technische universität
dortmund

A data driven framework for evolutionary problems in solid mechanics

Klaas Poelstra, Thorsten Bartel, Ben Schweizer

Preprint 2021-02 November 2021

Fakultät für Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund

tu-dortmund.de/MathPreprints

A data driven framework for evolutionary problems in solid mechanics

K. Poelstra[1], T. Bartel[2], and B. Schweizer[3]

November 18, 2021

Abstract: Data driven schemes introduced a new perspective in elasticity: While certain physical principles are regarded as invariable, material models for the relation between strain and stress are replaced by data clouds of admissible pairs of these variables. A data driven approach is of particular interest for plasticity problems, since the material modelling is even more unclear in this field. Unfortunately, so far, data driven approaches to evolutionary problems are much less understood. We try to contribute in this area and propose an evolutionary data driven scheme. We present a first analysis of the scheme regarding existence and data convergence. Encouraging numerical tests are also included.

MSC: 74C99, 35Q74, 74B20

Keywords: Plasticity, data driven, history surrogate, neural network

1 Introduction

Solid mechanics describes deformations of extended bodies. A large variety of models exists to describe small, large, or plastic deformations. The common basis was laid by Euler and Cauchy and can be described as follows: Let $\Omega \subset \mathbb{R}^n$ be the volume at rest and let $u : \Omega \to \mathbb{R}^n$ describe the deformed state due to applied loads $f : \Omega \to \mathbb{R}^n$. Then there is a symmetric stress tensor field $\sigma : \Omega \to \mathbb{R}^{n\times n}$ such that $-\nabla \cdot \sigma = f$ (balance of forces). The system is closed by a material law that relates the strain tensor $\epsilon = \nabla u : \Omega \to \mathbb{R}^{n\times n}$ and the stress tensor, we write $\sigma = \mathcal{G}(\epsilon)$ for the material law. Both, Piola-Kirchhoff finite elasticity and linearized elasticity are of this form. In order to treat a plastic evolution problem, we may interpret the material law as a map $\mathcal{G}$ that maps an evolution of strains, $\epsilon : [0, T] \to \mathbb{R}^{n\times n}$, to an evolution of stresses, $\sigma = \mathcal{G}(\epsilon) : [0, T] \to \mathbb{R}^{n\times n}$. When we interpret the material law $\mathcal{G}$ in this form, most plasticity models also have the above form.

[1]Fakultät für Mathematik, TU Dortmund, Vogelspothsweg 87, D-44227 Dortmund, klaas.poelstra@tu-dortmund.de

[2]Fakultät für Maschinenbau, TU Dortmund, Eulerstraße 5, D-44227 Dortmund, thorsten.bartel@tu-dortmund.de

[3]Fakultät für Mathematik, TU Dortmund, Vogelspothsweg 87, D-44227 Dortmund, ben.schweizer@tu-dortmund.de

In the data driven approach to solid mechanics, initiated in [7] and [8], one distinguishes sharply the two types of laws. The fundamental laws are accepted as invariable: The set $(\epsilon, \sigma) \in \mathcal{E} \subset L^2(\Omega, \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n})$ comprises functions that satisfy balance of forces and the constitutive relation $\epsilon = \nabla u$ for some admissible u. On the other hand, the material law $\sigma = \mathcal{G}(\epsilon)$ comes with uncertainties and is therefore replaced by a cloud of data points. The data points ideally come from measurements; they define a second set $(\epsilon, \sigma) \in \mathcal{D} \subset L^2(\Omega, \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n})$. In a static problem, the data driven approach is to seek a pair $(\bar{\epsilon}, \bar{\sigma}) \in \mathcal{E}$ that (approximately) minimizes the distance $\mathrm{dist}((\bar{\epsilon}, \bar{\sigma}), \mathcal{D})$ to the data set $\mathcal{D}$.

The data driven approach was outlined and illustrated with numerical tests in [7, 8]. The mathematical foundation in the setting of linear elasticity was laid in [3], finite elasticity was treated in [2]. Besides relaxation results for non-monotone laws and data convergence results, [3] establishes the topologies that are appropriate in data driven elasticity. Loosely speaking, the right notion of convergence is one in which strains and stresses converge weakly in L^2-spaces and, at the same time, differences of $(\epsilon, \sigma) \in \mathcal{E}$ and $(\epsilon', \sigma') \in \mathcal{D}$ converge strongly in L^2. The corresponding Kuratowski-limits are explained and analyzed in [3].

In [2], the authors establish an existence and consistency result for finite plasticity. Let us mention the notational differences in the two settings: In finite plasticity, one often writes Φ, $F = \nabla\Phi$, and P, for deformation, gradient, and stress; the relations to the quantities u, ϵ, and σ of linearized elasticity are given by $u = \Phi - \mathrm{id}$, $\epsilon = \frac{1}{2}(F + F^T) - \mathrm{id}$, $\sigma = P$. The geometrical background is that the manifold of rotations $SO(n)$ is replaced by its tangent space in the identity, which is the space $\mathbb{R}^{n\times n}_s$ of symmetric $n \times n$-matrices. Accordingly, in linearized elasticity (and in plasticity), one typically uses $\mathbb{R}^{n\times n}_s$ instead of $\mathbb{R}^{n\times n}$ as a target space for both ϵ and σ, and one sets $\epsilon = \nabla^s u := \frac{1}{2}(\nabla u + (\nabla u)^T)$. Mathematically, the main difference between linearized and finite elasticity is that the monotonicity of the map $\mathcal{G} : \mathbb{R}^{n\times n}_s \to \mathbb{R}^{n\times n}_s$ can reasonably be assumed in linear elasticity (positivity of Lamé constants), while the frame indifference of $\mathcal{G}$ makes every monotone law $\mathcal{G}$ unphysical in finite elasticity.

The aim of this article is to discuss a data driven framework for evolutionary problems, in particular, for the description of plastic deformations. As noted above, a history dependent map $\mathcal{G}$ can be used to describe plastic deformations. We discuss the question of how to adapt the data driven ideas to the evolutionary setting. One of the crucial mathematical observations is that strong L^2-convergence must be used at most places where weak L^2-convergence was the right tool in the stationary problem. The reason is that we need a good control of the convergence of histories in order to find convergence of actual states.

