity simulations (e.g., direct numerical simulations of microstructure or crystal plasticity), one can construct surrogate constitutive models that accurately approximate the macroscale response. These models can be integrated directly into finite element simulations, significantly reducing computational cost while retaining fidelity to the underlying physics. For example, neural operators have been used to capture history-dependent behavior learned from microstructural simulations, enabling real-time macroscale predictions in dynamic impact problems [21, 19]. Such efforts demonstrate the potential of machine learning to bridge scales and serve as a conduit for embedding fine-scale physics into macroscale models.

A new method for constitutive modeling of history dependent materials is recurrent neural operators (RNOs) [21, 3, 7, 2, 23]. Unlike traditional feedforward neural networks that map inputs to outputs in a single pass, RNOs are designed to model mappings between time-dependent function spaces. They maintain an evolving hidden state that encodes the material's memory, making them naturally suited for history-dependent phenomena such as viscoelasticity [7, 3], viscoplasticity [21, 2], or damage[23]. At each time step, the RNO takes in the current strain (or strain rate) and updates the hidden state, which in turn determines the stress response. This architecture enables the model to learn complex temporal dependencies without the need to predefine internal variables or their evolution laws. In this work, we aim to use a similar RNO

architecture trained using the PDE-constrained framework.

In Part III of our series, we develop a framework for learning history-dependent constitutive models using recurrent neural operators (RNOs) embedded within a PDE-constrained optimization setting. Building on our previous work, we replace the classical parametric constitutive relation with a data-driven model that implicitly captures material memory through a learned recurrent structure. The RNO is trained by minimizing the discrepancy between observed measurable data and the output of forward simulations that solve the governing PDEs, ensuring that the learned model is both predictive and physically consistent. Unlike PINNs, which penalize residuals without solving the PDEs, our approach explicitly enforces the governing equations through numerical simulation. We demonstrate this method using synthetic dynamic experiments, where the ground truth material response exhibits strong path dependence. Our results show that the RNO-based model is able to accurately recover the underlying mechanical behavior, outperforming classical models in scenarios involving complex loading histories. This work highlights the potential of neural operators to serve as general-purpose, structure-aware surrogates for history-dependent constitutive behavior.

The paper is organized as follows. We present the methodology in Section 4.2 with details provided in the appendix. A recurrent neural operator based constitutive model is assumed in section 4.2.1 and the framework for parameter recovery is derived in section 4.2.2. We present the results for training the model parameters in section 4.3.1 and validate the method in section 4.3.2. We conclude with some comments in Section 4.4.

4.2 Methodology

In this section we provide the details of the framework for a RNO-based constitutive model. The formulation is derived from our previous work [1]. We formulate the problem of material behavior identification from experimental data as a PDE-constrained optimization. In this section, we describe the formulation for a viscoplastic constitutive law. We begin by introducing a class of constitutive models that incorporate history dependence through internal variables. This is followed by a presentation of the governing equations, including the momentum balance and the evolution laws for the internal variables. We then pose the inverse problem: given partial observations from experiments,

determine the set of material parameters that produces a state trajectory—via the evolution laws—that best matches the measured data. This leads naturally to a PDE-constrained optimization formulation. To efficiently compute gradients of the objective with respect to the parameters under these constraints, we introduce the adjoint method, which yields the total derivative of the objective while satisfying the governing equations. We then outline the general solution procedure for the resulting optimization problem. Finally, we comment on the scalability of the approach with respect to the dimensionality of the parameterization.

4.2.1 Recurrent neural operator as a constitutive model

Viscoplastic materials are inherently history-dependent—their current stress state is influenced not only by the instantaneous deformation but also by the entire sequence of prior loading. Traditional constitutive models encode this memory through internal variables governed by phenomenological evolution laws. However, designing such models requires substantial physical insight and may not capture complex or emergent behavior present in novel materials. Neural networks offer a flexible alternative, and among them, recurrent architectures are particularly well-suited for capturing history dependence due to their natural ability to process sequential data. Recurrent neural operators (RNOs) have recently been proposed as a powerful and general framework for learning mappings between function spaces, particularly in the context of PDE-constrained problems [21, 3, 7]. Furthermore, universal approximation theorems has been proved for RNOs ability to capture viscoelastic and viscoplastic behavior [2, 21]. Their ability to model temporal evolution makes them an attractive choice for representing constitutive relations in history-dependent materials. In this work, we incorporate RNOs into our PDE-constrained optimization framework to model both stress response and internal state evolution directly from data, without prescribing explicit evolution laws.

Constitutive model and governing equations. We consider an open, bounded domain $\Omega \subset \mathbb{R}^n$ occupied by a deformable solid over the time interval $(0, T)$. The body has material density ρ and is governed by a constitutive model parameterized by a set of material parameters $P = \{P_{\mathbb{C}}, P_{\text{RNO},\sigma}, P_{\text{RNO},\xi}\}$. Additionally, the material exhibits history dependence captured by a set of in-

ternal variables $\xi := \{\xi_i\},\ i = 1, \ldots, N_\xi$. Assuming a Perzyna-type kinetic formulation, the constitutive behavior is expressed as

$$\begin{aligned} \sigma &= \mathbb{C}\varepsilon \ + \ S_{\mathrm{NN}}(\nabla u, \xi; P_{\mathrm{RNO},\sigma}), \quad \varepsilon = (\nabla u + \nabla u^T)/2 \\ \dot{\xi} &= G_{\mathrm{NN}}(\nabla u, \xi; P_{\mathrm{RNO},\xi}), \end{aligned} \tag{4.1}$$

where σ is the first Piola-Kirchhoff stress and u is the displacement field. $\mathbb{C}$ is the elastic modulus tensor and ε is the strain tensor. $S_{\mathrm{NN}}, G_{\mathrm{NN}}$ are neural networks and $\{P_{\mathrm{RNO},\sigma}, P_{\mathrm{RNO},\xi}\}$ represents the corresponding sets of weights and biases. The functions S_{NN} and G_{NN} describe the non-linear stress response and internal variable evolution, respectively. The classical material parameters are $P_{\mathbb{C}}$, which is the elastic modulus. The constitutive model combines classical linear elasticity with a recurrent neural operator (RNO) component. This hybrid approach balances physical fidelity with expressive power: the classical model captures well-established elastic behavior observed in a wide range of materials, while the RNO component augments the model's ability to capture complex, uncharacterized, or nonstandard material responses. In doing so, the RNO acts as a data-driven correction or extension to the phenomenological framework.

The dynamic evolution of the body is governed by

$$\begin{aligned} \rho\ddot{u} - \nabla \cdot \sigma &= b && \text{in } \Omega, \\ \dot{\xi} &= G_{\mathrm{NN}}(\nabla u, \xi; P) && \text{in } \Omega, \\ \sigma n &= f && \text{on } \partial\Omega, \\ u|_{t=0} = \dot{u}|_{t=0} &= \xi|_{t=0} = 0, \end{aligned} \tag{4.2}$$

where b and f denote prescribed body and surface loads. The first equation enforces conservation of linear momentum, while the second and third specify the constitutive law and internal variable evolution. We assume quiescent initial conditions.

4.2.2 Formulation of the optimization problem

We assume access to partial experimental data denoted by $D^{\mathrm{exp}} := \{u^{\mathrm{exp}}, M^{\mathrm{exp}}\}$, where u^{exp} represents displacements recorded at select time instances and locations—typically only on parts of the boundary—and M^{exp} denotes additional macroscopic observables, such as net forces or spatially averaged strains. The goal is to identify the parameter set P such that the predicted displacement

and internal variable evolution from the model in (4.2) yield outputs that match the measured data D^{exp}. This inverse problem is posed as a PDE-constrained optimization problem:

$$\begin{aligned} &\inf_{P \in \mathcal{P}} \ \mathcal{O}(P, u, \xi, D^{\text{exp}}) \\ &\text{Subject to: evolution of } \{u, \xi\} \text{ governed by (4.2),} \end{aligned} \tag{4.3}$$

where $\mathcal{P}$ defines the admissible range of physical parameters. The objective function $\mathcal{O}(P, u, \xi, D^{\text{exp}})$ quantifies the mismatch between the model predictions and experimental observations. Its structure can significantly affect both the solution and the convergence behavior of the optimization; we explore these aspects in Section 4.3.

Since each evaluation of the objective requires solving the full dynamic system in (4.2), sampling-based approaches such as Bayesian inference or Monte Carlo methods become prohibitively expensive—especially as the dimension of P increases. These methods scale poorly, with costs growing exponentially in the number of parameters. Therefore, we adopt a gradient-based optimization framework, where cost scales primarily with the complexity of the forward model rather than the parameter space dimension. This necessitates computing the *sensitivities*—the total derivative of the objective with respect to P, constrained by the governing PDEs. We use the adjoint method [26], a classical and efficient approach that provides these sensitivities at a computational cost comparable to a single forward solve. The sensitivities are then used to iteratively update P until convergence.

Adjoint method. We perform the adjoint method to obtain expressions for the sensitivities. Without loss of generality, we consider an objective of integral form

$$\mathcal{O}(P, u, \xi, D^{exp}) := \int_0^T \int_\Omega o(P, u, \xi, D^{exp}) \, d\Omega \, dt. \tag{4.4}$$

Then, the sensitivity is $\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P}$ subject to the constraint that the evolution laws in (4.2) are satisfied. We may compute this by applying standard chain rule. However, this would require expressions for $\frac{\mathrm{d}u}{\mathrm{d}P}$ and $\frac{\mathrm{d}\xi}{\mathrm{d}P}$. These, resulting from perturbations on the dynamic trajectory, are difficult to compute. Thus, we instead apply the adjoint method to circumvent the need to compute them.

We consider adding zero to the objective through the assumed satisfied weak form of the evolution

$$\begin{aligned}\mathcal{O} = &\int_0^T \int_\Omega o(P, u, \xi, D^{exp})\ d\Omega\ dt \\ &+ \int_0^T \Big\{ \int_\Omega \left[-\rho \ddot{u} \cdot v - \sigma \cdot \nabla v + b \cdot v - \phi(\dot{\xi} - G_{\mathrm{NN}}) \right]\ d\Omega \\ &+ \int_{\partial\Omega} f \cdot v\ dS \Big\}\ dt,\end{aligned} \tag{4.5}$$

where v and ϕ are test functions associated with momentum balance and internal variable evolution. Then, differentiating with respect to P gives

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg[&\frac{\partial o}{\partial P} + \frac{\partial o}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}P} + \frac{\partial o}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} - \rho \frac{\mathrm{d}\ddot{u}}{\mathrm{d}P} \cdot v \\ &- \left(\frac{\partial \sigma}{\partial P} + \frac{\partial \sigma}{\partial \nabla u} \cdot \nabla \frac{\mathrm{d}u}{\mathrm{d}P} + \frac{\partial \sigma}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} \right) \cdot \nabla v \\ &- \phi \left(\frac{\mathrm{d}\dot{\xi}}{\mathrm{d}P} - \frac{\partial G_{\mathrm{NN}}}{\partial P} - \frac{\partial G_{\mathrm{NN}}}{\partial \nabla u} \cdot \nabla \frac{\mathrm{d}u}{\mathrm{d}P} - \frac{\partial G_{\mathrm{NN}}}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}P} \right) \Bigg]\ d\Omega\ dt.\end{aligned} \tag{4.6}$$

After grouping terms, applying integration by parts and Gauss divergence theorem, we eliminate the dependence on $\frac{\mathrm{d}u}{\mathrm{d}P}$ and $\frac{\mathrm{d}\xi}{\mathrm{d}P}$ by choosing a particular v and ϕ. Substituting σ from equation 4.1, the expression for the sensitivities reduces to

$$\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \left[\frac{\partial o}{\partial P} - \left(\frac{\partial \mathbb{C}}{\partial P} + \frac{\partial S_{\mathrm{NN}}}{\partial P} \right) \cdot \nabla v + \phi \frac{\partial G_{\mathrm{NN}}}{\partial P} \right]\ d\Omega\ dt \tag{4.7}$$

if the adjoint variables v and ϕ satisfy the evolution

$$\begin{aligned}&\rho \ddot{v} - \nabla \cdot \left(\nabla v \cdot \mathbb{C}\ \frac{\partial \varepsilon}{\partial \nabla u} + \nabla v \cdot \frac{\partial S_{\mathrm{NN}}}{\partial \nabla u} - \phi \frac{\partial G_{\mathrm{NN}}}{\partial \nabla u} \right) = \frac{\partial o}{\partial u} && \text{on } \Omega, \\ &\dot{\phi} = \nabla v \cdot \frac{\partial S_{\mathrm{NN}}}{\partial \xi} - \frac{\partial G_{\mathrm{NN}}}{\partial \xi}\phi - \frac{\partial o}{\partial \xi} && \text{on } \Omega, \\ &\left(\nabla v \cdot \mathbb{C}\ \frac{\partial \varepsilon}{\partial \nabla u} + \nabla v \cdot \frac{\partial S_{\mathrm{NN}}}{\partial \nabla u} - \phi \frac{\partial G_{\mathrm{NN}}}{\partial \nabla u} \right) n = 0 && \text{on } \partial\Omega, \\ &v|_{t=T} = \dot{v}|_{t=T} = 0,\ \ \phi|_{t=T} = 0.\end{aligned} \tag{4.8}$$

We note that the initial conditions for the adjoint variables are given at the end time $t = T$. Thus, the adjoint system should be solved backwards in time starting from $t = T$ and moving to $t = 0$.

The general procedure is to start with an initial value for the parameter set P. First, we solve the forward problem (4.2) for $u(t),\ \xi(t)$. Then, the adjoint

problem (4.8) is solved backwards in time for $v(t)$, $\phi(t)$. These are used in (4.7) to compute the sensitivities. Finally, the sensitivities are used to update the parameter set through a gradient-based optimization algorithm and the process is repeated until convergence. This formulation is general and flexible, enabling the recovery of material behavior across a broad spectrum of constitutive responses. The inversion can be carried out over the full parameter set $P = \{P_{\mathbb{C}}, P_{\text{RNO},\sigma}, P_{\text{RNO},\xi}\}$, encompassing both physically motivated and neural network-based components.

Scalability. A key advantage of the proposed framework is its scalability with respect to the number of parameters. Since our method relies on adjoint-based gradient computation, the optimization cost scales linearly with the number of parameters, regardless of whether they are physical parameters in a classical constitutive law or weights and biases in a neural network. As a result, the framework remains computationally tractable even when incorporating high-dimensional parameterizations such as those arising from RNOs, making it well-suited for complex, data-driven constitutive modeling.

Numerics. We discuss the discretizations and numerical schemes we use to solve for the deformation and internal variables in C.2. The spatial discretization for the displacement field u is through standard $P = 1$ Largrange polynomial finite elements. The internal variables ξ are spatially discretized at quadrature points.

4.3 Demonstration using synthetic data

We assess the performance of the neural operator-based constitutive modeling framework introduced in Section 4.2.1 by applying it to synthetic dynamic experiments. These tests simulate high-rate compressive loading on thin cylindrical specimens and serve to validate whether the recurrent neural operator (RNO) can accurately capture history-dependent plasticity from sparse yet informative data.

We consider specimens of the form $\Omega := \overline{\Omega} \times (0, h)$, where $\overline{\Omega} \subset \mathbb{R}^2$ is the in-plane cross-section and $h \ll \text{dimension}(\overline{\Omega})$ is the thickness. During testing, the top and bottom faces are subjected to prescribed motion representative of

an impact experiment:

$$u_3(X_1, X_2, 0, t) = 0, \quad u_3(X_1, X_2, h, t) = (\lambda(t) - 1)h, \tag{4.9}$$

where $\lambda(t)$ is the nominal axial stretch. We assume uniform axial strain across the thickness, so the kinematics simplify to:

$$u(X_1, X_2, X_3, t) = \bar{u}(X_1, X_2, t) + (\lambda(t) - 1)X_3 \; e_3, \tag{4.10}$$

where $\bar{u}(X_1, X_2, t)$ is the in-plane displacement. This allows us to reduce the problem to two spatial dimensions (see Appendix C.1 for details). The synthetic dataset D^{exp} consists of:

- The final in-plane displacement field $\bar{u}^{\text{exp}}(X_1, X_2, T)$, and
- The time history of the axial reaction force $f_R^{\text{exp}}(t)$.

While dynamic tests typically record only reaction forces, we assume additional full-field displacement data is available—similar to that obtained using digital image correlation (DIC) techniques applied before and after the experiment. This richer dataset enables us to impose stronger constraints on the learning process.

We simulate n dynamic tests with varied loading histories using a known J2 elasto-viscoplastic model. The material parameters used to generate the data are listed in Table 4.1. These simulations form the ground-truth dataset for training. To recover the RNO weights and biases P, we minimize the following objective:

$$\begin{aligned} \mathcal{O}(P) := \sum_{i=1}^{n} \Bigg(& \frac{\alpha_u}{L^4} \int_{\bar{\Omega}} |\bar{u}^i(X_1, X_2, T) - \bar{u}^{\text{exp},i}(X_1, X_2, T)|^2 \; d\bar{\Omega} \\ & + \frac{\alpha_f}{T_i \mu^2 L^4} \int_0^T |f_u^i - f^{\text{exp},i} R|^2 \; dt \Bigg), \\ f_u^i := & \int_{\bar{\Omega}} \sigma_{33} \; dS, \\ & \text{where } \{\bar{u}^i, \xi^i\} \text{ satisfy (4.2) for } i = 1, \ldots, n, \end{aligned} \tag{4.11}$$

where $\{\bar{u}^i, \xi^i\}$ are the state variables for the i-th simulation, evolved according to the neural constitutive model via the governing equations (4.2). The weights α_u and α_f balance the contributions from displacement and force data, and L, μ, and T_i are characteristic length, shear modulus, and time scales, respectively. The recovery can be performed on all parameters in the constitutive law 4.1 (on $P_{\mathbb{C}}, P_{RNO,\sigma}$ and $P_{RNO,\xi}$).

Synthetic Elasto-Viscoplastic Material Parameters								
	ρ (g/cm^3)	λ (GPa)	μ (GPa)	σ_y/μ	ε_0^p	n	$\dot{\varepsilon}_0^p$ (s^{-1})	m
P^{gen}	20	15.0	10.0	0.001	0.15	2.0	1.00×10^3	1.00

Table 4.1: The parameters used to generate the synthetic data. The constitutive model used is a J2-isotropic viscoplastic model. λ and μ are the lame parameters and σ_y is the yield stress. ε_0^p, $\dot{\varepsilon}_0^p$ are the reference plastic strain and strain rate, respectively. n, m represent the exponents of strain hardening and rate hardening, respectively.

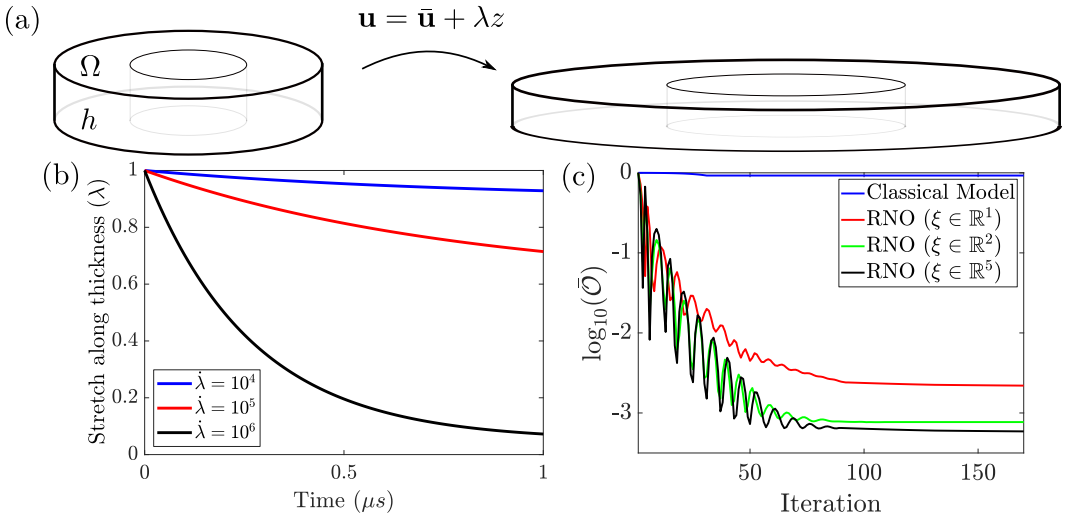


Figure 4.1: Results for the dynamic compression of circular annulus. (a) Geometry and deformation. (b) Applied stretch along the annulus thickness (c) Normalized objective vs iteration number.

4.3.1 Learning using compression data

We apply our method to a specimen whose cross-section is a circular annulus shown in Figure 4.1(a). We generate synthetic data for $n = 3$ tests with imposed thickness strains shown in Figure 4.1(b) that correspond to initial strain rate in the range $\dot{\varepsilon} = 10^4 - 10^6$ s^{-1} with the parameters P^{gen}. α_u and α_f are chosen such that the contribution of both parts of the objective are approximately equal, $\alpha_f/\alpha_u = 10$. In general, we have noticed that the RNO architecture is more efficient in learning the force data rather than the displacement data, which is why we fix α_f to be greater than α_u. This is because the experimental setting does not offer rich displacement data set for the RNO to learn. The forces arising from compression of thin structures offer more in-

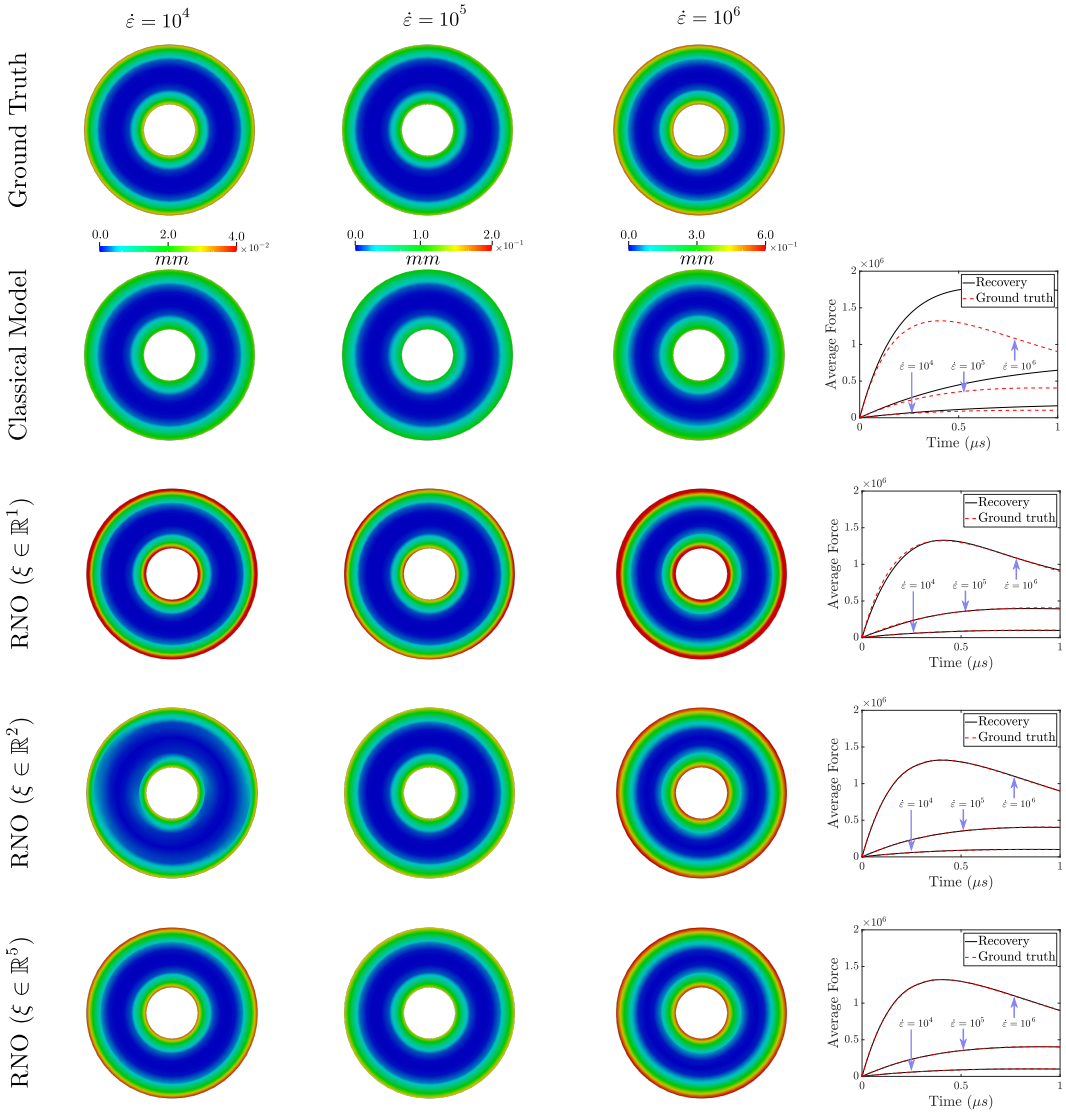


Figure 4.2: Force of compression versus time and displacements

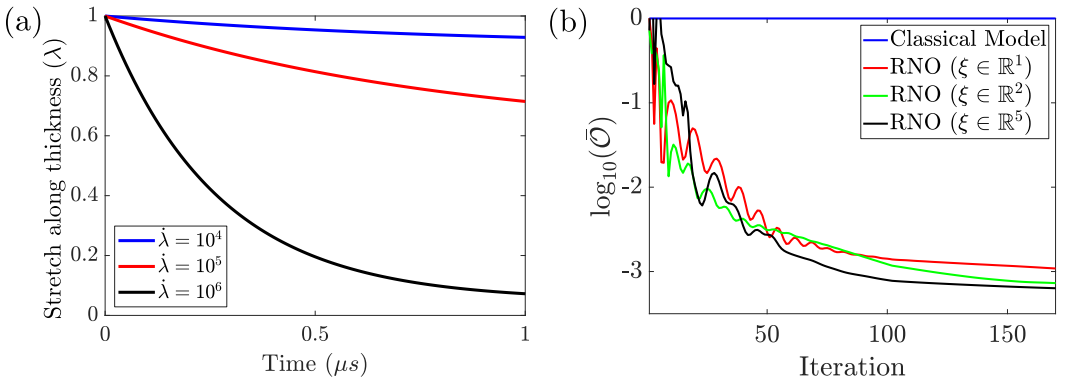


Figure 4.3: Results for the dynamic compression of flower hole configuration. (a) Applied stretch along the annulus thickness (c) Normalized objective vs iteration number.

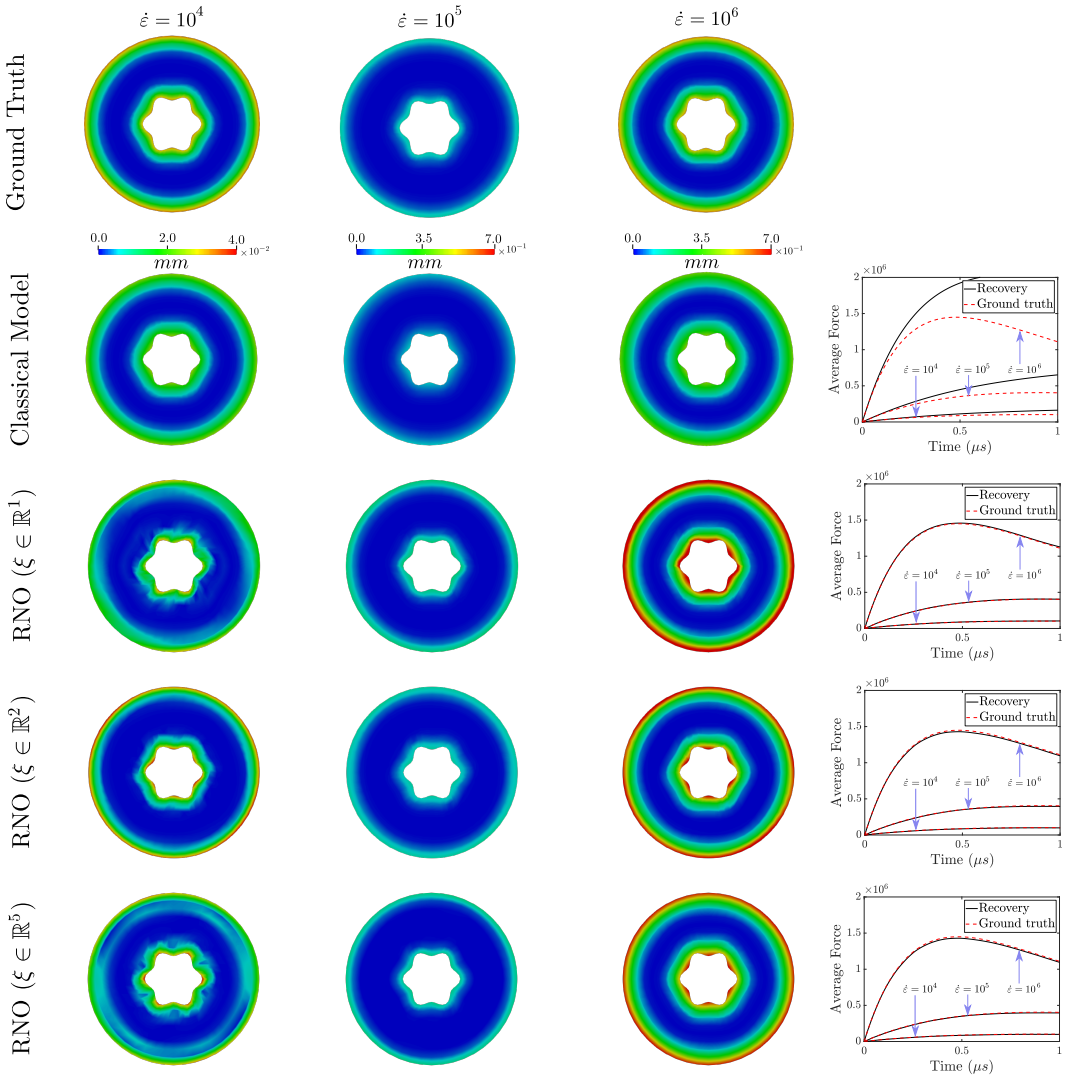


Figure 4.4: Force of compression versus time and displacements

formation, facilitating easier tuning of the weights and biases [1]. Figure 4.1(c) shows the evolution of the objective with iteration. We compare the inversion for three cases: (i) a linear model (classical model, optimizing over $P_{\mathbb{C}}$) , (ii) an RNO model with a single internal variable (RNO $\xi \in \mathbb{R}^1$, optimizing over $P_{\text{RNO},\sigma}, P_{\text{RNO},\xi}$), (iii) an RNO model with 2 internal variables (RNO $\xi \in \mathbb{R}^2$) and (iv) an RNO model with 5 internal variables (RNO $\xi \in \mathbb{R}^5$). Clearly, the RNO models outperform the classical linear constitutive law and show signs of better inversion with higher dimension of ξ. In classical plasticity theory, it is often assumed that five internal variables are needed to describe the full range

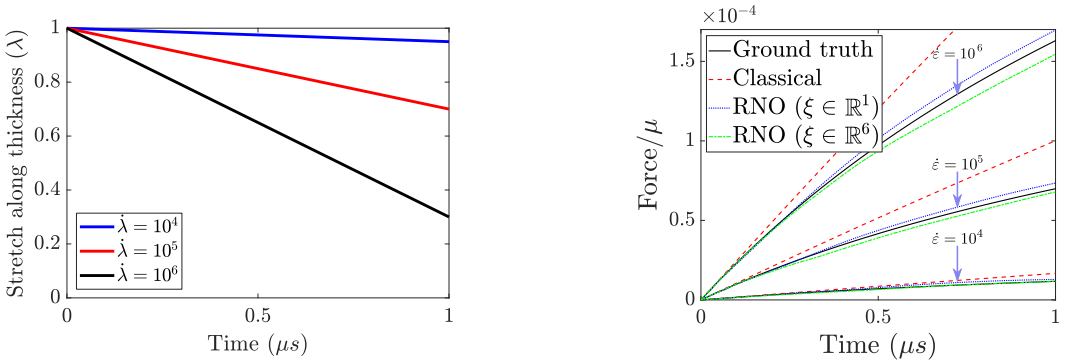
[1] We observed similar phenomenon in our previous study [1]. A configuration with a higher surface area produces higher compressive forces, enabling a better recovery.

of inelastic behavior in two-dimensional materials. This assumption guides our choice to test an RNO model with $\xi = 5$. However, our results reveal that such a high-dimensional internal state may not be strictly necessary for accurate modeling. Specifically, we observe that a model with only $\xi = 2$ internal variables is already capable of reproducing the experimental force and displacement data with high fidelity. This suggests a form of degeneracy in the optimization problem, where lower-dimensional representations can still effectively capture the essential features of the material response. More results with other dimensions of ξ can be found in appendix C.3. The objective decreases rapidly at first and then stabilizes at a value with a factor 10^3 than the initial value. Despite the fact that the objective is very small, the experimental observations agree well, as shown in Figures 4.2. Figure 4.2 compares the experimental reaction force with the one simulated with the recovered parameters, and we observe excellent agreement. We also show the magnitude of displacement seen for all the three models at the end time compared to the displacements of the generated data. Again, we see good agreement.

We repeat the inversion for a specimen with a flower-shaped hole, shown in figures 4.3 and 4.4, using the same set of strain rate conditions and weighting parameters. The results closely mirror those of the circular case. The RNO models outperform the classical linear model, with improved accuracy for higher-dimensional internal states. The objective decreases rapidly and stabilizes at a low value, and the simulated forces and displacements again show excellent agreement with the generated data.

4.3.2 Validation

To validate the performance and generalizability of the RNO models, we conduct a compression test on a square specimen without any internal holes or geometric complexities. This test serves as an out-of-distribution validation case, as it differs from the configurations used during training. Specifically, the loading profiles in this test are linear in time, in contrast to the exponential loading conditions used to generate the training data. This allows us to evaluate how well the RNO models extrapolate to loading paths not seen during training. The validation results are shown in Figure 4.5. The stretch along the thickness direction as a function of time is presented in Figure 4.5a, while the corresponding compression force response is shown in Figure 4.5b. In both



(a) Linear stretch profiles for the compression of square configuration for validation.

(b) Compression force versus time for validation.

Figure 4.5: Validation of the recovered models

measures, the RNO-based models significantly outperform the classical linear constitutive model. In particular, they are able to closely track the ground truth compression force, while the classical model exhibits large deviations. These results demonstrate that the RNO models are capable of capturing the underlying material behavior more accurately, even under loading protocols that were not explicitly part of the training dataset.

While the RNO models perform well within the scope of this experiment, it is important to acknowledge their limitations. In particular, neural networks are known to struggle with extrapolation beyond the domain represented in the training data. This limitation also applies to our RNO models: although they generalize across variations in geometry and loading profiles within the same experimental regime (i.e., compression of thin structures), they do not reliably predict material response in fundamentally different settings. For example, if we query the trained RNO with a strain trajectory outside the space of compressive behavior—such as torsion, bending, or multiaxial tension—the predictions become unstable or unphysical. This is not a shortcoming of our specific model, but a general limitation of data-driven neural networks. A common approach to improve generalization is to initialize the neural network using pre-training on a known constitutive model. In this context, the RNO architecture can be pretrained on stress-strain pairs generated from simulations using classical constitutive laws (e.g., hyperelasticity or viscoplasticity). Such a strategy would embed physical priors into the model from the outset and reduce the burden on the data. We propose exploring this hybrid strategy

in future work as a means to extend the utility of RNO models beyond their current scope.

4.4 Conclusion

Conventional methods for material characterization rely on converting measurable experimental data into stress-strain pairs, allowing the recovery of constitutive parameters. To facilitate this conversion, experiments are typically designed under highly controlled and homogeneous conditions where traction and displacement fields can be cleanly mapped to stress and strain. In prior work, we developed a framework that bypasses this intermediate step, recovering constitutive parameters directly from raw measurable data. This is achieved by posing the problem as a PDE-constrained optimization, where sensitivities are efficiently computed via the adjoint method. However, a key limitation of that approach is the need to assume a specific functional form for the constitutive law. With the rapid expansion of material systems and the increasing complexity of material behavior, prescribing such a form a priori has become increasingly difficult. Moreover, classical constitutive models are often phenomenological, lack a foundation in physics, and require significant domain expertise to define.

To address these limitations, we extend our earlier framework to incorporate a neural network-based constitutive law. Specifically, we adopt a recurrent neural operator (RNO) architecture due to its demonstrated ability to represent history-dependent behavior in a compact and generalizable manner. We validate this framework using synthetically generated data of dynamic compression on a thin annular specimen. Across a range of test cases, the RNO significantly outperforms classical linear models in recovering the underlying material response. A limitation of the present work is that the trained RNO model is specialized to the experimental configuration on which it was trained—compression of thin structures—and does not generalize to markedly different loading or geometrical scenarios. Future efforts could address this by pretraining the neural architecture on synthetic data generated from known constitutive models, thereby embedding physical priors and extending the utility of the framework beyond the original training distribution.

References

[1] A. Akerson, A. Rajan, and K. Bhattacharya. "Learning constitutive relations from experiments: 1. PDE constrained optimization." In: *arXiv preprint arXiv:2412.02864* (2024).

[2] K. Bhattacharya, L. Cao, G. Stepaniants, A. Stuart, and M. Trautner. "Learning memory and material dependent constitutive laws." In: *arXiv preprint arXiv:2502.05463* (2025).

[3] K. Bhattacharya, B. Liu, A. Stuart, and M. Trautner. "Learning Markovian homogenized models in viscoelasticity." In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 641–679.

[4] E. Billington, A. Tate, and W. Williams. *The Physics of Deformation and Flow.* 1982.

[5] R. Car and M. Parrinello. "Unified approach for molecular dynamics and density-functional theory." In: *Physical review letters* 55.22 (1985), p. 2471.

[6] D. Casem and E. Retzlaff. "A Kolsky bar for high-rate indentation." In: *Journal of Dynamic Behavior of Materials* 9.3 (2023), pp. 300–314.

[7] G. Chen. "Recurrent neural networks (RNNs) learn the constitutive law of viscoelasticity." In: *Computational Mechanics* 67.3 (2021), pp. 1009–1019.

[8] A. Dekhovich, M. H. Sluiter, D. M. Tax, and M. A. Bessa. "iPINNs: Incremental learning for Physics-informed neural networks." In: *Engineering with Computers* (2024), pp. 1–14.

[9] J. Dornheim, L. Morand, H. J. Nallani, and D. Helm. "Neural networks for constitutive modeling: From universal function approximators to advanced models and the integration of physics." In: *Archives of computational methods in engineering* 31.2 (2024), pp. 1097–1127.

[10] A. Eghtesad, J. Tan, J. N. Fuhg, and N. Bouklas. "NN-EVP: A physics informed neural network-based elasto-viscoplastic framework for predictions of grain size-aware flow response." In: *International Journal of Plasticity* 181 (2024), p. 104072.

[11] F. Feyel and J.-L. Chaboche. "Multi-scale non-linear FE2 analysis of composite structures: damage and fiber size effects." In: *Revue européenne des éléments finis* 10.2-4 (2001), pp. 449–472.

[12] M. E. Gurtin. *An introduction to continuum mechanics.* Vol. 158. Academic Press, 1982.

[13] C. M. Hamel, K. N. Long, and S. L. Kramer. "Calibrating constitutive models with full-field data via physics informed neural networks." In: *Strain* 59.2 (2023), e12431.

[14] A. G. Holzapfel. *Nonlinear solid mechanics II.* John Wiley & Sons, Inc., 2000.

[15] A. Hussain, A. H. Sakhaei, and M. Shafiee. "Machine learning-based constitutive modelling for material non-linearity: A review." In: *Mechanics of Advanced Materials and Structures* (2024), pp. 1–19.

[16] H. B. Khaniki, M. H. Ghayesh, R. Chin, and M. Amabili. "Hyperelastic structures: A review on the mechanics and biomechanics." In: *International Journal of Non-Linear Mechanics* 148 (2023), p. 104275.

[17] B. Koeppel and G. Subhash. "An experimental technique to investigate the dynamic indentation hardness of materials." In: *Experimental Techniques* 21.3 (1997), pp. 16–18.

[18] B. J. Koeppel. *Dynamic indentation hardness of materials.* Michigan Technological University, 1997.

[19] N. Kovachki, B. Liu, X. Sun, H. Zhou, K. Bhattacharya, M. Ortiz, and A. Stuart. "Multiscale modeling of materials: Computing, data science, uncertainty and goal-oriented optimization." In: *Mechanics of Materials* 165 (2022), p. 104156.

[20] D.-w. Li, J.-x. Liu, Y.-s. Fan, X.-g. Yang, and W.-q. Huang. "A preliminary discussion about the application of machine learning in the field of constitutive modeling focusing on alloys." In: *Journal of Alloys and Compounds* 976 (2024), p. 173210.

[21] B. Liu, E. Ocegueda, M. Trautner, A. M. Stuart, and K. Bhattacharya. "Learning macroscopic internal variables and history dependence from microscopic models." In: *Journal of the Mechanics and Physics of Solids* 178 (2023), p. 105329.

[22] S. K. Melly, L. Liu, Y. Liu, and J. Leng. "A review on material models for isotropic hyperelasticity." In: *International Journal of Mechanical System Dynamics* 1.1 (2021), pp. 71–88.

[23] R. G. Nascimento and F. A. Viana. "Cumulative damage modeling with recurrent neural networks." In: *AIAA Journal* 58.12 (2020), pp. 5459–5471.

[24] M. Nilsson. *Dynamic hardness testing using a split-Hopkinson pressure bar apparatus, methodology report, Swedish deference Research Agency.* Tech. rep. FOI, 2002.

[25] E. Ocegueda and K. Bhattacharya. "Interaction between deformation twinning and dislocation slip in polycrystalline solids." In: *arXiv preprint arXiv:2202.02908* (2022).

[26] R.-E. Plessix. "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications." In: *Geophysical Journal International* 167.2 (2006), pp. 495–503.

[27] B. Si, Z. Li, Y. Yang, L. Qiao, E. Liu, G. Xiao, and X. Shu. "Dynamic indentation testing and characterization of metals based on the split Hopkinson pressure bar (SHPB) device." In: *Mechanics of Materials* 177 (2023), p. 104550.

[28] E. B. Tadmor, M. Ortiz, and R. Phillips. "Quasicontinuum analysis of defects in solids." In: *Philosophical magazine A* 73.6 (1996), pp. 1529–1563.

[29] K. Upadhyay, J. N. Fuhg, N. Bouklas, and K. Ramesh. "Physics-informed data-driven discovery of constitutive models with application to strain-rate-sensitive soft materials." In: *Computational Mechanics* (2024), pp. 1–30.

[30] E. Van Der Giessen, P. A. Schultz, N. Bertin, V. V. Bulatov, W. Cai, G. Csányi, S. M. Foiles, M. G. Geers, C. González, M. Hütter, et al. "Roadmap on multiscale materials modeling." In: *Modelling and Simulation in Materials Science and Engineering* 28.4 (2020), p. 043001.

[31] E. Weinan. *Principles of multiscale modeling*. Cambridge University Press, 2011.

[32] A. Wihardja and K. Bhattacharya. "Constitutive relations from images." In: *arXiv preprint arXiv:2504.00898* (2025).

[33] H. Zhou and K. Bhattacharya. "Accelerated computational micromechanics and its application to polydomain liquid crystal elastomers." In: *Journal of the Mechanics and Physics of Solids* 153 (2021), p. 104470.

Chapter 5

LEARNING HOMOGENIZATION FOR ELLIPTIC OPERATORS

Bhattacharya, K., Kovachki, N.B., Rajan, A., Stuart, A.M. and Trautner, M., 2024. "Learning homogenization for elliptic operators." *SIAM Journal on Numerical Analysis*, 62(4), pp.1844-1873.

Contributions: A.R. developed the code for a portion of the numerical results. N.B.K. and M.T. took the lead in the mathematical proofs, developed the code, performed simuations. A.M.S. and K.B. supervised the mathematical proofs. N.B.K., M.T., A.M.S., and K.B. wrote the paper

---

5.1 Introduction

Homogenization theory is a well-established methodology that aims to eliminate fast-scale dependence in partial differential equations (PDEs) to obtain homogenized PDEs which produce a good approximate solution of the problem with fast scales while being more computationally tractable. In continuum mechanics, this methodology is of great practical importance as the constitutive laws derived from physical principles are governed by material behavior at small scales, but the quantities of interest are often relevant on larger scales. These homogenized constitutive laws often do not have a closed analytic form and may have new features not present in the microscale laws. Consequently, there has been a recent surge of interest in employing data-driven methods to learn homogenized constitutive laws.

The goal of this paper is to study the learnability of homogenized constitutive laws in the context of one of the canonical model problems of homogenization, namely the divergence form elliptic PDE. One significant challenge in applications of homogenization in material science arises from the presence of discontinuities and corner interfaces in the underlying material. This leads to a lack of smoothness in the coefficients and solutions of the associated equations,

a phenomenon extensively studied in numerical methods for PDEs. Addressing this challenge in the context of learning remains largely unexplored and is the focus of our work. We develop underlying theory and provide accompanying numerical studies to address learnability in this context.

In Subsection 5.1.1 we establish the mathematical framework and notation for the problem of interest, state the three main contributions of the paper, and overview the contents of each section of the paper. In Subsection 5.1.2 we provide a detailed literature review. Subsection 5.1.3 states the stability estimates that are key for the approximation theory developed in the paper and discusses the remainder of the paper in the context of these estimates.

5.1.1 Problem formulation

Consider the following linear multiscale elliptic equation on a bounded domain $\Omega \subset \mathbb{R}^d$:

$$-\nabla_x \cdot (A^\epsilon \nabla_x u^\epsilon) = f \quad x \in \Omega, \tag{5.1a}$$

$$u^\epsilon = 0 \quad x \in \partial\Omega. \tag{5.1b}$$

Here $A^\epsilon(x) = A\left(\frac{x}{\epsilon}\right)$ for $A(\cdot)$ which is 1-periodic and positive definite: $A : \mathbb{T}^d \to \mathbb{R}^{d\times d}_{\text{sym},\succ 0}$, a condition which holds throughout this work. Assume further that $f \in L^2(\Omega;\mathbb{R})$ and has no microscale variation with respect to x/ϵ.

Our focus is on linking this multiscale problem to the homogenized form of equation (5.1), which is

$$-\nabla_x \cdot \left(\overline{A}\nabla_x u\right) = f \quad x \in \Omega, \tag{5.2a}$$

$$u = 0 \quad x \in \partial\Omega, \tag{5.2b}$$

where $\overline{A}$ is given by

$$\overline{A} = \int_{\mathbb{T}^d} \left(A(y) + A(y)\nabla\chi(y)^T\right) \, \mathtt{d}y, \tag{5.3}$$

and $\chi : \mathbb{T}^d \to \mathbb{R}^d$ solves the cell problem

$$-\nabla \cdot (\nabla\chi A) = \nabla \cdot A, \quad \chi \text{ is 1-periodic.} \tag{5.4}$$

All of the preceding PDEs are to be interpreted as holding in the weak sense. For $0 < \epsilon \ll 1$, the solution u^ϵ of (5.1) is approximated by the solution u of

(5.2), and the error converges to zero as $\epsilon \to 0$ in various topologies [4, 8, 50]. We assume that

$$\|A\|_{L^\infty} := \sup_{y \in \mathbb{T}^d} |A(y)|_F < \infty$$

where $|\cdot|_F$ is the Frobenius norm. Hence $A \in L^\infty\left(\mathbb{T}^d; \mathbb{R}^{d\times d}\right)$ and $A^\epsilon \in L^\infty\left(\Omega; \mathbb{R}^{d\times d}\right)$. Similarly, for $A \in L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$, we define

$$\|A\|_{L^2}^2 := \int_{\mathbb{T}^d} |A(y)|_F^2 \, \mathsf{d}y.$$

Also, for given $\beta \geq \alpha > 0$, we define the following subset of 1-periodic, positive-definite, symmetric matrix fields in $L^\infty\left(\mathbb{T}^d; \mathbb{R}^{d\times d}\right)$ by

$$\mathsf{PD}_{\alpha,\beta} = \{A \in L^\infty(\mathbb{T}^d; \mathbb{R}^{d\times d}) : \ \forall (y,\xi) \in \mathbb{T}^d \times \mathbb{R}^d, \ \alpha|\xi|^2 \leq \langle \xi, A(y)\xi\rangle \leq \beta|\xi|^2\}.$$

For open set $\Omega \subset \mathbb{R}^d$, we denote the *variation* of a function $u \in L^1_{\text{loc}}(\Omega)$ by

$$V(u,\Omega) = \sup\left\{ \sum_{i=1}^d \int_\Omega \frac{\partial \Phi_i}{\partial x_i} u \, \mathsf{d}x : \ \Phi \in C_0^\infty(\Omega; \mathbb{R}^d), \ \|\Phi\|_{L^\infty(\Omega;\mathbb{R}^d)} \leq 1 \right\}$$

and the set of functions of bounded variation on $\mathbb{T}^d$ as

$$\mathsf{BV} = \{u \in L^1(\mathbb{T}^d) : \ V(u, \mathbb{T}^d) < \infty\}.$$

For further information on BV, we refer to [32]. Finally, we often work in the Sobolev space H^1 restricted to spatially mean-zero periodic functions, denoted

$$\dot{H}^1 := \left\{ v \in W^{1,2}(\mathbb{T}^d) \;\middle|\; v \text{ is 1-periodic}, \int_{\mathbb{T}^d} v \ \mathsf{d}y = 0 \right\};$$

the norm on this space is defined by

$$\|g\|_{\dot{H}^1} := \|\nabla g\|_{L^2}. \tag{5.5}$$

Numerically solving (5.1) is far more computationally expensive than solving the homogenized equation (5.2), motivating the wish to find the homogenized coefficient $\overline{A}$ defining equation (5.2). The difficult part of obtaining the equation (5.2) is solving the cell problem (5.4). Although explicit solutions exist in the one-dimensional setting for piecewise constant A [7] and in the two-dimensional setting where A is a layered material [50], in general a closed form solution is not available and the cell problem must be solved numerically. Note that in general the action of the divergence $\nabla\cdot$ on terms involving A in the cell problem necessitates the use of weak solutions for $A \notin C^1(\mathbb{T}^d, \mathbb{R}^{d\times d})$;

this is a commonly occurring situation in applications such as those arising from porous medium flow, or to vector-valued generalizations of the setting here to elasticity, rendering the numerical solution non-trivial. For this reason, it is potentially valuable to approximate the solution map

$$G: \ A \mapsto \chi, \tag{5.6}$$

defined by the cell problem, using a map defined by a neural operator. More generally it is foundational to the broader program of learning homogenized constitutive models from data to thoroughly study this issue for the divergence form elliptic equation as the insights gained will be important for understanding the learning of more complex parameterized homogenized models, such as those arising in nonlinear elasticity, viscoelasticity, and plasticity.

The full map from A to the homogenized tensor $\overline{A}$ is expressed by $A \mapsto (\chi, A) \mapsto \overline{A}$, and one could instead learn the map

$$F : A \mapsto \overline{A}. \tag{5.7}$$

Since the map $(\chi, A) \mapsto \overline{A}$ is is defined by a quadrature, we focus on the approximation of $A \mapsto \chi$ and state equivalent results for the map $A \mapsto \overline{A}$ that emerge as consequences of the approximation of χ. In this paper we make the following contributions:

1. We state and prove universal approximation theorems for the map G defined by (5.4) and (5.6), and map F defined by (5.3), (5.4), and (5.7).
2. We provide explicit examples of microstructures which satisfy the hypotheses of our theorems; these include microstructures generated by probability measures which generate discontinuous functions in BV.
3. We provide numerical experiments to demonstrate the ability of neural operators to approximate the solution map on four different classes of material parameters A, all covered by our theoretical setting.

In Subsection 5.1.2 we provide an overview of the literature, followed in Subsection 5.1.3 by a discussion of stability estimates for (5.4), with respect to variations in A; these are at the heart of the analysis of universal approximation. The main body of the text then commences with Section 5.2, which

characterizes the microstructures of interest to us in the context of continuum mechanics. Section 5.3 states universal approximation theorems for $G(\cdot)$ and $F(\cdot)$, using the Fourier neural operator. In Section 5.4 we give numerical experiments illustrating the approximation of the map G defined by (5.6) on microstructures of interest in continuum mechanics. Details of the stability estimates, the proofs of universal approximation theorems, properties of the microstructures, and details of numerical experiments are given in Appendices D.1, D.2, D.3, and D.4, respectively.

5.1.2 Literature review

Homogenization aims to derive macroscopic equations that describe the effective properties and behavior of solutions to problems at larger scales given a system that exhibits behaviour at (possibly multiple) smaller scales. Although it is developed for the various cases of random, statistically stationary, and periodic small-scale structures, we work here entirely in the periodic setting. The underlying assumption of periodic homogenization theory is that the coefficient is periodic in the small-scale variable, and that the scale separation is large compared to the macroscopic scales of interest. Convergence of the solution of the multiscale problem to the homogenized solution is well-studied; see [1, 12]. We refer to the texts [4, 8, 50] for more comprehensive citations to the literature. Homogenization has found extensive application in the setting of continuum mechanics [18] where, for many multiscale materials, the scale-separation assumption is natural. In this work, we are motivated in part by learning constitutive models for solid materials, where crystalline microstructure renders the material parameters discontinuous and may include corner interfaces. This difficulty has been explored extensively in the context of numerical methods for PDEs, particularly with adaptive finite element methods [25, 9, 47, 49].

There is a significant body of work on the approximation theory associated with parametrically dependent solutions of PDEs, including viewing these solution as a map between the function space of the parameter and the function space of the solution, especially for problems possessing holomorphic regularity [13, 14, 11]. This work could potentially be used to study the cell problem for homogenization that is our focus here. However, there has been recent interest in taking a data-driven approach to solving PDEs via machine learning because

of its flexibility and ease of implementation. A particular approach to learning solutions to PDEs is operator learning, a machine learning methodology where the map to be learned is viewed as an operator acting between infinite-dimensional function spaces rather than between finite-dimensional spaces [6, 33, 40, 46, 29]. Determining whether, and then when, operator learning models have advantages over classical numerical methods in solving PDEs remains an active area of research [3]. The paper [43] makes a contribution to this area, in the context of the divergence form elliptic PDE and the map from coefficient to solution when the coefficient is analytic over its domain; the authors prove that ϵ error is achievable for a DeepONet [40] of size only polylogarithmic in ϵ, leveraging the exponential convergence of spectral collocation methods for boundary value problems with analytic solutions. However, in the setting of learning homogenized constitutive laws in material science, discontinuous coefficients form a natural focus and indeed form the focus of this paper. A few characteristics make operator learning a promising option in this context. First, machine learning has been groundbreaking in application settings with no clear underlying equations, such as computer vision and language models [21, 10]. In constitutive modeling, though the microscale constitutive laws are known, the homogenized equations are generally unknown and can incorporate dependencies that are not present on the microscale, such as history dependence, anisotropy, and slip-stick behavior [51, 5]. Thus, constitutive models lie in a partially equation-free setting where data-driven methods could be useful. Second, machine learned models as surrogates for expensive computation can be valuable when the cost of producing data and training the model can be amortized over many forward uses of the trained model. Since the same materials are often used for fabrication over long time periods, this can be a setting where the upfront cost of data production and model training is justified.

Other work has already begun to explore the use of data-driven methods for constitutive modeling; a general review of the problem and its challenges, in the context of constitutive modeling of composite materials, may be found in [38]. Several works use the popular framework of physics-informed machine learning to approach the problem [17, 52, 41, 19]. In [2], physical constraints are enforced on the network architecture while learning nonlinear elastic constitutive laws. In [34], the model is given access to additional problem-specific physical knowledge. Similarly, the work of [53] predicts the Cholesky factor of the tangent stiffness matrix from which the stress may be calculated; this

method enforces certain physical criteria. The paper [26] studies approximation error and uncertainty quantification for this learning problem. In [20], a derivative-free approach is taken to learning homogenized solutions where regularity of the material coefficient is assumed. The work of [35] illustrates the potential of operator learning methodology to model constitutive laws with history dependence, such as those that arise in crystal plasticity. Finally, a number of further works demonstrate empirically the potential of learning constitutive models, including [44, 39, 54, 37].

However, the underlying theory behind operator learning for constitutive models lags behind its empirical application. In [7], approximation theories are developed to justify the use of a recurrent Markovian architecture that performs well in application settings with history dependence. This architecture is further explored in [36] with more complex microstructures. Universal approximation results are a first step in developing theory for learning because they guarantee that there exists an ϵ-approximate operator within the operator approximation class, which is consistent with an assumed true model underlying the data [15, 31, 29, 28]. In addition to universal approximation, further insight may be gained by seeking to quantify the data or model size required to obtain a given level of accuracy; the papers [31, 28, 42] also contain work in this direction, as do the papers [22, 48], which build on the analysis developed in [13, 14, 11] referred to above. In our work we leverage an existing universal approximation theorem for Fourier neural operators (FNOs), a particular practically useful architecture from within the neural operator (NO) class [28]. We take two different approaches to proving approximation theorems based on separate PDE solution stability results in pursuit of a more robust understanding of the learning problem. Since the state of the field is in its infancy, it is valuable to have different approaches to these analysis problems. Finally, we perform numerical experiments on various microstructures to understand the practical effects of non-smooth PDE coefficients in learning solutions. We highlight the fact that in this paper we do not tackle issues related to the non-convex optimization problem at the heart of training neural networks; we simply use state of the art stochastic gradient descent for training, noting that theory explaining its excellent empirical behaviour is lacking.

Throughout this paper we focus on equation (5.1), which describes a conductivity equation in a heterogeneous medium; a natural generalization of interest

is to the constitutive law of linear elasticity, in which the solution is vector-valued and the coefficient is a fourth order tensor. Though it is a linear elliptic equation, we echo the sentiment of Blanc and Le Bris [8] with their warning "do not underestimate the difficulty of equation (5.1)." There are many effects to be understood in this setting, and resolving learning challenges is a key step towards understanding similar questions for the learning of parametric dependence in more complex homogenized constitutive laws where machine-learning may prove particularly useful.

5.1.3 Stability estimates

At the heart of universal approximation theorems is stability of the solution map (5.6); in particular continuity of the map for certain classes of A. In this subsection, we present three key stability results that are used to prove the approximation theorems in Section 5.3. The proofs of the following stability estimates may all be found in Appendix D.1.

A first strike at the stability of the solution map (5.6) is a modification of the classic L^∞/H^1 Lipschitz continuity result for dependence of the solution of elliptic PDEs on the coefficient; here generalization is necessary because the coefficient also appears on the right-hand side of the equation defining $G(\cdot)$:

Proposition 1. *Consider the cell problem defined by equation* (5.4)*. The following hold:*

1. *If $A \in \mathsf{PD}_{\alpha,\beta}$, then* (5.4) *has a unique solution $\chi \in \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ and*

$$\|\chi\|_{\dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)} \leq \frac{\sqrt{d}\beta}{\alpha}.$$

2. *For $\chi^{(1)}$ and $\chi^{(2)}$ solutions to the cell problem in equation* (5.4) *associated with coefficients $A^{(1)}, A^{(2)} \in \mathsf{PD}_{\alpha,\beta}$, respectively, it follows that*

$$\|\chi^{(2)} - \chi^{(1)}\|_{\dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)} \leq \frac{\sqrt{d}}{\alpha}\left(1 + \frac{\beta}{\alpha}\right)\|A^{(1)} - A^{(2)}\|_{L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})}. \quad (5.8)$$

However, this perturbation result is insufficient for approximation theory because the space L^∞ is not separable and it is not natural to develop approximation theory in such spaces [16, Chapter 9]. While it is possible to define the problem on a separable subspace of L^∞, see Lemma 1, such spaces are not

particularly useful in applications to micromechanics. Many natural models for realistic microstructures work with classes of discontinuous functions in which the boundary of material discontinuity can occur anywhere in the domain. Such functions cannot be contained in any separable subspace of L^∞; see Lemma 2. To deal with this issue it is desirable to establish continuity from L^q to $\dot{H}^1$ for some $q \in [2, \infty)$. To this end, we provide two additional stability results. The first stability result gives continuity, but not Lipschitz continuity, from L^2 to $\dot{H}^1$. The second stability result gives Lipschitz continuity from L^q to $\dot{H}^1$, some $q \in (2, \infty)$.

Proposition 2. *Endow $\mathsf{PD}_{\alpha,\beta}$ with the $L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$ induced topology and let $K \subset \mathsf{PD}_{\alpha,\beta}$ be a closed set. Define the mapping $G : K \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$ by $A \mapsto \chi$ as given by* (5.4)*. Then there exists a bounded continuous mapping*

$$\mathcal{G} \in C(L^2(\mathbb{T}^d;\ \mathbb{R}^{d\times d}); \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d))$$

such that $\mathcal{G}(A) = G(A)$ for any $A \in K$.

The preceding L^2 continuity proposition is used to prove the approximation results for the FNO in Theorems 7 and 8. While not necessary for the approximation theory proofs, the following proposition on Lipschitz continuity from L^q to $\dot{H}^1$ establishes a more concrete bound on the approximation error, which allows for additional analysis such as providing rough bounds on grid error as discussed in Subsection 5.4.3.

Proposition 3. *There exists $q_0 \in (2, \infty)$ such that, for all q satisfying $q \in (q_0, \infty]$, the following holds. Endow $\mathsf{PD}_{\alpha,\beta}$ with the $L^q(\mathbb{T}^d; \mathbb{R}^{d\times d})$ topology and let $K \subset \mathsf{PD}_{\alpha,\beta}$ be a closed set. Define the mapping $G : K \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$ by $A \mapsto \chi$ as given by* (5.4)*. Then there exists a bounded Lipschitz-continuous mapping*

$$\mathcal{G} :\ L^q(\mathbb{T}^d; \mathbb{R}^{d\times d}) \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$$

such that $\mathcal{G}(A) = G(A)$ for any $A \in K$.

Remark 4. Explicit upper bounds for q_0 in Proposition 3 exist and are discussed in Remark 12.

5.2 Microstructures

The main application area of this work is constitutive modeling. In this section we describe various classes of microstructures that our theory covers. In particular, we describe four classes of microstructures in two dimensions:

1. Smooth microstructures generated via truncated, rescaled log-normal random fields.
2. Discontinuous microstructures with smooth interfaces generated by Lipschitz star-shaped inclusions.
3. Discontinuous microstructures with square inclusions.
4. Voronoi crystal microstructures.

Visualizations of examples of these microstructures may be found in Figure 5.1. We emphasize that all four examples lead to functions in BV, a fact that we exploit in Section 5.4 when showing that our abstract analysis from Section 5.3 applies to them all.

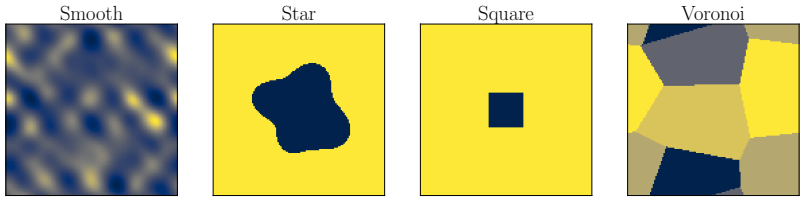


Figure 5.1: Microstructure Examples.

Smooth microstructures The smooth microstructures are generated by exponentiating a rescaled Gaussian random field. A is symmetric and coercive everywhere in the domain with a bounded eigenvalue ratio. Furthermore, the smooth function A and its derivatives are Lipschitz. Our theory is developed specifically to analyze non-smooth microstructures, so this example is used mainly as a point of comparison.

Star inclusions For the star inclusion microstructure, A is taken to be constant inside and outside the star-shaped boundary. The boundary function is smooth and Lipschitz in each of its derivatives. A is positive and coercive in

both regions with a bounded eigenvalue ratio. This microstructure introduces discontinuities, but the boundary remains smooth.

Square inclusions For the square inclusion microstructure, A is taken to be constant inside and outside the square boundary. Since we assume periodicity, without loss of generality the square inclusion is centered. The size of the square inclusion within the cell is varied between samples as are the constant values of A. This microstructure builds on the complexity of the star inclusion microstructure by adding corners to the inclusion boundary.

Voronoi interfaces The Voronoi crystal microstructures are generated by assuming a random Voronoi tessellation and letting A be piecewise-constant taking a single value on each Voronoi cell. The values of A on the cells and locations of the cell centers may be varied. This is the most complex microstructure among our examples and is a primary motivation for this work as Voronoi tessellations are a common model for crystal structure in materials.

5.3 Universal approximation results

In this section we state the two approximation theorems for learning solution operators to the cell problem. Theorem 7 concerns learning the map $A \to \chi$ in equation (5.4), and Theorem 8 concerns learning the map $A \to \overline{A}$ described by the combination of equations (5.4) and (5.3). Theorems 7 and 8 are specific to learning a Fourier neural operator (FNO), which is a subclass of the general neural operator. The proofs of the theorems in this section may be found in Appendix D.2.

5.3.1 Definitions of neural operators

First, we define a general neural operator (NO) and the Fourier neural operator (FNO). The definitions are largely taken from [29], and we refer to this work for a more in-depth understanding of these operators. In this work, we restrict the domain to the torus.

Definition 5 (**General Neural Operator**). *Let $\mathcal{A}$ and $\mathcal{U}$ be two Banach spaces of real vector-valued functions over domain $\mathbb{T}^d$. Assume input functions*

$a \in \mathcal{A}$ are $\mathbb{R}^{d_a}$-valued while the output functions $u \in \mathcal{U}$ are $\mathbb{R}^{d_u}$-valued. The neural operator architecture $\mathcal{G}_\theta : \mathcal{A} \to \mathcal{U}$ is

$$\begin{aligned}\mathcal{G}_\theta &= \mathcal{Q} \circ \mathsf{L}_{T-1} \circ \cdots \circ \mathsf{L}_0 \circ \mathcal{P}, \\ v_{t+1} &= \mathsf{L}_t v_t = \sigma_t(W_t v_t + \mathcal{K}_t v_t + b_t), \quad t = 0, 1, \ldots, T-1\end{aligned}$$

with $v_0 = \mathcal{P}(a)$, $u = \mathcal{Q}(v_T)$ and $\mathcal{G}_\theta(a) = u$. Here, $\mathcal{P} : \mathbb{R}^{d_a} \to \mathbb{R}^{d_{v_0}}$ is a local lifting map, $\mathcal{Q} : \mathbb{R}^{d_{v_T}} \to \mathbb{R}^{d_u}$ is a local projection map and the σ_t are fixed non-linear activation functions acting locally as maps $\mathbb{R}^{d_{v_{t+1}}} \to \mathbb{R}^{d_{v_{t+1}}}$ in each layer (with all of $\mathcal{P}$, $\mathcal{Q}$ and the σ_t viewed as operators acting pointwise, or point-wise almost everywhere, over the domain $\mathbb{T}^d$), $W_t \in \mathbb{R}^{d_{v_{t+1}} \times d_{v_t}}$ are matrices, $\mathcal{K}_t : \{v_t : \mathbb{T}^d \to \mathbb{R}^{d_{v_t}}\} \to \{v_{t+1} : \mathbb{T}^d \to \mathbb{R}^{d_{v_{t+1}}}\}$ are integral kernel operators and $b_t : \mathbb{T}^d \to \mathbb{R}^{d_{v_{t+1}}}$ are bias functions. For any $m \in \mathbb{N}_0$, the activation functions σ_t are restricted to the set of continuous $\mathbb{R} \to \mathbb{R}$ maps which make real-valued, feed-forward neural networks dense in $C^m(\mathbb{R}^d)$ on compact sets for any fixed network depth.[1] The integral kernel operators $\mathcal{K}_t$ are defined as

$$(\mathcal{K}_t v_t)(x) = \int_{\mathbb{T}^d} \kappa_t(x, y) v_t(y)\, dy$$

with standard multi-layered perceptrons (MLP) $\kappa_t : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{R}^{d_{v_{t+1}} \times d_{v_t}}$. We denote by θ the collection of parameters that specify $\mathcal{G}_\theta$, which include the weights W_t, biases b_t, parameters of the kernels κ_t, and the parameters describing the lifting and projection maps $\mathcal{P}$ and $\mathcal{Q}$ (usually also MLPs).

The FNO is a subclass of the NO.

Definition 6 (**Fourier Neural Operator**)**.** *The FNO inherits the structure and definition of the NO in Definition 5, together with some specific design choices. We fix $d_{v_t} = d_v$ for all t, where d_v is referred to as the number of channels, or model width, of the FNO. We fix $\sigma_t = \sigma$ to be a globally Lipschitz, non-polynomial, C^∞ function.[2] Finally, the kernel operators $\mathcal{K}_t$ are parameterized in the Fourier domain in the following manner. Let*

$$\psi_k(x) = e^{2\pi i \langle k, x \rangle}, \quad x \in \mathbb{T}^d,\ k \in \mathbb{Z}^d,$$

[1] We note that all globally Lipschitz, non-polynomial, $C^m(\mathbb{R})$ functions belong to this class.

[2] In this work in all numerical experiments we use the GeLU activation function as in [33].

denote the Fourier basis for $L^2(\mathbb{T}^d;\mathbb{C})$ *where* $i = \sqrt{-1}$ *is the imaginary unit. Then, for each* t*, the kernel operator* $\mathcal{K}_t$ *is parameterized by*

$$(\mathcal{K}_t v_t)_l(x) = \sum_{\substack{k\in\mathbb{Z}^d \\ |k|\le k_{\max}}} \left(\sum_{j=1}^{d_v} P_{lj}^k \langle (v_t)_j, \overline{\psi}_k \rangle_{L^2(\mathbb{T}^d;\mathbb{C})} \right) \psi_k(x).$$

Here, $l = 1, \dots, d_v$ *and each* $P^k \in \mathbb{C}^{d_v \times d_v}$ *constitute the learnable parameters of the integral operator.*

From the definition of the FNO, we note that parameterizing the kernels in the Fourier domain allows for efficient computation using the FFT. We refer to [29, 33] for additional details.

Finally we observe that in numerous applications, an example being learning of the map $A \mapsto \bar{A}$ (5.3), (5.4), it is desirable to modify the FNO so that the output space is simply a Euclidean space, and not a function space; this generalization is explored in [27]. An alternative approach, exemplified by Theorem 8 in the next subsection, is to allow the FNO output to be a function that may be evaluated at any point in the domain to yield an approximation of the point in Euclidean space.

5.3.2 Main theorems

These two theorems guarantee the existence of an FNO approximating the maps $A \mapsto \chi$ and $A \mapsto \overline{A}$ and are based on the stability estimate for continuity from $L^2 \to \dot{H}^1$ obtained in Proposition 2. Both theorems are proved in Appendix D.2.

Theorem 7. *Let* $K \subset \mathsf{PD}_{\alpha,\beta}$ *and define the mapping* $G : K \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ *by* $A \mapsto \chi$ *as given by* (5.4)*. Assume in addition that* K *is compact in* $L^2(\mathbb{T}^d;\mathbb{R}^{d\times d})$*. Then, for any* $\epsilon > 0$*, there exists an FNO* $\Psi : K \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ *such that*

$$\sup_{A\in K} \|G(A) - \Psi(A)\|_{\dot{H}^1} < \epsilon.$$

Theorem 8. *Let* $K \subset \mathsf{PD}_{\alpha,\beta}$ *and define the mapping* $F : K \to \mathbb{R}^{d\times d}$ *by* $A \mapsto \bar{A}$ *as given by* (5.3), (5.4)*. Assume in addition that* K *is compact in* $L^2(\mathbb{T}^d;\mathbb{R}^{d\times d})$*. Then, for any* $\epsilon > 0$*, there exists an FNO* $\Phi : K \to L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$ *such that*

$$\sup_{A\in K} \sup_{x\in\mathbb{T}^d} |F(A) - \Phi(A)(x)|_F < \epsilon.$$

The above approximation results can also be formulated to hold, on average, over any probability measure with a finite second moment that is supported on $\mathsf{PD}_{\alpha,\beta}$. In particular, if we let μ be such a probability measure then there exists an FNO or a neural operator Ψ such that

$$\mathbb{E}_{A\sim\mu}\|G(A) - \Psi(A)\|_{\dot{H}^1} < \epsilon. \tag{5.9}$$

This follows by applying Theorem 18 from [28] in the respective proofs instead of Theorem 5 from the same work. We do not carry out the full details here. While this allows approximation over the non-compact set $\mathsf{PD}_{\alpha,\beta}$, the error can only be controlled on average instead of uniformly. In Section 5.4, inputs are generated via probability measures supported on compact subsets of L^2; thus both the approximation Theorem 7, and its analog in the form (5.9), are relevant.

5.4 Numerical experiments

In this section, we show that it is possible to find good operator approximations of the homogenization map (5.6), defined by (5.4), in practice. We focus on use of the FNO and note that, while Theorems 7 and 8 assert the existence of desirable operator approximations, they are not constructive and do not come equipped with error estimates. We find approximations using standard empirical loss minimization techniques and, by means of numerical experiments, quantify the complexity with respect to volume of data and with respect to size of parametric approximation.

We work with the microstructures from Section 5.2. In this context we note that Theorems 7 and 8 apply. To demonstrate this it is necessary to establish that the subsets of coefficient functions employed are compact in L^2. We achieve this by noting that all our sets of coefficient functions are contained in $\mathsf{PD}_{\alpha,\beta} \cap \mathsf{BV}$, as defined in Subsection 5.1.1. Then we use Lemma 11 to establish compactness of these subsets of coefficient functions in L^2. The smooth microstructure example serves as a comparison case for examining the impact of discontinuous coefficients on the learning accuracy. The remaining three examples present different approximation theoretic challenges including curved boundaries (star inclusions), corners (square inclusions) and junctions of several domains (Voronoi).

The experiments are all conducted using an FNO with a fixed number $T = 4$ of hidden layers. The two remaining parameters to vary are the channel width d_v and the number of Fourier modes k_{max}. For implementation details, see Appendix D.4. We make the following observations based on the numerical experiments.

1. The effective $\overline{A}$ tensors computed from the model predicted solutions exhibit relative error under 1% for all examples; the effective $\overline{A}$ is computed from the learned cell problem solution χ using equation (5.3).

2. The error in the learned χ is significantly higher along discontinuous material boundaries and corner interfaces, as expected. However, the FNO operator approximation is able to approximate the solution with reasonable relative error even for the most complex case; this most complex case concerns the set of input functions with varying Voronoi geometry and varying microstructural properties within the domain.

3. In comparison with the smooth microstructure case, learning the map for the Voronoi microstructure requires substantially more data to avoid training a model which plateaus at a poor level of accuracy.

4. When compared with the smooth microstructure case, the error for the Voronoi microstructure decreases more slowly with respect to increasing model width, but shows more favourable response with respect to increasing the number of Fourier modes.

5. Models trained at one discretization may be evaluated at different discretizations for both the smooth and Voronoi microstructures as is characteristic of the FNO. The Voronoi microstructure exhibits, empirically, greater robustness to changes in discretization.

We first describe implementation details of each of the microstructures in Subsection 5.4.1. Then we show outcomes of the numerical experiments in Subsection 5.4.2, discussing them in Subsection 5.4.3.

5.4.1 Microstructure implementation

For each microstructure, two positive eigenvalues and three components of the two eigenvectors are randomly generated, and the final eigenvector component

is chosen to enforce symmetry. All eigenvalue ratios are at most e^2 by construction. In this manner, A is symmetric and coercive and has a bounded eigenvalue ratio.

Smooth Microstructures The smooth microstructures are generated by exponentiating a rescaled approximation of a Gaussian random field. The random field used to generate the eigenvalues and three eigenvector components of $A(x)$ is as follows:

$$\widehat{\lambda}_i(x) = \sum_{k_1,k_2=1}^{4} \xi^{(1)}_{k_1,k_2} \sin(2\pi k_1 x_1)\cos(2\pi k_2 x_2) + \xi^{(2)}_{k_1,k_2} \cos(2\pi k_1 x_1)\sin(2\pi k_2 x_2),$$

$$\lambda_i(x) = \exp\left(\frac{\widehat{\lambda}_i(x)}{\max_{x'\in[0,1]^2} |\widehat{\lambda}_i(x')|}\right),$$

where $\xi^{(j)}_{k_1,k_2}$ are i.i.d. normal Gaussian random variables.