We are interested in time-dependent problems, but will restrict ourselfs to a time discrete setting. From now on, $0 < T \in \mathbb{R}$, a number $K \in \mathbb{N}$, and a family of time instances $0 = t_0 < t_1 < ... < t_K = T$ are fixed. We use $\mathbb{N}_K := \{0, ..., K\}$ as an index set and seek solution vectors $q = (q_0, ..., q_K)$, where $q_k = (\epsilon_k, \sigma_k)$ for every $k \in \mathbb{N}_K$. Let $\Omega \subset \mathbb{R}^n$ for $n \in \mathbb{N}$ be a bounded Lipschitz domain, and let $\Gamma \subset \partial\Omega$ a non-empty and (relative) open subset of the boundary. We assume that, for every $k \in \mathbb{N}_K$, Dirichlet boundary data on Γ are prescribed by a function $U_k \in H^1(\Omega, \mathbb{R}^n)$.

On deformations $(u_k)_{k\in\mathbb{N}_K}$ we impose, for every $k \in \mathbb{N}_K$, that

$$u_k - U_k \in H^1_\Gamma(\Omega) := \left\{ v \in H^1(\Omega) \,\middle|\, v|_\Gamma = 0 \text{ in the sense of traces} \right\} .$$

Loads are given by a family $(f_k)_{k\in\mathbb{N}_K}$ such that $f_k \in (H^1_\Gamma(\Omega))'$ for every $k \in \mathbb{N}_K$. On solutions $(q_k)_{k\in\mathbb{N}_K}$, we demand $q_k = (\epsilon_k, \sigma_k) \in \mathcal{E}^k_*$ for every k, where

$$\mathcal{E}^k_* := \left\{ (\epsilon,\sigma) \in L^2(\Omega;\mathbb{R}^{n\times n}_s \times \mathbb{R}^{n\times n}_s) \middle| -\nabla\cdot\sigma = f_k,\ \exists u_k :\ \epsilon = \nabla^s u_k \right\} , \tag{1.1}$$

and it is understood that the function u_k satisfies $u_k - U_k \in H^1_\Gamma(\Omega)$.

The data set $\mathcal{D}$ has a quite different nature than $\mathcal{E}^k_*$. The set $\mathcal{E}^k_*$ describes physically admissible states at a single time instance. The subscript star is used here to indicate that functions in this space have no time dependence. By contrast, in the material law, the stress σ_k at time instance k depends on the whole history of strains, i.e., on $\epsilon_0, ..., \epsilon_k$. The data set is therefore a subset of the space of evolutions,

$$\mathcal{D} \subset Z := \ell^2(\mathbb{N}_K; L^2(\Omega;\mathbb{R}^{n\times n}_s \times \mathbb{R}^{n\times n}_s)) . \tag{1.2}$$

Usually, the material law is local in nature. If the law is, additionally, independent of the position x (homogeneous material), then the data set is given by a subset

$$\mathcal{D}_{\text{loc}} \subset Z_{\text{loc}} := \ell^2(\mathbb{N}_K; \mathbb{R}^{n\times n}_s \times \mathbb{R}^{n\times n}_s) . \tag{1.3}$$

Even in this simplest case, the data set is contained in the space of evolutions. The local material laws may also depend on x, in which case we write $\mathcal{D}_{\text{loc}}(x)$. The relation between local and global space is given by

$$\mathcal{D} = \{(\epsilon,\sigma) \in Z | (\epsilon,\sigma)(x) \in \mathcal{D}_{\text{loc}}(x) \text{ for a.e. } x \in \Omega\} . \tag{1.4}$$

Our aim is to define a data driven scheme corresponding to spaces $\mathcal{E}^k_*$ and $\mathcal{D}$ as above. The scheme should provide solutions $(q_k)_{k\in\mathbb{N}_K}$ with $q_k \in \mathcal{E}^k_*$ for all k, such that some kind of distance to $\mathcal{D}$ is minimized (or: approximately minimized) in the space of evolutions. We demand that the new scheme is local in time. By that we mean that $q_k = (\epsilon_k, \sigma_k)$ can be calculated from some information on the history $\epsilon_0, ..., \epsilon_{k-1}$, and with two spaces $\mathcal{E}^k_*$ and $\mathcal{D}_*$ in the spirit of a stationary data driven scheme. An appropriate space $\mathcal{D}_*$ for a single time step needs to be constructed. We will perform such a construction by compressing the information contained in an evolution $\epsilon_0, ..., \epsilon_{k-1}$ into a new variable η_{k-1}. We refer to η as the *history surrogate*. The construction uses a new function which updates the history surrogate in each time step; the function is called the *propagator* and we use the letter H for this function.

Literature. We have already described the fundamental articles [7, 8, 3, 2]. Regarding further theoretical analysis in terms of relaxation of data sets, we mention [13]. Regarding the practical use of data driven algorithms, many recent contributions are available, as we will see in the following.

An article that is closely related to the research presented here is [4]. It also treats time dependent problems in a data driven perspective. The authors describe the fundamental problems of this setting, we refer here to their Equation (16): The

data set in time instance k is constrained to the past local history. Later on, in equation (26), an internal variable q is used; it corresponds to our η. Furthermore, a propagator P_e is used in their equation (28); it corresponds to our propagator H_{loc} of (1.8). Beginning with Section 3.4 of [4], the perspective becomes different to ours. The aim in [4] is to develop a differential formalism. That formalism is used to find relations to calculate solution updates for the variables ϵ, σ, and q. Our approach somehow remains in the spirit of global minimization: We still want to find the variables in the new time instance by solving a minimization problem. We therefore focus, in contrast to [4], on the following two aspects in the general description: (i) Creating a single data set that can be used for all time instances. (ii) Using a distance functional in a space that incorporates also the history variable.

Time dependent problems are also considered in [9], but not in the sense of history dependent material laws.

A classical finite element scheme is used in [14] to calculate the deformation of a plate with a hole. The results are used to extract data points at the different points of the geometry. Adaptations to nonlinear elasticity models is the interest in [11]. Brittle fracture mechanics is considered in [1].

In [5], the authors develop a technique to improve data driven schemes in the case of sparse data sets. The approach is based on the construction of local tangent spaces in the data set, obtained through the machine learning tool of tensor voting. Since this technique applies to the original data driven scheme, but also to the maximum entropy scheme, it is an interesting question whether or not the technique can be applied also in the evolutionary setting. The contribution [10] points in a somewhat similar direction. The aim there is to improve data sets with a model-free procedure and with methods that are related to data driven solvers.

Fundamental observations are contained in the short paper [6]. The essential solution step in the data driven approach is to minimize a distance functional. This is usually done with iterated projections: An approximate solution $q \in \mathcal{D}$ is projected onto the set $\mathcal{E}$, the result is then projected onto $\mathcal{D}$. This is repeated until a fixed point of this iteration is found. If $\mathcal{E}$ and $\mathcal{D}$ were two orthogonal affine subspaces with non-empty intersection, then this iteration would converge and provide a minimizer of the distance functional (actually: the iteration would find the minimizer in one step and the fixed point would not only be a minimizer of the distance, but a point in the intersection of the two spaces). Since $\mathcal{D}$ is not even a manifold, it is by no means clear why the iteration process should provide a solution. Kanno describes in [6] with intuitive simple examples that, in general, the iteration process indeed does not provide a minimum. Another contribution of [6] is the description of the minimization problem as a mixed-integer programming task. In that form, global minimizers can be found with branch-and-bound algorithms. We emphasize that this approach allows only a very limited numbers of unknowns.

Also [12] deals with a reformulation of the minimization task as a constrained optimization problem.

Two naive data driven evolution schemes. The problems in the construction of an evolutionary scheme can be conceived best when we investigate a naive extension of the static approach. Given a history $\hat{q}_{k-1} := (q_0, ..., q_{k-1})$ in a point x, let us consider the time independent data set that consists of those elements q_k that

possess a data-wise permitted extension: For every x,

$$\mathcal{D}_{\mathrm{loc},k}(\hat{q}_{k-1}) := \{q_k \,|\, \exists q_{k+1}, ..., q_K \text{ such that } (q_0, ..., q_{k-1}, q_k, q_{k+1}, ..., q_k) \in \mathcal{D}_{\mathrm{loc}}\} \,.$$

For a given history $\hat{q}_{k-1}$ and the corresponding data set $\mathcal{D}_k(\hat{q}_{k-1})$, we can then consider the task

$$\inf_{\bar{q}\in\mathcal{E}_*^k} \mathrm{dist}^2(\bar{q}, \mathcal{D}_k(\hat{q}_{k-1}))\,. \tag{1.5}$$

We claim that Scheme (1.5) is not useful in a practical application: Let us assume that the data set consists of discrete points, which means that its elements are of the form $((\epsilon_0^i, \sigma_0^i), ..., (\epsilon_K^i, \sigma_K^i))$ with i running in a finite index set. When a specific pair $(\epsilon_k^i, \sigma_k^i)$ is chosen at level k, there is typically only one index i that has exactly this pair at entry k. The above construction then demands the scheme to follow the data pair with index i for the rest of the process, no matter what the process demands in terms of loads or boundary conditions.

The above scheme can easily be improved in the spirit of the data driven approach. To be less rigid in the choice of q_k, we need not demand that the history, as chosen up to this point, is continued. Instead, we allow to switch to other histories as long as the distance is not too large. With

$$\tilde{\mathcal{D}}_{\mathrm{loc},k} := \{(q_0, ..., q_k) \,|\, \exists q_{k+1}, ..., q_K \text{ such that } (q_0, ..., q_K) \in \mathcal{D}_{\mathrm{loc}}\}\,,$$

we may solve, in each time step,

$$\inf_{\bar{q}\in\mathcal{E}_*} \mathrm{dist}^2((q_0, ..., q_{k-1}, \bar{q}), \tilde{\mathcal{D}}_k)\,. \tag{1.6}$$

The idea of this scheme is that we search a pair $\bar{q} = (\bar{\epsilon}_k, \bar{\sigma}_k)$ which is close to a pair of time level k in the data set, but we only consider those pairs in the data set that have a history that is similar to $(q_0, ..., q_{k-1})$.

This scheme certainly improves (1.5), since now a new data index i can be chosen at time instance k. Nevertheless, the scheme still does not seem to be practical: One has to choose a distance functional in the high dimensional space of evolutions. How important is a difference in the history? If the distance punishes deviations in the history too much, then the scheme will try to follow a single data entry, just as scheme (1.5) did. It is not clear how to adjust the scheme such that only relevant informations on the histories are used.

The new scheme with history surrogates

We start from a data set of evolutions $\mathcal{D} \subset Z$. Our aim is to compress the relevant information about a history $\hat{q}_{k-1} = (q_0, ..., q_{k-1})$ in a new variable η_{k-1}. In every spatial point $x \in \Omega$, the variable η_{k-1} should be an element of $\mathbb{R}^m$, for some fixed $m \in \mathbb{N}$. Accordingly, the global variable is $\eta \in L^2(\Omega, \mathbb{R}^m)$. A notable compression is achieved when the dimension m is much smaller than the dimension of the space for $\hat{q}_{k-1}$, which is $2 \cdot k \cdot n(n+1)/2$.

Our assumption is that the history surrogate η can be computed recursively with a propagator H:

$$\begin{aligned} H : L^2(\Omega; \mathbb{R}_s^{n\times n}) \times L^2(\Omega; \mathbb{R}^m) &\to L^2(\Omega; \mathbb{R}^m)\,, \\ (\epsilon_k, \eta_{k-1}) &\mapsto \eta_k\,. \end{aligned} \tag{1.7}$$

Usually, the function H will be given again by local functions,

$$\begin{aligned} H_{\mathrm{loc}}(\,.\,;x) : \mathbb{R}^{n\times n}_s \times \mathbb{R}^m \to \mathbb{R}^m\,, \\ (\epsilon_k(x), \eta_{k-1}(x)) \mapsto \eta_k(x)\,, \end{aligned} \tag{1.8}$$

and H is defined by the pointwise application of H_{loc}. We always assume that H is a continuous function.