Star-Shaped Inclusions The star-shaped inclusions are generated by defining a random Lipschitz polar boundary function as

$$r(\theta) = \mathsf{a} + \mathsf{b}\sum_{k=1}^{5} \xi_k \sin(k\theta)$$

where ξ_k are i.i.d. uniform random variables $U[-1,1]$, and a and b are constants that guarantee $0 < \epsilon < r < 0.5 - \epsilon$ for some fixed $\epsilon > 0$. Then $A(x)$ is constant inside and outside the boundary. We randomly sample eigenvalues for A on each domain via $\lambda_i \sim U[e^{-1}, e]$. The three components of the eigenvectors are i.i.d. normal random variables.

Square Inclusions The radius of the square is randomly generated via

$$r = \mathsf{a} + \mathsf{b}\zeta$$

where ζ is a uniform random variable on $[0,1]$ and a and b are positive constants that guarantee $0 < \epsilon < r < 0.5 - \epsilon$ for some fixed $\epsilon > 0$. The values of A on each of the constant domains are chosen in the same manner as in the star-shaped inclusion case.

Voronoi Interfaces The Voronoi crystal microstructure has constant A on each Voronoi cell and is chosen uniformly at random in the same manner as for the star inclusions. Voronoi tessellations are a common model for crystal structure in materials. In one Voronoi example, we fix the geometry for all data, and in a second Voronoi example we vary the geometry by randomly sampling five cell centers from a uniform distribution on the unit square.

5.4.2 Results

Each FNO model is trained using the empirical estimate of the mean squared H^1 norm:

$$\text{Loss}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\left(\|\chi^{(n)} - \widehat{\chi}^{(n)}\|_{L^2}^2 + \|\nabla\chi^{(n)} - \nabla\widehat{\chi}^{(n)}\|_{L^2}^2\right) \tag{5.10}$$

where n is the sample index, χ is the true solution, and $\widehat{\chi}$ is the FNO approximation of the solution parameterized by θ. In the analysis, we examine several different measures of error, including the following relative H^1 and relative $W^{1,10}$ errors.

$$\text{Relative } H^1 \text{ Error (RHE)} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\|\chi^{(n)} - \widehat{\chi}^{(n)}\|_{L^2}^2 + \|\nabla\chi^{(n)} - \nabla\widehat{\chi}^{(n)}\|_{L^2}^2}{\|\chi^{(n)}\|_{L^2}^2 + \|\nabla\chi^{(n)}\|_{L^2}^2}\right)^{\frac{1}{2}} \tag{5.11a}$$

$$\text{Relative } W^{1,10} \text{ Error (RWE)} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\|\chi^{(n)} - \widehat{\chi}^{(n)}\|_{L^{10}}^{10} + \|\nabla\chi^{(n)} - \nabla\widehat{\chi}^{(n)}\|_{L^{10}}^{10}}{\|\chi^{(n)}\|_{L^{10}}^{10} + \|\nabla\chi^{(n)}\|_{L^{10}}^{10}}\right)^{\frac{1}{10}} \tag{5.11b}$$

The $W^{1,10}$ norm gives a sense of the higher errors that occur at interfaces, corners and functions. We could have used $W^{1,p}$ for any p large enough.

Finally, we also look at error in $\overline{A}$, which we scale by the difference between the arithmetic and harmonic mean of A. Any effective $\overline{A}$ should have a norm in this range; these are known in mechanics as Voigt-Reuss bounds and have a physical interpretation as bounds obtained via energy principles by ignoring equilibrium for the upper bound (arithmetic mean) and ignoring compatibility for the lower bound (harmonic mean) [23]. The resulting error measure is given by

$$\text{Relative } \overline{A} \text{ Error (RAE)} \ = \frac{\|\overline{A} - \widehat{\overline{A}}\|_F}{a_m - a_h} \tag{5.12}$$

where the arithmetic mean a_m and harmonic mean a_h are given by

$$a_m = \left\| \int_{\mathbb{T}^2} A(x) \, \mathsf{d}x \right\|_F$$
$$a_h = \left\| \left(\int_{\mathbb{T}^2} A^{-1}(x) \, \mathsf{d}x \right)^{-1} \right\|_F .$$

We note that using $a_m - a_h$ rather than $\|\overline{A}\|_F$ as a scaling factor in equation (5.12) leads to a larger error value, so achieving low error in this measure of distance is harder.

We train models on five different datasets. Visualizations of the median-error test samples for each example may be viewed in Figure 5.2, and the numerical errors are shown in Figure 5.3. Each of these models is trained on 9500 data samples generated using an FE solver on a triangular mesh with the solution interpolated to a 128×128 grid. Additional model details may be found in Appendix D.4.

We perform an experiment to test the discretization-robustness of the FNO model, results of which are shown in Figure 5.4. The models are trained with data from the resolution 128×128 and evaluated on test data with different resolution. We emphasize that evaluating the FNO on different resolution is trivial in implementation by design.

We also investigate the effects of the number of training data and the model size on the error for the smooth and Voronoi microstructures; similar experiments, for different operator learning problems, are presented in [24]. A plot of error versus training data may be found in Figure 5.5, and plots of error versus the number of Fourier modes for fixed total model size, as measured by (model width) $\times$ (number Fourier modes), may be found in Figure 5.6. Figure 5.6 addresses the question of how to optimally distribute computational budget through different parameterizations to achieve minimum error at given cost as measured by number of parameters; it should be compared to similar experiments in [30].

5.4.3 Discussion

As can be seen from the data in Figure 5.3, the microstructures exhibiting discontinuities lead to higher model error than the smooth microstructure, and the introduction of corner interfaces leads to further increase in error. The

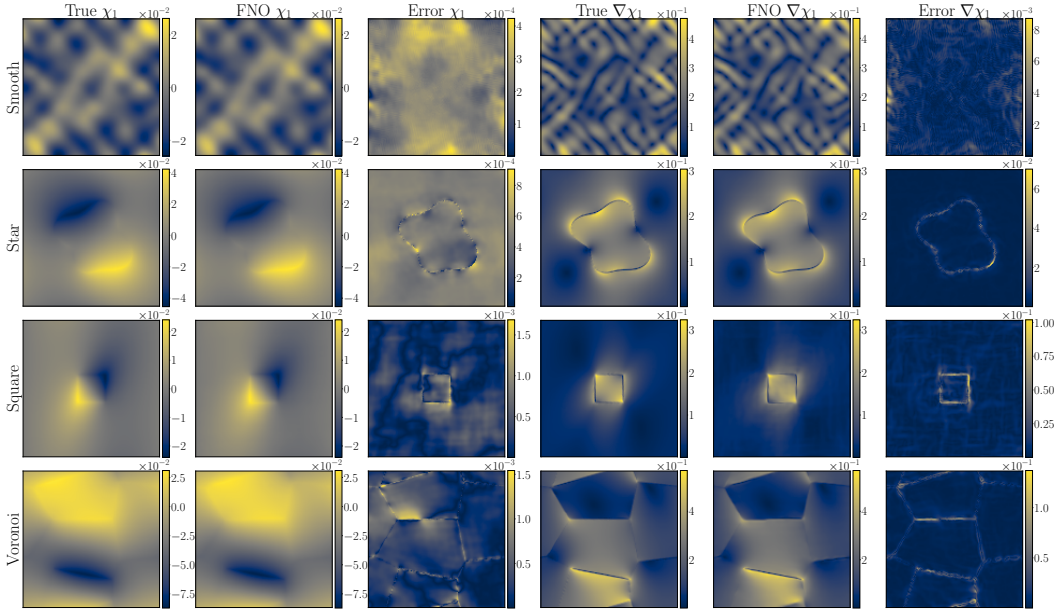


Figure 5.2: Visualization of the trained models evaluated on test samples that gave median relative H^1 error for each microstructure. The microstructure inputs of each row correspond to those of Figure 5.1. The first shows the true χ_1, the second shows the FNO predicted χ_1, and the third shows the absolute value of the error between the true and predicted χ_1. The fourth column shows the 2-norm of the gradient of the true χ_1, and the fifth shows the 2-norm of the gradient of the predicted χ_1. The last column shows the 2-norm of the difference between the two gradients.

Microstructure	Mean RHE	Mean RWE	Median RAE
Smooth	$0.0062 \pm 1 \cdot 10^{-4}$	$0.0091 \pm 1 \cdot 10^{-4}$	$0.0007 \pm 1 \cdot 10^{-5}$
Star	$0.0313 \pm 1 \cdot 10^{-4}$	$0.1318 \pm 5 \cdot 10^{-4}$	$0.0014 \pm 3 \cdot 10^{-5}$
Square	$0.1012 \pm 5 \cdot 10^{-4}$	$0.2741 \pm 2 \cdot 10^{-3}$	$0.0047 \pm 1 \cdot 10^{-4}$
Voronoi	$0.0565 \pm 4 \cdot 10^{-4}$	$0.2129 \pm 3 \cdot 10^{-3}$	$0.0027 \pm 8 \cdot 10^{-5}$
Voronoi (Fixed Geometry)	$0.0073 \pm 3 \cdot 10^{-5}$	$0.0140 \pm 3 \cdot 10^{-4}$	$0.0007 \pm 2 \cdot 10^{-5}$

Figure 5.3: Errors for each each numerical experiment; five sample models are trained for each microstructure. The expressions for the RHE (Relative H^1 Error), RWE (Relative $W^{1,10}$ Error) and RAE (Relative $\overline{A}$ Error) may be found in equations (5.11) and (5.12). The errors are evaluated over a test set of size 500. All examples have varying geometry except the second Voronoi example.

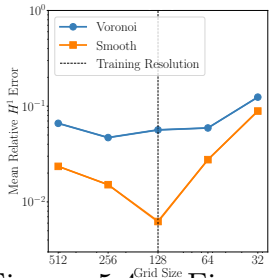


Figure 5.4: Five sample models trained on Smooth and Voronoi data at 128×128 grid resolution evaluated at different resolutions.

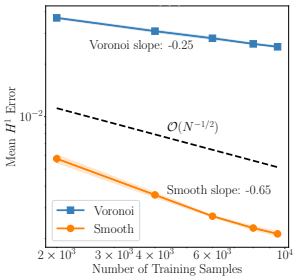


Figure 5.5: A comparison of test error for different amounts of training data for models trained on Voronoi and Smooth data. Five sample models are used for each data point.

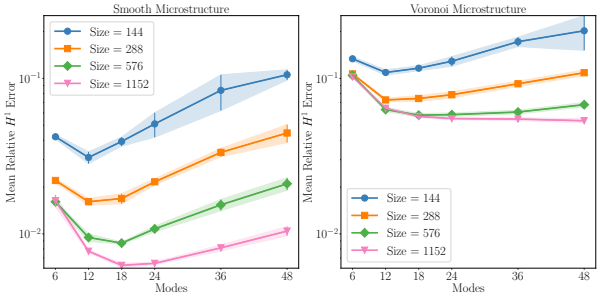


Figure 5.6: Relative H^1 error versus model size for the smooth and Voronoi examples with varying geometry. The number of Fourier modes in each direction and the model width were varied. Each line indicates a constant product of modes×width. Training data size was fixed at 9500 samples, and five samples were used for each data point.

visualizations of the median-error test samples in Figure 5.2 give some intuition; error is an order of magnitude higher along discontinuous boundaries; this is most apparent in the gradient. The true solution gradient often takes its most extreme values along the discontinuities, and the RWE gives an indication of how well the model captures the most extreme values in the solution. Unsurprisingly, this error is much higher than the RHE, but we note that it is confined to a small area of the domain along discontinuous boundaries and corner interfaces.

In the discretization-robustness experiment described in Figure 5.4, we observe that the Voronoi model exhibits greater robustness to changes in discretization. We hypothesize that, in the direction of decreasing resolution, the smaller error increase for the Voronoi model, in comparison with the smooth model, could be due to the piecewise-constant nature of the Voronoi microstructure on faces; improved resolution here does not help. On the other hand, for larger grid sizes, increased resolution on corners and discontinuities can help, which could explain the decrease in error from grid edge size of 128 to 256 for the Voronoi model while the smooth model increases in error. One could fine-tune the trained models with small amounts of data from different resolutions, but we leave this transfer learning exploration to future work.

We also examine the effect of the number of training data samples and the FNO size on model accuracy for the smooth and Voronoi microstructures. For

data size dependence, we observe in Figure 5.5 that for these two microstructures, the test error scales $\approx N^{-0.65}$ and $\approx N^{-0.25}$, respectively, where N is the number of training data. In theory, we do not expect to beat the Monte Carlo error decay of $\frac{1}{\sqrt{N}}$ [45]. We note that this is comparable to the behavior during training over 400 epochs; the test error for the smooth microstructure continues to decrease over the entire training periodic, but the test error for the Voronoi microstructure plateaus by around 100 epochs. The model size also presents a qualitatively different effect on error for the smooth and Voronoi microstructures. In Figure 5.6, we see the tradeoff between the number of Fourier modes and the model width for approximately constant model size, measured as the product of the width and number of modes. The Voronoi example benefits from additional Fourier modes, whereas the smooth example worsens. On the other hand, the smooth model benefits more from an increase in model width. We refer to [24, 30] for in-depth numerical studies of errors, choice of hyperparameters, and parameter distributions for FNO; here we highlight only the qualitative differences between the model behavior for different microstructures.

We also note that a significant portion of the model error may be attributed to grid ambiguity; with a 128×128 grid, the FNO does not know where between gridpoints a discontinuity may fall. This may be quantified empirically in the case of the square microstructure. We perform an experiment in which we create data of square microstructure inclusions whose boundary falls exactly on the gridpoints. One dataset treats the boundary as open, and the other treats the boundary as closed; the input grid points that fall on the boundary differ between the two datasets. We quantify grid ambiguity error by the difference in the outputs of a model given both the open square data and the closed square data. We find that the absolute H^1 norm of the difference between these two outputs is 0.041, which is slightly under twice the absolute H^1 norm of the output compared to the true solution, which has a value of 0.025. We hypothesize that the model learns to assume the boundary falls near the middle of the grid square, which explains why the output difference between the two datasets is roughly twice the true error. From a theory standpoint, one could bound the Lipschitz constant of the FNO and compare it to the Lipschitz constant of the true map described by Proposition 3. However, we leave the theoretical estimates of error rates to future work.

Finally, we compare the error in the effective $\overline{A}$ defined in (5.3). This error is scaled by a difference between the Frobenius norms of the arithmetic and harmonic means of the true A because the Frobenius norm of the true $\overline{A}$ should fall within that range. For this reason, in the case where the arithmetic and harmonic means are very close, as is frequently the case for the square and star inclusions, it is not valuable to learn the true $\overline{A}$. On the other hand, the varying-geometry Voronoi microstructure example on average has about 100 times greater difference between the arithmetic and harmonic means, in comparison with the star and square microstructure examples. This characteristic of the Voronoi microstructure further underscores the value of learning in this setting.

5.5 Conclusions

In this work, we establish approximation theory for learning the solution operator arising from the elliptic homogenization cell problem (5.4), viewed as a mapping from the coefficient to the solution; the theory allows for discontinuous coefficients. We also perform numerical experiments that validate the theory, explore qualitative differences between various microstructures, and quantify error/cost trade-offs in the approximation. We provide two different stability results for the underlying solutions that build understanding of the underlying map. These stability results, when combined with existing universal approximation results for neural operators, result in rigorous approximation theory for learning in this problem setting. On the empirical side we provide, and then study numerically, examples of various microstructures that satisfy the conditions of the approximation theory. We observe that model error is dominated by error along discontinuous and corner interfaces, and that discontinuous microstructures give rise to qualitatively different learning behavior. Finally, we remark that the learned effective properties are highly accurate, especially in the case of the Voronoi microstructure that we regard as the most complex. Since discontinuous microstructures arise naturally in solid mechanics, understanding learning behavior in this context is an important prerequisite for using machine learning for applications. In this area and others, numerous questions remain which address the rigor necessary for use of machine learning in scientific applications.

We have confined our studies to one of the canonical model problems of ho-

mogenization theory, the divergence form elliptic setting with periodic microstructure, to obtain deeper understanding of the learning constitutive laws. One interesting potential extension of this work is the setting in which the material coefficient A is not periodic but random with respect to the microstructure. Another is where it is only locally periodic and has dependence on the macroscale variable as well; thus $A^\epsilon = A(x, \frac{x}{\epsilon})$. In this case, the form of the cell problem (5.4) and homogenized coefficient (5.3) remain the same, but A and χ both have parametric dependence on x. The approximation theory and the empirical learning problem would grow in complexity in comparison to what is developed here, but the resulting methodology could be useful and foundational for understanding more complex constitutive models in which the force balance equation couples to other variables. Indeed, the need for efficient learning of constitutive models is particularly pressing in complex settings such as crystal plasticity. We anticipate that the potential use of machine learning to determine parametric dependence of constitutive models defined by homogenization will be for these more complex problems. The work described in this paper provides an underpinning conceptual approach, foundational analysis and set of numerical experiments that serve to underpin more applied work in this field.

References

[1] G. Allaire. "Homogenization and two-scale convergence." In: *SIAM Journal on Mathematical Analysis* 23.6 (1992), pp. 1482–1518.

[2] F. As' ad, P. Avery, and C. Farhat. "A mechanics-informed artificial neural network approach in data-driven constitutive modeling." In: *International Journal for Numerical Methods in Engineering* 123.12 (2022), pp. 2738–2759.

[3] P. Batlle, M. Darcy, B. Hosseini, and H. Owhadi. "Kernel methods are competitive for operator learning." In: *arXiv preprint arXiv:2304.13202* (2023).

[4] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures.* Vol. 374. American Mathematical Soc., 2011.

[5] K. Bhattacharya. "Phase boundary propagation in a heterogeneous body." In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 455.1982 (1999), pp. 757–766.

[6] K. Bhattacharya, B. Hosseini, N. B. Kovachki, and A. M. Stuart. "Model reduction and neural networks for parametric PDEs." In: *The SMAI Journal of Computational Mathematics* 7 (2021), pp. 121–157.

[7] K. Bhattacharya, B. Liu, A. Stuart, and M. Trautner. "Learning Markovian homogenized models in viscoelasticity." In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 641–679.

[8] X. Blanc and C. Le Bris. *Homogenization theory for multiscale problems: An introduction.* Vol. 21. Springer Nature, 2023.

[9] A. Bonito, R. A. DeVore, and R. H. Nochetto. "Adaptive finite element methods for elliptic problems with discontinuous coefficients." In: *SIAM Journal on Numerical Analysis* 51.6 (2013), pp. 3106–3134.

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. "Language models are few-shot learners." In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1877–1901.

[11] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab. "Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs." In: *ESAIM: Mathematical Modelling and Numerical Analysis* 47.1 (2013), pp. 253–280.

[12] D. Cioranescu and P. Donato. *An Introduction To Homogenization.* Vol. 17. Oxford University Press, 1999.

[13] A. Cohen, R. DeVore, and C. Schwab. "Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs." In: *Foundations of Computational Mathematics* 10.6 (2010), pp. 615–646.

[14] A. Cohen, R. Devore, and C. Schwab. "Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's." In: *Analysis and Applications* 9.01 (2011), pp. 11–47.

[15] G. Cybenko. "Approximation by superpositions of a sigmoidal function." In: *Mathematics of Control, Signals and Systems* 2.4 (1989), pp. 303–314.

[16] R. DeVore and G. Lorentz. *Constructive approximation.* Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1993.

[17] J. N. Fuhg and N. Bouklas. "On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling." In: *Computer Methods in Applied Mechanics and Engineering* 394 (2022).

[18] M. E. Gurtin. *An Introduction To Continuum Mechanics.* Academic press, 1982.

[19] E. Haghighat, S. Abouali, and R. Vaziri. "Constitutive model characterization and discovery using physics-informed deep learning." In: *Engineering Applications of Artificial Intelligence* 120 (2023), p. 105828.

[20] J. Han and Y. Lee. "A neural network approach for homogenization of multiscale problems." In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 716–734.

[21] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.

[22] L. Herrmann, C. Schwab, and J. Zech. "Neural and gpc operator surrogates: Construction and expression rate bounds." In: *arXiv preprint arXiv:2207.04950* (2022).

[23] R. Hill. "The elastic behaviour of a crystalline aggregate." In: *Proceedings of the Physical Society. Section A* 65.5 (1952), p. 349.

[24] M. V. de Hoop, Z. Huang, E. Qian, and A. M. Stuart. "The cost-accuracy trade-off in operator learning with neural networks." In: *Journal of Machine Learning* 1.3 (2022), pp. 299–341. ISSN: 2790-2048.

[25] T. Y. Hou and X.-H. Wu. "A multiscale finite element method for elliptic problems in composite materials and porous media." In: *Journal of Computational Physics* 134.1 (1997), pp. 169–189.

[26] D. Z. Huang, K. Xu, C. Farhat, and E. Darve. "Learning constitutive relations from indirect observations using deep neural networks." In: *Journal of Computational Physics* 416 (2020), p. 109491.

[27] D. Z. Huang, N. H. Nelsen, and M. Trautner. "An operator learning perspective on parameter-to-observable maps." In: *arXiv preprint arXiv:2402.06031* (2024).

[28] N. Kovachki, S. Lanthaler, and S. Mishra. "On universal approximation and error bounds for Fourier Neural Operators." In: *The Journal of Machine Learning Research* 22.290 (2021), pp. 1–76.

[29] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. "Neural operator: Learning maps between function spaces with applications to PDEs." In: *Journal of Machine Learning Research* 24.89 (2023), pp. 1–97.

[30] S. Lanthaler, Z. Li, and A. M. Stuart. "The nonlocal neural operator: Universal approximation." In: *arXiv preprint arXiv:2304.13221* (2023).

[31] S. Lanthaler, S. Mishra, and G. E. Karniadakis. "Error estimates for deeponets: A deep learning framework in infinite dimensions." In: *Transactions of Mathematics and Its Applications* 6.1 (2022).

[32] G. Leoni. *A first course in sobolev spaces.* American Mathematical Soc., 2017.

[33] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. "Fourier neural operator for parametric partial differential equations." In: *International Conference on Learning Representations* (2021).

[34] K. Linka, M. Hillgärtner, K. P. Abdolazizi, R. C. Aydin, M. Itskov, and C. J. Cyron. "Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning." In: *Journal of Computational Physics* 429 (2021), p. 110010.

[35] B. Liu, N. Kovachki, Z. Li, K. Azizzadenesheli, A. Anandkumar, A. M. Stuart, and K. Bhattacharya. "A learning-based multiscale method and its application to inelastic impact problems." In: *Journal of the Mechanics and Physics of Solids* 158 (2022).

[36] B. Liu, E. Ocegueda, M. Trautner, A. M. Stuart, and K. Bhattacharya. "Learning macroscopic internal variables and history dependence from microscopic models." In: *Journal of the Mechanics and Physics of Solids* (2023).

[37] X. Liu, F. Tao, and W. Yu. "A neural network enhanced system for learning nonlinear constitutive law and failure initiation criterion of composites using indirectly measurable data." In: *Composite Structures* 252 (2020), p. 112658.

[38] X. Liu, S. Tian, F. Tao, and W. Yu. "A review of artificial neural networks in the constitutive modeling of composite materials." In: *Composites Part B: Engineering* 224 (2021), p. 109152.

[39] H. J. Logarzo, G. Capuano, and J. J. Rimoli. "Smart constitutive laws: Inelastic homogenization through machine learning." In: *Computer Methods in Applied Mechanics and Engineering* 373 (2021), p. 113482.

[40] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators." In: *Nature Machine Intelligence* 3.3 (2021), pp. 218–229.

[41] P. Ma, P. Y. Chen, B. Deng, J. B. Tenenbaum, T. Du, C. Gan, and W. Matusik. "Learning neural constitutive laws from motion observations for generalizable PDE dynamics." In: *arXiv preprint arXiv:2304.14369* (2023).

[42] C. Marcati, J. A. Opschoor, P. C. Petersen, and C. Schwab. "Exponential relu neural network approximation rates for point and edge singularities." In: *Foundations of Computational Mathematics* (2022), pp. 1–85.

[43] C. Marcati and C. Schwab. "Exponential convergence of deep operator networks for elliptic partial differential equations." In: *SIAM Journal on Numerical Analysis* 61.3 (2023), pp. 1513–1545.

[44] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, J. Cao, and M. Bessa. "Deep learning predicts path-dependent plasticity." In: *Proceedings of the National Academy of Sciences* 116.52 (2019), pp. 26414–26420.

[45] K. P. Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.

[46] N. H. Nelsen and A. M. Stuart. "The random feature model for input-output maps between banach spaces." In: *SIAM Journal on Scientific Computing* 43.5 (2021), A3212–A3243.

[47] R. H. Nochetto, K. G. Siebert, and A. Veeser. "Theory of adaptive finite element methods: An introduction." In: *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*. Springer. 2009, pp. 409–542.

[48] J. A. Opschoor, C. Schwab, and J. Zech. "Exponential ReLU DNN expression of holomorphic maps in high dimension." In: *Constructive Approximation* 55.1 (2022), pp. 537–582.

[49] H. Owhadi and L. Zhang. "Numerical homogenization of the acoustic wave equations with a continuum of scales." In: *Computer Methods in Applied Mechanics and Engineering* 198.3 (2008), pp. 397–406. ISSN: 0045-7825. DOI: https://doi.org/10.1016/j.cma.2008.08.012. URL: https://www.sciencedirect.com/science/article/pii/S0045782508003083.

[50] G. Pavliotis and A. Stuart. *Multiscale methods: Averaging and homogenization.* Springer Science & Business Media, 2008.

[51] R. Phillips. *Crystals, defects and microstructures: Modeling across scales.* Cambridge University Press, 2001.

[52] A. M. Tartakovsky, C. O. Marrero, P. Perdikaris, G. D. Tartakovsky, and D. Barajas-Solano. "Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems." In: *Water Resources Research* 56.5 (2020), e2019WR026731.

[53] K. Xu, D. Z. Huang, and E. Darve. "Learning constitutive relations using symmetric positive definite neural networks." In: *Journal of Computational Physics* 428 (2021), p. 110072.

[54] H. You, Y. Yu, S. Silling, and M. D'Elia. "Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws." In: *CEUR-WS* 2964.177 (2021).

Chapter 6

PCA-NET SURROGATE FOR TOPOLOGY OPTIMIZATION WITH A POINT-LOAD SINGULARITY

Hertlein, N., Rajan, A., Chen, J., Bhattacharya, K., and Buskohl, P., 2025. "PCA-Net Surrogate for Topology Optimization with a Point-load Singularity." In preparation.

Contributions: N. H. developed the code for topology optimization. The neural network code for learning sensitivities and singular solutions was written by A. R., N. H. and J. C. . P. B. and K. B. conceptualized the methodology and supervised the project. The manuscript was written collectively by everyone.

6.1 Introduction

6.1.1 Background

Topology optimization (TO) has become a popular and valuable tool for approaching materials and design problems in various disciplines. Since its introduction, it has been applied in mechanics [32], thermodynamics [10], metamaterials [26], compliant mechanisms [13], and others (a thorough review has been previously published [11]) including situations with uncertainties [35, 2]. Because it generally involves iterative simulations in a gradient-based optimization loop, computational expense is a persistent challenge [27]. And because objective functions optimized in TO are usually nonconvex, susceptibility to local optimality [33] is another limitation across most applications.

Recent efforts have attempted to mitigate these challenges through various techniques from machine learning (ML). For example, frameworks based on neural networks (NNs) have been used to predict optimal topologies directly, given a problem definition. Yu et al. [39] trained a convolutional neural network (CNN) for this purpose, and after the CNN had been trained, the computational cost was negligible for any new problem that comes from the same general class as the training data. The same high level approach has been

taken elsewhere [21], including with generative adversarial networks (GANs) [31, 28, 16] variational autoencoders (VAEs) [9] and diffusion models [14]. In some cases, these models have generated designs that outperform conventional TO designs, especially when the problem is more constrained, for example, to improve manufacturability [17]. Additionally, in this configuration, GANs have shown promising capacity for fast, natural interaction with designers during the early stages of design, including by providing reduced dimensional design representations that can ease exploration of high-performing geometries in a collaboration between the model and the human designer [36].

Others integrate ML more centrally into the TO routine by training surrogate models for design sensitivities, which can then be called each optimizer iteration and remove the need for solving the linear elasticity partial differential equation (PDE). These efforts often involve retraining the surrogate successively after batches of several iterations that are interwoven with traditional iterations [8, 40, 6], a technique most valuable when either training is extremely efficient or the particular TO formulation's sensitivity analysis is especially costly. Other efforts have called the surrogate throughout the entire optimization run, which required careful selection of training data from past problems' convergence histories and operated only on very basic cantilever beam problems [30].

These approaches of ML within the TO routine, along with additional approaches [7, 34], are discussed in greater detail in a recent review [38]. Despite their advances in post-training computational savings and in discovering some design candidate that outperform TO structurally, challenges still remain. Overall computational savings are limited to cases where the upfront training expense can be justified. Situations like this include when multiple TO runs of a similar class is anticipated [16] or when a designer desires to interact with generated topologies to explore alternative possibilities throughout the design space [31, 36]. Very large neural networks are often used, with parameter counts frequently in the millions, which increases training cost. NNs also tend to be limited to a pre-specified output dimensionality, limiting usage to TO problems of only that particular size.

Neural operators are a generalization of traditional neural networks designed to approximate the underlying operators that map between infinite-dimensional

spaces [23]. These operators can represent complex relationships, such as the solution operator of partial differential equations (PDEs) that map input parameters to the solution function [5]. While prior strategies aim to discretize the input and output spaces of PDEs and train neural networks to map between these finite-dimensional spaces, neural operators aim to learn the mapping between function spaces themselves. This provides the added benefit of mesh-independence, where unlike with many NN approaches, accuracy may remain even as resolution increases [5]. The benefits can be seen in solving the Navier-Stokes equation, for example, which has been approximated at arbitrary resolutions in climate science [22], superesolution of low-resolution images has been demonstrated in computer vision [37], and elastic waves simulations have been approximated in 3D [24].

One specific type of neural operator is PCA-Net [5]. PCA-Net combines principal component analysis (PCA) to reduce the dimensionality of both the input and output spaces with a NN that approximates the mapping between the resulting finite-dimensional latent spaces. The NN effectively maps the PCA-reduced inputs to the coefficients of the outputs' PCA projections, and a decoding step reconstructs the outputs from its PCA coefficients. The approximate solution obtained by PCA-Net can be seen as a linear combination of basis functions derived from the PCA of the output data. While this linearity may limit accuracy particularly in the presence of non-smooth data, it represents an efficient and simple approach that is amenable to analysis. Neural operators like PCA-Net can be seen as an alternative to traditional methods which focus on learning mappings between functions spaces directly from data.

6.1.2 PCA-net and topology optimization

Neural operators in the context of topology optimization remain less explored, leaving questions about resulting performance, its optimal situation within the overall TO routine, and whether the general approach is compatible with the TO problem. Initial indications are promising, with Liang et al. [25] demonstrating a Fourier neural operator for level set TO, which generalizes to output dimensionalities outside of its training data. They further show that, after an initial training, their framework could be transferred to a new

class of boundary conditions with relatively low re-training cost, which seems to be enabled in part by the neural operator's flexibility and low parameter count relative to convolutional approaches. In density-based TO, Erzmann et al [12] presented a neural operator that calculates design sensitivities from displacement fields to inform a gradient-based TO loop. Performance at first was poor, but improved dramatically when the loss function was engineered to encourage equivariance.

Looking to the example architecture of PCA-Net [5], the present work expands on these efforts by considering both direct prediction of the optimal design (referred to as *direct design*) and prediction of design sensitivities to be queried sequentially within a gradient descent loop (referred to as *sensitivity prediction*) on the same class of TO problems (Fig. 6.1). While the PCA-Net architecture may be less expressive compared to other neural operators like FNO or Deep-O-Net, the architecture is simple and significantly faster in evaluation once trained. Further, its efficacy in learning elliptic problems like those that arise in TO has been established [5]. In direct design, PCA-Net gives excellent fidelity to ground truth designs in both per-pixel accuracy and mechanical performance. In sensitivity prediction, performance is reasonable but weaker, motivating further study which revealed that the singularity presented by the mechanical point force was not directly compatible with PCA-Net. Two alternative methods for addressing this are presented, including a softening of the singularity by applying a modest spatial distribution to the point load and removal of the singularity from the PCA-Net training data via a Green's function analytical approximation. Both methods significantly improve the prediction results, and provide the opportunity to explore relationships between PCA-reduced representations of sensitivity data and the resulting framework performance.

6.2 Methods

This section will describe the TO formulation, the class of problems considered, the PCA-based neural operator studied, and the two configurations for approximating TO with PCA-Net.

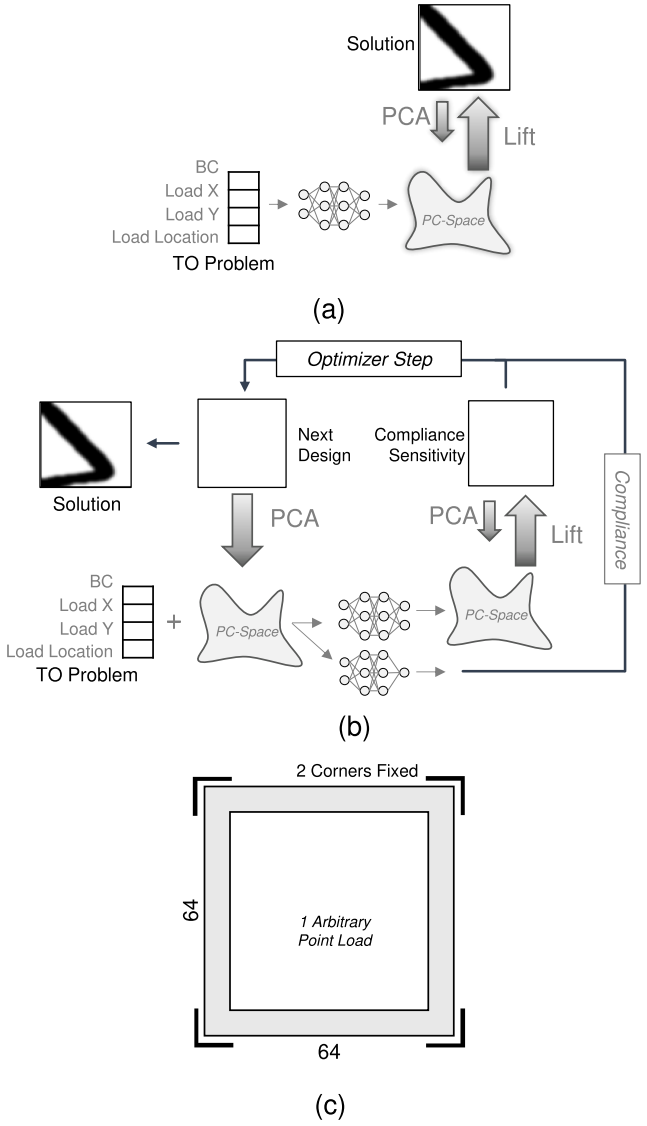


Figure 6.1: Overview of PCA-Net surrogate implementations for topology optimization. (a) The direct design configuration in which PCA-Net directly predicts the optimal design given a problem definition. The problem definition is represented as a vector and passed directly into a feedforward neural network, whose output is lifted into the image space using PCA parameters calculated from the training data. (b) The sensitivity prediction routine, in which the problem definition is appended to a reduced-dimensional version of the current iteration's design to use as input. One neural network predicts the compliance sensitivity values while another predicts the global compliance scalar, enabling a design update which is repeated until convergence. (c) Schematic of the TO design domain and boundary condition parameter space. Training samples are selected randomly by varying the load position within the white domain and selecting two corners for a fixed boundary condition.

6.2.1 A class of topology optimization problems

We consider the classical problem of compliance optimization [4]. We consider a design domain that is fixed on part of the boundary, and subject to an applied load on another part of the boundary. We seek the design that minimizes the compliance (maximizes the stiffness) with respect to the applied load subject to a constraint on the overall volume fraction of the domain. Assuming a finite element discretization, we can write the objective as [32, 3]

$$\min_{\mathbf{x}} \quad c(\mathbf{x}) = \mathbf{U}^{\mathbf{T}}\mathbf{K}\mathbf{U} = \sum_{c=1}^{N}(x_e)^p \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \tag{6.1}$$

where c is the global structural compliance, $\mathbf{x}$ is the design variable with $x_i \in [0, 1]$ is the material density in each element, K is the stiffness matrix and U the displacement field obtained by solving the equations of the linear elasticity. Note that we have used solid isotropic material with penalization (SIMP) interpolation for the stiffness with the usual value of $p = 3$. We assume that the volume fraction is constrained to 30%. The design sensitivities can then be calculated using the adjoint method, and a gradient based optimization iteratively until convergence to an optimum, $\mathbf{x}^*$. In this work, this is implemented using the Python version of the 99-line code [32, 1], including its mesh independency filter with the radius set to 3 elements.

In the sensitivity prediction configuration, PCA-Net will be trained to predict $\frac{\partial c}{\partial x_e}$ and a second model will predict $c(\mathbf{x})$. This will occur iteratively until convergence, such that up to hundreds of calls to PCA-Net will be required for a given problem. In the direct design configuration, the target prediction of PCA-Net will be the final design x^*.

In this work, we focus on the class of problems shown in Fig. 6.1(c). A 64 element square design domain is considered, with a single point load that can be located anywhere greater than 5 nodes from the boundary, at any angle. Two corners are assigned a fixed boundary condition stretching 5 nodes out from the corner node in both directions.

6.2.2 Learning problems

In the previous section, we have described a class of problems. While domain is fixed, the two corners that are clamped, the point of application of applied force and the direction of the applied force are unspecified. We could solve each

of these design problems by the standard method described above. However, this is computationally expensive. So, the goal of this work is to see if we can solve a few of these problems, and use this as data to train a neural operator to act as a surrogate for solver. We specifically consider PCA-Net as the neural operator for three problems.

Direct design In direct design, we seek to learn the map from the inputs consisting of the corners that are fixed, the point of application of the point load and the angle of the point load to the output that is the optimal design. The input is encoded in a vector as follows. The first four elements of the vector represent the fixed boundary conditions. These elements correspond to the existence (value 1) or absence (value 0) of fixed boundary conditions on the top left, top right, bottom left, and bottom right of the design. The fifth and sixth elements represent the force in the x- and y-directions, the seventh and eighth elements represent the location of the load on the x- and y-axis. The output is the optimal design that is a function discretized into 64^2 in this case.