At least two motivations can be given to use a function H of the recursive form as above. One motivation is the form of classical plasticity models. These use a plastic strain p as an additional variable, possibly also other hidden variables, e.g., to model isotropic hardening. The relevant information on the history is stored in these variables. The models have the property that the updates are given by the old values and the new strain, for example: p_k can be calculated from p_{k-1} and ϵ_k. When p_k contains all relevant information on the history, then $\eta_k := p_k$ can be used as a history surrogate and a propagator function H as above is defined.

Another motivation has to do with the structure of recurrent neural networks. With our assumption on H, the system can be regarded as a recurrent neural network which can be trained with a data set $\mathcal{D}_{\mathrm{loc}}$ to learn a useful map H.

We note that the map H allows to calculate the evolution $(\eta_j)_{j\le k}$ for every strain history $(\epsilon_j)_{j\le k}$. More precisely, for a strain history $\hat{\epsilon}_{k-1} = (\epsilon_0, ..., \epsilon_{k-1})$, we obtain η_k with the memory function

$$\eta_k = M(\epsilon_0, ..., \epsilon_{k-1}) := H(\epsilon_{k-1}, H(\epsilon_{k-2}, H(...H(\epsilon_0, 0)...)\,. \tag{1.9}$$

We assumed here that we always start with trivial initial data for the surrogate function: $\eta_{-1} := 0$. Note that we leave the number of arguments of M free; in that sense the mathematical terminology M_k would be more appropriate, but we do not see the danger of misunderstandings.

With the memory map M we now define the reduced data set. We use the collection of all triples (ϵ, σ, η) such that (ϵ, σ) occurs in the data set for some time instance k and η is the corresponding history surrogate,

$$\mathcal{D}^M_{*,\mathrm{loc}}(x) := \{(\epsilon_k, \sigma_k, M(\hat{\epsilon}_{k-1}; x)) \,|\, (\epsilon, \sigma) \in \mathcal{D}_{\mathrm{loc}}(x),\, k \le K\} \subset \mathbb{R}^{n\times n}_s \times \mathbb{R}^{n\times n}_s \times \mathbb{R}^m\,. \tag{1.10}$$

We emphasize that the space $\mathcal{D}^M_{*,\mathrm{loc}}$ uses the information of *all* data points and of *all* time instances k. At this point, we exploit that the system is time invariant.

The local spaces may also depend explicitly on the spatial position x. In any case, the global space consists of functions,

$$\mathcal{D}^M_* := \left\{(\epsilon, \sigma, \eta) \in L^2(\Omega) \,\middle|\, (\epsilon(x), \sigma(x), \eta(x)) \in \mathcal{D}^M_{*,\mathrm{loc}}(x)\ \forall x \in \Omega\right\}\,. \tag{1.11}$$

Our scheme will be based on the spaces $\mathcal{E}^k_*$ and $\mathcal{D}^M_*$.

Scheme for data driven evolutionary problems. We can now formulate the scheme that is proposed and analyzed in this paper. In time step k we seek for $\bar{q} = (\bar{\epsilon}, \bar{\sigma}) \in \mathcal{E}^k_*$ that is close to some pair (ϵ_l, σ_l) in the data set. Additionally, we want to make sure that the history surrogate η_{k-1} of the calculated evolution

is close to the history surrogate that corresponds to the pair (ϵ_l, σ_l). In order to achieve both aims, we consider the minimization problem

$$\inf_{\bar{q}\in\mathcal{E}_*^k} \operatorname{dist}^2\left((\bar{q}, \eta_{k-1}), \mathcal{D}_*^M\right) . \tag{1.12}$$

An adequate distance function must be chosen. Here, we use the $L^2(\Omega)$-norm in space and an ℓ^2-type distance in the finite dimensional target space $\mathbb{R}_s^{n\times n} \times \mathbb{R}_s^{n\times n} \times \mathbb{R}^m$. Our analysis of scheme (1.12) for data driven evolution problems embraces an existence result and a data convergence result in the context of plasticity.

We will use also the following notation that highlights the analogy to the stationary data driven framework:

$$(\epsilon, \sigma) \in \mathcal{D}_{*,\mathrm{loc}}^M(\eta; x) \ :\Longleftrightarrow \ \ (\epsilon, \sigma, \eta) \in \mathcal{D}_{*,\mathrm{loc}}^M(x) . \tag{1.13}$$

The corresponding (global) set $\mathcal{D}_*^M(\eta)$ (of functions) is defined as in (1.11).

2 Admissible data sets and propagators H

This section has two goals. One is to define properties of the set $\mathcal{D}_{*,\mathrm{loc}}^M(\eta)$ that admit, e.g., results on the existence of solutions. The second goal is to provide an example. More precisely, we cast a classical plasticity problem in the framework of sets $\mathcal{D}$ and $\mathcal{D}_*^M$ and show that $\mathcal{D}_*^M$ satisfies all the desired properties, see Definition 2.2 and Lemma 2.3.

In order to make the overall procedure clear, let us list three options to describe a time-discrete evolutionary system:

1. We are given a set of equations that provides, for every strain history $\epsilon_1, ..., \epsilon_k$ a corresponding stress evolution σ_k. Typically, the system will define such a mapping with the help of internal variables, say, p_k.

2. A data set $\mathcal{D}$ contains all admissible paths $(\epsilon_1, \sigma_1), ..., (\epsilon_K, \sigma_K)$. The data set might come from measurements, it might as well come from equations as in 1.; in the latter case, there is an infinite number of paths.

3. A data set $\mathcal{D}$ can be compressed with a propagator function H to a set $\mathcal{D}_*^M$ as in (1.10). The evolutionary system is entirely described by the two objects $\mathcal{D}_*^M$ and H.

We use the three concepts as follows: With equations as in 1. we generate data sets $\mathcal{D}$ as in 2. Equations as in 1. also define a propagator H that we can use to compress the data set as in 3. Important is the following: When a data set $\mathcal{D}$ was generated by measurements, we have no knowledge about an appropriate propagator H. In our numerical tests we will therefore also work with propagators H that are not those of the data generation.

Our considerations imply the following: Generating $\mathcal{D}$ from well-defined equations and applying the "correct" propagator H produces sets $\mathcal{D}_*^M$ with good properties. Applying a "wrong" propagator H can have two very different results: The "wrong" history surrogate η might still contain all the necessary information about

the history; in this case, the compressed data set $\mathcal{D}_*^M$ might still have all the desired properties. On the other hand, the "wrong" history surrogate η might miss relevant information; in this case, we cannot expect that ϵ_k and η_{k-1} allow to calculate σ_k and $\mathcal{D}_*^M$ will not have the desired properties.