The NN's output target is the first 500 PCs of the optimal design, which is then lifted to the desired output dimensionality, which is 64^2 in this case. Loss is assessed as the relative L2 norm between the result $\hat{\mathbf{x}}^*$ and $\mathbf{x}^*$, the TO solution for the same problem.

Sensitivity prediction In this problem, we seek to learn the sensitivity of a particular design. So, the inputs are a combination of the problem (the corners that are fixed, the point of application of the point load and the angle of the point load) as well as a design $\mathbf{x}$. The latter is a function, but discretized into 64^2. The output is the sensitivity $\partial c/\partial \mathbf{x}$. This is again a function, but discretized into 64^2.

Compliance prediction We seek to predict the global compliance of any particular design. The inputs are the same as the sensitivity prediction, problem specification and design, and the output is the global compliance.

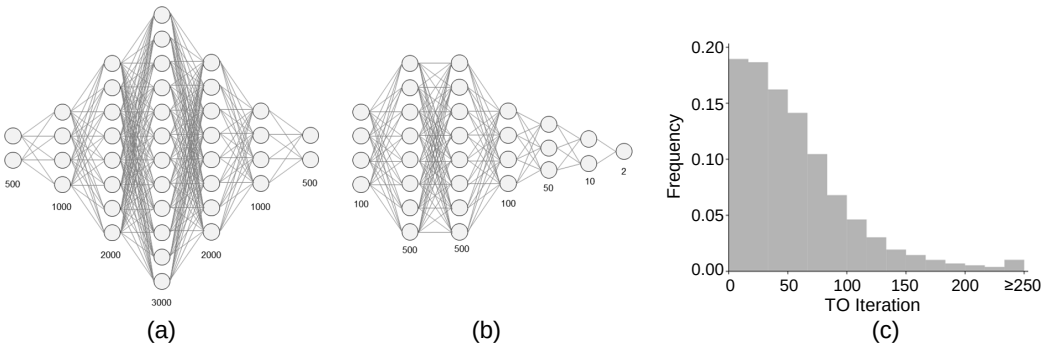


Figure 6.2: PCA-Net architectures and their training data selection. (a) Schematic of the fully connected neural network used for predicting the reduced dimensional images, including both the final design (for direct design) and the compliance sensitivity (for sensitivity prediction). Numbers beneath each layer indicate the number of neurons. (b) The network used for predicting global compliance. (c) For sensitivity prediction, the distribution of samples selected per TO iteration. This biasing is roughly proportional to the rate of change of design as a function of iteration, as discussed more fully in our previous work [18].

6.2.3 PCA-net and architecture

The learning problems described above involve functions as inputs and outputs, and therefore the maps from the input to output are operators. Even though we work with discretized data, we would like the learning to be independent of discretization to some degree. This requires neural operators, neural approximations that maps functions to functions. Specifically, we use PCA-Net in this work [5]. A PCA-Net is a composition of a projection operator (that projects input functions to finite dimensions), a multi-layer neural network that maps the projected input to a finite dimensional vector and a final lifting operator to function space. The projection operator as well as the lifting operator are both based on principal component analysis (PCA) of the input and output data. In this work, functions are represented as 64^2 vectors, and the PCA is use to project/lift them into 500 dimensional vectors in the direct design and sensitivity prediction, and 100 dimensional vectors in the compliance prediction.

The particular architecture used for the various cases are shown in Figure 6.2 with a RELU activation function [20].

6.2.4 Data and training

Direct design requires predicting the optimal design. We sample the inputs (clamped corners, load location and load direction) randomly, and for each input, we solve the topology optimization problem (Sec. 6.2.1) iteratively until convergence. The data consists of the inputs and the final design $\mathbf{x}^*$ as output. Fifty thousand samples are used for training, with an additional five thousand used for validation and ten thousand for testing.

We propose to use sensitivity prediction in the iterative design loop. In a typical TO iteration, the nature of the inputs and outputs in the early iteration differ from those close to the optimal. To ensure good performance throughout, convergence trajectories of TO examples should be sampled accordingly. Previous work has employed re-training for a given problem every several iterations [8, 40], but that carries significant online computational cost. Sampling only early portions of trajectories has also been proposed [30]. In the present work, following the best results from a previous comparison of approaches [18], 4 ± 2 iterations are selected from each problem, according to a bias curve loosely based on the rate of change of $\mathbf{x}_i$ as a function of i. The actual distribution in an example collection is shown in Fig. 6.2(c). Also as in previous work [18], compliance sensitivities were linearly scaled to the range of $[-1, -0.01]$ and compliance values were scaled according to the linear mapping going from $[0, 150]$ to $[0, 1]$. Fifty thousand total samples are collected in this manner for the training dataset, with an additional five thousand and ten thousand for validation and testing, respectively.

For all problems, training was conducted using the Adam optimizer [19] with a batch size of 32. The learning rate was scheduled by cosine annealing with a minimum of 10^{-6}. The models were implemented in PyTorch [29] and executed on an NVIDIA V100 GPU.

6.3 Results

6.3.1 Direct design

For the direct design configuration, training and validation loss curves are shown in Fig. 6.3. Learning was smooth and stable across $\sim$4.7 hours wall-clock time, with some overfitting suggesting fewer epochs would have been sufficient. After training, across the 10 thousand test problems, PCA-Net in-

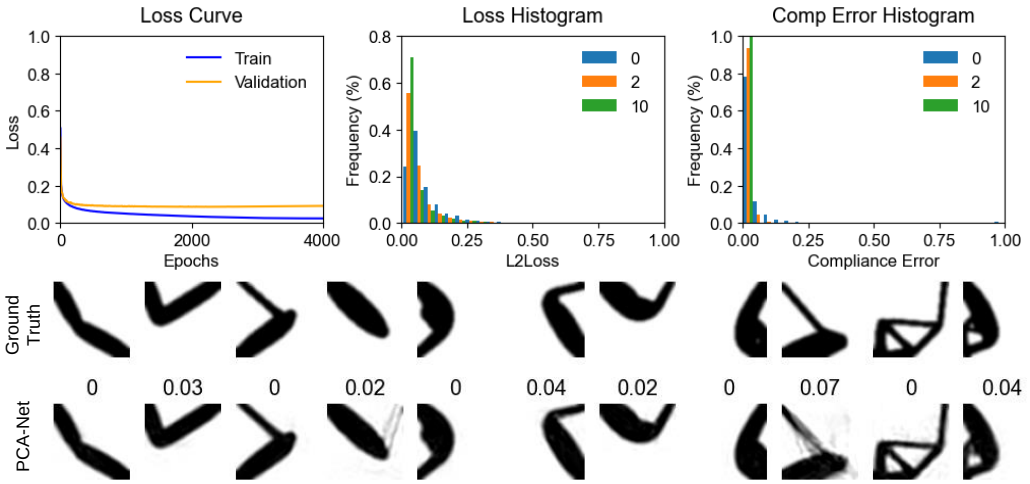


Figure 6.3: PCA-Net is an effective direct design TO surrogate. The graph in the first column shows the relative mean squared error loss during training & validation, as calculated in the PC-space. Learning is generally stable, with some minor overfitting. The second column represents the same error measurement taken on the final output in image space, for the test set. The final column is the relative error in compliance calculated on the final output according to the relevant problem definition, clipped at 0 and 1. Histogram legends indicate the number of traditional TO iterations performed on PCA-Net's output to improve performance, which led to considerable improvement in performance. Overall, performance remains similar across load distributions. Randomly selected examples are shown below the graphs, including the ground truth and the raw PCA-Net outputs. Numbers above the example outputs represent relative compliance error.

ference time averaged roughly 0.001 second per problem. The first histogram shows the same relative L2 loss measure in pixel-space, that is, after lifting the predicted PCs to the intended $\hat{\mathbf{x}}^*$ dimensionality. The second histogram shows relative compliance error of PCA-Net predictions, evaluated as the difference in compliance of $\hat{\mathbf{x}}^*$ from PCA-Net and $\mathbf{x}^*$ from TO, normalized by the compliance of $\mathbf{x}^*$. Randomly selected example designs are shown below. Despite presence of some shortcomings typical of ML approaches for TO, such as blurriness, predictions demonstrate excellent accuracy relative to the ground truth TO designs. Median relative compliance error was 0.8%. Some accuracy improvement was obtained by post-processing raw PCA-Net outputs with a few iterations of traditional TO as a finishing operation. Results after 2 and 10 iterations are included in the histograms.

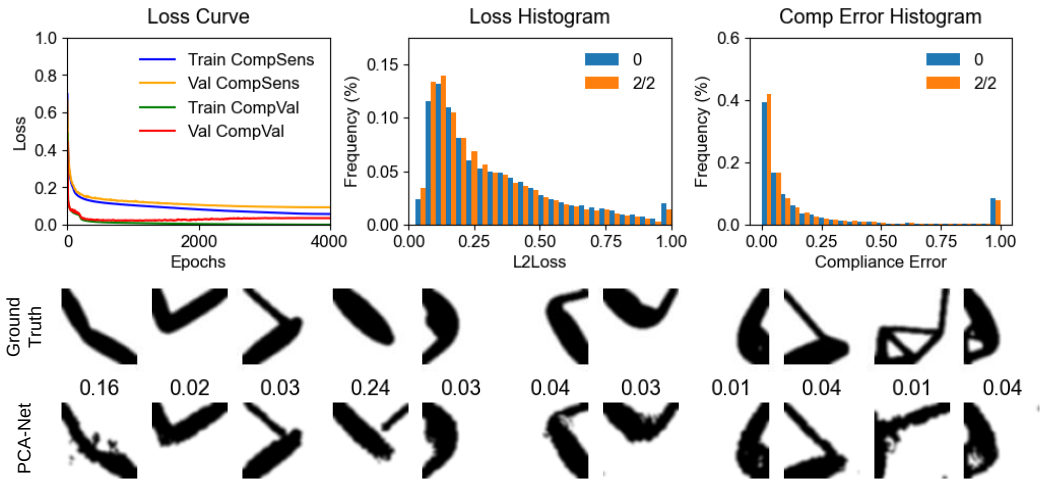


Figure 6.4: PCA-Net surrogate for compliance sensitivity analysis exhibits higher training error and topology prediction error than the direct design surrogate. Formatted in the same manner as Fig. 3, results for sensitivity analysis-based PCA-Net performance of TO indicate reasonable outcomes when considering the final optimum. Histogram legends indicate the number of traditional TO iterations performed before and after the remaining iterations performed using PCA-Net. Results show that overall accuracy on the optimum is lower than in the direct design configuration. This appears to be driven by blotchy design patterns combined with some thinned strut features which in some cases are structurally disconnected.

6.3.2 Sensitivity prediction

Results from the sensitivity prediction configuration are shown in Fig. 6.4, in the same format as the previous figure. Additional loss curves are included for the NN that predicted global compliance $c(\mathbf{x})$. Training time was around 4.9 hours. Test problems solved with PCA-Net alone are represented in the histogram and example designs, with the histograms also including results from problems where the initial and final 2 iterations were solved using traditional sensitivity analysis to improve accuracy. All outliers were assigned a maximum error value of 1.

Relative to the direct design results, histograms here show lower performance, with a considerable number of outliers (approaching 10% of problems). Median relative compliance error increased over the direct design results to 6.2%. The average time to solve a single problem was significantly higher at ∼3.2 seconds per problem, due to the requirement for iterative calls to the surrogate as well as additional calculations in the optimizer in between. Geometric trends

common in poor-performing designs include thinly-connected or disconnected strut features, wavy edges, and extraneous material, indicating difficulty across TO problems and geometries.

Causes could be multiple. Because PCA-Net operates across all iterations in this configuration, it must act as an effective surrogate for a wide range of pixel intensity distributions. This may be a fundamentally more challenging task than always operating on PC-representations of almost entirely binary data. Small error may also compound through the optimization trajectory, as PCA-Net predictions are used to suggest designs for the next PCA-Net prediction–a pattern not applicable in the direct design case. Significant challenges also may be associated with the nature of design sensitivity fields, which often include a sharp gradient at the location of the load, due to the point load singularity. Mitigating the impact of this singularity will be addressed in the following section.

6.4 Singularity correction approaches

A key difference between the two problems studied in Sec. 6.3 is the presence of a singularity in sensitivity field that occurs due to the point load. In direct design , no analog exists because the target outputs $\mathbf{x}^*$ are inherently smooth due to the TO neighborhood filter. In sensitivity prediction, sharp valleys generally exist in $\frac{\partial c}{\partial x_e}$ around the load due to the use of a point load, which is a common practice in TO. This may present a particular challenge for neural operators like PCA-Net, because linear dimensionality reduction techniques such as PCA may struggle to capture it in the reduced dimensional space. To understand the extent to which this effect hampers performance in the TO context, two methods for removing the singularity are investigated.

6.4.1 Analytical removal via a green's functions

The underlying singularity occurs in the displacement field due to a point load. From linear elasticity theory, the displacement field behaves as

$$\mathbf{U}(r) \sim \mathbf{F} \log(r), \tag{6.2}$$

where $\mathbf{F}$ is the point load and r is the radial distance from loading location [15]. It follows that the behaviour of design sensitivity is given by

$$\frac{\partial c}{\partial x_e} \sim \frac{1}{r^2}. \tag{6.3}$$

This poses an issue for PCA based neural networks to learn the sensitivities. Singular functions create outliers that may skew the principal components, leading to poor reconstruction of compliance sensitivity in the rest of the region. Furthermore, the NN predictions might also be heavily inaccurate around the singularity due to its architecture. Therefore, we decompose the domain into a singular and non-singular region and treat them separately during training. An analytical solution is used to model the behaviour around the singular region while the PCA-based NN is used to model the non-singular region.

The compliance sensitivity is given by

$$\frac{\partial c}{\partial x_e} = \left(\frac{\partial c}{\partial x_e}\right)_s + \left(\frac{\partial c}{\partial x_e}\right)_{ns}, \tag{6.4}$$

where the terms on the right hand side represent the singular and non-singular components. We can obtain the analytical form of the singular component, given by

$$\left(\frac{\partial c}{\partial x_e}\right)_s = c_s(\mathbf{F})\ \frac{1}{r^2}, \tag{6.5}$$

where c_s is a constant dependent on the point load. By definition, the non-singular component is a smooth function which can be learned by a PCA-based NN. Therefore, in order to compute $\frac{\partial c}{\partial x_e}$ from the input conditions, $c(\mathbf{F})$ and $\left(\frac{\partial c}{\partial x_e}\right)_{ns}$ must be predicted. An overview of the framework has been represented in Fig. 6.5.

For complete details of this framework, please refer to the supplementary information.

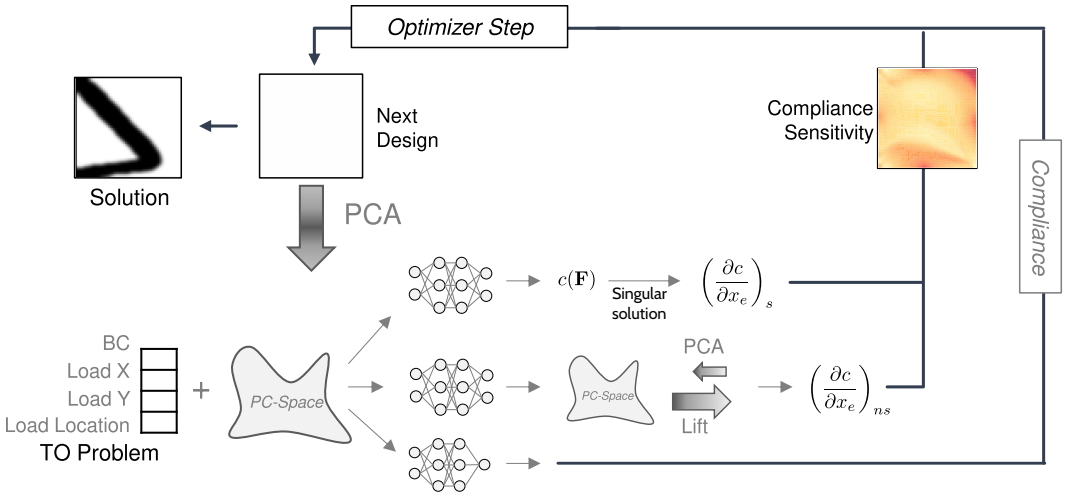


Figure 6.5: Modified PCA based neural network to mitigate errors arising due to singularities.

6.4.2 Load distribution

The second approach entailed spreading the point force according to a smoothing distribution to effectively eliminate $(\frac{\partial c}{\partial x_e})_s$ mentioned in Sec. 6.4.1. This distribution is defined as

$$\chi_0(r) = \begin{cases} 0, & r \leq r_{min} \\ 1, & r \geq r_{min} \\ 1 + C_2 \left(\frac{\sin\left(C_3(r_{min} - r)\right)}{C_3^2} + \frac{(r - r_{min})}{C_3} \right), & r_{min} < r < r_{max} \end{cases} \tag{6.6}$$

where the constants are defined as

$$C_1 = (r_{max} - r_{min}), \quad C_2 = \frac{-2\pi}{(r_{max} - r_{min})^2}, \quad C_3 = \frac{2\pi}{(r_{max} - r_{min})}. \tag{6.7}$$

This gives a smooth distribution which is then scaled linearly to ensure the resultant force retains a magnitude of unity. The parameters r_{min} and r_{max} were selected as 1 and 3, respectively, to ensure the load was spread approximately as widely as the removal of $(\frac{\partial c}{\partial x_e})_s$ from Sec. 6.4.1.

With the same class of TO problem definitions and the same iteration sampling strategy, a new training dataset was collected using this distributed load. Inherently, by smoothing out the load, this changes the actual TO problem being solved. However, the neighborhood filter in TO already effectively ensures a spread of material around the point load, such that smoothing the load to this modest degree bears little effect on the final geometry. Apart from usage

of this distributed load data and a larger input vector that captures all components of the distributed load (100 components to capture the load instead of 4), PCA-Net training and testing was performed in the same manner as before.

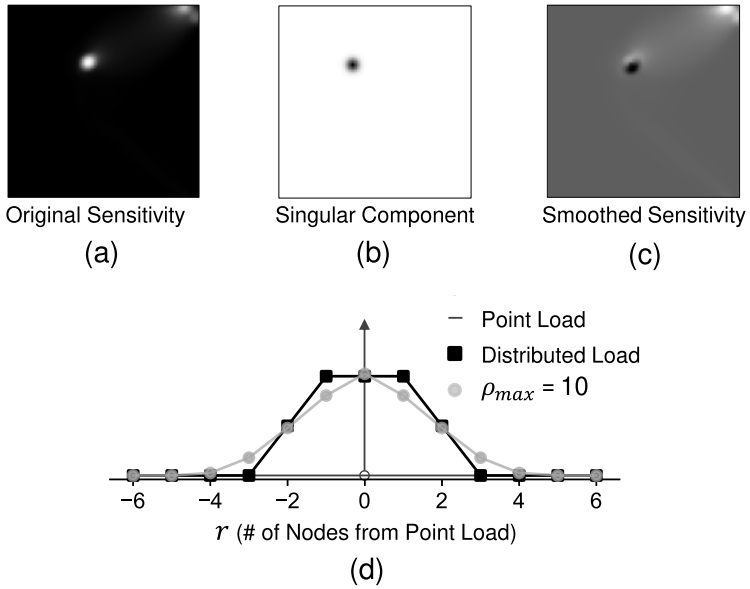


Figure 6.6: Methods of smoothing the problem for more effective sensitivity predictions. Two methods are explored. When performing an analytical removal of the singularity, the singular component (b) of the original compliance sensitivity (a) is subtracted, resulting in a smoother image (c) that PCA-Net is required to predict. The region in which the singularity is removed is defined according to an exponential, as shown in the case of $\rho_{max} = 10$ in panel (d). This graph also shows the relative load distribution explored as the second method of smoothing the problem. For comparison, the original point load is illustrated as a Dirac delta function.

6.4.3 Singularity correction results

With loss curves shown in Fig. 6.7 for both methods of removing the singularity at the point load, significant performance improvement was achieved with both methods. Although distributions of relative L2 error remain similar, median relative compliance error was reduced to 4.6% and 3.9% for the analytical removal and load distribution, respectively. This suggests a reduction in spatially small errors that give large mechanical performance issues, such as disconnected struts. Review of the examples included in Fig. 6.8 indicates some structural defects remain present in some examples, which may be the result of other complicating factors inherent in the sensitivity prediction con-

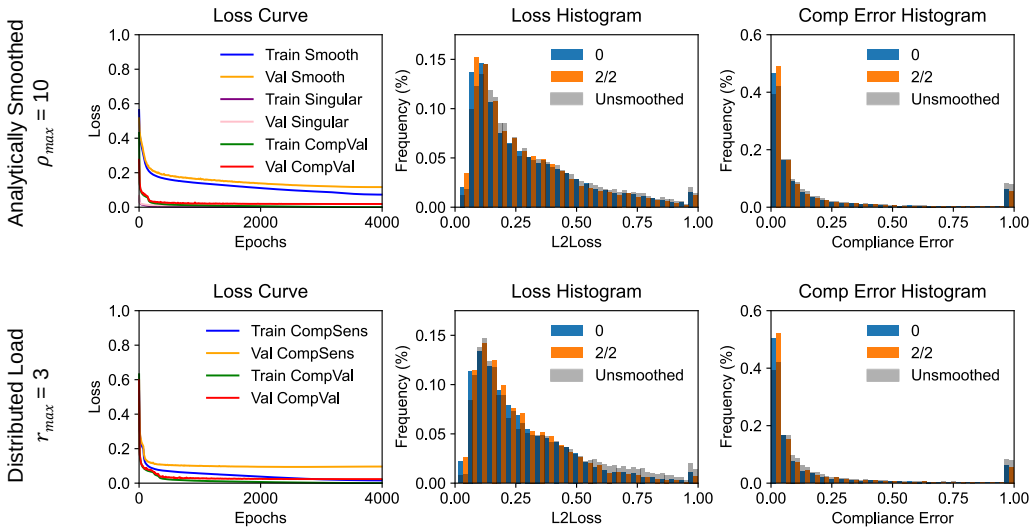


Figure 6.7: PCA-Net surrogate in the compliance sensitivity configuration improves with load smoothing. The first row corresponds to analytical removal of the singularity, with the second row representing the case of the distributed load. Columns are the same as Figs. 3 and 4. Note: PCA-Net prediction error decreases with removal or smoothing of the point load singularity. Histogram error distribution also decrease and consolidate with load smoothing. Two TO pre-processing iteration and two TO post-processing iterations further promote this trend. The loss histogram and compliance error histogram from point load in fig. 4 is overlaid in gray for comparison.

figuration as mentioned in Sec. 6.3.2. Despite this, overall improvement with removal of the singularity at the load is noticeable. As before, pre- and post-processing with a small number of TO iterations provides additional gains, while increasing per-problem runtimes from about 3.2 seconds to about 3.8 seconds.

6.5 Discussion

Of the two problems studied, direct design gave significantly better performance, in both per-pixel accuracy and structural compliance measured relative to designs from traditional TO. From a purely performance-driven standpoint, this suggests the direct design configuration could be a more appropriate fit for neural operators such as PCA-Net. However, this may depend on additional practical factors driven by the use case. For example, a single training sample for this configuration requires an entire TO run, while the same run could provide several training samples for sensitivity prediction. After training how-

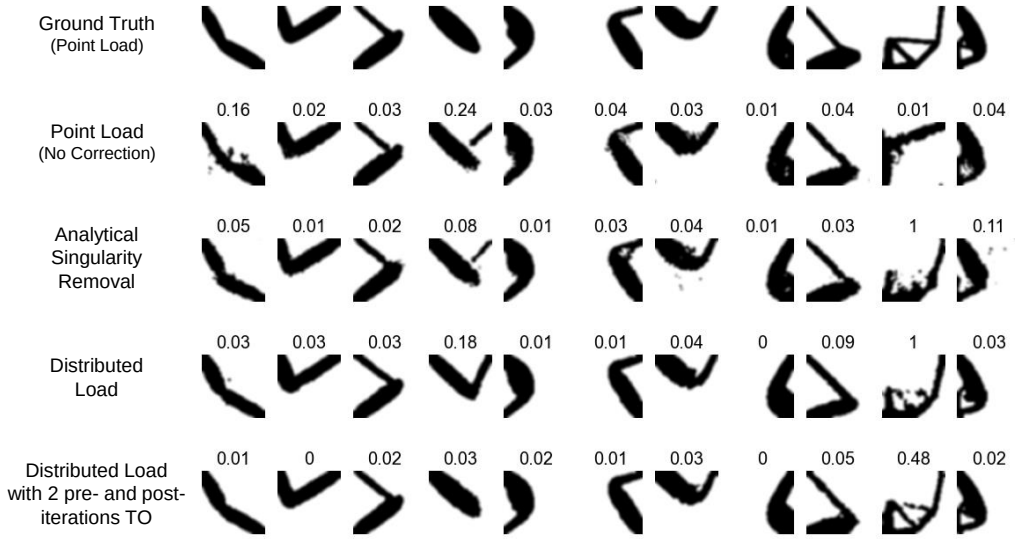


Figure 6.8: TO predictions when PCA-Net used as a compliance sensitivity surrogate are effected by point load singularity. Array of representative topology solutions based (row 1) on ground truth TO runs with point load boundary conditions and PCA-net surrogates with various modifications applied to the point load, including removal of the singularity and spatially soothing the load. Performance improved with both methods of smoothing, with further improvements using a modest number ($\sim$2) of pre- and post-processing TO iterations. Annotations denote relative compliance error.

ever, in this work, direct design runtimes per problem were around 3 orders of magnitude lower than sensitivity prediction. In practice, these factors must be balanced.

Several effects may be limiting performance in sensitivity prediction, such as those mentioned in Sec. 6.3.2, but both methods of smoothing the problem yielded improvements. Additional analysis was conducted to investigate the effects of these smoothing techniques, especially within the PC representation within the neural operator architecture.

For both ground truth sensitivities from TO and predicted sensitivities, the first two PCs are plotted in Fig. 6.9. The distributed load resulted in a more consolidated distribution of samples in these PCs. However, this shift in distribution was not associated with dramatically improved performance relative to the analytically smoothed data, suggesting that PCs beyond the first are important in determining overall TO outcomes.

The amount of variance captured per component was also plotted, in Fig. 6.10, for final designs in the direct design case and sensitivities in the sensitivity pre-

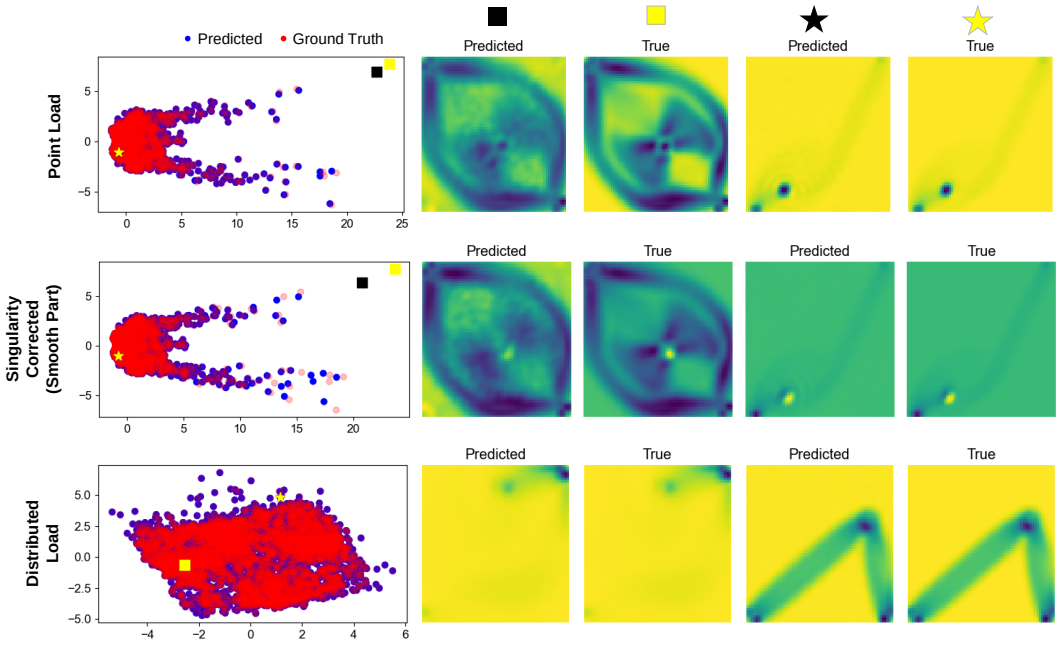


Figure 6.9: Higher variance observed in point load PCA projection. Left Column: Projection of the compliance sensitivity into the first two principal components for the ground truth (red) and PCA-Net prediction (blue), under the different load smoothing conditions. Note the consolidated distribution with load smoothing. Right Columns: plots of the compliance sensitivity field in the physical domain for representative examples denoted by square and star annotations.

diction case. For direct design, regardless of whether the load was distributed or not, most of the variance is captured very quickly, at roughly the same rate. The first 100 components captured 97.5% of the variance. For sensitivities, variance is captured relatively slowly with the raw unsmoothed data (83.8% captured by component 100), with low and intermediate PCs capturing much more variance in both forms of smoothed data (92.6% and 98.3% for analytical smoothing and distributed load, respectively). This trend is correlated with performance of the three approaches (raw, analytical singularity removal, load distribution), further suggesting that relatively early components above the first two are important in final TO outcomes. It also suggests that other, potentially nonlinear, dimensionality reduction techniques that can capture sensitivity data variance faster than PCA may result in further improved performance in the sensitivity prediction configuration.

To understand how iterations are distributed in the PCA representation of sensitivities, samples from iterations ranging from 1 to $\geq$8 are plotted Fig.

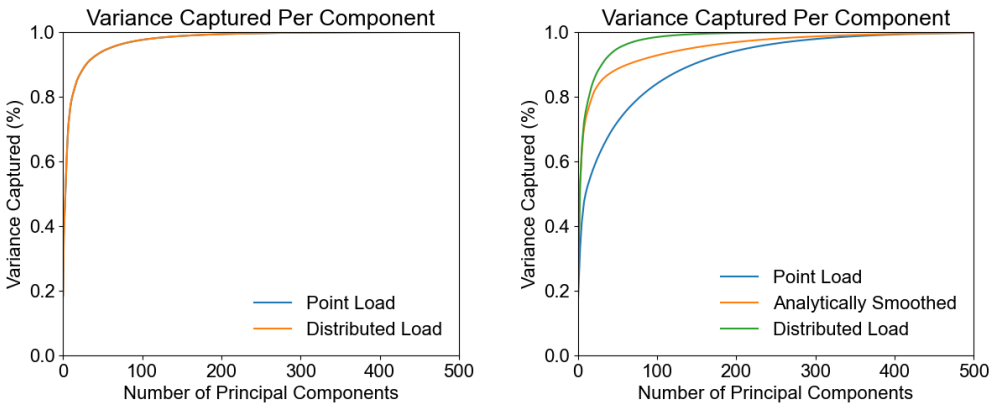


Figure 6.10: Compliance sensitivity variance decreases with smoothing. Comparison of explained variance (left column) and PCA-net prediction errors (right columns) for Direct Design (top row) and Sensitivity Prediction (bottom row). Strong dependence of the number of components needed to capture the variance on the smoothing of the load.

6.11, for both the original and distributed load. In both cases, the overall distribution starts relatively compact, spreads out increasingly through the first several iterations, and then becomes relatively compact again by iteration 8. The effect is more noticeable for the point load, but is present for both. By enabling PCs to be calculated on more compact data and training a neural network on a more consistent distribution, training a PCA-Net-based sensitivity prediction framework only on iterations $\geq$8, while leaving initial iterations to traditional TO, could give further improvements in performance.

6.6 Conclusion

In this work, the PCA-Net neural operator was investigated on minimum compliance SIMP TO in multiple configurations. In the direct design configuration, performance was excellent both for its per-pixel accuracy and its structural compliance, with median relative compliance error of 0.8%. When the surrogate was alternatively trained to perform only the sensitivity analysis portion of the TO routine within the iterative design update loop, structural performance declined to 6.2% median error, in part due to the presence of a singularity in the design sensitivities at the point force. Two smoothing methods were introduced, and both improved performance to around 4-4.5% median error. Analysis of PC spaces showed that PCA's ability to capture variance more quickly in smoother data is linked to better TO outcomes. In

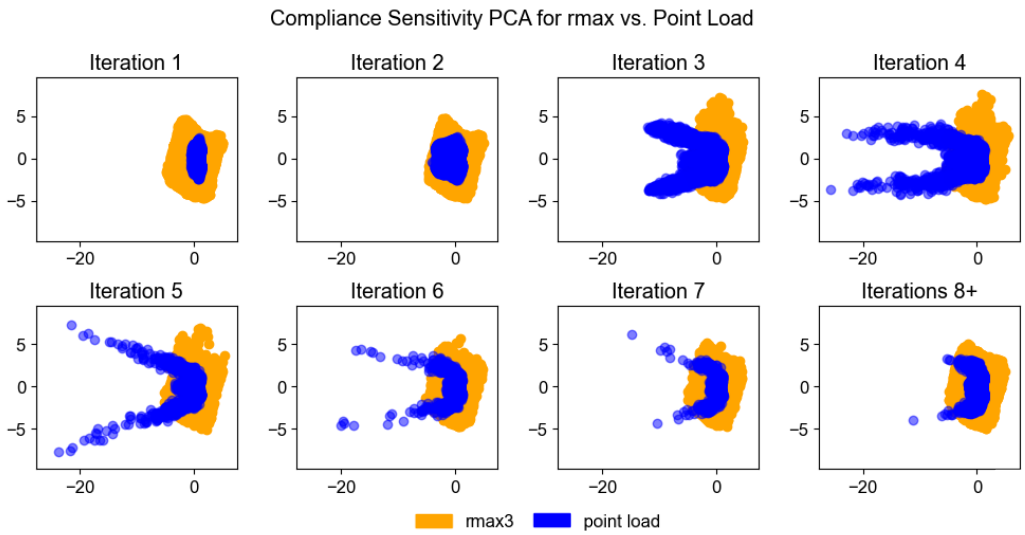


Figure 6.11: Highest variance observed in the PCA of the compliance sensitivity between TO iterations 2-7. Projections of the compliance sensitivity into the first two principal components for the rmax=7 data set. Note: all TO runs assume a uniform density initial guess and are only differentiated by their problem boundary conditions. Motivates training a PCA-Net sensitivity surrogate just on iterations beyond 8.

addition to showing PCA-Net to be an effective surrogate for direct design and sensitivity analysis when smoothing is applied, this effort helps lay the groundwork for further exploration of neural operators for TO that consider factors such as generalization, capacity for arbitrarily scalable output, and training data minimization.

References

[1] N. Aage. *Topology optimization codes written in Python*. Apr. 2019. URL: https://www.topopt.mek.dtu.dk/apps-and-software/topology-optimization-codes-written-in-python.

[2] M. Allen and K. Maute. "Reliability-based design optimization of aeroelastic structures." In: *Structural and Multidisciplinary Optimization* 27.4 (June 2004), pp. 228–242. ISSN: 1615147X. DOI: 10.1007/s00158-004-0384-1.

[3] E. Andreassen, A. Clausen, M. Schevenels, B. S. Lazarov, and O. Sigmund. "Efficient topology optimization in MATLAB using 88 lines of code." In: *Structural and Multidisciplinary Optimization* 43.1 (Jan. 2011), pp. 1–16. ISSN: 1615147X. DOI: 10.1007/s00158-010-0594-7.

[4] M. P. Bendsoe. "Optimal shape design as a material distribution problem." In: *Structural Optimization* 1 (1989), pp. 193–202.

[5] K. Bhattacharya, B. Hosseini, N. B. Kovachki, and A. M. Stuart. *Model Reduction And Neural Networks For Parametric PDEs*. Tech. rep. 2021, pp. 121–157.

[6] S. Bi, J. Zhang, and G. Zhang. "Scalable deep-learning-accelerated topology optimization for additively manufactured materials." In: *NeurIPS Workshop on ML for Engineering Modeling* (Nov. 2020). URL: http://arxiv.org/abs/2011.14177.

[7] N. K. Brown, A. P. Garland, G. M. Fadel, and G. Li. "Deep reinforcement learning for engineering design through topology optimization of elementally discretized design domains." In: *Materials and Design* 218 (June 2022). ISSN: 18734197. DOI: 10.1016/j.matdes.2022.110672.

[8] H. Chi, Y. Zhang, T. L. E. Tang, L. Mirabella, L. Dalloro, L. Song, and G. H. Paulino. "Universal machine learning for topology optimization." In: *Computer Methods in Applied Mechanics and Engineering* 375 (Mar. 2021). ISSN: 00457825. DOI: 10.1016/j.cma.2019.112739.

[9] C. Chu, A. Leichner, F. Wenz, and H. Andrä. "Exploring VAE-driven implicit parametric unit cells for multiscale topology optimization." In: *Materials and Design* 244 (Aug. 2024). ISSN: 18734197. DOI: 10.1016/j.matdes.2024.113087.

[10] S. Das and A. Sutradhar. "Multi-physics topology optimization of functionally graded controllable porous structures: Application to heat dissipating problems." In: *Materials and Design* 193 (Aug. 2020). ISSN: 18734197. DOI: 10.1016/j.matdes.2020.108775.

[11] J. D. Deaton and R. V. Grandhi. *A survey of structural and multidisciplinary continuum topology optimization: Post 2000*. Jan. 2014. DOI: 10.1007/s00158-013-0956-z.

[12] D. Erzmann and S. Dittmer. "Equivariant neural operators for gradient-consistent topology optimization." In: *Journal of Computational Design and Engineering* 11.3 (June 2024), pp. 91–100. ISSN: 22885048. DOI: 10.1093/jcde/qwae039.

[13] A. T. Gaynor, N. A. Meisel, C. B. Williams, and J. K. Guest. "Multiple-material topology optimization of compliant mechanisms created via polyJet three-dimensional printing." In: *Journal of Manufacturing Science and Engineering, Transactions of the ASME* 136.6 (Dec. 2014). ISSN: 15288935. DOI: 10.1115/1.4028439.