Let us now collect the fundamental assumptions on the limit problem, which is assumed to be described by $\mathcal{D}_*^M$ and H as in 3. We essentially demand a graph property and some kind of monotonicity on $\mathcal{D}_*^M$. Later on, we turn to data convergence results. There, we will demand that the limit model is approximated fine and uniformly when more and better data points are collected. We emphasize that each approximate model (discrete data set) is also described by some set $\mathcal{D}_*^M$ and some propagator H, but that discrete set $\mathcal{D}_*^M$ will not satisfy the assumptions that are collected below.

Assumption 2.1 (Properties of a limiting data set $\mathcal{D}_*^M$)**.** *We consider the following properties of the limiting material data set $\mathcal{D}_*^M$, defined by a family of local sets $\mathcal{D}_{*,\mathrm{loc}}^M$.*

1. Monotonicity. *There exists a constant $\gamma \geq 0$ such that the following holds: For every $\eta \in \mathbb{R}^m$ and every $x \in \Omega$,*

$$\mathcal{D}_{*,\mathrm{loc}}^M(\eta; x) \subset \mathbb{R}_s^{n\times n} \times \mathbb{R}_s^{n\times n} \tag{2.1}$$

is the graph of a γ-monotone map. More precisely: For every $\epsilon \in \mathbb{R}_s^{n\times n}$, there exists a uniquely determined $\sigma \in \mathbb{R}_s^{n\times n}$ such that $(\epsilon, \sigma) \in \mathcal{D}_{*,\mathrm{loc}}^M(\eta; x)$. The graph is monotone with constant γ in the sense that

$$(\sigma - \sigma') \cdot (\epsilon - \epsilon') \geq \gamma |\epsilon - \epsilon'|^2 \tag{2.2}$$

for all $(\epsilon, \sigma), (\epsilon', \sigma') \in \mathcal{D}_{*,\mathrm{loc}}^M(\eta; x)$.

2. Growth. *For some constant $\gamma_0 > 0$, independent of η and x, and a constant $C_g = C_g(\eta) \geq 0$, independent of x, the data set $\mathcal{D}_{*,\mathrm{loc}}^M(\eta, x)$ satisfies the growth assumptions*

$$\gamma_0 |\sigma|^2 \leq |\epsilon|^2 + C_g(\eta) \quad \text{and} \quad \epsilon \cdot \sigma \geq \gamma_0 |\epsilon|^2 - C_g(\eta) \quad \forall (\epsilon, \sigma) \in \mathcal{D}_{*,\mathrm{loc}}^M(\eta; x)\,. \tag{2.3}$$

We assume that the map $\eta \mapsto C_g(\eta)$ has at most quadratic growth.

3. Continuity. *For every x and every sequence $(q^h, \eta^h) \in \mathcal{D}_{*,\mathrm{loc}}^M(x)$ with $(q^h, \eta^h) \to (q, \eta)$ as $h \to 0$, there holds $(q, \eta) \in \mathcal{D}_{*,\mathrm{loc}}^M(x)$.*

In the remainder of this section we describe one set of equations such that the corresponding set $\mathcal{D}_{*,\mathrm{loc}}^M$ satisfies all the above properties. Regarding the properties of the dual convex functions Ψ and Ψ^* we recall that, in a rate independent system, Ψ is the characteristic function of a convex set and Ψ^* is 1-homogeneous.

Definition 2.2 (Plasticity with kinematic hardening)**.** *We consider the local data set $\mathcal{D}_{\mathrm{loc}}$ for a specific continuous model, plasticity with kinematic hardening. The model relies on two positive symmetric linear maps $A, B : \mathbb{R}_s^{n\times n} \to \mathbb{R}_s^{n\times n}$, and a convex function $\Psi : \mathbb{R}_s^{n\times n} \to \bar{\mathbb{R}}$ with conjugate convex function Ψ^*. We assume continuity of Ψ^* and $\Psi^* \geq 0$, $\Psi^*(0) = 0$. The plastic deformation is measured with*

a plastic strain $p \in \mathbb{R}^{n\times n}_s$. Given p_{k-1} and the new strain ϵ_k, the other two variables in times instance k are defined with the two equations

$$\sigma_k = A(\epsilon_k - p_k)\,, \tag{2.4}$$
$$A(\epsilon_k - p_k) \in \partial\Psi^*(p_k - p_{k-1}) + Bp_k\,. \tag{2.5}$$

The first equation is a Hooke's law between stress and elastic strain. The second equation can equivalently be written as $p_k - p_{k-1} \in \partial\Psi(A(\epsilon_k - p_k) - Bp_k)$ and expresses the flow rule for the plastic strain.

The data set $\mathcal{D}_{\mathrm{loc}}$ is the set of all families $(\epsilon_k, \sigma_k)_{k\in\mathbb{N}_K}$ where the matrices $\epsilon_k \in \mathbb{R}^{n\times n}_s$ are arbitrary and all $\sigma_k, p_k \in \mathbb{R}^{n\times n}_s$ are determined by (2.4)–(2.5). The corresponding propagator is $H : (\epsilon_k, p_{k-1}) \mapsto p_k$.

The plasticity system (2.4)–(2.5) can be solved with variational methods, for notational simplicity we use here $f_k = 0$. With norms defined by $|\epsilon|^2_A := \epsilon \cdot A\epsilon$ and $|p|^2_B := p \cdot Bp$, one considers

$$E_{p_{k-1}}(\epsilon_k, p_k) := \frac{1}{2}|\epsilon_k - p_k|^2_A + \Psi^*(p_k - p_{k-1}) + \frac{1}{2}|p_k|^2_B \tag{2.6}$$

and the reduced functional

$$I_k(\epsilon_k) := \inf_{p_k} E_{p_{k-1}}(\epsilon_k, p_k)\,. \tag{2.7}$$

The system (2.4)–(2.5) is obtained from the partial derivatives $\partial_{\epsilon_k} E_{p_{k-1}}(\epsilon_k, p_k) = A(\epsilon_k - p_k)$ and $\partial_{p_k} E_{p_{k-1}}(\epsilon_k, p_k) = -A(\epsilon_k - p_k) + \partial\Psi^*(p_k - p_{k-1}) + Bp_k$.

Lemma 2.3 (Properties of the plasticity data set)**.** *Let $\mathcal{D}_{\mathrm{loc}}$ be given by plasticity with kinematic hardening as in Definition 2.2, let H be the corresponding map $H(\epsilon_k, p_{k-1}) = p_k$. Then $\mathcal{D}^M_{*,\mathrm{loc}}$ satisfies the properties of Assumption 2.1 with $\gamma > 0$. The propagator H satisfies a linear growth condition.*

Proof. Since the propagator H produces, in every time instance, the correct p_k of the plasticity equations, we have to analyze the set

$$\mathcal{D}^M_{*,\mathrm{loc}}(p_{k-1}) = \{(\epsilon_k, \sigma_k) \,|\, \exists p_k : (\sigma_k, p_k) \text{ solves (2.4)–(2.5) for } \epsilon_k\} \subset \mathbb{R}^{n\times n}_s \times \mathbb{R}^{n\times n}_s\,. \tag{2.8}$$

This set is defined by a map in the sense that all $\epsilon_k \in \mathbb{R}^{n\times n}_s$ are admitted and a solution $\sigma_k \in \mathbb{R}^{n\times n}_s$ exists for all $\epsilon_k \in \mathbb{R}^{n\times n}_s$.

Regarding monotonicity, we calculate with the inverse map $C = A^{-1}$. We write $|.|_C$ and $|.|_B$ for the norms that are induced by C and B. We consider two solutions $(\epsilon_k, \sigma_k, p_k)$ and $(\epsilon'_k, \sigma'_k, p'_k)$ to the same p_{k-1} and calculate

$$\begin{aligned}
(\sigma_k - \sigma'_k) \cdot (\epsilon_k - \epsilon'_k) &= (\sigma_k - \sigma'_k) \cdot C(\sigma_k - \sigma'_k) + (\sigma_k - \sigma'_k) \cdot (p_k - p'_k) \\
&= |\sigma_k - \sigma'_k|^2_C + (\sigma_k - Bp_k - \sigma'_k + Bp'_k) \cdot (p_k - p'_k) + |p_k - p'_k|^2_B \\
&\in |\sigma_k - \sigma'_k|^2_C + |p_k - p'_k|^2_B + [\partial\Psi^*(p_k - p_{k-1}) - \partial\Psi^*(p'_k - p_{k-1})] \cdot (p_k - p'_k) \\
&\geq |\sigma_k - \sigma'_k|^2_C + |p_k - p'_k|^2_B\,,
\end{aligned}$$

where we used monotonicity of the subdifferential $\partial\Psi^*$ in the last inequality. This shows monotonicity for some $\gamma > 0$ and Item 1.

We turn to the growth properties of Item 2. We set $\eta := p_{k-1}$ and consider a special solution of (2.4)–(2.5), namely $p' = \eta$, $\sigma' = B\eta$, $\epsilon' = \eta + A^{-1}B\eta$. One readily verifies that this yields a solution, since $A(\epsilon' - p') = B\eta = \sigma'$ and $\sigma' - Bp' = 0 \in \partial\Psi^*(0) = \partial\Psi^*(p' - \eta)$. Using this special solution in the monotonicity formula $(\sigma - \sigma') \cdot (\epsilon - \epsilon') \geq |\sigma - \sigma'|_C^2 + |p - p'|_B^2$ we find, with $D = \mathrm{id} + A^{-1}B$

$$(\sigma - B\eta) \cdot (\epsilon - D\eta) \geq |\sigma - B\eta|_C^2 + |p - \eta|_B^2 \,. \tag{2.9}$$

Upon rearrangement, using $\sigma = A(\epsilon - p)$ and Young's inequality, we find the condition $\epsilon \cdot \sigma \geq \gamma(|\sigma|^2 + |p|^2) - C_g(\eta)$ for some $\gamma > 0$ and with C_g of quadratic growth. This, in turn, implies also the growth condition $\epsilon \cdot \sigma \geq \gamma_0|\epsilon|^2 - C_g(\eta)$.

Relation (2.9) yields also the first growth condition, $\gamma_0|\sigma|^2 \leq |\epsilon|^2 + C_g(\eta)$, and the linear growth condition for the propagator $H : (\epsilon, \eta) \mapsto p$.

The continuity of Item 3 follows from the continuous dependence of solutions. We consider sequences $\eta^h \to \eta$ and $\epsilon^h \to \epsilon$ and solutions $(\epsilon^h, \sigma^h, p^h)$ of system (2.4)–(2.5), where p_{k-1} is replaced by η^h. The solutions are bounded by the above growth estimates. Upon choosing a subsequence, we may therefore also assume the convergences $p^h \to p$ and $\sigma^h \to \sigma$. The limits are again a solution of the time step plasticity system: In order to verify this fact, we write relation (2.5), $\sigma - Bp \in \partial\Psi^*(p - \eta)$, in an equivalent variational form: $\Psi^*(y) \leq \Psi^*(p - \eta) + (y - (p - \eta)) \cdot (\sigma - Bp)$ $\forall y$. Continuity of Ψ^* provides that limits can be taken in this relation. We recall that, here, we are analyzing the local situation (no dependence of a spatial variable) such that all convergences are strong convergences in finite dimensional vector spaces. □

3 Existence results

In this section, we analyze scheme (1.12) for general data sets. We recall that the domain Ω is bounded.

Existence for the limiting data set

Lemma 3.1 (Existence of time-discrete solutions for limit data sets)**.** *Let $\mathcal{D}_*^M$ be a (limiting) data set that satisfies Assumption 2.1 and let H be a propagator. Then, for every $k \leq K$ and every η_{k-1}, the scheme* (1.12)*,*

$$\inf_{\bar{q} \in \mathcal{E}_*^k} \mathrm{dist}^2 \left((\bar{q}, \eta_{k-1}), \mathcal{D}_*^M \right) \,, \tag{3.1}$$

possesses a solution $\bar{q} = (\epsilon_k, \sigma_k)$. Upon setting $\eta_{-1} = 0$ and $\eta_k = H(\epsilon_k, \eta_{k-1})$ for every $k \in \mathbb{N}_K$, the scheme defines a solution sequence $(\epsilon_k, \sigma_k)_{k \in \mathbb{N}_K}$. The sequence satisfies, for every $k \in \mathbb{N}_K$,

$$(\epsilon_k, \sigma_k)(., t) \in \mathcal{E}_*^k \,, \quad \text{and} \quad (\epsilon_k, \sigma_k, \eta_{k-1}) \in \mathcal{D}_*^M \,. \tag{3.2}$$

We note that (3.2) implies that the infimum in (3.1) is actually a minimum and that the minimal distance vanishes.