[14] G. Giannone and F. Ahmed. "Diffusing the optimal topology: A generative optimization approach." In: (Mar. 2023). URL: http://arxiv.org/abs/2303.09760.

[15] M. E. Gurtin. "The linear theory of elasticity." In: *Mechanics of Solids*. Ed. by C. Truesdell. Vol. II. Springer, 1973, pp. 1–295.

[16] N. Hertlein, P. R. Buskohl, A. Gillman, K. Vemaganti, and S. Anand. "Generative adversarial network for early-stage design flexibility in topology optimization for additive manufacturing". In: *Journal of Manufacturing Systems* 59 (Apr. 2021), pp. 675–685. ISSN: 02786125. DOI: 10.1016/j.jmsy.2021.04.007.

[17] N. Hertlein, A. Gillman, and P. R. Buskohl. "Generative adversarial design analysis of non-convexity in topology optimization." In: *International design engineering technical conferences and computers and information in engineering conference* (2022). DOI: 10.1115/DETC2022-89997.

[18] N. Hertlein, A. Gillman, and P. R. Buskohl. "{TBD}". In: *{TBD}* (2025).

[19] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *International conference on learning representations*. Dec. 2015. URL: http://arxiv.org/abs/1412.6980.

[20] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. *Self-Normalizing Neural Networks*. Tech. rep.

[21] H. T. Kollmann, D. W. Abueidda, S. Koric, E. Guleryuz, and N. A. Sobh. "Deep learning for topology optimization of 2D metamaterials." In: *Materials and Design* 196 (Nov. 2020). ISSN: 18734197. DOI: 10.1016/j.matdes.2020.109098.

[22] T. Kurth, S. Subramanian, P. Harrington, J. Pathak, M. Mardani, D. Hall, A. Miele, K. Kashinath, and A. Anandkumar. "FourCastNet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators". In: *Proceedings of the Platform for Advanced Scientific Computing Conference, PASC 2023*. Association for Computing Machinery, Inc, June 2023, pp. 1–11. DOI: 10.1145/3592979.3593412.

[23] S. Lanthaler. *Operator learning with PCA-Net: upper and lower complexity bounds*. Tech. rep. 2023, pp. 1–67. URL: http://jmlr.org/papers/v24/23-0478.html..

[24] F. Lehmann, F. Gatti, M. Bertin, and D. Clouteau. "3D elastic wave propagation with a Factorized Fourier Neural Operator (F-FNO)." In: *Computer Methods in Applied Mechanics and Engineering* 420 (2024). DOI: 10.57745/LAI6YU.

[25] K. Liang, D. Zhu, and F. Li. "A Fourier neural operator-based lightweight machine learning framework for topology optimization." In: *Applied Mathematical Modelling* 129 (May 2024), pp. 714–732. ISSN: 0307904X. DOI: 10.1016/j.apm.2024.02.011.

[26] Q. Liu, R. Xu, Y. Zhou, J. Ge, S. Yuan, Y. Long, and T. Shi. "Metamaterials mapped lightweight structures by principal stress lines and topology optimization: Methodology, additive manufacturing, ductile failure and tests". In: *Materials and Design* 212 (Dec. 2021). ISSN: 18734197. DOI: 10.1016/j.matdes.2021.110192.

[27] S. Mukherjee, D. Lu, B. Raghavan, P. Breitkopf, S. Dutta, M. Xiao, and W. Zhang. "Accelerating large-scale topology optimization: state-of-the-art and challenges." In: *Archives of Computational Methods in Engineering* 28.7 (Dec. 2021), pp. 4549–4571. ISSN: 18861784. DOI: 10.1007/s11831-021-09544-3.

[28] Z. Nie, T. Lin, L. Burak Kara, A. Preprint Zhenguo Nie, and H. Jiang. *TopologyGAN: Topology Optimization Using Generative Adversarial Networks Based on Physical Fields Over the Initial Domain*. Tech. rep. 2020. URL: https://github.com/zhenguonie/TopologyGAN.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury Google, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. K. Xamla, E. Yang, Z. Devito, M. Raison Nabla, A. Tejani, S. Chilamkurthy, Q. Ai, B. Steiner, L. F. Facebook, J. B. Facebook, and S. Chintala. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Tech. rep. 2019.

[30] C. Qian and W. Ye. "Accelerating gradient-based topology optimization design with dual-model artificial neural networks". In: *Structural and Multidisciplinary Optimization* 63.4 (Apr. 2021), pp. 1687–1707. ISSN: 16151488. DOI: 10.1007/s00158-020-02770-6.

[31] C. Sharpe and C. C. Seepersad. "Topology design with conditional generative adversarial networks." In: *IDETC*. Anaheim, CA, Aug. 2019. URL: https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2019/59186/V02AT03A062/6453126/v02at03a062-detc2019-97833.pdf.

[32] O. Sigmund. "A 99 line topology optimization code written in Matlab." In: *Struct Multidisc Optim* 21 (2001), pp. 120–127. URL: http://www.topopt.dtu.dk..

[33] O. Sigmund and J. Petersson. *Numerical instabilities in topology optimization: A survey on procedures dealing with checkerboards, mesh-dependencies and local minima*. Tech. rep. 1998, pp. 68–75.

[34] Z. Sun, Y. Wang, P. Liu, and Y. Luo. "Topological dimensionality reduction-based machine learning for efficient gradient-free 3D topology optimization." In: *Materials and Design* 220 (Aug. 2022). ISSN: 18734197. DOI: 10.1016/j.matdes.2022.110885.

[35] A. P. Torres, J. E. Warner, M. A. Aguiló, and J. K. Guest. "Robust topology optimization under loading uncertainties via stochastic reduced order models." In: *International Journal for Numerical Methods in Engineering* 122.20 (Oct. 2021), pp. 5718–5743. ISSN: 10970207. DOI: 10.1002/nme.6770.

[36] S. Valdez, N. Rodriguez, and C. Seepersad. "Latent variable representations for interactive structural design exploration." In: *IDETC-CIE* (2022). DOI: 10.1115/DETC2022-91076.

[37] M. Wei and X. Zhang. *Super-Resolution Neural Operator*. Tech. rep. URL: https://github.com/.

[38] R. V. Woldseth, N. Aage, J. A. Bærentzen, and O. Sigmund. *On the use of artificial neural networks in topology optimisation.* Oct. 2022. DOI: 10.1007/s00158-022-03347-1.

[39] Y. Yu, T. Hur, J. Jung, and I. G. Jang. "Deep learning for determining a near-optimal topological design without any iteration." In: *Structural and Multidisciplinary Optimization* 59.3 (Mar. 2019), pp. 787–799. ISSN: 16151488. DOI: 10.1007/s00158-018-2101-5.

[40] Y. Zhang, H. Chi, B. Chen, T. L. Elaine Tang, L. Mirabella, L. Song, and G. H. Paulino. "Speeding up computational morphogenesis with online neural synthetic gradients." In: *Proceedings of the International Joint Conference on Neural Networks*. Vol. 2021-July. Institute of Electrical and Electronics Engineers Inc., July 2021. DOI: 10.1109/IJCNN52387.2021.9533789.

Chapter 7

CONCLUSION AND FUTURE WORK

In this thesis, we have addressed two challenges: developing a novel method for material characterization, and exploring the use of neural networks to solve complex mechanics problems.

The first two chapters focus on the formulation and application of a new approach for recovering material properties directly from experimental data. Traditional techniques typically require highly constrained experimental conditions—such as inducing homogeneous deformation fields or relying on simplified loading configurations—which limits their applicability to real-world, complex scenarios. In contrast, our method enables the inference of material parameters from general experimental setups directly from measurable quantities such as full-field deformation and force data. Assuming a constitutive model, we simulate the boundary value problem corresponding to the experimental configuration and compute the resulting fields. An objective function is then defined to quantify the discrepancy between these simulated fields and the actual measurements. To minimize this objective, we employ a gradient-based optimization method, where sensitivities are efficiently computed using the adjoint method. This approach, known as PDE-constrained optimization, allows for scalable and accurate parameter identification even in high-dimensional settings.

We demonstrate the method on synthetic datasets under both quasistatic and dynamic loading conditions. Importantly, because the framework does not rely on specific governing equations, it generalizes naturally to more complex scenarios, such as dynamic contact problems. Building on these synthetic validations, we apply the methodology to real experimental data, recovering material properties for rolled homogeneous armor (RHA) steel and polycrystalline aluminum alloys from instrumented indentation tests. These case studies highlight the method's potential for practical use in characterizing materials under complex loading and boundary conditions.

Classical constitutive models, while successful in modeling a wide range of materials often rely on specific assumptions and require significant *a priori*

knowledge of the underlying physical mechanisms. As the catalog of engineered materials continues to expand, including complex alloys and metamaterials, these traditional approaches may fall short, either due to the absence of established models or the difficulty of parameter identification. In the third chapter, we address this challenge by introducing a data-driven constitutive model based on a recurrent neural operator (RNO). RNOs are well-suited for modeling systems with temporal dependencies, making them a natural choice for capturing elasto-viscoplastic behavior without the need for hand-crafted constitutive laws. We integrate the RNO within a PDE-constrained optimization framework, where the network parameters (weights and biases) are optimized directly from experimental data. As a demonstration, we consider a dynamically compressed thin annular specimen, and use displacement and force measurements to train the RNO model. The trained model is then validated on an unseen loading configuration to assess its generalization. Results show that the RNO not only captures the observed behavior accurately but also outperforms a classical elasto-viscoplastic model in predictive performance. This study illustrates the potential of neural operator-based constitutive models as flexible and robust alternatives to classical approaches, especially in scenarios where first-principles modeling is impractical.

While the RNO framework enables data-driven modeling of constitutive behavior at the macroscopic continuum scale, it operates independently of the physical mechanisms that give rise to this behavior. In reality, continuum-scale responses often originate from physics at smaller scales, governed by their own PDEs. Capturing this relationship requires solving the governing equations at both macro and microscale: the macroscale provides boundary conditions to the microscale model, which in turn returns effective material behavior back to the larger scale. This bidirectional coupling forms the basis of multiscale modeling.

Although powerful, this approach is computationally intensive, as it demands repeated microscale PDE solves across the domain during simulation. To make this process tractable, the fourth chapter introduces neural operators as efficient surrogates for these lower-scale solvers. Rather than resolving the microscale physics directly, we train a neural operator to map microstructural parameters and loading conditions to approximate PDE solutions. We focus on elliptic equations relevant to equilibrium problems in heterogeneous media

and evaluate the model on a variety of microstructure types. Results show that neural operators can provide fast, reliable approximations, paving the way for scalable multiscale simulations without sacrificing fidelity.

Taken together, the two chapters of this thesis show that neural operators can be powerful tools in computational mechanics, playing two distinct roles: (i) as constitutive models that are trained directly from experimental data without requiring explicit assumptions about the material's functional form, and (ii) as surrogates for lower-scale PDE solvers, enabling efficient multiscale simulations. These findings open the door to more flexible and computationally feasible approaches for modeling complex material behavior.

In the fifth chapter, we extend the application of neural operators to tackle an engineering design problem. Topology optimization is a well-established approach for solving inverse design problems, which involves the repeated solution of governing equations to compute sensitivities and update the design iteratively. However, this process is computationally expensive due to the high number of PDE solves required. To accelerate the optimization, we employ a PCA-based neural operator to approximate the PDE solutions efficiently. Because the solution fields are high-dimensional, we first project them onto a lower-dimensional principal component (PCA) space. The neural operator is then trained to predict the solution in this reduced space, enabling fast and accurate reconstruction of the full solution field. This dimensionality reduction improves computational efficiency, facilitates easier training and inference.

7.1 Future directions

The PDE-constrained optimization technique for material characterization relies critically on the assumption that the experimental data contain sufficient information to enable accurate identification of material properties. However, this is a nontrivial assumption and presents an area for further research. Designing experiments that produce data rich enough to reliably recover constitutive parameters remains a significant challenge. Moreover, different parameters influence different regions of the data, suggesting that certain experimental configurations may be more informative for specific aspects of the material behavior. Therefore, the development of a formal framework for intelligent experiment design—aimed at maximizing the informativeness of measurements for parameter identification—is a necessary direction for future work.

The RNO-based constitutive model presented in Chapter 4 exhibits limited generalization when applied to settings beyond those encountered during training. For this reason, validation was restricted to strain rates and length scales similar to those used during training, and there no guarantee the model will give accurate predictions outside of this regime. This limitation can potentially be addressed by pretraining the RNO on data generated from a known classical constitutive model, followed by transfer learning within the PDE-constrained framework. Such offline pretraining introduces minimal additional computational cost but can significantly enhance the robustness of the model. Another key limitation of the RNO is its lack of built-in physical constraints—it does not inherently enforce principles such as material symmetry, frame indifference, or thermodynamic consistency. There is, however, a growing body of work focused on embedding these physical principles into neural network architectures from the outset [1, 2], which presents a promising direction for enhancing interpretability, reliability and generalization of data-driven constitutive models.

In Chapter 5, we introduced a neural operator for approximating solutions to elliptic PDEs at the microscale. While the results are promising, several open challenges remain. One major difficulty is that plasticity at the microscale can induce localized singularities in the solution field, which are notoriously difficult for neural operators to capture. Similarly, sharp variations in material properties across the domain can lead to discontinuities or steep gradients, further complicating learning. Even though we present a method to learn singular solutions using neural operators, it relies on the assumption that the local solution around the singularity is known. This might not be generalizable to all problems. Another important limitation is that current neural operator surrogates typically approximate only the solution fields—not their gradients. This restricts their applicability in quasistatic problems, where accurate stress and strain gradients are essential for computing internal forces. As a result, such surrogates are more suited to dynamic problems where gradient accuracy is less critical.

References

[1] D. K. Klein, M. Fernández, R. J. Martin, P. Neff, and O. Weeger. "Polyconvex anisotropic hyperelasticity with neural networks." In: *Journal of the Mechanics and Physics of Solids* 159 (2022), p. 104703.

[2] F. Masi, I. Stefanou, P. Vannucci, and V. Maffi-Berthier. "Thermodynamics-based artificial neural networks for constitutive modeling." In: *Journal of the Mechanics and Physics of Solids* 147 (2021), p. 104277.

A p p e n d i x A

APPENDIX FOR CHAPTER 2

A.1 Adjoint problem in elasto-viscoplascitity

We consider an objective of integral from

$$\mathcal{O}(u, q, \varepsilon^p, P) = \int_0^T \int_\Omega o(u, q, \varepsilon^p, P)\, d\Omega. \tag{A.1}$$

We follow Section 2.2.1 to use the adjoint method to compute the sensitivity of $\mathcal{O}$ with respect to the parameters P. So we augment to objective using the governing equations

$$\begin{aligned}\mathcal{O} = \int_0^T \int_\Omega \Bigg\{ & o + \rho\ddot{u}\cdot v + \frac{\partial W^e}{\partial \varepsilon}\cdot\nabla v - b\cdot v \\ & + \gamma\dot{q}\left[\sigma_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}}\right] + \zeta\cdot(\dot{\varepsilon}^p - \dot{q}M)\Bigg\}\, d\Omega dt \\ & + \int_0^T \int_{\partial\Omega} (f\cdot v)\; dS\, dt,\end{aligned} \tag{A.2}$$

where the fields v, γ and ζ which correspond to the displacement, plastic hardening and plastic strain, respectively, are to be determined. Also, we have used the Kuhn-Tucker condition to replace $(2.16)_2$ to include the irreversibility of the accumulated plastic strain. We differentiate this augmented objective with respect to the parameters P,

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg\{ & \frac{\partial o}{\partial P} + \frac{\partial \rho}{\partial P}\ddot{u}\cdot v + \frac{\partial^2 W^e}{\partial\varepsilon\partial P}\cdot\nabla v + \gamma\dot{q}\left(\frac{\partial\bar{\sigma}_M}{\partial P} - \frac{\partial^2 W^p}{\partial q\partial P} - \frac{\partial^2 g^*}{\partial\dot{q}\partial P}\right) \\ & + \frac{\partial o}{\partial u}\delta_P u + \rho v\cdot\delta_P\ddot{u} + \left(\nabla v\cdot\frac{\partial^2 W^e}{\partial\varepsilon\partial\varepsilon} + \gamma\dot{q}\frac{\partial\bar{\sigma}_M}{\partial\varepsilon} - \dot{q}\zeta\cdot\frac{\partial M}{\partial\varepsilon}\right)\cdot\nabla\delta_P u \\ & + \left(\frac{\partial o}{\partial q} - \gamma\dot{q}\frac{\partial^2 W^p}{\partial q^2}\right)\delta_P q \\ & + \left(-\gamma\dot{q}\frac{\partial^2\bar{g}^*}{\partial\dot{q}^2} + \gamma\left[\sigma_M - \frac{\partial W^p}{\partial q} - \frac{\partial\bar{g}^*}{\partial\dot{q}}\right] - \zeta\cdot M\right)\delta_P\dot{q} \\ & + \zeta\cdot\delta_P\dot{\varepsilon}^p \\ & + \left(\frac{\partial o}{\partial\varepsilon^p} + \nabla v\cdot\frac{\partial^2 W^e}{\partial\varepsilon\partial\varepsilon^p} + \gamma\dot{q}\frac{\partial\bar{\sigma}_M}{\partial\varepsilon^p} - \dot{q}\zeta\cdot\frac{\partial M}{\partial\varepsilon^p}\right)\cdot\delta_P\varepsilon^p\Bigg\}\; d\Omega\; dt.\end{aligned} \tag{A.3}$$

We then integrate by parts, and enforce quiescent conditions on v at the final time T to remove the boundary terms. Then, localizing gives the sensitivities as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg\{ & \frac{\partial o}{\partial P} + \frac{\partial \rho}{\partial P}\ddot{u}\cdot v + \frac{\partial^2 W^e}{\partial \varepsilon \partial P}\cdot \nabla v \\ & + \gamma \dot{q}\left(\frac{\partial \bar{\sigma}_M}{\partial P} - \frac{\partial^2 W^p}{\partial q \partial P} - \frac{\partial^2 g^*}{\partial \dot{q}\partial P}\right)\Bigg\} d\Omega\, dt,\end{aligned} \tag{A.4}$$

if the adjoint variables satisfy the evolution

$$\begin{aligned}
0 = \int_\Omega \Bigg[& \rho \ddot{v}\cdot \delta_P u + \frac{\partial o}{\partial u}\cdot \delta_P u \\
& + \left(\nabla v \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon} + \gamma \dot{q}\frac{\partial \bar{\sigma}_M}{\partial \varepsilon} - \dot{q}\zeta \cdot \frac{\partial M}{\partial \varepsilon}\right)\cdot \nabla \delta_P u\Bigg]\, d\Omega \qquad \forall \delta_P u \in \mathcal{K}_0, \\
& \frac{\mathrm{d}}{\mathrm{d}t}\left[\gamma\left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}}\right) - \gamma \dot{q}\frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2} - \zeta \cdot M\right] = \frac{\partial o}{\partial q} - \gamma \dot{q}\frac{\partial^2 W^p}{\partial q^2} \quad \text{on } \Omega, \\
& \frac{\mathrm{d}\zeta}{\mathrm{d}t} = \frac{\partial o}{\partial \varepsilon^p} + \nabla v \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p} + \gamma \dot{q}\frac{\partial \bar{\sigma}_M}{\partial \varepsilon^p} - \dot{q}\zeta \cdot \frac{\partial M}{\partial \varepsilon^p} \qquad \text{on } \Omega, \\
& v|_{t=T} = 0, \quad \dot{v}|_{t=T} = 0, \quad \gamma|_{t=T} = 0, \quad \zeta|_{t=T} = 0,
\end{aligned} \tag{A.5}$$

where $\mathcal{K}_0 := \{\varphi \in H^1(\Omega),\ \varphi = 0 \text{ on } \partial_u \Omega\}$ is the space of kinematically admissible displacement variations. This is the weak form of (2.18).

A.2 Numerical method in elasto-viscoplasticity

We discuss the finite element discretization and numerical method to solve the forward and adjoint problems for both the quasistatic and dynamic setting. We consider a spatial discretization with standard $p = 1$ Lagrange polynomial shape functions for the displacements

$$u = \sum_{i=1}^{n_u} u_i N_i(x), \tag{A.6}$$

where $N_i : \Omega \mapsto \mathbb{R}^n$ are the standard vector valued shape functions with compact support. The fields q and ε^p are discretized at quadrature points as

$$q(x_g) = q_g, \qquad \varepsilon^p(x_g) = \varepsilon^p_g, \tag{A.7}$$

for some Gauss point x_g, $g = 1, \ldots, n_g$.

A.2.1 Quasistatic forward problem

The governing relations in this setting are

$$\begin{aligned}
0 &= \int_\Omega \left[\mathbb{C}\varepsilon^e \cdot \nabla\delta u - b\cdot\delta u\right]\, d\Omega - \int_{\partial_f\Omega} f\cdot\delta u\, dS && \forall \delta u \in \mathcal{K}_0,\\
0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot q) && \text{on } \Omega,\\
\dot\varepsilon^p &= \dot q M && \text{on } \Omega,\\
q|_{t=0} &= 0,\ \varepsilon^p|_{t=0} = 0
\end{aligned} \tag{A.8}$$

where $\mathcal{K}_0 := \{\varphi \in H^1(\Omega),\ \varphi = 0 \text{ on } \partial_u\Omega\}$ is the space of kinematically admissible displacement variations.

We consider a fully implicit approach, with a backwards Euler approximation of the temporal derivatives of the plastic variables. We examine this from the n to $n+1$ time step. That is, we look to solve for $\{u^{n+1}, q^{n+1}, \varepsilon^{p,n+1}\}$ given $\{u^n, q^n, \varepsilon^{p,n}\}$ assuming the discretizations in (A.6) and (A.7). Thus, we look to solve

$$\begin{aligned}
0 = R_i := \int_\Omega \Big[&\mathbb{C}\left(\varepsilon(u^{n+1}) - \varepsilon^{p,n+1}\right)\cdot\nabla N_i\\
&- b\cdot N_i\Big]\, d\Omega - \int_{\partial_f\Omega} f\cdot N_i\, dS && i = 1,\dots,n_u\\
0 \in \Big[&\sigma_M(\nabla u^{n+1}, \varepsilon^{p,n+1})\\
&- \frac{\partial W^p}{\partial q} - \partial\psi\left(\frac{q^{n+1}-q^n}{\Delta t}\right)\Big]_{x_g} && g = 1,\dots,n_g\\
\left[\frac{\varepsilon^{p,n+1}-\varepsilon^{p,n}}{\Delta t}\right]_{x_g} &= \left[\frac{q^{n+1}-q^n}{\Delta t} M(\nabla u^{n+1}, \varepsilon^{p,n+1})\right]_{x_g} && g = 1,\dots,n_g.
\end{aligned} \tag{A.9}$$

We solve this through a nested Newton-Raphson approach. From the unidirectional nature of M, we have $M(\nabla u^{n+1}, \varepsilon^{p,n+1}) = M(\nabla u^{n+1}, \varepsilon^{p,n})$. Then, from the last line (A.9), we may explicitly write $\varepsilon^{p,n+1} = \varepsilon^{p,n+1}(\nabla u^{n+1}, q^{n+1}, \varepsilon^{p,n}, q^n)$. We may then reduce the plastic updates to a single scalar yield equation from the second line of (A.9),

$$\begin{aligned}
0 \in \Big[&\sigma_M(\nabla u^{n+1}, \varepsilon^{p,n+1}(\nabla u^{n+1}, q^{n+1}, \varepsilon^{p,n}, q^n))\\
&- \frac{\partial W^p}{\partial q}\,(q^{n+1}) - \partial\psi\left(\frac{q^{n+1}-q^n}{\Delta t}\right)\Big]_{x_g}
\end{aligned} \tag{A.10}$$

for $g = 1, \ldots, n_g$. Then, given ∇u^{n+1}, $\varepsilon^{p,n}$, and q^n, the above equation is a scalar relation for q^{n+1} at each quadrature point. We solve this through a standard Newton-Raphson method. Then, from the solution of the above, we may write $q^{n+1} = q^{n+1}(\nabla u^{n+1}, \varepsilon^{p,n}, q^n)$ and therefore $\varepsilon^{p,n+1} = \varepsilon^{p,n+1}(\nabla u^{n+1}, \varepsilon^{p,n}, q^n)$. This allows us to reduce the entire system to

$$\begin{aligned} 0 = R_i := \int_\Omega \left[\mathbb{C} \left(\varepsilon(u^{n+1}) - \varepsilon^{p,n+1}(\nabla u^{n+1}, \varepsilon^{p,n}, q^n) \right) \cdot \nabla N_i - b \cdot N_i \right] \, d\Omega \\ - \int_{\partial_f \Omega} f \cdot N_i \, dS \end{aligned} \tag{A.11}$$

for $i = 1, \ldots, n_u$, where the plastic update relations are accounted for through the dependence of $\varepsilon^{p,n+1}$ on ∇u^{n+1}, $\varepsilon^{p,n}$ and q^n. We solve this system through a Newton-Raphson method. Thus, it is necessary to compute the stiffness matrix

$$K_{ij} := \int_\Omega \left[\mathbb{C} \left(\nabla N_j - \frac{\partial \varepsilon^{p,n+1}}{\partial \nabla u} \cdot \nabla N_j \right) \cdot \nabla N_i \right]] \, d\Omega \tag{A.12}$$

where

$$\frac{\partial \varepsilon^{p,n+1}}{\partial \nabla u} = M(\nabla u^{n+1}, \varepsilon^{p,n}) \otimes \frac{\partial q^{n+1}}{\partial \nabla u} + (q^{n+1} - q^n) \frac{\partial M}{\partial \nabla u}. \tag{A.13}$$

Here, the derivative $\frac{\partial q^{n+1}}{\partial \nabla u}$ is found through an implicit differentiation of (A.10) giving

$$\frac{\partial q^{n+1}}{\partial \nabla u} = \frac{2\mu}{\frac{\partial^2 W^p}{\partial q^2} + \frac{1}{\Delta t} \frac{\partial^2 \psi}{\partial \dot{q}^2} + 3\mu} M(\nabla u^{n+1}, \varepsilon^{p,n}). \tag{A.14}$$

Then, starting from $u^{n+1,1} = u^n$ and $p = 1$, we conduct Newton Raphson iterations over p in the form

$$\begin{aligned} K_{ij}(u^{n+1,p}) \Delta u_j &= -R_i(u^{n+1,p}), \qquad i = 1, \ldots, n_u, \ j = 1, \ldots, n_u \\ u_j^{n+1,p+1} &= \Delta u_j + u_j^{n+1,p}, \qquad j = 1, \ldots, n_u. \end{aligned} \tag{A.15}$$

until $|R(u^{n+1,p+1})| < tol$, and set $u^{n+1} = u^{n+1,p+1}$ and $q^{n+1} = q^{n+1}(\nabla u^{n+1}, \varepsilon^{p,n}, q^n)$ and $\varepsilon^{p,n+1} = \varepsilon^{p,n+1}(\nabla u^{n+1}, \varepsilon^{p,n}, q^n)$.

A.2.2 Dynamic forward problem

We specialize to the dynamic compression test where a disc (possibly with a hole) is compressed axially. See Section 2.3.2) and Appendix (A.4). The governing equations are (A.28).

We use an explicit central difference scheme to update the displacement field. The plasticity updates q and ε^p are then updated implicitly with a backwards Euler update. For the n to $n+1$ time-step the displacement updates are

$$\begin{aligned} \ddot{\bar{u}}_i^n &= M_{ij}^{-1} F_j^n(\bar{u}^n, \varepsilon^{p,n}, q^n, t^n), \\ \dot{\bar{u}}_i^{n+1/2} &= \dot{\bar{u}}_i^{n-1/2} + \Delta t^n \, \ddot{\bar{u}}_i^n, \\ \bar{u}_i^{n+1} &= \bar{u}_i^n + \Delta t^{n+1/2} \, \dot{\bar{u}}_i^{n+1/2}, \end{aligned} \tag{A.16}$$

where

$$M_{ij} = \int_\Omega \rho(x) N_i \cdot N_i \, d\Omega, \qquad F_j^n = \int_\Omega \left[-\sigma(\varepsilon^n, \varepsilon^{p,n}, \alpha^n, \eta) \cdot \nabla N_j + b \cdot N_j \right] d\Omega. \tag{A.17}$$

In standard fashion, the integrals above are approximated with Gauss quadrature. We then update the plasticity variables through an implicit backwards Euler discretization. For this, we employ a predictor-corrector scheme [8] to solve point-wise at each quadrature point,

$$\begin{aligned} &0 \in \bar{\sigma}_M(\varepsilon^{n+1}|_{x_g}, \varepsilon_g^{p,(n+1)}, \eta(x_g)) - \frac{\partial W^p}{\partial q} - \partial g^* \left(\frac{q_g^{n+1} - q_g^n}{\Delta t}, \eta(x_g) \right), \\ &\varepsilon_g^{p,(n+1)} = \varepsilon_g^{p,n} + \Delta q M(\varepsilon_g^{n+1}, \varepsilon_g^{p,(n+1)}). \end{aligned} \tag{A.18}$$

A.2.3 Quasistatic adjoint problem

The governing set of quasistatic adjoint relations are

$$\begin{aligned} 0 = \int_\Omega \Bigg[&\frac{\partial o}{\partial u} \cdot \delta_P u \\ &+ \left(\nabla v \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon} + \gamma \dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon} - \dot{q} \zeta \cdot \frac{\partial M}{\partial \varepsilon} \right) \cdot \nabla \delta_P u \Bigg] \, d\Omega \qquad && \forall \delta_P u \in \mathcal{U}, \\ &\frac{\mathrm{d}}{\mathrm{d}t} \left[\gamma \left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} \right) - \gamma \dot{q} \frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2} - \zeta \cdot M \right] = \frac{\partial o}{\partial q} - \gamma \dot{q} \frac{\partial^2 W^p}{\partial q^2} \qquad && \text{on } \Omega, \\ &\frac{\mathrm{d}\zeta}{\mathrm{d}t} = \frac{\partial o}{\partial \varepsilon^p} + \nabla v \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p} + \gamma \dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon^p} - \dot{q} \zeta \cdot \frac{\partial M}{\partial \varepsilon^p} \qquad && \text{on } \Omega, \\ &\gamma|_{t=T} = 0, \quad \zeta|_{t=T} = 0. \end{aligned} \tag{A.19}$$

As the boundary conditions are found at the final time $t = T$, we solve this system backwards in time. We consider a fully implicit approach, with a backward Euler approximation of the temporal derivatives of the plastic variables. We examine this from the $n+1$ to n time step. That is, we look to solve for $\{v^n, \gamma^n, \zeta^n\}$ given $\{v^{n+1}, \gamma^{n+1}, \zeta^{n+1}\}$ assuming the discretizations in (A.6)

and (A.7) given the complete set of forward-problem solution variables. Thus, we look to solve

$$\begin{aligned}
0 = R_i^{adj} = &\int_\Omega \left[\left.\frac{\partial o}{\partial u}\right|_{t_n} \cdot N_i + \left(\nabla v^n \cdot \left.\frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon}\right|_{t_n} \right.\right. \\
&\left.\left. + \gamma^n \left(\dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon} \right)_{t_n} - \zeta^n \cdot \left(\dot{q} \frac{\partial M}{\partial \varepsilon} \right)_{t_n} \right) \cdot \nabla N_i \right] d\Omega \\
\left.\frac{\partial o}{\partial q}\right|_{t_n} - \gamma^n \left(\dot{q} \frac{\partial^2 W^p}{\partial q^2} \right)_{t_n} = &\frac{1}{\Delta t} \left[\gamma \left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} \right) - \gamma \dot{q} \frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2} - \zeta \cdot M \right]_{t_n}^{t_{n+1}} \\
\frac{\zeta^{n+1} - \zeta_n}{\Delta t} = &\left.\frac{\partial o}{\partial \varepsilon^p}\right|_{t_n} + \nabla v^n \cdot \left.\frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p}\right|_{t_n} + \gamma^n \left(\dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon^p} \right)_{t_n} \\
&- \zeta^n \cdot \left(\dot{q} \frac{\partial M}{\partial \varepsilon^p} \right)_{t_n},
\end{aligned} \tag{A.20}$$

This is a linear set of equations for the $\{v^n, \gamma^n, \zeta^n\}$ and can be solved through direct inversion.

A.2.4 Dynamic adjoint problem

Similar to the forward problem, we apply an explicit central difference scheme to update the adjoint field v. The other adjoint variables, namely γ and ζ are solved locally using then updated implicitly with a backwards Euler update. For the n to $n+1$ time-step the displacement updates are

$$\begin{aligned}
\ddot{\bar{v}}_i^n &= M_{ij}^{-1} H_j^n(\bar{u}^n, \varepsilon^{p,n}, q^n, t^n), \\
\dot{\bar{v}}_i^{n+1/2} &= \dot{\bar{v}}_i^{n-1/2} + \Delta t^n \, \ddot{\bar{v}}_i^n, \\
\bar{v}_i^{n+1} &= \bar{v}_i^n + \Delta t^{n+1/2} \, \dot{\bar{v}}_i^{n+1/2},
\end{aligned} \tag{A.21}$$

where

$$M_{ij} = \int_\Omega \rho(x) N_i \cdot N_i \, d\Omega,$$

$$H_j^n = \int_\Omega \left[- \left(\nabla v^n \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon} + \gamma^n \dot{q}^n \frac{\partial \bar{\sigma}_M}{\partial \varepsilon} - \dot{q}^n \zeta^n \cdot \frac{\partial M}{\partial \varepsilon} \right) \cdot \nabla N_j \right. \tag{A.22}$$

$$\left. - \frac{\partial o}{\partial u} \cdot N_j \right] d\Omega. \tag{A.23}$$

In standard fashion, the integrals above are approximated with Gauss quadrature. We then update the adjoint variables γ and ζ through an implicit

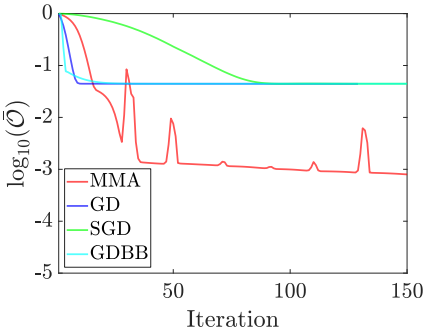


Figure A.1: Comparison of the performance of the method of moving asymptotes (MMA), gradient descent (GD), stochastic gradient descent (SG) and the Barzilai-Borwein method (GDBB) for the parameter update.

backwards Euler discretization. For this, we solve a linear set of equations point-wise at each quadrature point,

$$\begin{aligned}
&\gamma^{n+1}\left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} - \dot{q}\frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2}\right) - \zeta^{n+1}\cdot M = \gamma^n\left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} - \dot{q}\frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2}\right),\\
&-\zeta^n\cdot M + \Delta t\frac{\partial o}{\partial q} - \Delta t\gamma^n\dot{q}\frac{\partial^2 W^p}{\partial q^2},\\
&\zeta^{n+1} = \zeta^n + \frac{\partial o}{\partial \varepsilon^p} + \nabla v\cdot\frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p} + \gamma^n\dot{q}\frac{\partial \bar{\sigma}_M}{\partial \varepsilon^p} - \dot{q}\zeta^n\cdot\frac{\partial M}{\partial \varepsilon^p}.
\end{aligned} \tag{A.24}$$

A.3 Optimization

The gradient of the objective function is computed using the adjoint method. However, the parameter update is performed using method of moving asymptotes (MMA) as it allows for upper and lower bounds to be applied to the parameters. Furthermore, MMA automatically adjusts step sizes for each parameter using asymptotes, which prevents large and unstable updates. However, other methods like a gradient descent or stochastic gradient could also be used. We compare the results of MMA with gradient-descent, stochastic gradient descent and the Barzilai-Borwein method in Figure A.1. Both gradient descent techniques have a power-law decay of the step-size. The stochastic gradient-decent has a 5% noise added to the gradients at every step. The Barzilai-Borwein method is similar to the gradient-descent, with the distinction being that the step-size is chosen based on the previous iterations. We observe that MMA outperforms the other two methods, as the other methods get “stuck” in local minima.

A.4 Dimension reduction for annular specimen

The annulus ring being thin allows us to perform computations on a reduced dimensional space enabling quicker algorithm. Reiterating our approximation of uniform axial strain along the e_3 direction,

$$u(X_1, X_2, X_3, t) = \bar{u}(X_1, X_2, t) + (\lambda(t) - 1)X_3\ e_3, \tag{A.25}$$

where $\bar{u} := \bar{\Omega} \mapsto \mathbb{R}^2$ is the in-plane displacement. The corresponding deformation gradient is given by

$$\nabla u = \begin{pmatrix} \bar{\nabla}\bar{u} & 0 \\ 0 & \lambda \end{pmatrix}, \tag{A.26}$$

where $\bar{\nabla}$ is the gradient computed along the in-plane directions $\{e_1, e_2\}$. The governing equations are,

$$\begin{aligned} 0 &= \int_\Omega \left[\rho\ddot{u} \cdot \delta u + \sigma \cdot \nabla\delta u - b \cdot \delta u\right]\ d\Omega - \int_{\partial_f\Omega} f \cdot \delta u\ dS && \forall \delta u \in \mathcal{K}_0, \\ 0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot{q}) && \text{on } \Omega, \\ \dot{\varepsilon}^p &= \dot{q}M && \text{on } \Omega. \end{aligned} \tag{A.27}$$

Substituting the approximation of the displacement A.25 into the above equations, we obtain the simplified form

$$\begin{aligned} 0 &= -\frac{h^2||\Omega||}{3}\rho\ddot{\lambda} - \int_\Omega \sigma_{33} + \int_\Omega\ f \cdot e_3\ ds\ \Big|_{X_3=h}, \\ 0 &= \int_\Omega \left[\rho\ddot{\bar{u}} \cdot \delta\bar{u} + \sigma \cdot \nabla\delta\bar{u} - b \cdot \delta\bar{u}\right]\ d\Omega && \forall \delta\bar{u} \in \mathcal{K}_0, \\ 0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot{q}) && \text{on } \Omega, \\ \dot{\varepsilon}^p &= \dot{q}M && \text{on } \Omega. \end{aligned} \tag{A.28}$$

Since $\bar{u}$ exists in a 2-dimensional space, the computational cost for solving the above equations are cheaper than solving for a 3-dimensional system of equations. The dynamic boundary condition involves specifying $\lambda(t)$; therefor, the first equation is not necessary. The net force on the surface $X_3 = h$ can be calculated using

$$f = \int_\Omega \sigma_{33}\ d\Omega\ \Big|_{X_3=h}. \tag{A.29}$$

Our experimental data consists of the final in plane displacement $\bar{u}^{\exp}(X_1, X_2, T)$ and the net axial force $f_R^{\exp}$, which are compared with the inplane-displacements obtained from equation A.28 and the surface forces computed using A.29.