Proof. By Assumption 2.1, Item 1, the set $\mathcal{D}^M_{*,\mathrm{loc}}(\eta_{k-1}(x);x)$ is the graph of a monotone map $S_{k,x} : \mathbb{R}^{n\times n}_s \to \mathbb{R}^{n\times n}_s$. Our aim is to conclude with the Theorem of Browder and Minty the existence of a function $\bar{q} \in \mathcal{E}^k_* \cap \mathcal{D}^M_*$.

The maps $S_{k,x}$ allow to consider also the corresponding map on function spaces. We recall that $U_k \in H^1(\Omega,\mathbb{R}^n)$ prescribes boundary values on $\Gamma \subset \partial\Omega$ in time step k. With the homogeneous space $H^1_\Gamma(\Omega)$ we therefore consider functions $u = u_k$ such that $v = u - U_k \in H^1_\Gamma(\Omega)$. A nonlinear map $F = F_k$ can be defined by

$$F : H^1_\Gamma(\Omega) \to (H^1_\Gamma(\Omega))', \qquad F(v)(\varphi) := \int_\Omega S_k(\nabla^s(U_k+v)) \cdot \nabla^s\varphi\,, \tag{3.3}$$

where we use the natural notation $(S_k(w))(x) = S_{k,x}(w(x))$. By the first growth assumption on $\mathcal{D}^M_*(.,\eta_{k-1})$, the map S_k has at most linear growth. More precisely, we find the estimate

$$\gamma_0 \int_\Omega |S_k(\nabla^s(U_k+v))|^2 \le \int_\Omega |\nabla^s(U_k+v)(x)|^2 + C_g(\eta_{k-1}(x))\,dx\,,$$

and the quadratic growth of C_g provides that F of (3.3) is well-defined. The monotonicity of the set $\mathcal{D}^M_*(.,\eta_{k-1})$ yields the monotonicity of the maps $S_{k,x} : \mathbb{R}^{n\times n}_s \to \mathbb{R}^{n\times n}_s$ and, in turn, the monotonicity of F which reads $\langle F(w)-F(v), w-v\rangle \ge 0$.

The second growth assumption on $\mathcal{D}^M_*(.,\eta_{k-1})$ provides

$$\begin{aligned}
\langle F(v), v\rangle &= \int_\Omega S_k(\nabla^s(U_k+v)) \cdot \nabla^s(U_k+v) - \int_\Omega S_k(\nabla^s(U_k+v)) \cdot \nabla^s U_k \\
&\ge \gamma_0 \int_\Omega |\nabla^s(U_k+v))|^2 - \int_\Omega C_g(\eta_{k-1}) - \int_\Omega S_k(\nabla^s(U_k+v)) \cdot \nabla^s U_k \\
&\ge \frac{\gamma_0}{2} \int_\Omega |\nabla^s(U_k+v))|^2 - C(\eta_{k-1}) \ge \frac{\gamma_0}{4} \int_\Omega |\nabla^s v|^2 - C'(\eta_{k-1})\,.
\end{aligned}$$

Together with Korn's inequality for given boundary data, we obtain the coercivity of the map F, which requires that $\|v_j\|_{H^1_\Gamma(\Omega)} \to \infty$ implies $\langle F(v_j), v_j\rangle \to \infty$.

We finally have to check that F is continuous on finite-dimensional subspaces of $H^1_\Gamma(\Omega)$. Assumption 2.1, Item 3, provides the continuity of almost every map $S_{k,x}$ (those x are permitted where $|\eta_{k-1}(x)| < \infty$ is satisfied). As a consequence, F is continuous on finite-dimensional subspaces.

The existence theorem of Browder and Minty on monotone maps can now be applied. It yields, for $f_k \in (H^1_\Gamma(\Omega))'$, the existence of a solution v of the equation $F(v) = f_k$.

From the solution v we construct $u_k = U_k + v$ and, furthermore, $\epsilon_k := \nabla^s u_k \in L^2(\Omega)$ and $\sigma_k := S_k(\nabla^s u_k) \in L^2(\Omega)$. By choice of S_k, the solution $q_k = (\epsilon_k, \sigma_k)$ is in the surrogate data set $\mathcal{D}^M_*(.,\eta_{k-1})$. The definition of F and the set-up of the minimization problem provides $(\epsilon_k,\sigma_k) \in \mathcal{E}^k_*$. In particular, it realizes the distance 0 in the infimum problem (3.1). This yields the existence result for each time step.

With a finite number of iterations, we obtain the desired solution sequence: Given $(\epsilon_0,\sigma_0,\eta_0),...,(\epsilon_{k-1},\sigma_{k-1},\eta_{k-1})$, we solve the infimum problem (3.1) for k to obtain (ϵ_k,σ_k) and set $\eta_k = H(\epsilon_k,\eta_{k-1})$. □

Existence and boundedness of approximate solutions

We next study scheme (3.1) for data sets that need not satisfy Assumption 2.1. For finite data sets we cannot expect that the scheme can be solved with the distance 0. The situation is even worse: It is not even clear whether or not the infimum is attained (typically, for space continuous problems, there is no minimizer, for spatially discrete problems, there is a minimizer). We therefore investigate for general data sets (not necessarily a map, not necessarily monotone) approximate solutions to scheme (3.1).

An existence result for approximate solutions can be obtained for general sets $\mathcal{D}_*^M \subset \mathbb{R}_s^{n\times n} \times \mathbb{R}_s^{n\times n} \times \mathbb{R}^m$ and general propagators $H : \mathbb{R}_s^{n\times n} \times \mathbb{R}^m \to \mathbb{R}^m$. In order to study sequences, let $(h_l)_{l\in\mathbb{N}}$, $h = h_l \searrow 0$. We study reduced data sets $\mathcal{D}_{*,h}^M$ and propagators H_h.