A p p e n d i x B

APPENDIX FOR CHAPTER 3

B.1 Numerical method for the governing equations

We discuss the finite element discretization and numerical schemes used to solve the forward problem for the cases of quasi-statics, dynamics, and dynamic rigid contact. For all of these, we consider a spatial discretization with standard $p = 1$ Lagrange polynomial shape functions for the displacements

$$u = \sum_{i=1}^{n_u} u_i N_i(x), \tag{B.1}$$

where $N_i : \Omega \mapsto \mathbb{R}^n$ are the standard vector valued shape functions with compact support. The fields q and ε^p are discretized at quadrature points as

$$q(x_g) = q_g, \qquad \varepsilon^p(x_g) = \varepsilon^p_g, \tag{B.2}$$

for some Gauss point x_g, $g = 1, \ldots, n_g$. Stating the evolution equations in equation 3.7, we need to solve the equations

$$\begin{aligned}
0 &= \int_\Omega \left[\rho \ddot{u} \cdot \delta u + \mathbb{C}\varepsilon^e \cdot \nabla \delta u - b \cdot \delta u\right] \, d\Omega - \int_{\partial_f \Omega} f \cdot \delta u \, dS \\
&\quad - \int_{\partial\Omega} \lambda \frac{\partial \mathcal{C}_I}{\partial u} \cdot \delta u \, dS && \forall \delta u \in \mathcal{K}_0, \\
0 &\in \sigma_M - \frac{\partial W^p}{\partial q} - \partial\psi(\dot{q}) && \text{on } \Omega, \\
\dot{\varepsilon}^p &= \dot{q} M && \text{on } \Omega, \\
0 &= \mathcal{C}_I(X + u, q_I) - \ell^2 && \text{on } \partial\Omega, \\
u|_{t=0} &= \dot{u}|_{t=0} = 0, \; q|_{t=0} = 0, \; \varepsilon^p|_{t=0} = 0.
\end{aligned} \tag{B.3}$$

The first equation is the dynamic evolution of the displacement field u. The second and third equations represent the kinetic relations for the plastic variables q, ε^p. We use an explicit central difference scheme to update the displacement field. The plasticity updates q and ε^p are then updated implicitly with a backwards Euler update. For the n to $n+1$ time-step the displacement updates are

$$\begin{aligned} \ddot{u}_i^n &= M_{ij}^{-1} F_j^n(u^n, \varepsilon^{p,n}, q^n, t^n), \\ \dot{u}_i^{n+1/2} &= \dot{u}_i^{n-1/2} + \Delta t^n \, \ddot{u}_i^n, \\ u_i^{n+1} &= u_i^n + \Delta t^{n+1/2} \, \dot{u}_i^{n+1/2}, \end{aligned} \tag{B.4}$$

where

$$M_{ij} = \int_\Omega \rho(x) N_i \cdot N_i \, d\Omega. \tag{B.5}$$

F_j^n is the term containing internal forces, body forces, tractions and contact forces, which is computed with a staggered predictor-corrector algorithm to compute F_j^n. First we assume contact forces are absent and predict the displacement field $u_i^{n,pre}$,

$$\begin{aligned} F_j^n &= \int_\Omega \left[-\mathbb{C}\varepsilon^e(u^n) \cdot \nabla N_j + b \cdot N_j \right] d\Omega + \int_{\partial_f \Omega} f \cdot N_j \, \partial S, \\ \ddot{u}_i^{n,pre} &= M_{ij}^{-1} F_j^n(u^n, \varepsilon^{p,n}, q^n, t^n), \\ \dot{u}_i^{n+1/2,pre} &= \dot{u}_i^{n-1/2} + \Delta t^n \, \ddot{u}_i^{n,pre}, \\ u_i^{n+1,pre} &= u_i^n + \Delta t^{n+1/2} \, \dot{u}_i^{n+1/2,pre}. \end{aligned} \tag{B.6}$$

If the contact condition is satisfied throughout $\partial\Omega$, the predicted displacement field is accepted. However, if the contact condition is not satisfied and penetration occurs, we need a correction in the displacement field. We first compute the penetration depth δ at each node,

$$\begin{aligned} x_i^{n+1,pre} &= x_i^n + u_i^{n+1,pre}, \\ \delta(x_i) &= \arg\min_\delta \left(\mathcal{C}_I(x_i^{n+1,pre} + \delta, q_I) - \ell^2 \right). \end{aligned} \tag{B.7}$$

The contact force attached with this penetration depth is given by

$$\begin{aligned} F_j^n &= \int_\Omega \left[-\mathbb{C}\varepsilon^e(u^n) \cdot \nabla N_j + b \cdot N_j \right] d\Omega + \int_{\partial_f \Omega} f \cdot N_j \, \partial S \\ &\quad + \int_{\partial\Omega} \rho(x) N_i \cdot N_i \frac{\delta(x)}{\Delta t^2} \cdot N_j \, \partial S, \\ \ddot{u}_i^n &= M_{ij}^{-1} F_j^n(u^n, \varepsilon^{p,n}, q^n, t^n), \\ \dot{u}_i^{n+1/2} &= \dot{u}_i^{n-1/2} + \Delta t^n \, \ddot{u}_i^n, \\ u_i^{n+1} &= u_i^n + \Delta t^{n+1/2} \, \dot{u}_i^{n+1/2}. \end{aligned} \tag{B.8}$$

The plastic variables are solved in standard fashion where the integrals approximated with Gauss quadrature. We then update the plasticity variables

through an implicit backwards Euler discretization. For this, we employ a predictor-corrector scheme [9] to solve point-wise at each quadrature point,

$$\begin{aligned} 0 &\in \bar{\sigma}_M(\varepsilon^{n+1}|_{x_g}, \varepsilon_g^{p,(n+1)}, \eta(x_g)) - \frac{\partial W^p}{\partial q} - \partial g^* \left(\frac{q_g^{n+1} - q_g^n}{\Delta t}, \eta(x_g) \right), \\ \varepsilon_g^{p,(n+1)} &= \varepsilon_g^{p,n} + \Delta q M(\varepsilon_g^{n+1}, \varepsilon_g^{p,(n+1)}). \end{aligned} \tag{B.9}$$

B.2 Optimization and adjoint formulation

B.2.1 Dynamic rigid contact

We consider an objective of integral from

$$\mathcal{O}(u, q, \varepsilon^p, \lambda, P) = \int_0^T \int_\Omega o(u, q, \varepsilon^p, \lambda, P)\, d\Omega. \tag{B.10}$$

We follow Section 3.2.2 to use the adjoint method to compute the sensitivity of $\mathcal{O}$ with respect to the parameters P. So we augment to objective using the governing equations

$$\begin{aligned} \mathcal{O} = \int_0^T \int_\Omega \Bigg\{ & o + \rho \ddot{u} \cdot v + \mathbb{C}\varepsilon^e \cdot \nabla v - b \cdot v \\ & + \gamma \dot{q} \left[\sigma_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} \right] + \zeta \cdot (\dot{\varepsilon}^p - \dot{q} M) \Bigg\} d\Omega dt \\ & + \int_0^T \int_{\partial\Omega_f} (f \cdot v)\ dS\, dt - \int_0^T \int_{\partial\Omega} \lambda \frac{\partial \mathcal{C}_I}{\partial u} \cdot v\ dS + \int_0^T \int_{\partial\Omega} \tau\lambda \left(\mathcal{C}_I - \ell^2 \right)\ dS dt, \end{aligned} \tag{B.11}$$

where the fields v, γ, ζ and τ which correspond to the displacement, plastic hardening, plastic strain and lagrange multiplier for contact, respectively, are to be determined. Also, we have used the Kuhn-Tucker condition to replace $(3.7)_2$ to include the irreversibility of the accumulated plastic strain. We differentiate this augmented objective with respect to the parameters P,

$$\begin{aligned} \frac{\mathrm{d}\mathcal{O}}{\mathrm{d}P} = \int_0^T \int_\Omega \Bigg\{ & \frac{\partial o}{\partial P} + \frac{\partial \rho}{\partial P} \ddot{u} \cdot v + \frac{\partial \mathbb{C}}{\partial P} \varepsilon^e \cdot \nabla v + \gamma \dot{q} \left(\frac{\partial \bar{\sigma}_M}{\partial P} - \frac{\partial^2 W^p}{\partial q \partial P} - \frac{\partial^2 g^*}{\partial \dot{q} \partial P} \right) \\ & + \frac{\partial o}{\partial u} \delta_P u + \rho v \cdot \delta_P \ddot{u} + \left(\nabla v \cdot \mathbb{C} + \gamma \dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon} - \dot{q} \zeta \cdot \frac{\partial M}{\partial \varepsilon} \right) \cdot \nabla \delta_P u \\ & + \left(\frac{\partial o}{\partial q} - \gamma \dot{q} \frac{\partial^2 W^p}{\partial q^2} \right) \delta_P q + \left(-\gamma \dot{q} \frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2} + \gamma \left[\sigma_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} \right] - \zeta \cdot M \right) \delta_P \dot{q} \\ & + \zeta \cdot \delta_P \dot{\varepsilon}^p \\ & + \left(\frac{\partial o}{\partial \varepsilon^p} + \nabla v \cdot \frac{\partial^2 W^e}{\partial \varepsilon \partial \varepsilon^p} + \gamma \dot{q} \frac{\partial \bar{\sigma}_M}{\partial \varepsilon^p} - \dot{q} \zeta \cdot \frac{\partial M}{\partial \varepsilon^p} \right) \cdot \delta_P \varepsilon^p \Bigg\}\ d\Omega\ dt \\ & + \int_0^T \int_{\partial\Omega} \left(\tau\lambda \frac{\partial \mathcal{C}_I}{\partial u} - \lambda \frac{\partial^2 \mathcal{C}_I}{\partial u \partial u} \right) \cdot \delta_P u\ dS\ dt + \int_0^T \int_{\partial\Omega_{\lambda \neq 0}} \left(\frac{\partial \mathcal{C}_I}{\partial u} \cdot v - \frac{\partial o}{\partial \lambda} \right) \cdot \delta_P \lambda\ dS\ dt. \end{aligned} \tag{B.12}$$

Following this, we integrate by parts, and enforce quiescent conditions on v at the final time T to remove boundary terms. This leads to the adjoint problem

$$
\begin{aligned}
0 = &\int_\Omega \left[\rho\ddot{v}\cdot\delta u + \left(\mathbb{C}\nabla v + \gamma\dot{q}\frac{\partial\sigma_M}{\partial\varepsilon} - \dot{q}\mu\cdot\frac{\partial M}{\partial\varepsilon}\right)\cdot\nabla\delta u \right.\\
&\left. + \frac{\partial o}{\partial u}\cdot\delta u\right] d\Omega + \int_{\partial\Omega}\left(b\lambda\frac{\partial\mathcal{C}_I}{\partial u} - \lambda v\cdot\frac{\partial^2\mathcal{C}_I}{\partial u\partial u}\right)\cdot\delta u\; dS \quad \forall\delta u\in\mathcal{K}_0\\
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\gamma\left(\sigma_M - \sigma_0 - \frac{\partial g^*}{\partial\dot{q}}\right) - \gamma\dot{q}\frac{\partial^2 g^*}{\partial\dot{q}^2} - \mu\cdot M\right] = -\gamma\dot{q}\frac{\partial\sigma_0}{\partial q} \qquad \text{on } \Omega\\
&\frac{\mathrm{d}\mu}{\mathrm{d}t} = \nabla\xi\cdot\frac{\partial^2 W^e}{\partial\varepsilon\partial\varepsilon^p} + \gamma\dot{q}\frac{\partial\sigma_M}{\partial\varepsilon^p} - \dot{q}\mu\cdot\frac{\partial M}{\partial\varepsilon^p} \qquad \text{on } \Omega\\
&0 = \frac{\partial C_I}{\partial u}\cdot v - \frac{\partial o}{\partial\lambda} \qquad \text{on } \partial\Omega_{\lambda\neq 0}\\
&v|_{t=T} = \dot{v}|_{t=T} = 0,\ \gamma|_{t=T} = 0, \quad \mu|_{t=T} = 0,
\end{aligned}
$$

where $\mathcal{K}_0 := \{\varphi\in H^1(\Omega),\ \varphi = 0 \text{ on } \partial_u\Omega\}$ is the space of kinematically admissible displacement variations.

B.3 Numerical method for the adjoint problem

B.3.1 Dynamic contact

Since the boundary conditions for the adjoint variables are specified at final time T, we solve this system backwards. The adjoint variable v is updated using explicitly and γ, ζ are updated implicitly. For the n to $n+1$ time-step the adjoint updates are

$$
\begin{aligned}
\ddot{v}_i^n &= M_{ij}^{-1}H_j^n(u^n,\varepsilon^{p,n},q^n,t^n),\\
\dot{v}_i^{n+1/2} &= \dot{v}_i^{n-1/2} + \Delta t^n\,\ddot{v}_i^n,\\
v_i^{n+1} &= v_i^n + \Delta t^{n+1/2}\,\dot{v}_i^{n+1/2},
\end{aligned}
\tag{B.13}
$$

where

$$
M_{ij} = \int_\Omega \rho(x) N_i\cdot N_i d\Omega. \tag{B.14}
$$

We then update the adjoint variables γ and ζ through an implicit backwards Euler discretization. For this, we solve a linear set of equations point-wise at each quadrature point,

$$
\begin{aligned}
&\gamma^{n+1}\left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} - \dot{q}\frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2}\right) - \zeta^{n+1}\cdot M = \gamma^n\left(\bar{\sigma}_M - \frac{\partial W^p}{\partial q} - \frac{\partial \bar{g}^*}{\partial \dot{q}} - \dot{q}\frac{\partial^2 \bar{g}^*}{\partial \dot{q}^2}\right),\\
&-\zeta^n\cdot M + \Delta t\frac{\partial o}{\partial q} - \Delta t\gamma^n\dot{q}\frac{\partial^2 W^p}{\partial q^2},\\
&\zeta^{n+1} = \zeta^n + \frac{\partial o}{\partial \varepsilon^p} + \nabla v\cdot\frac{\partial^2 W^e}{\partial\varepsilon\partial\varepsilon^p} + \gamma^n\dot{q}\frac{\partial\bar{\sigma}_M}{\partial\varepsilon^p} - \dot{q}\zeta^n\cdot\frac{\partial M}{\partial\varepsilon^p}.
\end{aligned}
\tag{B.15}
$$

H_j^n is the equation B.13 is computed with a staggered predictor-corrector algorithm. First we assume contact forces are absent and predict the displacement field $u_i^{n,pre}$,

$$
\begin{aligned}
H_j^n &= \int_\Omega\left[-\left(\nabla v^n\cdot\mathbb{C} + \gamma^n\dot{q}^n\frac{\partial\bar{\sigma}_M}{\partial\varepsilon} - \dot{q}^n\zeta^n\cdot\frac{\partial M}{\partial\varepsilon}\right)\cdot\nabla N_j - \frac{\partial o}{\partial u}\cdot N_j\right]d\Omega\\
&\quad + \int_{\partial\Omega}\left(-\lambda^n v^n\cdot\frac{\partial^2\mathcal{C}_I}{\partial u\partial u}\right)\cdot N_j\ dS\\
\ddot{v}_i^{n,pre} &= M_{ij}^{-1}H_j^n(u^n,\varepsilon^{p,n},q^n,t^n),\\
\dot{v}_i^{n+1/2,pre} &= \dot{v}_i^{n-1/2} + \Delta t^n\,\ddot{v}_i^{n,pre},\\
v_i^{n+1,pre} &= u_i^n + \Delta t^{n+1/2}\,\dot{v}_i^{n+1/2,pre}.
\end{aligned}
\tag{B.16}
$$

Following this we check if the condition,

$$
0 = \frac{\partial C_I}{\partial u}\cdot v - \frac{\partial o}{\partial\lambda},
\tag{B.17}
$$

is satisfied on $\partial\Omega_{\lambda\neq 0}$. If the constraint is not satisfied, we correct H_j^n by solving for the lagrange multiplier $b\lambda$,

$$
\upsilon(x_i) = \arg\min_\delta\ \left(\frac{\partial C_I}{\partial u}\cdot v - \frac{\partial o}{\partial\lambda}\right).
\tag{B.18}
$$

The contact force attached with this penetration depth is given by

$$
\begin{aligned}
H_j^n &= \int_\Omega\left[-\left(\nabla v^n\cdot\mathbb{C} + \gamma^n\dot{q}^n\frac{\partial\bar{\sigma}_M}{\partial\varepsilon} - \dot{q}^n\zeta^n\cdot\frac{\partial M}{\partial\varepsilon}\right)\cdot\nabla N_j - \frac{\partial o}{\partial u}\cdot N_j\right]d\Omega\\
&\quad + \int_{\partial\Omega}\left(-\lambda^n v^n\cdot\frac{\partial^2\mathcal{C}_I}{\partial u\partial u}\right)\cdot N_j\ dS + \int_{\partial\Omega_{\lambda\neq 0}}\upsilon(x_i)\frac{\partial\mathcal{C}_I}{\partial u}\cdot N_j\ dS\\
\ddot{u}_i^n &= M_{ij}^{-1}F_j^n(u^n,\varepsilon^{p,n},q^n,t^n),\\
\dot{u}_i^{n+1/2} &= \dot{u}_i^{n-1/2} + \Delta t^n\,\ddot{u}_i^n,\\
u_i^{n+1} &= u_i^n + \Delta t^{n+1/2}\,\dot{u}_i^{n+1/2}.
\end{aligned}
\tag{B.19}
$$

In standard fashion, the integrals above are approximated with Gauss quadrature.

A p p e n d i x C

APPENDIX FOR CHAPTER 4

C.1 Dimension reduction for annular specimen

The annulus ring being thin allows us to perform computations on a reduced dimensional space enabling quicker algorithm. Reiterating our approximation of uniform axial strain along the e_3 direction,

$$u(X_1, X_2, X_3, t) = \bar{u}(X_1, X_2, t) + (\lambda(t) - 1)X_3 \; e_3, \tag{C.1}$$

where $\bar{u} := \bar{\Omega} \mapsto \mathbb{R}^2$ is the in-plane displacement. The corresponding deformation gradient is given by

$$\nabla u = \begin{pmatrix} \bar{\nabla}\bar{u} & 0 \\ 0 & \lambda \end{pmatrix}, \tag{C.2}$$

where $\bar{\nabla}$ is the gradient computed along the in-plane directions $\{e_1, e_2\}$. The governing equations are,

$$\begin{aligned} 0 &= \int_\Omega \left[\rho\ddot{u} \cdot \delta u + \sigma \cdot \nabla\delta u - b \cdot \delta u\right] \; d\Omega - \int_{\partial_f\Omega} f \cdot \delta u \; dS \qquad && \forall \delta u \in \mathcal{K}_0, \\ \dot{\xi} &= G(\nabla u, \xi) && \text{on } \Omega. \end{aligned} \tag{C.3}$$

Substituting the approximation of the displacement C.1 into the above equations, we obtain the simplified form

$$\begin{aligned} 0 &= -\frac{h^2||\Omega||}{3}\rho\ddot{\lambda} - \int_\Omega \sigma_{33} + \int_\Omega \; f \cdot e_3 \; ds \; \Big|_{X_3=h}, \\ 0 &= \int_\Omega \left[\rho\ddot{\bar{u}} \cdot \delta\bar{u} + \sigma \cdot \nabla\delta\bar{u} - b \cdot \delta\bar{u}\right] \; d\Omega \qquad && \forall \delta\bar{u} \in \mathcal{K}_0, \\ \dot{\xi} &= G(\nabla u, \xi) && \text{on } \Omega. \end{aligned} \tag{C.4}$$

Since $\bar{u}$ exists in a 2-dimensional space, the computational cost to solve the above equations is cheaper than solving a 3-dimensional system of equations. The dynamic boundary condition involves specifying $\lambda(t)$; therefore, the first equation is not necessary. The net force on the surface $X_3 = h$ can be calculated using

$$f = \int_\Omega \sigma_{33} \; d\Omega \; \Big|_{X_3=h}. \tag{C.5}$$

Our experimental data consists of the final in plane displacement $\bar{u}^{\exp}(X_1, X_2, T)$ and the net axial force $f_R^{\exp}$, which are compared with the inplane-displacements obtained from equation C.4 and the surface forces computed using C.5.

C.2 Numerical method

We discuss the finite element discretization and the numerical method to solve the forward and adjoint problems for both the quasistatic and dynamic setting. We consider a spatial discretization with standard $p = 1$ Lagrange polynomial shape functions for the displacements

$$\bar{u} = \sum_{i=1}^{n_u} \bar{u}_i N_i(x), \tag{C.6}$$

where $N_i : \Omega \mapsto \mathbb{R}^n$ are the standard vector valued shape functions with compact support. The internal variable field ξ is discretized at quadrature points as

$$\xi(x_g) = \xi_g, \tag{C.7}$$

for some Gauss point x_g, $g = 1, \ldots, n_g$.

C.2.1 Forward problem

The governing equations for the compression of the thin annulus are given in equation (C.4). We use an explicit central difference scheme to update the displacement field and the internal variables. For the n to $n+1$ time-step the displacement updates are

$$\begin{aligned} \ddot{\bar{u}}_i^n &= M_{ij}^{-1} F_j^n(\bar{u}^n, \xi^n, t^n), \\ \dot{\bar{u}}_i^{n+1/2} &= \dot{\bar{u}}_i^{n-1/2} + \Delta t^n \, \ddot{\bar{u}}_i^n, \\ \bar{u}_i^{n+1} &= \bar{u}_i^n + \Delta t^{n+1/2} \, \dot{\bar{u}}_i^{n+1/2}, \end{aligned} \tag{C.8}$$

where

$$M_{ij} = \int_\Omega \rho(x) N_i \cdot N_i \, d\Omega, \qquad F_j^n = \int_\Omega \left[-\sigma(\nabla u^n, \varepsilon^{p,n}, \xi) \cdot \nabla N_j + b \cdot N_j \right] d\Omega. \tag{C.9}$$

In standard fashion, the integrals above are approximated with Gauss quadrature. We then update the internal variables through an explicit backwards Euler discretization,

$$\xi_g^{n+1} = \xi_g^n + \Delta t \; G_{\mathrm{NN}} \left(\nabla u(x_g)^n, \xi_g^n \right). \tag{C.10}$$

C.2.2 Adjoint problem

Similar to the forward problem, we apply an explicit central difference scheme to update the adjoint field v. The other adjoint variable, namely ϕ, is solved locally using a backward Euler update. For the n to $n+1$ time-step the displacement updates are

$$\begin{aligned} \ddot{\bar{v}}_i^n &= M_{ij}^{-1} H_j^n(\bar{u}^n, \xi^n, \phi^n, t^n), \\ \dot{\bar{v}}_i^{n+1/2} &= \dot{\bar{v}}_i^{n-1/2} + \Delta t^n \, \ddot{\bar{v}}_i^n, \\ \bar{v}_i^{n+1} &= \bar{v}_i^n + \Delta t^{n+1/2} \, \dot{\bar{v}}_i^{n+1/2}, \end{aligned} \tag{C.11}$$

where

$$\begin{aligned} M_{ij} &= \int_\Omega \rho(x) N_i \cdot N_i \, d\Omega, \\ H_j^n &= \int_\Omega \left[- \left(\nabla v^n \cdot \left(\frac{\partial \sigma}{\partial \nabla u} \right)^n - \phi^n \left(\frac{\partial G}{\partial \nabla u} \right)^n \right) \cdot \nabla N_j - \frac{\partial o}{\partial u} \cdot N_j \right] d\Omega. \end{aligned} \tag{C.12}$$

In standard fashion, the integrals above are approximated with Gauss quadrature. We then update the adjoint variables ϕ through an explicit backwards Euler discretization point-wise at each quadrature point,

$$\phi_g^{n+1} = \phi_g^n + \Delta t \, \left(\nabla v(x_g)^n \cdot \left(\frac{\partial \sigma}{\partial \xi} \right)_g^n - \left(\frac{\partial G_{\text{NN}}}{\partial \xi} \right)_g^n - \frac{\partial o}{\partial \xi} \right). \tag{C.13}$$

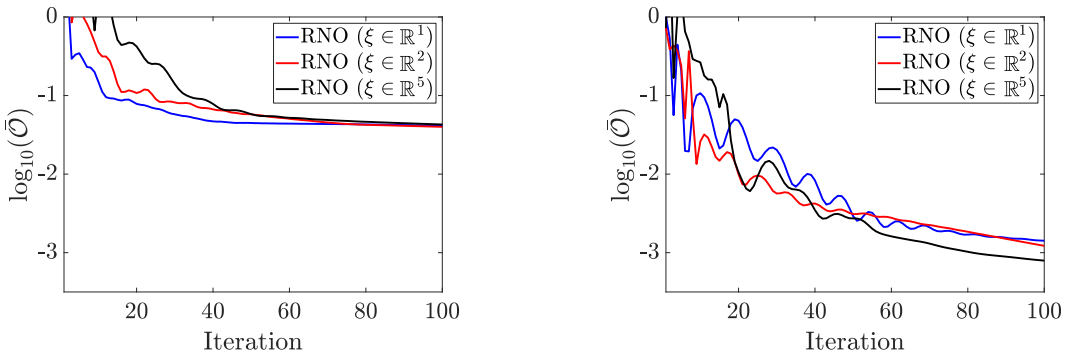
The derivatives for σ for the RNO constitutive (refer to equation 4.1) is given by

$$\frac{\partial \sigma}{\partial \nabla u} = \mathbb{C} I + \frac{\partial S_{\text{NN}}}{\partial \nabla u}, \quad \frac{\partial \sigma}{\partial \xi} = \frac{\partial S_{\text{NN}}}{\partial \xi}. \tag{C.14}$$

The derivatives of the neural networks S_{NN} and G_{NN} are computed using Pytorch's autograd function.

C.3 Extra set of results for elliptic configuration

We present additional results for the elliptical configuration. Interestingly, we observe that the choice of the objective weighting ratio α_f/α_u (as seen in equation 4.11) plays a critical role in the model's ability to benefit from increased internal state dimensionality. Specifically, when $\alpha_f/\alpha_u = 0.1$, the final objective remains nearly unchanged as the dimension of ξ increases, suggesting that the displacement data is hard to learn and saturates the inversion process (refer to figure C.1). In contrast, when $\alpha_f/\alpha_u = 10.0$, the objective decreases



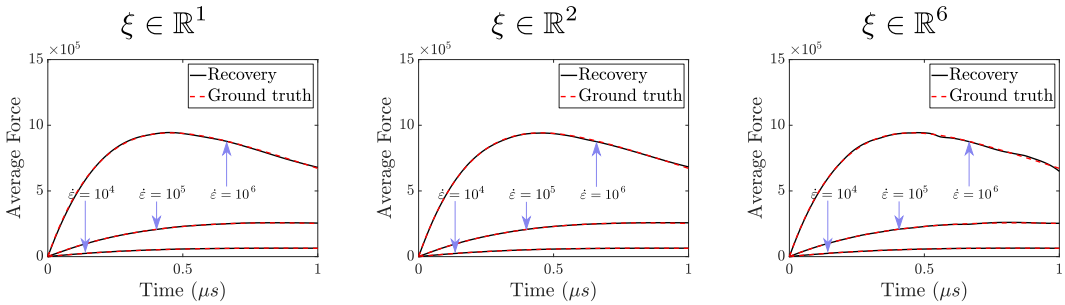
(a) $\alpha_f/\alpha_u = 0.1$ in equation 4.11

(b) $\alpha_f/\alpha_u = 10.0$ in equation 4.11

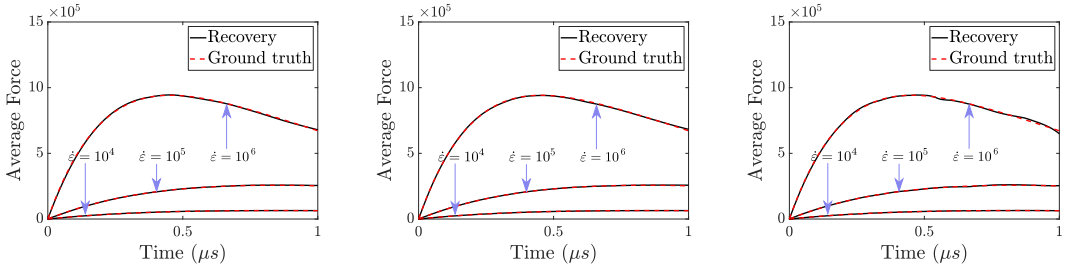
Figure C.1: Objective versus iteration for three different RNOs for the dynamic compression of elliptical hole configuration.

noticeably with higher-dimensional internal states, indicating that the force data is more informative for learning the latent structure. This highlights the importance of balancing the data modalities to effectively utilize the model's capacity.

While the current experimental setup offers a useful testbed for evaluating the RNO framework, it is not ideal for full material characterization. Due to the axisymmetry of the loading and geometry, the in-plane displacement fields obtained from DIC-like measurements are relatively low-rank and lack the richness needed to constrain the constitutive response fully. As a result, the displacement data does not significantly improve recovery of the internal material parameters, especially in configurations where force data is weighted less heavily. This limitation underscores the need for more informative two-dimensional deformation patterns—such as those arising from shear-dominated or asymmetric loading conditions—which can more effectively excite and probe the underlying constitutive behavior. Developing such 2D experimental configurations is a promising direction for future work.



(a) $\alpha_f/\alpha_u = 0.1$ in equation 4.11



(b) $\alpha_f/\alpha_u = 10.0$ in equation 4.11

Figure C.2: Force versus time for three different RNOs for the dynamic compression of elliptical hole configuration.

Appendix D

APPENDIX FOR CHAPTER 5

D.1 Proofs of stability estimates

In this section, we prove the stability estimates stated in Section 5.1.3. The following lemma is a modification of the standard estimate for parametric dependence of elliptic equations on their coefficient. We include it here for completeness.

Proposition 1. *Consider the cell problem defined by equation* (5.4)*. The following hold:*

1. *If $A \in \mathsf{PD}_{\alpha,\beta}$, then* (5.4) *has a unique solution $\chi \in \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ and*
$$\|\chi\|_{\dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)} \leq \frac{\sqrt{d}\beta}{\alpha}.$$
2. *For $\chi^{(1)}$ and $\chi^{(2)}$ solutions to the cell problem in equation* (5.4) *associated with coefficients $A^{(1)}, A^{(2)} \in \mathsf{PD}_{\alpha,\beta}$, respectively, it follows that*
$$\|\chi^{(2)} - \chi^{(1)}\|_{\dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)} \leq \frac{\sqrt{d}}{\alpha}\left(1 + \frac{\beta}{\alpha}\right)\|A^{(1)} - A^{(2)}\|_{L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})}. \tag{5.8}$$

Proof. For existence and uniqueness of the solution to the cell problem using Lax-Milgram, we refer to the texts [3, 10]; we simply derive the bounds and stability estimate. First, note that (5.4) decouples, in particular,

$$-\nabla\cdot(A\nabla\chi_\ell) = \nabla\cdot Ae_\ell, \qquad y \in \mathbb{T}^d \tag{D.1}$$

for $l = 1,\dots,d$ where e_ℓ is the ℓ-th standard basis vector of $\mathbb{R}^d$ and each $\chi_\ell \in \dot{H}^1(\mathbb{T}^d;\mathbb{R})$. Multiplying by χ_ℓ and integrating by parts shows

$$\begin{aligned}
\alpha\|\nabla\chi_\ell\|_{L^2}^2 &\leq \int_{\mathbb{T}^d}\langle A\nabla\chi_\ell, \nabla\chi_\ell\rangle \,\mathsf{d}y \\
&= -\int_{\mathbb{T}^d}\langle Ae_\ell, \nabla\chi_\ell\rangle \,\mathsf{d}y \\
&\leq \int_{\mathbb{T}^d}|Ae_\ell||\nabla\chi_\ell| \,\mathsf{d}y \\
&\leq \left(\int_{\mathbb{T}^d}|Ae_\ell|^2 \,\mathsf{d}y\right)^{\frac{1}{2}}\left(\int_{\mathbb{T}^d}|\nabla\chi_\ell|^2 \,\mathsf{d}y\right)^{\frac{1}{2}} \\
&\leq \|A\|_{L^\infty}\|\nabla\chi_\ell\|_{L^2}.
\end{aligned}$$

Therefore

$$\|\nabla\chi\|_{L^2}^2 = \sum_{l=1}^{d} \|\nabla\chi_\ell\|_{L^2}^2 \leq \frac{d\|A\|_{L^\infty}^2}{\alpha^2} \leq \frac{d\beta^2}{\alpha^2},$$

which implies the first result.

To prove the second result, we denote the right hand side of D.1 by $f_\ell^{(i)} = \nabla \cdot A^{(i)} e_\ell$ in what follows. For any $v \in \dot{H}^1(\mathbb{T}^d;\mathbb{R})$, we have that

$$-\int_{\mathbb{T}^d} \nabla\cdot(A^{(1)}\nabla\chi_\ell^{(1)})v \ \mathsf{d}y = \int_{\mathbb{T}^d} f_\ell^{(1)} v \ \mathsf{d}y$$
$$-\int_{\partial\mathbb{T}^d} vA^{(1)}\nabla\chi_\ell^{(1)}\cdot\widehat{n} \ \mathsf{d}y + \int_{\mathbb{T}^d} \nabla v\cdot A^{(1)}\nabla\chi_\ell^{(1)} \ \mathsf{d}y = \int_{\mathbb{T}^d} f_\ell^{(1)} v \ \mathsf{d}y.$$

Since v, $A^{(1)}$, and the solution $\chi_\ell^{(1)}$ are all periodic on $\mathbb{T}^d$, the first term is 0. Combining with the equation for $\chi_\ell^{(2)}$, we get

$$\int_{\mathbb{T}^d} \nabla v\cdot\left(A^{(1)}-A^{(2)}\right)\nabla\chi_\ell^{(1)} \ \mathsf{d}y =$$
$$= \int_{\mathbb{T}^d} (f_\ell^{(1)} - f_\ell^{(2)})v + \nabla v\cdot\left(A^{(2)}\left(\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\right)\right) \mathsf{d}y.$$

Setting $v = \chi_\ell^{(2)} - \chi_\ell^{(1)}$, we have

$$\int_{\mathbb{T}^d}\left(\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\right)\cdot\left(\left(A^{(1)}-A^{(2)}\right)\nabla\chi_\ell^{(1)}\right) \mathsf{d}y = \int_{\mathbb{T}^d}(f_\ell^{(1)}-f_\ell^{(2)})\left(\chi_\ell^{(2)}-\chi_\ell^{(1)}\right) \mathsf{d}y$$
$$+\int_{\mathbb{T}^d}\left(\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\right)\cdot\left(A^{(2)}\left(\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\right)\right) \mathsf{d}y,$$
$$\alpha\|\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\|_{L^2}^2 \leq \|A^{(1)}-A^{(2)}\|_{L^\infty}\|\nabla\chi_\ell^{(1)}\|_{L^2}\|\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\|_{L^2}$$
$$+\|f_\ell^{(1)}-f_\ell^{(2)}\|_{\dot{H}^{-1}}\|\nabla\chi_\ell^{(2)}-\nabla\chi_\ell^{(1)}\|_{L^2},$$

$$\|\chi_\ell^{(2)}-\chi_\ell^{(1)}\|_{\dot{H}^1} \leq \frac{1}{\alpha}\left(\|A^{(1)}-A^{(2)}\|_{L^\infty}\|\nabla\chi_\ell^{(1)}\|_{L^2} + \|f_\ell^{(1)}-f_\ell^{(2)}\|_{\dot{H}^{-1}}\right). \quad \text{(D.2)}$$

Evaluating,

$$\|f_\ell^{(1)}-f_\ell^{(2)}\|_{\dot{H}^{-1}} = \|\nabla\cdot A^{(1)}e_\ell - \nabla\cdot A^{(2)}e_\ell\|_{\dot{H}^{-1}}, \quad \text{(D.3)}$$
$$= \sup_{\|\xi\|_{\dot{H}^1}=1}\int_{\mathbb{T}^d}\xi\nabla\cdot(A^{(1)}-A^{(2)})e_\ell \ \mathsf{d}y, \quad \text{(D.4)}$$
$$\leq \sup_{\|\xi\|_{\dot{H}^1}=1}\|(A^{(1)}-A^{(2)})e_\ell\|_{L^2}\|\nabla\xi\|_{L^2}, \quad \text{(D.5)}$$
$$\leq \|A^{(1)}-A^{(2)}\|_{L^2} \leq \|A^{(1)}-A^{(2)}\|_{L^\infty} \quad \text{(D.6)}$$

since our domain is $\mathbb{T}^d$. Combining this with (D.2) and the bound of $\|\nabla\chi_\ell\|_{L^2} \leq \frac{\beta}{\alpha}$ obtained in the first part of this proposition, we have

$$\|\chi_\ell^{(2)}-\chi_\ell^{(1)}\|_{\dot{H}^1} \leq \frac{1}{\alpha}\left(1+\frac{\beta}{\alpha}\right)\|A^{(1)}-A^{(2)}\|_{L^\infty}. \quad \text{(D.7)}$$

Returning to d vector components yields the result.

The following result shows that the mapping $A \mapsto \bar{A}$ is continuous on separable subspaces of $L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$.

Lemma 1. *Let $\mathcal{A} \subset L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$ be a separable subspace and $K \subset \mathcal{A}\cap \mathsf{PD}_{\alpha,\beta}$ a closed set in L^∞. Define the mapping $F : K \to \mathbb{R}^{d\times d}$ by $A \mapsto \bar{A}$ as given by (5.3). Then there exists a continuous mapping $\mathcal{F} \in C(\mathcal{A};\mathbb{R}^{d\times d})$ such that $\mathcal{F}(A) = F(A)$ for any $A \in K$.*

Proof. Let $A^{(1)}, A^{(2)} \in K$ then, by Proposition 1,

$$\begin{aligned}
\left|F(A^{(1)}) - F(A^{(2)})\right|_F &\le \int_{\mathbb{T}^d} |A^{(1)} - A^{(2)}|_F\big(1 + |\nabla\chi^{(1)}|_F\big)\,\mathsf{d}y \\
&\quad + \int_{\mathbb{T}^d} |A^{(2)}|_F |\nabla\chi^{(1)} - \nabla\chi^{(2)}|_F\,\mathsf{d}y \\
&\le \|A^{(1)} - A^{(2)}\|_{L^\infty}\big(1 + \|\nabla\chi^{(1)}\|_{L^2}\big) + \|A^{(2)}\|_{L^\infty}\|\nabla\chi^{(1)} - \nabla\chi^{(2)}\|_{L^2} \\
&\le \left(1 + \frac{\sqrt{d}}{\alpha}\left(\|A^{(1)}\|_{L^\infty} + \|A^{(2)}\|_{L^\infty}\left(\frac{\min\left(\|A^{(1)}\|_{L^\infty}, \|A^{(2)}\|_{L^\infty}\right)}{\alpha} + 1\right)\right)\right) \\
&\quad \cdot \|A^{(1)} - A^{(2)}\|_{L^\infty},
\end{aligned}$$

hence $F \in C(K;\mathbb{R}^{d\times d})$. Applying the Tietze extension theorem [5] to F implies the existence of $\mathcal{F}$.

The following lemma shows that, unfortunately, separable subspaces of $L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$ are not very useful. Indeed, in the desired area of application of continuum mechanics, we ought to be able to place a boundary of material discontinuity anywhere in the domain. The following result shows that doing so is impossible for a subset of $\mathsf{PD}_{\alpha,\beta}$ which lies only in a separable subspace of $L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$.

Lemma 2. *For any $t \in [0,1]$ define $c_t : [0,1] \to \mathbb{R}$ by*

$$c_t(x) = \begin{cases} 1, & x \le t \\ 0, & x > t \end{cases}, \qquad \forall\, x \in [0,1].$$

Define $E = \{c_t : t \in [0,1]\} \subset L^\infty([0,1])$. There exists no separable subspace $\mathcal{A} \subset L^\infty([0,1])$ such that $E \subseteq \mathcal{A}$.

Proof. Suppose otherwise. Since $(\mathcal{A}, \|\cdot\|_{L^\infty})$ is a separable metric space, $(E, \|\cdot\|_{L^\infty})$ must be separable since $E \subseteq \mathcal{A}$; this is a contradiction since $(E, \|\cdot\|_{L^\infty})$ is not separable. To see this, let $\{c_{t_j}\}_{j=1}^\infty$ be an arbitrary countable susbset of E. Then for any $t \notin \{t_j\}_{j=1}^\infty$, we have,

$$\inf_{\{t_j\}_{j=1}^\infty} \|c_t - c_{t_j}\|_{L^\infty} = 1.$$

Hence no countable subset can be dense.

Instead of working on a compact subset of a separable subspace of $L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$, we may instead try to find a suitable probability measure which contains the discontinous functions of interest. The following remarks makes clear why such an approch would still be problematic for the purposes of approximation.

Remark 9 (Gaussian Threshholding). Let μ be a Gaussian measure on $L^2([0,1])$. Define

$$T(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \qquad \forall\, x \in [0,1]$$

and consider the corresponding Nemytskii operator $N_T : L^2([0,1]) \to L^\infty([0,1])$. Then, working with the definitions in Lemma 2, it is easy to see that $E \subset \text{supp } {N_T}^\sharp\mu$. Therefore there exists no separable subspace of $L^\infty([0,1])$ which contains supp ${N_T}^\sharp\mu$.

We therefore abandon L^∞ and instead show continuity and Lipschitz continuity for some L^q with $q < \infty$ to $\dot{H}^1$. The following lemma is a general result for convergence of sequences in metric spaces which is used in a more specific context in the next lemma.

Lemma 3. *Let (M,d) be a metric space and $(a_n) \subset M$ a sequence. If every subsequence $(a_{n_k}) \subset (a_n)$ contains a subsequence $(a_{n_{k_l}}) \subset (a_{n_k})$ such that $(a_{n_{k_l}}) \to a \in M$ then $(a_n) \to a$.*

Proof. Suppose otherwise. Then, there exists some $\epsilon > 0$ such that, for every $N \in \mathbb{Z}^+$, there exists some $n = n(N) > N$ such that

$$d(a_n, a) \geq \epsilon.$$

Then we can construct a subsequence $(a_{n_j}) \subset (a_n)$ such that $d(a_{n_j}, a) \geq \epsilon\ \forall n_j$. Therefore a_{n_j} does not have a subsequence converging to a, which is a contradiction.

The following lemma proves existence of a limit in $L^2(D;\mathbb{R}^d)$ of a sequence of outputs of operators in $L^\infty(D;\mathbb{R}^{d\times d})$.

Lemma 4. *Let $D \subseteq \mathbb{R}^d$ be an open set and $(A_n) \subset L^\infty(D; \mathbb{R}^{d\times d})$ a sequence satisfying the following:*

1. *$A_n \in \mathsf{PD}_{\alpha,\beta}$ for all n,*
2. *There exists $A \in L^\infty(D; \mathbb{R}^{d\times d})$ such that $(A_n) \to A$ in $L^2(D; \mathbb{R}^{d\times d})$.*

Then, for any $g \in L^2(D; \mathbb{R}^d)$, we have that $(A_n g) \to Ag$ in $L^2(D; \mathbb{R}^d)$.

Proof. We have

$$\|A_n g\|_{L^2} \leq \beta \|g\|_{L^2}$$

hence $(A_n g) \subset L^2(D; \mathbb{R}^d)$ and, similarly, by finite-dimensional norm equivalence, there is a constant $C_1 > 0$ such that

$$\|Ag\|_{L^2} \leq C_1 \|A\|_{L^\infty} \|g\|_{L^2}$$

hence $Ag \in L^2(D; \mathbb{R}^d)$. Again, by finite-dimensional norm equivalence, we have that there exists a constant $C_2 > 0$ such that, for $j \in \{1, \dots, d\}$ and almost every $y \in D$, we have

$$(A_n g)_j(y)^2 \leq |A_n^{(j)}(y)|^2 |g(y)|^2 \leq C_2 \beta^2 |g(y)|^2$$

where $A_n^{(j)}(y)$ denotes the j-th row of $A_n^{(j)}(y)$. In particular,

$$|(A_n g)_j(y)| \leq \sqrt{C_2}\beta |g(y)|.$$

Let $(A_{n_k}) \subset (A_n)$ be an arbitrary subsequence. Since $(A_n) \to A$, we have that $(A_{n_k}) \to A$ in $L^2(D; \mathbb{R}^{d\times d})$. Therefore, there exists a subsequence $(A_{n_{k_l}}) \subset (A_{n_k})$ such that $A_{n_{k_l}}(y) \to A(y)$ for almost every $y \in D$. Then $A_{n_{k_l}}(y) g(y) \to A(y)g(y)$ for almost every $y \in D$. Since $|g| \in L^2(\mathbb{R}^d)$, we have, by the dominated convergence theorem, that $(A_{n_{k_l}} g)_j \to (Ag)_j$ in $L^2(D)$ for every $j \in \{1, \dots, d\}$. Therefore $(A_{n_{k_l}} g) \to Ag$ in $L^2(D; \mathbb{R}^d)$. Since the subsequence (A_{n_k}) was arbitrary, Lemma 3 implies the result.

Finally, we may prove Proposition 2.

Proposition 2. *Endow $\mathsf{PD}_{\alpha,\beta}$ with the $L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$ induced topology and let $K \subset \mathsf{PD}_{\alpha,\beta}$ be a closed set. Define the mapping $G : K \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$ by $A \mapsto \chi$ as given by (5.4). Then there exists a bounded continuous mapping*

$$\mathcal{G} \in C(L^2(\mathbb{T}^d;\ \mathbb{R}^{d\times d}); \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d))$$

such that $\mathcal{G}(A) = G(A)$ for any $A \in K$.

Proof. Consider the PDE

$$-\nabla \cdot (A\nabla u) = \nabla \cdot Ae, \qquad y \in \mathbb{T}^d \tag{D.8}$$

where e is some standard basis vector of $\mathbb{R}^d$. Let $(A_n) \subset K$ be a sequence such that $(A_n) \to A \in K$ in $L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$. Denote by $u_n \in \dot{H}^1(\mathbb{T}^d)$ the solution to (D.8) corresponding to each A_n and by $u \in \dot{H}^1(\mathbb{T}^d)$ the solution corresponding to the limiting A. A similar calculation as in the proof of Proposition 1 shows

$$\begin{aligned}\alpha\|u_n - u\|_{\dot{H}^1}^2 &\leq \int_{\mathbb{T}^d} \langle (A - A_n)(\nabla u + e), \nabla u_n - \nabla u\rangle \, \mathsf{d}y \\ &\leq \|u_n - u\|_{\dot{H}^1} \|(A_n - A)(\nabla u + e)\|_{L^2}.\end{aligned}$$

Since $\nabla u + e \in L^2(\mathbb{T}^d; \mathbb{R}^d)$, by Lemma 4, $\Big(A_n(\nabla u + e)\Big) \to A(\nabla u + e)$ in $L^2(\mathbb{T}^d; \mathbb{R}^d)$ hence $(u_n) \to u$ in $\dot{H}^1(\mathbb{T}^d)$. In particular, the mapping $A \mapsto u$ defined by (D.8) is continuous. Since the problem (5.4) decouples as shown by (D.1), we have that each component mapping $G_l : K \to \dot{H}^1(\mathbb{T}^d)$ defined by $A \mapsto \chi_\ell$ is continuous thus G is continuous. Applying the Tietze extension theorem [5] to G implies the existence of $\mathcal{G}$.

The following is a straightforward consequence of Proposition 2 that establishes continuity of the map $A \mapsto \overline{A}$ defined in (5.3) as well.

Lemma 5. *Endow $\mathsf{PD}_{\alpha,\beta}$ with the $L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$ induced topology and let $K \subset \mathsf{PD}_{\alpha,\beta}$ be a closed set. Define the mapping $F : K \to \mathbb{R}^{d\times d}$ by $A \mapsto \bar{A}$ as given by* (5.3)*. Then there exists a bounded continuous mapping $\mathcal{F} \in C\Big(L^2(\mathbb{T}^d; \mathbb{R}^{d\times d}); \mathbb{R}^{d\times d}\Big)$ such that $\mathcal{F}(A) = F(A)$ for any $A \in K$.*

Proof. Since $\nabla : \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d) \to L^2(\mathbb{T}^d; \mathbb{R}^{d\times d})$ is a bounded operator, Lemma 2 implies that the mapping $A \mapsto A + A\nabla\chi^T$ is continuous as compositions, sums, and products of continuous functions are continuous. Now let $A \in \mathsf{PD}_{\alpha,\beta}$ then $A \in L^1(\mathbb{T}^d; \mathbb{R}^{d\times d})$ since $A \in L^\infty(\mathbb{T}^d; \mathbb{R}^{d\times d})$. Thus

$$\left|\int_{\mathbb{T}^d} A \, \mathsf{d}y\right|_F \leq \int_{\mathbb{T}^d} |A|_F \, \mathsf{d}y \leq \|A\|_{L^2}$$

by Hölder's inequality and the fact that $\int_{\mathbb{T}^d} \mathsf{d}y = 1$. Hence $F \in C(K; \mathbb{R}^{d\times d})$ as a composition of continuous maps. Again applying the Tietze extension theorem [5] to F implies the existence of $\mathcal{F}$.

To prove Proposition 3, we need to establish Lipschitz continuity. We first establish the following result, which is similar to the one proved in [4] in Theorem 2.1. We show it again here both for completeness and because we specialize to the case of the cell problem (5.4) with periodic boundary conditions rather than the system (5.1) with Dirichlet boundary conditions.

Lemma 6. *Let $A^{(1)}, A^{(2)} \in \mathsf{PD}_{\alpha,\beta}$ and let $\chi^{(1)}, \chi^{(2)}$ be the corresponding solutions to* (5.4)*. Then*

$$\|\chi^{(1)} - \chi^{(2)}\|_{\dot{H}^1} \leq \frac{\sqrt{d}}{\alpha} \left(\|A^{(2)} - A^{(1)}\|_{L^2} + \|\nabla \chi^{(2)}\|_{L^p} \|A^{(2)} - A^{(1)}\|_{L^q} \right) \quad \text{(D.9)}$$

for $p \geq 2$ and $q = \frac{2p}{p-2}$.

Proof. As in the proof of Proposition 1, we denote $f^{(i)} = \nabla \cdot A^{(i)}$ for $i \in \{1, 2\}$ for simplicity of notation and to be easily comparable to the proof of Theorem 2.1 in [4]. Since both sides of the cell problem equation (5.4) depend on $A^{(i)}$, we introduce $\widetilde{\chi}$ as the solution of

$$-\nabla \cdot \left(\nabla \widetilde{\chi} A^{(2)} \right) = \nabla \cdot A^{(1)}, \quad \widetilde{\chi} \in \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d) \quad \text{(D.10)}$$

as an intermediate function. We obtain bounds using $\widetilde{\chi}$ and apply the triangle inequality to

$$\|(\chi^{(1)} - \widetilde{\chi}) + (\widetilde{\chi} - \chi^{(2)})\|_{\dot{H}^1}$$

to obtain a bound on $\|\chi^{(1)} - \chi^{(2)}\|_{\dot{H}^1}$. From the naïve perturbation bound in (D.2) we have

$$\|\widetilde{\chi}_\ell - \chi_\ell^{(2)}\|_{\dot{H}^1} \leq \frac{1}{\alpha} \|f_\ell^{(1)} - f_\ell^{(2)}\|_{\dot{H}^{-1}},$$

so we are left to bound $\|\chi_\ell^{(1)} - \widetilde{\chi}_\ell\|_{\dot{H}^1}$. We note that

$$\nabla \cdot \left(A^{(2)} \nabla \widetilde{\chi}_\ell \right) = \nabla \cdot \left(A^{(1)} \nabla \chi_\ell^{(1)} \right)$$
$$\int_{\mathbb{T}^d} A^{(2)} \nabla \widetilde{\chi}_\ell \cdot \nabla v \ \mathsf{d}y = \int_{\mathbb{T}^d} A^{(1)} \nabla \chi_\ell^{(1)} \cdot \nabla v \, \mathsf{d}y \quad \forall v \in \dot{H}^1(\mathbb{T}^d; \mathbb{R}).$$

Letting $v = \chi_\ell^{(1)} - \widetilde{\chi}_\ell$,

$$\int_{\mathbb{T}^d} A^{(2)} \nabla \widetilde{\chi}_\ell \cdot \left(\nabla \chi_\ell^{(1)} - \nabla \widetilde{\chi}_\ell \right) \mathsf{d}y = \int_{\mathbb{T}^d} A^{(1)} \nabla \chi_\ell^{(1)} \cdot \left(\nabla \chi_\ell^{(1)} - \nabla \widetilde{\chi}_\ell \right) \mathsf{d}y$$
$$\int_{\mathbb{T}^d} A^{(2)} \left(\nabla \widetilde{\chi}_\ell - \nabla \chi_\ell^{(1)} \right) \cdot \left(\nabla \widetilde{\chi}_\ell - \nabla \chi_\ell^{(1)} \right) \mathsf{d}y$$
$$= \int_{\mathbb{T}^d} \left(A^{(2)} - A^{(1)} \right) \nabla \chi_\ell^{(1)} \cdot \left(\nabla \chi_\ell^{(1)} - \nabla \widetilde{\chi}_\ell \right) \mathsf{d}y$$
$$\alpha \|\widetilde{\chi}_\ell - \chi_\ell^{(1)}\|_{\dot{H}^1} \leq \|(A^{(2)} - A^{(1)})(\nabla \chi_\ell^{(1)})\|_{L^2}.$$

Applying Hölder for L^2, we get

$$\|\widetilde{\chi}_\ell - \chi_\ell^{(1)}\|_{\dot{H}^1} \leq \frac{1}{\alpha}\|\nabla\chi_\ell^{(1)}\|_{L^p}\|A^{(2)} - A^{(1)}\|_{L^q} \tag{D.11}$$

for $q = \frac{2p}{p-2}$ where $p \in [2, \infty]$. Putting the two parts together, we have that

$$\begin{aligned}\|\chi_\ell^{(2)} - \chi_\ell^{(1)}\|_{\dot{H}^1} &\leq \frac{1}{\alpha}\|\nabla\cdot A^{(2)}e_\ell - \nabla\cdot A^{(1)}e_\ell\|_{\dot{H}^{-1}} + \frac{1}{\alpha}\|\nabla\chi_\ell^{(1)}\|_{L^p}\|A^{(2)} - A^{(1)}\|_{L^q} \\ &\leq \frac{1}{\alpha}\|A^{(2)} - A^{(1)}\|_{L^2} + \frac{1}{\alpha}\|\nabla\chi_\ell^{(1)}\|_{L^p}\|A^{(2)} - A^{(1)}\|_{L^q}.\end{aligned}$$

Combining bounds for all d dimensions yields the result.

Remark 10. Since $L^q(\Omega) \hookrightarrow L^2(\Omega)$ for bounded $\Omega \subset \mathbb{R}^d$ and $q \geq 2$, we could also write the bound of Lemma 6 as

$$\|\chi_\ell^{(2)} - \chi_\ell^{(1)}\|_{\dot{H}^1} \leq \frac{1}{\alpha}\left(C + \|\nabla\chi_\ell^{(1)}\|_{L^p}\right)\|A^{(2)} - A^{(1)}\|_{L^q}$$

for some C dependent only on q and Ω.

The result of Lemma 6 is unhelpful if $\|\nabla\chi\|_{L^p}$ is unbounded. In this setting, it is not possible for Lemma 6 to result in Lipschitz continuity as a map from L^2 to $\dot{H}^1$. Instead, we seek to bound $\|\nabla\chi\|_{L^p}$ for some p satisfying $2 < p < \infty$.

Before continuing, we establish a bound on the gradient of the solution to the Poisson equation on the torus. This follows the strategy of [4] for the Dirichlet problem. In order to avoid extra factors of 2π in all formulae, we work on the rescaled torus denoted $\mathbb{Y}^d = [0, 2\pi]^d$ with opposite faces identified for the following result of Lemma 7. As we work on the torus, it is useful to first set up notation for the function spaces of interest. Let

$$\mathcal{D}(\mathbb{Y}^d) = C_c^\infty(\mathbb{Y}^d) = C^\infty(\mathbb{Y}^d)$$

be the space of test functions where the last equality follows from compactness of the torus. Functions can be either $\mathbb{R}$ or $\mathbb{C}$ valued hence we do not explicitly specify the range. We equip $\mathcal{D}(\mathbb{Y}^d)$ with a locally convex topology generated by an appropriate family of semi-norms, see, for example, [11, Section 3.2.1]. Any function $g \in \mathcal{D}(\mathbb{Y}^d)$ can be represented by its Fourier series

$$g(x) = \sum_{k\in\mathbb{Z}^d} \widehat{g}(k)\mathrm{e}^{ix\cdot k}$$

where $\widehat{g}$ denotes the Fourier transform of g and convergence of the right-hand side sum is with respect to the topology of $\mathcal{D}(\mathbb{Y}^d)$, and i denotes the imaginary unit. It holds that $\widehat{g} \in \mathcal{S}(\mathbb{Z}^d)$, the Schwartz space of rapidly decreasing functions on the integer lattice, so we have

$$|\widehat{g}(k)| \leq c_m(1+|k|)^{-m}, \quad m = 0, 1, \ldots$$

for some constants c_m. We may then define the topological (continuous) dual space of $\mathcal{D}(\mathbb{Y}^d)$, the space of distributions, denoted $\mathcal{D}'(\mathbb{Y}^d)$, which can be described as follows: the condition that $f \in \mathcal{D}'(\mathbb{Y}^d)$ is characterized by the property

$$|\widehat{f}(k)| \leq b_m(1+|k|)^{m}, \quad m = 0, 1, \ldots$$

for some constants b_m. We take the weak-* topology on $\mathcal{D}'(\mathbb{Y}^d)$ and generally use the prime notation for any such dual space. For any $-\infty < s < \infty$, we define the fractional Laplacian as

$$(-\Delta)^s f = \sum_{k \in \mathbb{Z}^d \setminus \{0\}} |k|^{2s} \widehat{f}(k) \mathrm{e}^{ik\cdot x} \tag{D.12}$$

where the right-hand side sum converges in the topology of $\mathcal{D}'(\mathbb{Y}^d)$. It is easy to see that $(-\Delta)^s : \mathcal{D}'(\mathbb{Y}^d) \to \mathcal{D}'(\mathbb{Y}^d)$ is continuous. Furthermore, for any $j \in \{1, \ldots, d\}$, we define the family of operators $\tilde{R}_j : \mathcal{D}'(\mathbb{Y}^d) \to \mathcal{D}'(\mathbb{Y}^d)$, defining periodic Riesz transforms, by

$$\tilde{R}_j f = \sum_{k \in \mathbb{Z}^d} -\frac{ik_j}{|k|} \widehat{f}(k) \mathrm{e}^{ik\cdot x} \tag{D.13}$$

where we identify $\frac{k_j}{|k|}|_{k=0} = \lim_{|k|\to 0} \frac{k_j}{|k|} = 0$. Again, we stress that convergence of the right-hand side sum is in the topology of $\mathcal{D}'(\mathbb{Y}^d)$. Lastly, we denote by $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}'(\mathbb{R}^d)$ the Schwartz space and the space of tempered distributions on $\mathbb{R}^d$, respectively; see, for example, [12, Chapter 1] for the precise definitions.

The following lemma establishes boundedness of the periodic Riesz transform on $L^p(\mathbb{Y}^d)$. It is essential in proving boundedness of the gradient to the solution of the Poisson equation on the torus. The result is essentially proven in [12]. We include it here, in our specific torus setting, giving the full argument for completeness.

Lemma 7. *There exists a constant $c = c(d,p) > 0$ such that, for any $j \in \{1, \ldots, d\}$ and any $f \in L^p(\mathbb{Y}^d)$ for some $2 \leq p < \infty$, we have*

$$\|\tilde{R}_j f\|_{L^p(\mathbb{Y}^d)} \leq c \|f\|_{L^p(\mathbb{Y}^d)}.$$

Proof. Let $g \in L^2(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ for some $1 < p < \infty$. For any $j \in \{1, \ldots, d\}$, define the family of operators R_j by

$$(R_j g)(x) = \lim_{\delta^{-1}, \epsilon \to 0^+} \int_{\delta \geq |t| \geq \epsilon} g(x-t) K_j(t)\, dt,$$

where

$$K_j(t) = \frac{\Gamma\Big((d+1)/2\Big) t_j}{\pi^{(d+1)/2} |t|^{d+1}}$$

and Γ denotes the Euler-Gamma function. By [12, Chapter 4, Theorem 4.5], $K_j \in \mathcal{S}'(\mathbb{R}^d)$ and its Fourier transform satisfies

$$\widehat{K_j}(t) = -\frac{i t_j}{|t|}.$$

Therefore, for any $\phi \in \mathcal{S}(\mathbb{R}^d)$, we have

$$(K_j * \phi)\widehat{\ }(t) = -\frac{i t_j}{|t|} \widehat{\phi}(t)$$

where $*$ denotes convolution, see, for example, [12, Chapter 1, Theorem 3.18]. Since $g \in L^2(\mathbb{R}^d)$, we therefore find that, by [12, Chapter 6, Theorem 2.6],

$$(R_j g)\widehat{\ }(x) = -\frac{i x_j}{|x|} \widehat{g}(x) \tag{D.14}$$

for Lebesgue almost every $x \in \mathbb{R}^d$. The result [12, Chapter 6, Theorem 2.6] further shows that there exists a constant $c = c(d, p) > 0$ such that

$$\|R_j g\|_{L^p(\mathbb{R}^d)} \leq c \|g\|_{L^p(\mathbb{R}^d)}.$$

We note from (D.14) and the definition (D.13) that $\tilde{R}_j$ may be viewed as R_j with the restriction of the Fourier multiplier $-\frac{i x_j}{|x|}$ to the lattice $\mathbb{Z}^d$. We can therefore use the transference theory of [12] to establish boundedness of $\tilde{R}_j$ from the boundedness of R_j. In particular, note that the mapping $x \mapsto -\frac{i x_j}{|x|}$ is continuous at all $x \in \mathbb{R}^d$ except $x = 0$. However, by symmetry, we have that, for all $\epsilon > 0$

$$\int_{|x| \leq \epsilon} -\frac{i x_j}{|x|}\, dx = 0.$$

Therefore we can apply [12, Chapter 7, Theorem 3.8, Corollary 3.16] to conclude that, since R_j is bounded from $L^p(\mathbb{R}^d)$ to $L^p(\mathbb{R}^d)$, $\tilde{R}_j$ is bounded from $L^p(\mathbb{Y}^d)$ to $L^p(\mathbb{Y}^d)$ with

$$\|\tilde{R}_j\|_{L^p(\mathbb{Y}^d) \to L^p(\mathbb{Y}^d)} \leq \|R_j\|_{L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)}.$$

This implies the desired result.

We define the Bessel potential spaces by

$$L^{s,p}(\mathbb{Y}^d) = \{u \in \mathcal{D}'(\mathbb{Y}^d) \mid \|u\|_{L^{s,p}(\mathbb{Y}^d)} := \|(I-\Delta)^{s/2}u\|_{L^p(\mathbb{Y}^d)} < \infty\}$$

for any $-\infty < s < \infty$ and $1 < p < \infty$. We also define the homogeneous version of these spaces, sometimes called the Riesz potential spaces, by

$$\dot{L}^{s,p}(\mathbb{Y}^d) = \{u \in \mathcal{D}'(\mathbb{Y}^d) \mid \|u\|_{\dot{L}^{s,p}(\mathbb{Y}^d)} := \|(-\Delta)^{s/2}u\|_{L^p(\mathbb{Y}^d)} < \infty, \int_{\mathbb{Y}^d} u \, \mathsf{d}y = 0\}.$$

It is clear that $\dot{L}^{s,p}(\mathbb{Y}^d) \subset L^{s,p}(\mathbb{Y}^d)$ is closed subspace. We then have the following result for the Poisson equation.

Lemma 8. *For each $f \in L^{s,p}(\mathbb{Y}^d)$, for $-\infty < s < \infty$ and $2 \le p < \infty$, the solution u of the equation*

$$-\Delta u = f, \quad u \text{ 1-periodic}, \int_{\mathbb{Y}^d} u \, \mathsf{d}y = 0 \tag{D.15}$$

satisfies

$$\|\nabla u\|_{\dot{L}^{s+1,p}(\mathbb{Y}^d)} \le K\|f\|_{\dot{L}^{s,p}(\mathbb{Y}^d)} \tag{D.16}$$

for some finite $K > 0$ depending only on p and d.

Proof. From the definitions (D.12) and (D.13), it is easy to see that the Riesz transform can be written as

$$\tilde{R}_j = -\partial_{x_j}(-\Delta)^{-1/2}$$

in the sense of distributions. Consider now equation (D.15) with $f \in L^{s,p}(\mathbb{Y}^d)$ for $2 \le p < \infty$. We have that

$$\begin{aligned}\|\partial_{x_j}u\|_{\dot{L}^{s+1,p}(\mathbb{Y}^d)} &= \|\partial_{x_j}(-\Delta)^{-1}f\|_{\dot{L}^{s+1,p}(\mathbb{Y}^d)} \\ &= \|\partial_{x_j}(-\Delta)^{-1/2}(-\Delta)^{s/2}f\|_{L^p(\mathbb{Y}^d)} \\ &= \|\tilde{R}_j(-\Delta)^{s/2}f\|_{L^p(\mathbb{Y}^d)}.\end{aligned}$$

It is clear that

$$\|(-\Delta)^{s/2}f\|_{L^p(\mathbb{Y}^d)} = \|f\|_{\dot{L}^{s,p}(\mathbb{Y}^d)} < \infty$$

hence $(-\Delta)^{s/2}f \in L^p(\mathbb{Y}^d)$. We can thus apply Lemma 7 to find a constant $c = c(d,p) > 0$ such that

$$\|\partial_{x_j}u\|_{\dot{L}^{s+1,p}(\mathbb{Y}^d)} \le c\|(-\Delta)^{s/2}f\|_{L^p(\mathbb{Y}^d)} = c\|f\|_{\dot{L}^{s,p}(\mathbb{Y}^d)}.$$

The result follows by finite-dimensional norm equivalence.

Next we define the homogeneous Sobolev spaces on the torus as

$$\dot{W}^{k,p}(\mathbb{T}^d) = \{u \in W^{k,p}(\mathbb{T}^d) \mid u \text{ is 1-periodic}, \int_{\mathbb{T}^d} u \, \mathsf{d}y = 0\} \tag{D.17}$$

for $k = 0, 1, \ldots,$ and $1 \leq p \leq \infty$ with the standard norm on $W^{k,p}$, see, for example [1].

Remark 11. By [11, Section 3.5.4], we have that, for any $k = 0, 1, \ldots$ and $1 < p < \infty$,

$$L^{k,p}(\mathbb{T}^d) = W^{k,p}(\mathbb{T}^d), \qquad \dot{L}^{k,p}(\mathbb{T}^d) = \dot{W}^{k,p}(\mathbb{T}^d).$$

Furthermore, by [11, Section 3.5.6],

$$\begin{aligned} W^{-k,p'}(\mathbb{T}^d) &= \left(W^{k,p}(\mathbb{T}^d)\right)' = \left(L^{k,p}(\mathbb{T}^d)\right)' = L^{-k,p'}(\mathbb{T}^d), \\ \dot{W}^{-k,p'}(\mathbb{T}^d) &= \left(\dot{W}^{k,p}(\mathbb{T}^d)\right)' = \left(\dot{L}^{k,p}(\mathbb{T}^d)\right)' = \dot{L}^{-k,p'}(\mathbb{T}^d) \end{aligned}$$

where p' is the Hölder conjugate of p, i.e., $1/p + 1/p' = 1$.

In the following, we use the notation

$$[K_0, K_1]_{\theta,q} \tag{D.18}$$

to denote the real interpolation between two Banach spaces continuously embedded in the same Hausdorff topological space, as described in [1]. We also need Lemma A1 from [6], which we have copied below as Lemma 9 to ease readability. Although this lemma was written only for $q = 2$, the result still holds for our $q > 2$ with a very similar proof.

Lemma 9. *Let $E_1 \subset E_0$ be two Banach spaces with E_1 continuously embedded in E_0. Let $T : E_j \to E_j$ be a bounded operator with closed range and assume that T is a projection, $j \in \{0, 1\}$. Denote by K_0 and K_1 the ranges of $T|_{E_0}$ and $T|_{E_1}$, respectively. Then the following two spaces coincide with equivalent norms:*

$$[K_0, K_1]_{\theta,q} = [E_0, E_1]_{\theta,q} \cap K_0 \quad \forall \theta \in (0, 1).$$

We now state the result for the bound on $\|\nabla\chi\|_{L^p}$ with a proof largely developed in [4].

Lemma 10. *Let χ solve* (5.4) *for $A \in \mathsf{PD}_{\alpha,\beta}$. Then*

$$\|\nabla\chi\|_{L^p} \leq \frac{K^{\eta(p)}}{1 - K^{\eta(p)}\left(1 - \frac{\alpha}{\beta}\right)} \tag{D.19}$$

for $2 \leq p < p^*\left(\frac{\alpha}{\beta}\right)$ *where*

$$p^*(t) := max\Big\{p \mid K^{-\eta(p)} \geq 1 - t,\ 2 < p < Q\Big\} \tag{D.20}$$

for $\eta(p) = \frac{1/2-1/p}{1/2-1/Q}$ *and* $K = K(d, Q)$ *is the constant in Lemma 8, for any choice of* $Q > p$.

Proof. The operator $T = -\Delta$ is invertible from H^{-1} to $\dot{H}^1$, and the inverse T^{-1} is bounded with norm 1 since the Poisson equation with periodic boundary conditions has a unique solution in $\dot{H}^1$ for $f \in H^{-1}$ with bound $\|u\|_{\dot{H}^1} \leq \|f\|_{H^{-1}}$. From Lemma 8 it is also bounded with norm $K = K(d, Q)$ from $W^{-1,Q}$ to $\dot{W}^{1,Q}$ for any $Q > 2$. By the real method of interpolation [1], for $2 < p < Q$ we have that

$$W^{1,p} = \Big[H^1, W^{1,Q}\Big]_{\eta(p),p} \tag{D.21}$$

using the notation of [1] where $\eta(p) = \frac{1/2-1/p}{1/2-1/Q}$. From the duality theorem (Theorem 3.7.1. of [2]), we have that

$$\Big[H^{-1}, W^{-1,Q}\Big]_{\eta(p),p} = \Big(\Big[H^1, W^{1,Q'}\Big]_{\eta(p),p'}\Big)', \tag{D.22}$$

From real interpolation, the right hand side equals $(W^{1,p'})' = W^{-1,p}$ in our notation. Therefore, we have the necessary dual statement that parallels (D.21):

$$W^{-1,p} = \Big[H^{-1}, W^{-1,Q}\Big]_{\eta(p),p}. \tag{D.23}$$

Next we restrict these spaces to functions with periodic boundary conditions. Using the projection onto the space of continuous, periodic functions on $\mathbb{T}^d$ and noticing that $W^{1,Q} \hookrightarrow H^1$, we apply Lemma 9 with $K_0 = \dot{H}^1$ and have

$$\dot{W}^{1,p} = [\dot{H}^1, \dot{W}^{1,Q}]_{\eta(p),p}. \tag{D.24}$$

Using the exact interpolation theorem, Theorem 7.23 of [1], T^{-1} is also a bounded map from $W^{-1,p}$ to $\dot{W}^{1,p}$ with norm $K^{\eta(p)}$:

$$\|T^{-1}f\|_{\dot{W}^{1,p}} \leq K^{\eta(p)}\|f\|_{W^{-1,p}}. \tag{D.25}$$

The remainder of the proof is identical to that of the proof of Proposition 1 in [4], but we state it here in our notation for completeness. Define S: $\dot{W}^{1,p} \to W^{-1,p}$ as the operator $Su = -\nabla \cdot \Big(\frac{1}{\beta} A \nabla u\Big)$. Let V be the perturbation

operator $V := T - S$. Since $A \in \mathsf{PD}_{\alpha,\beta}$, S and V are bounded operators from $\dot{W}^{1,p}$ to $W^{-1,p}$, with the operator norms $\|S\| \leq 1$ and $\|V\| \leq 1 - \frac{\alpha}{\beta}$. Therefore,

$$\|T^{-1}V\|_{\dot{W}^{1,p}\to\dot{W}^{1,p}} \leq \|T^{-1}\|_{W^{-1,p}\to\dot{W}^{1,p}}\|V\|_{\dot{W}^{1,p}\to W^{-1,p}} \leq K^{\eta(p)}\left(1-\frac{\alpha}{\beta}\right), \tag{D.26}$$

where the input and output spaces defining the operator norms are included for clarity. Since T is invertible, $S = T(I - T^{-1}V)$ is invertible provided $K^{\eta(p)}\left(1-\frac{\alpha}{\beta}\right) < 1$. Moreover, for S^{-1} as a mapping from $W^{-1,p}$ to $\dot{W}^{1,p}$,

$$\|S^{-1}\| \leq \|(I-T^{-1}V)^{-1}\|_{\dot{W}^{1,p}\to\dot{W}^{1,p}}\|T^{-1}\|_{W^{-1,p}\to\dot{W}^{1,p}} \leq \frac{K^{\eta(p)}}{1-K^{\eta(p)}\left(1-\frac{\alpha}{\beta}\right)}. \tag{D.27}$$

Therefore,

$$\|\nabla\chi\|_{L^p} = \|\chi\|_{\dot{W}^{1,p}} \leq \frac{1}{\beta}\|S^{-1}\|\|\nabla\cdot A\| \leq \frac{K^{\eta(p)}}{1-K^{\eta(p)}\left(1-\frac{\alpha}{\beta}\right)} \tag{D.28}$$

provided $K^{\eta(p)}\left(1-\frac{\alpha}{\beta}\right) < 1$. The bound and specified range of p follow.

Finally, we may prove Proposition 3

Proposition 3. *There exists $q_0 \in (2,\infty)$ such that, for all q satisfying $q \in (q_0,\infty]$, the following holds. Endow $\mathsf{PD}_{\alpha,\beta}$ with the $L^q(\mathbb{T}^d;\mathbb{R}^{d\times d})$ topology and let $K \subset \mathsf{PD}_{\alpha,\beta}$ be a closed set. Define the mapping $G : K \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ by $A \mapsto \chi$ as given by (5.4). Then there exists a bounded Lipschitz-continuous mapping*

$$\mathcal{G} :\ L^q(\mathbb{T}^d;\mathbb{R}^{d\times d}) \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$$

such that $\mathcal{G}(A) = G(A)$ for any $A \in K$.

Proof. Lemma 10 guarantees a $p_0 > 2$ such that $\|\nabla\chi^{(2)}\|_{L^p}$ in Lemma 6 is bounded above by a constant for $2 < p < p_0$. Then Lemma 6 gives Lipschitz continuity of the solution map from $L^q(\mathbb{T}^d) \mapsto \dot{H}^1(\mathbb{T}^d)$ for q satisfying $q_0 < q < \infty$ for some $q_0 > 2$.

Remark 12. From the results of Lemma 10 and Lemma 6, we have that we can take $q_0 = \frac{2p_0}{p_0-2}$ where

$$p_0 = \max\{p \mid K^{-\eta(p)} \geq 1-t,\ 2 < p < Q\}.$$

Therefore, bounds on p_0 may be inherited from bounds on K that appears in Lemma 8.

D.2 Proofs of approximation theorems

In this section we prove the approximation theorems stated in Section 5.3.

Theorem 7. *Let $K \subset \mathsf{PD}_{\alpha,\beta}$ and define the mapping $G : K \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ by $A \mapsto \chi$ as given by (5.4). Assume in addition that K is compact in $L^2(\mathbb{T}^d;\mathbb{R}^{d\times d})$. Then, for any $\epsilon > 0$, there exists an FNO $\Psi : K \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ such that*

$$\sup_{A\in K} \|G(A) - \Psi(A)\|_{\dot{H}^1} < \epsilon.$$

Proof. By Proposition 2, there exists a continuous map

$\mathcal{G} \in C(L^2(\mathbb{T}^d;\mathbb{R}^{d\times d}); \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d))$ such that $\mathcal{G}(A) = G(A)$ for any $A \in K$. By [7, Theorem 5], there exists a FNO $\Psi : L^2(\mathbb{T}^d;\mathbb{R}^{d\times d}) \to \dot{H}^1(\mathbb{T}^d;\mathbb{R}^d)$ such that

$$\sup_{A\in K} \|\mathcal{G}(A) - \Psi(A)\|_{\dot{H}^1} < \epsilon.$$

Therefore

$$\sup_{A\in K} \|G(A) - \Psi(A)\|_{\dot{H}^1} = \sup_{A\in K} \|\mathcal{G}(A) - \Psi(A)\|_{\dot{H}^1} < \epsilon$$

as desired.

Theorem 8. *Let $K \subset \mathsf{PD}_{\alpha,\beta}$ and define the mapping $F : K \to \mathbb{R}^{d\times d}$ by $A \mapsto \bar{A}$ as given by (5.3), (5.4). Assume in addition that K is compact in $L^2(\mathbb{T}^d;\mathbb{R}^{d\times d})$. Then, for any $\epsilon > 0$, there exists an FNO $\Phi : K \to L^\infty(\mathbb{T}^d;\mathbb{R}^{d\times d})$ such that*

$$\sup_{A\in K} \sup_{x\in\mathbb{T}^d} |F(A) - \Phi(A)(x)|_F < \epsilon.$$

Proof. The result follows as in Theorem 7 by applying Lemma 5 instead of Proposition 2.

D.3 Proofs for microstructure examples

The following lemma establishes the compactness of subsets of $\mathsf{PD}_{\alpha,\beta}$ generated by the probability measures from Section 5.4. As we are unaware of a proof in the literature, we have provided one below. The proof uses the L^1-Lipschitz spaces, which are defined as

$$\mathsf{Lip}_\alpha(L^1) = \{u \in L^1 : \ \exists M(u) > 0 : \ \omega(u,t)_1 \le Mt^\alpha\}$$

where $\omega(u,t)_1$ is the 1-modulus of continuity, defined via

$$\omega(u,t)_1 = \sup_{0\leq|h|\leq t} \|\tau_h u - u\|_{L^1(\mathbb{T}^d)}.$$

Lemma 11. *$\mathsf{BV}(\mathbb{T}^d) \cap L^\infty(\mathbb{T}^d)$ is compactly embedded in $L^2(\mathbb{T}^d)$.*

Proof. Let $u \in B$, where B is a bounded subset of $\mathsf{BV}(\mathbb{T}^d) \cap L^\infty(\mathbb{T}^d)$ with L^∞ norm and BV seminorm bounded by M, and let $\tau_h f$ denote the translation of f by h, i.e., $\tau_h f(x) = f(x-h)$. Then

$$\|\tau_h u - u\|_{L^2} \leq \|\tau_h u - u\|_{L^1}^{1/2} \|\tau_h u - u\|_{L^\infty}^{1/2}. \tag{D.29}$$

Since $\mathsf{BV}(\mathbb{T}^d) \equiv \mathsf{Lip}_1(L^1(\mathbb{T}^d))$, $\|\tau_h u - u\|_{L^1} \leq \|u\|_{\mathsf{BV}}|h|$. We have then

$$\|\tau_h u - u\|_{L^2} \leq \|u\|_{\mathsf{BV}}^{1/2} |h|^{1/2} (2M)^{1/2}.$$

By the Fréchet-Kolmogorov theorem [13], this equicontinuity result is sufficient for compactness of B in $L^2(\mathbb{T}^d)$.

Using the result of Lemma 11, we see that any set of microstructure coefficients bounded in $L^\infty(\mathbb{T}^d) \cap BV(\mathbb{T}^d)$ satisfies the compactness assumption of the Approximation Theorems in Section 5.3. It is clear that the method of construction of the examples in Subsection 5.4.1 leads to such sets.

D.4 Numerical implementation details

All FNO models are implemented in pytorch using python 3.9.7. Unless otherwise specified, the models have 18 modes in each dimension, a width of 64, and 4 hidden layers. The lifting layer is a linear transformation with trainable parameters, and the projecting layer is a pointwise multilayer perceptron with trainable parameters. The batch size is 20, the learning rate is 0.001, and the number of epochs is 400. These hyperparameters are chosen with a small grid search, but we emphasize that the FNO does not drastically change in performance unless these parameters are changed by an order of magnitude. For a model trained on 9500 data using these hyperparameters and accelerated with an Nvidia P100 GPU, the training time is approximately 7 hours. In Figures 5.4, 5.5, and 5.6, the error bars shown correspond to two standard deviations in each direction over the five samples. All code for this work may be found at github.com/mtrautner/LearningHomogenization/.

Appendix E

APPENDIX FOR CHAPTER 6

E.1 Exploration of designs and sensitivities in PCA distributions

Figure E.1 shows the first two principal components (PCs) of the ground truths in the point load direct design training dataset, which was composed of X thousand samples, with example designs shown surrounding the plot. Designs near each other in the first two PCs often tend to have similar boundary conditions leading to qualitatively similar topologies (e.g., green and gray, purple and blue). Figure E.2 shows the first two PCs of the ground truth compliance sensitivities in point load data, point load data with the singularity analytically removed, and the dataset with the distributed load. The same 12 examples are shown around the perimeter of each plot. The former two cases are arranged in a pitchfork-like pattern, where one side of the pitchfork tends to have samples with boundary conditions of North East and South West, while samples on the other side often have boundary conditions of North West and South East. The points that do not belong on either side of the pitchfork tend to have boundary conditions that are along the same side, such as North East and North West. The distributed load dataset is more compactly distributed, with a less-clear distribution of boundary conditions. Because PCA-Net performance is strong with both forms of smoothing yet the distributions in these first two PCs are quite different, review of higher-order PCs is of interest.

E.2 Exploration of higher-order principal components

For the sensitivity prediction configuration, Figure E.3 shows projections of the compliance sensitivity into the first eight principal components. The distributed load data set, the point load with analytical singularity removal, and the original point load data are included as separate colors. At low-order PCs (1 and 2), the analytically smoothed data and the original unsmoothed data show a split pattern with two distinct offshoot regions from the main cluster. The distributed load data shows the same general pattern by PCs 3 and 4, indicating that the same trend was captured at a different rate.

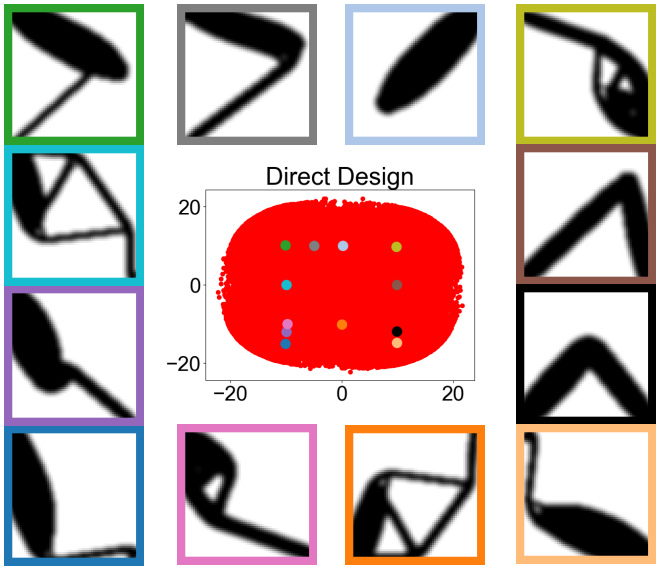


Figure E.1: First two dimensions of the PCA projection of the final designs in the direct design training set. Individual example designs are also shown, with the same designs indicated in the plot by the colors corresponding to their outlines.

The distributions for the analytically smoothed data (which performed well in TO) and the original unmodified data (which performed less well in TO) appear qualitatively similar in shape, with the smoothed data generally possessing a noticeably broader distribution of that shape. The distributed load (strong TO performance) resulted in a distinctly different distribution. Because the resulting performance in the TO loop is not directly correlated with the shape of PCA distributions, other factors appear more important, such as the ability to capture variance with low- and intermediate-order PCs as indicated in the main text's Figure 9.

E.3 Varying the degree of smoothing

In the case of distributing the point force to effectively smooth the sensitivity prediction problem, distributing the load across a relatively modest region resulted in a significant performance improvement as measured by structural compliance. In this work, there was no clear way to optimally select the exact amount of distribution applied. To understand whether more distribution might have given further improvement, the load distribution parameter r_{max} was increased from 3 shown in the main text to 7 — roughly doubling the size. In the same format as in Section 5, the results are shown in Figs. E.4 -

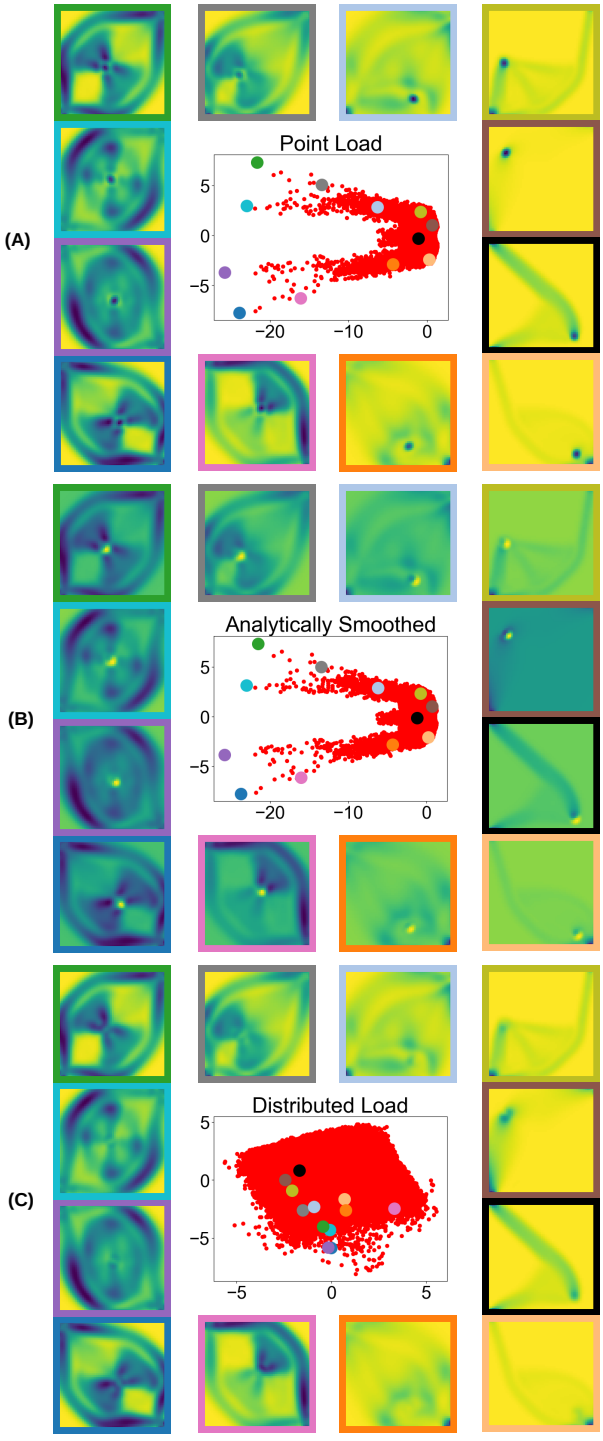


Figure E.2: First two dimensions of the PCA projection of the compliance sensitivities in the ground truth of the training set for a) point load b) point load with singularity analytically removed c) distributed load. Example designs are shown surrounding the plots, with the same 12 examples shown for each panel.

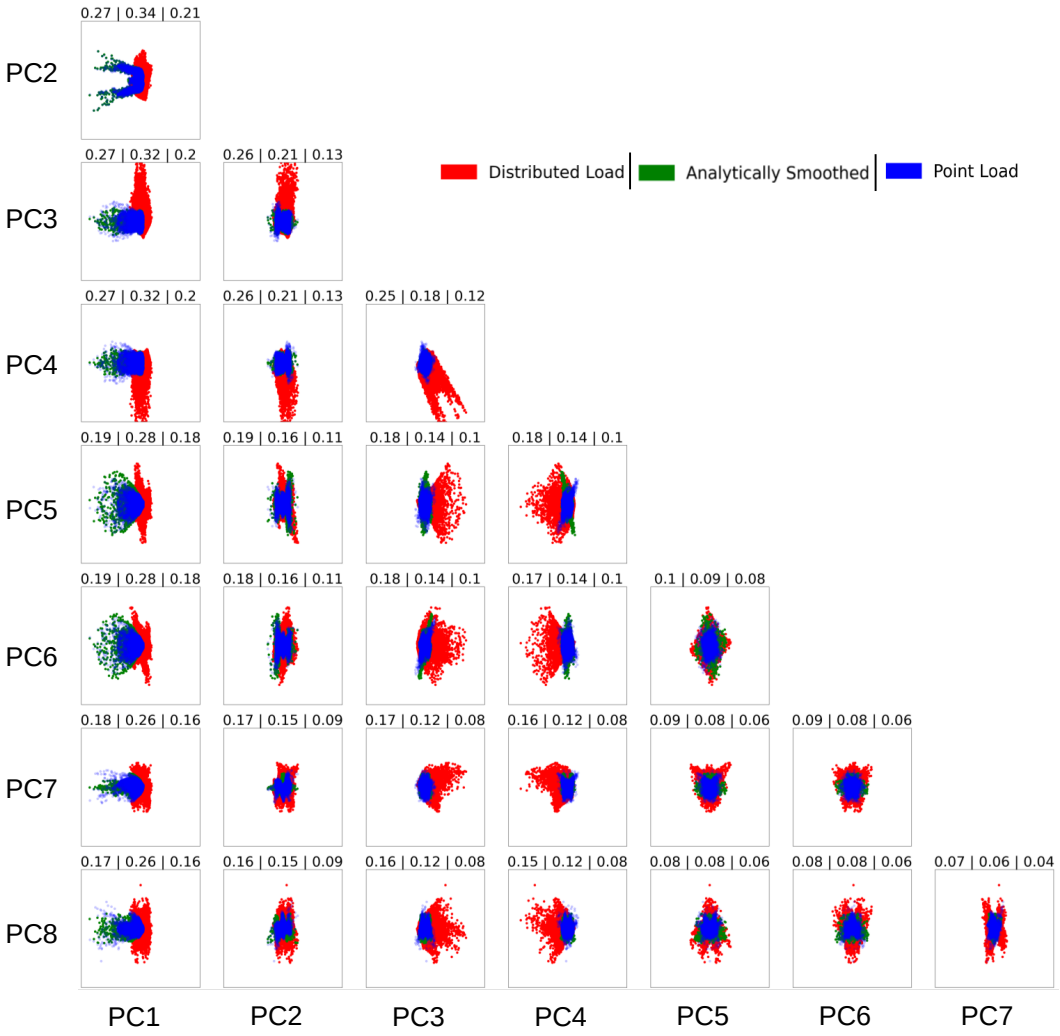


Figure E.3: Higher-order PCs for the sensitivity prediction datasets. Numbers above each subplot indicate the fraction of variance captured by the two PCs shown, for the distributed load, analytically smoothed, and original point load datasets (from left to right), which are included as separate colors.

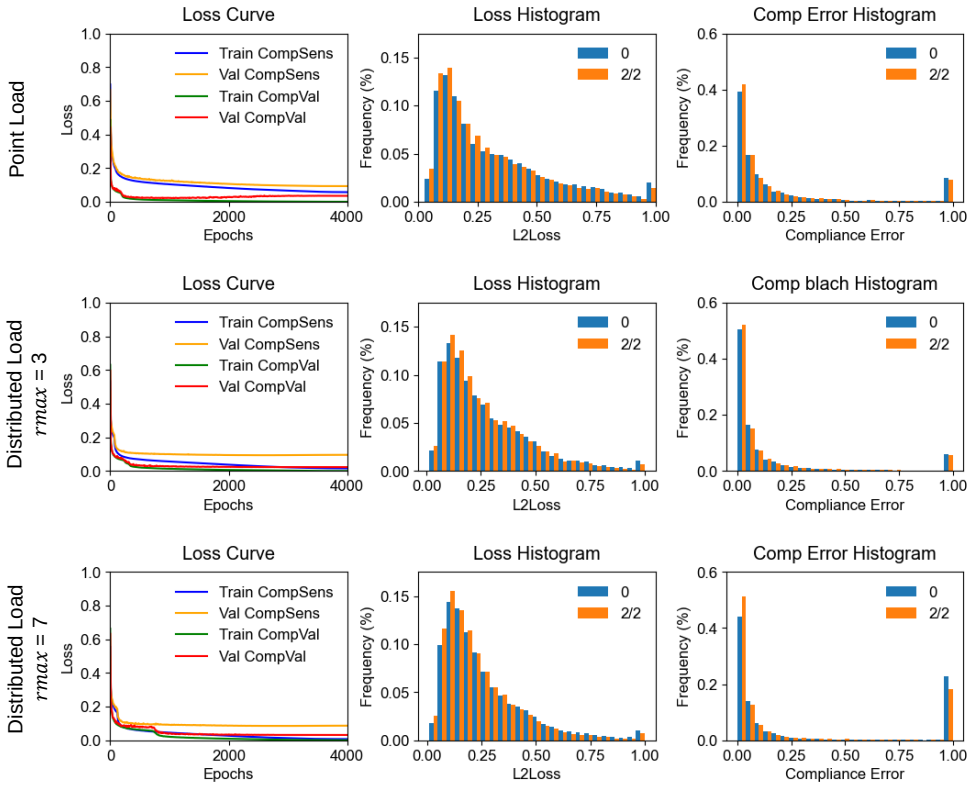


Figure E.4: PCA-Net surrogate in the compliance sensitivity configuration improves with load smoothing. The first row corresponds to point load, the second row corresponds to the case of the distributed load with r_{max} of 3, and the third row corresponds to the case of the distributed load with r_{max} of 7. The columns are the same as Figs. 4 and 6 in the main text.

E.6 below. The original point load and distributed load ($r_{max} = 3$) cases are reprinted here for comparison. Relative compliance error overall increased under the more broadly distributed load. In particular, there was a dramatic increase in the number of outliers (designs with relative compliance error greater than 1). Upon inspection of some these samples, loads were found to fall just slightly off of the structure, causing dramatically higher compliance values. This effect may have been largely avoided when r_{max} was set to 3 because the neighborhood filter included in the TO routine was similar at 3 elements, ensuring design features would generally have a thickness of at least 3 elements.

A similar activity was performed with the degree of distribution in the singularity removal process, testing parameters both narrower ($r_{max} = 8$) and wider ($r_{max} = 12$) than the original value shown in the main text. The results are shown in the same format in Figs. E.7 and E.8. Using an r_{max} of 8 when analytically removing the singularity leads to significantly worse performance

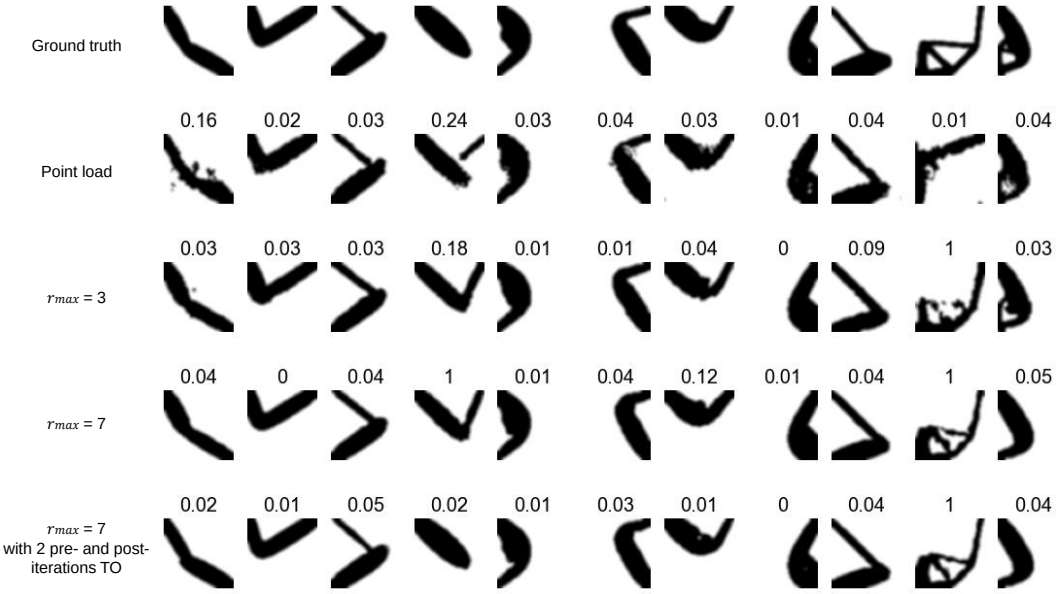


Figure E.5: Example topologies. Going down the are ground truth predictions, point load predictions, distributed load with r_{max} of 3, distributed load with r_{max} of 7, and distributed load with r_{max} of 7 with 2 TO pre-processing iterations and 2 TO post-processing iterations. The compliance errors compared with the compliance of the ground truth of each design are displayed above. Note: the compliance errors are measured using the ground truths of the respective load, not the point load.

	Training
Direct Design 7x7	4.7 hours
SensPred 7x7	4.7 hours

	Output Images	2 Post Iterations	5 Post Iterations	10 Post Iterations
Direct Design 7x7	0.00082 seconds	0.6 seconds	1 second	1.66 seconds

	Output Images	2 Pre- 2 Post-Iterations
SensPred 7x7	2.6 seconds	3.13 seconds

Figure E.6: Recorded time log for distributed load with r_{max} of 7 training and testing. The first table is the total time it took to train the distributed load of r_{max} 7 for both direct design and sensitivity prediction. The second table is the time per sample it took to make the predictions for each configuration of direct design. The third table is the time per sample it took to make the predictions for each configuration of sensitivity prediction.

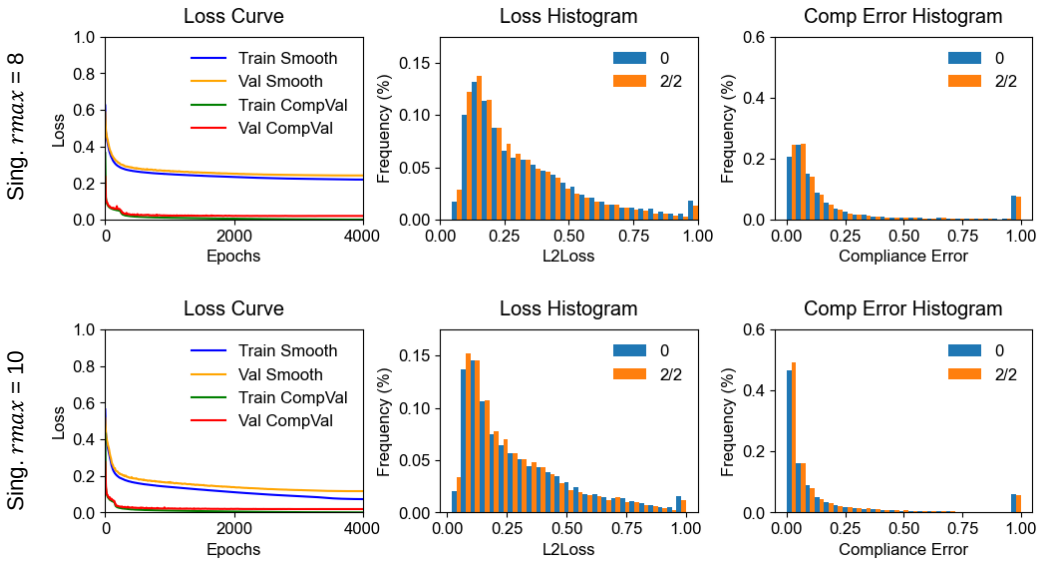


Figure E.7: Results of analytical singularity removal when done with an r_{max} of 8, 10, and 12. The first row corresponds to an r_{max} of 8, the second row corresponds to an r_{max} of 10 (used in the main text, shown here for comparison), The columns are the same as those used in Figs. 4 and 6 in the main paper.

compared to r_{max} of 10, suggesting that the entire singularity was not captured. As a result, the change may have simply complicated the problem, making it even more difficult than on the original unmodified dataset. Using an r_{max} of 12 resulted in predictions sufficiently poor that the optimizer often encountered numerical difficulties when determining the next design, so problems did not converge to any design at all. These attempts suggest that the selection of the region for singularity removal is an important factor, and an practical value for the problem set studied in this work is $r_{max} = 3$.

Collectively, because performance degraded with smoothing in both methods when the degree was too large, care must be taken in the setup and parameter selection, with the awareness that excessive smoothing may create additional unanticipated effects that outweigh any performance improvements.

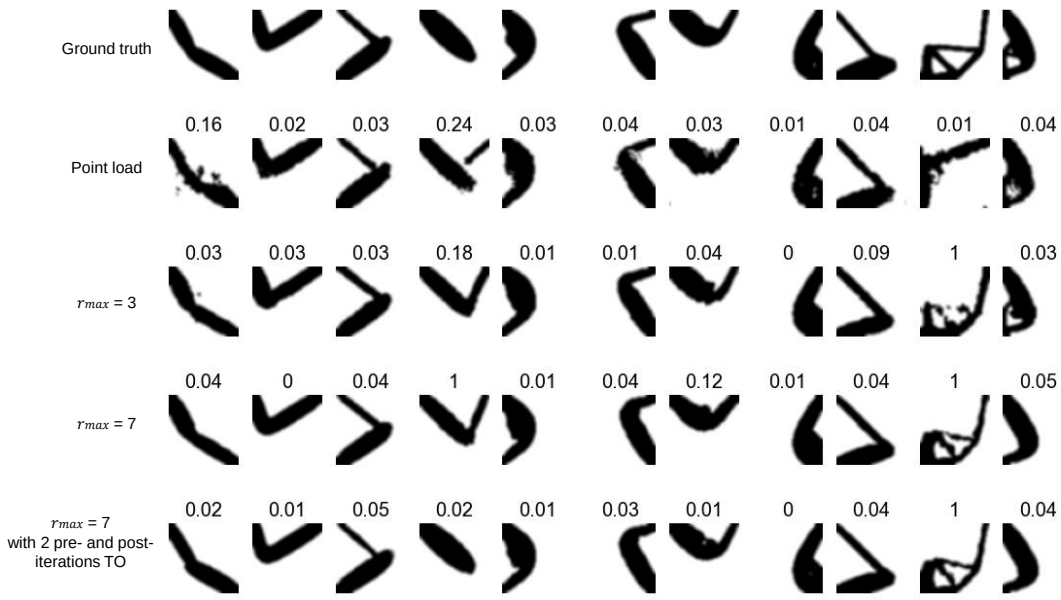


Figure E.8: Examples of predictions made using analytical singularity removal with differing r_{max} values. The first row is the ground truth, the second row are predictions when doing singularity removal with an r_{max} of 8. The third row is the same with 2 pre-processing TO iterations and 2 post-processing TO iterations. The fourth row is the predictions when doing singularity removal with an r_{max} of 10, which is also in the main text. The row after is the same with 2 pre-processing TO iterations and 2 post-processing TO iterations. No examples were created when doing removal using r_{max} of 12 due to failures of tests.

E.4 Analytical removal via a green's functions

The problem for topology optimization involves minimizing compliance subject to equilibirum,

$$\min_{\Omega} \ \mathcal{J}(u,\phi) = \int_{\Omega} \ \frac{1}{2}\nabla u \cdot \mathbb{C}\nabla u \ d\Omega, \tag{E.1}$$

$$\text{s. t. } \nabla \cdot \mathbb{C}\nabla u = f \quad \text{in } \Omega, \tag{E.2}$$

where ϕ is the design variables over which optimization is performed. f is a point load. We use adjoint method to solve this optimization problem. The compliance can be written as

$$\mathcal{J} = \int_{\Omega} \ \frac{1}{2}\nabla u \cdot \mathbb{C}\nabla u \ + v\left(\nabla \cdot \mathbb{C}\nabla u - f\right) d\Omega. \tag{E.3}$$

The directional derivative of the compliance along $\delta\phi$ is given as

$$\mathcal{J}_{,\phi}\delta\phi = \lim_{\epsilon\to 0}\frac{\mathcal{J}(\phi+\epsilon\delta\phi)-\mathcal{J}(\phi)}{\epsilon} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_\Omega \frac{1}{2}\nabla u(\phi+\epsilon\delta\phi)\cdot\mathbb{C}(\phi+\epsilon\delta\phi)\nabla u(\phi+\epsilon\delta\phi) \tag{E.4}$$

$$+ v\left(\nabla\cdot\mathbb{C}(\phi+\epsilon\delta\phi)\nabla u(\phi+\epsilon\delta\phi) - f(\phi+\epsilon\delta\phi)\right) \tag{E.5}$$

$$-\frac{1}{2}\nabla u(\phi)\cdot\mathbb{C}(\phi)\nabla u(\phi) \tag{E.6}$$

$$- v\left(\nabla\cdot\mathbb{C}(\phi)\nabla u(\phi) - f(\phi)\right) d\Omega, \tag{E.7}$$

$$= \int_\Omega \frac{1}{2}\nabla u\cdot\mathbb{C}_{,\phi}\nabla u + \nabla u_{,\phi}\cdot\mathbb{C}\nabla u \tag{E.8}$$

$$+ v\left(\nabla\cdot\mathbb{C}_{,\phi}\nabla u - f_{,\phi}\right) \tag{E.9}$$

$$+ v\left(\nabla\cdot\mathbb{C}\nabla u_{,\phi}\right)\ \delta\phi\ d\Omega, \tag{E.10}$$

where $\mathbb{C}_{,\phi}, f_{,\phi}, u_{,\phi}$ are given by

$$\mathbb{C}_{,\phi}\delta\phi = \lim_{\epsilon\to 0}\frac{\mathbb{C}(\phi+\epsilon\delta\phi)-\mathbb{C}(\phi)}{\epsilon}, \tag{E.11}$$

$$f_{,\phi}\delta\phi = \lim_{\epsilon\to 0}\frac{f(\phi+\epsilon\delta\phi)-f(\phi)}{\epsilon}, \tag{E.12}$$

$$u_{,\phi}\delta\phi = \lim_{\epsilon\to 0}\frac{u(\phi+\epsilon\delta\phi)-u(\phi)}{\epsilon}. \tag{E.13}$$

$$\tag{E.14}$$

Since v is a free variable, we can choose the value such that an adjoint equation is satisfied given by

$$0 = \int_\Omega \nabla u_{,\phi}\cdot\mathbb{C}\nabla u + v\left(\nabla\cdot\mathbb{C}\nabla u_{,\phi}\right)\ d\Omega, \tag{E.15}$$

$$= \int_\Omega \nabla u_{,\phi}\cdot\mathbb{C}\nabla u - \nabla v\cdot\mathbb{C}\nabla u_{,\phi}\ d\Omega, \tag{E.16}$$

$$= \int_\Omega \nabla u_{,\phi}\cdot\left(\mathbb{C}\nabla u - \nabla v\right)\ d\Omega, \tag{E.17}$$

$$= \nabla\cdot\mathbb{C}\nabla u - \nabla\cdot\mathbb{C}\nabla v. \tag{E.18}$$

Substituting the value of v back into derivative of the compliance, we obtain

$$\mathcal{J}_{,\phi}\delta\phi = \int_\Omega \frac{1}{2}\nabla u\cdot\mathbb{C}_{,\phi}\nabla u + v\left(\nabla\cdot\mathbb{C}_{,\phi}\nabla u - f_{,\phi}\right)\ \delta\phi\ d\Omega. \tag{E.19}$$

We know $f\in\mathbb{R}^2,\ f = f_0^i\hat{e}_i$, where $\hat{e}_i$ are unit vectors in each direction. The variables u, v contain a singular component and a smooth component. In order to deal with the singular component, we are going to look at a particular set of vector functions defined by G^i and s such that

$$\left(\mathbb{C}_{ijkl}G^i_{k,j}\right)_{,j} = \delta\hat{e}_i \quad (i \text{ no sum}) \tag{E.20}$$

$$s = \chi_0 G^i f_0^i \quad (\text{or } s_k = \chi_0 G^i_k f^i_o \text{ in index notation}) \tag{E.21}$$

where $\chi_0 \in C_0^\infty, \mathbb{R}$. The variable u is decomposed into two components containing the singular solution and a smooth integrable part (u_s)

$$u = s + u_s. \tag{E.22}$$

Substituting this decomposition into the forward problem, we obtain

$$\nabla \cdot \mathbb{C}\nabla u = \nabla \cdot \mathbb{C}\nabla s + \nabla \cdot \mathbb{C}\nabla u_s. \tag{E.23}$$

The second term on the right is smoothly integrable due to u_s. The first term can be simplified into

$$\nabla \cdot \mathbb{C}\nabla s = \left(\mathbb{C}_{ijkl}\ s_{k,l}\right)_{,j} = \left(\mathbb{C}_{ijkl}\ \chi_{0,j} G^m_k f^m_0\right)_{,j} + \left(\mathbb{C}_{ijkl}\ \chi_0 G^m_{k,l} f^m_0\right)_{,j} \tag{E.24}$$

$$= \left(\mathbb{C}_{ijkl}\ \chi_{0,j} G^m_k f^m_0\right)_{,j} + \left(\mathbb{C}_{ijkl}\chi_0\right)_{,j} G^m_{k,l} f^m_0 \tag{E.25}$$

$$+ \underbrace{\left(\mathbb{C}_{ijkl}\ G^m_{k,l} f^m_0\right)_{,j} \chi_0}_{f}. \tag{E.26}$$

We can see that the first tow terms on the right hand side are smooth. Therefore,

$$\nabla \cdot \mathbb{C}\nabla u_s = \nabla \cdot \mathbb{C}\nabla u - \nabla \cdot \mathbb{C}\nabla s, \tag{E.27}$$

is smoothly integrable. Since the adjoint problem and the forward problem are similar PDEs, we can decompose the adjoint variable in a similar way too

$$v = s' + v_s, \tag{E.28}$$

where s' is the singular component of the adjoint variable and v_s is smooth component. Furthermore,

$$\nabla \cdot \mathbb{C}\nabla v_s = \nabla \cdot \mathbb{C}\nabla v - \nabla \cdot \mathbb{C}\nabla s' \tag{E.29}$$

is smoothly integrable. Subtituting these into the compliance sensitivity, we obtain

$$\mathcal{J}_{,\phi}\delta\phi = \int_\Omega\ \frac{1}{2}\nabla u \cdot \mathbb{C}_{,\phi}\nabla u + v\left(\nabla \cdot \mathbb{C}_{,\phi}\nabla u - f_{,\phi}\right)\ \delta\phi\ d\Omega, \tag{E.30}$$

$$= \int_\Omega\ \frac{1}{2}\nabla u \cdot \mathbb{C}_{,\phi}\nabla u - \nabla v \cdot \mathbb{C}_{,\phi}\nabla u - \underbrace{v\ f_{,\phi}}_{=0}\ \delta\phi\ d\Omega. \tag{E.31}$$

Substituting the decomposition for u, v, we obtain

$$\mathcal{J}_{,\phi}\delta\phi = \int_\Omega\ \frac{1}{2}\nabla s \cdot \mathbb{C}_{,\phi}\nabla s + \nabla s \cdot \mathbb{C}_{,\phi}\nabla u_s + \frac{1}{2}\nabla u_s \cdot \mathbb{C}_{,\phi}\nabla u_s \tag{E.32}$$

$$- \nabla s' \cdot \mathbb{C}_{,\phi}\nabla s' - 2\nabla s' \cdot \mathbb{C}_{,\phi}\nabla v_s - \nabla v_s \cdot \mathbb{C}_{,\phi}\nabla v_s\ \delta\phi\ d\Omega, \tag{E.33}$$

where r is the distance from the point of load applied and c_1 is an arbitrary constant. The first and fourth term are not integrable. Therefore, we will look at the compliance sensitivity's projection perpendicular to these non-integrable terms, as these would be smooth and would be easier to learn through PCA-net. From Green's function solution, we know that the functions G^i have the form

$$G^i = c_1 \log(r)\ \hat{e}_i, \tag{E.34}$$

$$s = c_1\ \chi_0(r) \log(r) f_0^i\ \hat{e}_i. \tag{E.35}$$

Substituting this form into the sensitivity

$$\mathcal{J}_{,\phi}\delta\phi = \int_\Omega\ \frac{1}{2}\nabla s \cdot \mathbb{C}_{,\phi}\nabla s - \nabla s' \cdot \mathbb{C}_{,\phi}\nabla s'\ \delta\phi\ d\Omega \tag{E.36}$$

$$+ \int_\Omega \underbrace{\nabla s \cdot \mathbb{C}_{,\phi}\nabla u_s + \frac{1}{2}\nabla u_s \cdot \mathbb{C}_{,\phi}\nabla u_s - 2\nabla s' \cdot \mathbb{C}_{,\phi}\nabla v_s - \nabla v_s \cdot \mathbb{C}_{,\phi}\nabla v_s\ \delta\phi\ d\Omega}_{=\mathcal{J}_s\ \text{(smooth)}}, \tag{E.37}$$

$$= \left(\frac{1}{2}c_1^2 - {c'}_1^2\right) \int_\Omega \chi_0^2 \left(f_0^i \hat{e}_i \cdot \mathbb{C}_{,\phi} f_0^j \hat{e}_j\right) (1/r)^2\ d\Omega\ \delta\phi\ +\ \mathcal{J}_s\ \delta\phi. \tag{E.38}$$

$$\tag{E.39}$$

Therefore the singularity is of the form $\log(r)$. Since there is an arbitrary constant before the singular part, we can choose any value such that $\mathcal{J}_{,s}$ is smooth and learnable by a PCA Net. Therefore, we project $\mathcal{J}_{,\phi}$ onto $\log(r)$ and compute $\mathcal{J}_{,s}$.