We recall that, in an application, we are given data as a subset of evolutions,

$$\mathcal{D}_{h,\mathrm{loc}} = \left\{ (\tilde{\epsilon}^{j,h}, \tilde{\sigma}^{j,h}) \,\middle|\, j \le J(h) \right\} \subset Z_{\mathrm{loc}} \,. \tag{3.4}$$

With a propagator H_h, this data set is pre-processed to obtain $\mathcal{D}_{*,h}^M$. For a finite number of measurements $J(h)$, the set $\mathcal{D}_{*,h}^M$ contains at most $K \cdot J(h)$ triples, one triple for each time-instance and for each measured evolution.

The definition of an approximate solution requires some care since we have to distinguish between the elements $(\bar{\epsilon}_k, \bar{\sigma}_k)$ of $\mathcal{E}_*^k$, the elements (ϵ_k, σ_k) related to the data set, the elements $\tilde{\eta}_k = H_h(\bar{\epsilon}_k, \tilde{\eta}_{k-1})$ that are calculated in the iteration, and the elements η_k which are such that $(\epsilon_k, \sigma_k, \eta_{k-1})$ is indeed in the reduced data set.

Lemma 3.2 (Existence of time-discrete quasi-minimizers of (3.1))**.** *Let $h \searrow 0$ be a sequence of real numbers, let a reduced data set $\mathcal{D}_{*,h}^M \in L^2(\Omega; \mathbb{R}_s^{n\times n} \times \mathbb{R}_s^{n\times n} \times \mathbb{R}^m)$ and a propagator $H_h : L^2(\Omega;\mathbb{R}^m) \to L^2(\Omega;\mathbb{R}^m)$ be given for every h. Then there exist sequences $q^h = (\epsilon^h, \sigma^h)$, $\bar{q}^h = (\bar{\epsilon}^h, \bar{\sigma}^h) \in Z$ together with $\eta_k^h, \tilde{\eta}_k^h \in L^2(\Omega, \mathbb{R}^m)$ for every k such that:*

$$\bar{q}_k^h = (\bar{\epsilon}_k^h, \bar{\sigma}_k^h) \in \mathcal{E}_*^k \,, \quad q_k^h = (\epsilon_k^h, \sigma_k^h) \,, \ \eta_{k-1}^h \text{ with } \ (q_k^h, \eta_{k-1}^h) \in \mathcal{D}_{*,h}^M \qquad \forall k \,. \tag{3.5}$$

Setting $\tilde{\eta}_k^h := H_h(\bar{\epsilon}_k^h, \tilde{\eta}_{k-1}^h)$ for every $k \in \mathbb{N}_K$ and $\tilde{\eta}_{-1}^h := 0$, for every k holds

$$\mathrm{dist}^2((\bar{q}_k^h, \tilde{\eta}_{k-1}^h), (q_k^h, \eta_{k-1}^h)) \le h + \inf_{\bar{q}\in\mathcal{E}_*^k} \mathrm{dist}^2((\bar{q}, \tilde{\eta}_{k-1}^h), \mathcal{D}_{*,h}^M) \,. \tag{3.6}$$

Let the sets $\mathcal{D}_{*,h}^M$ satisfy the growth conditions (2.3) of Assumption 2.1 with a single value $\gamma_0 > 0$ and a single function C_g. Let furthermore the family H_h satisfy a uniform linear growth condition. Then the solution sequence is $L^2(\Omega)$-bounded, independent of h.

Proof. Step 1: Existence. For the existence result, we only have to use K times the definition of the infimum.

Let $\tilde{\eta}_{k-1}$ be given. The infimum on the right hand side of (3.6) defines some non-negative real number $A_k = \inf_{\bar{q}\in\mathcal{E}_*^k} \mathrm{dist}^2((\bar{q}, \tilde{\eta}_{k-1}), \mathcal{D}_{*,h}^M) \ge 0$. By definition of the infimum, there exists $\bar{q}_k^h = (\bar{\epsilon}_k^h, \bar{\sigma}_k^h) \in \mathcal{E}_*^k$ such that the infimum is realized up to a small error, $\mathrm{dist}^2((\bar{q}_k^h, \tilde{\eta}_{k-1}^h), \mathcal{D}_{*,h}^M) < A_k + h/2$.

The distance to a set is defined as the infimum over all distances; this implies that, given the tupel $(\bar{q}_k^h, \tilde{\eta}_{k-1}^h)$, there exists $(q_k^h, \eta_{k-1}^h) \in \mathcal{D}_{*,h}^M$ such that the distance is realized up to another small error, $\mathrm{dist}^2((\bar{q}_k^h, \tilde{\eta}_{k-1}^h), (q_k^h, \eta_{k-1}^h)) < A_k + h$. Together with $\tilde{\eta}_k^h := H_h(\bar{\epsilon}_k^h, \tilde{\eta}_{k-1}^h)$, this provides the desired sequence.

Step 2: Boundedness. Since $(\bar{\epsilon}_k^h, \bar{\sigma}_k^h)$ is in $\mathcal{E}_*$, there exists a potential $\bar{u}_k^h$ with $\bar{\epsilon}_k^h = \nabla^s \bar{u}_k^h$ and there holds $-\nabla \cdot \bar{\sigma}_k^h = f_k$. Multiplication of this equation with $u_k - U_k$ provides

$$\int_\Omega \bar{\sigma}_k^h \cdot \bar{\epsilon}_k^h = \int_\Omega f_k \, (u_k - U_k) + \int_\Omega \bar{\sigma}_k^h \cdot \nabla U_k \, .$$

The pair $(\bar{\epsilon}_k^h, \bar{\sigma}_k^h)$ is not in $\mathcal{D}_{*,h}^M$, we therefore have to transform the above relation into a relation containing $(\epsilon_k^h, \sigma_k^h)$. We rewrite the left hand side as

$$\int_\Omega \bar{\sigma}_k^h \cdot \bar{\epsilon}_k^h = \int_\Omega [\sigma_k^h + (\bar{\sigma}_k^h - \sigma_k^h)] \cdot [\epsilon_k^h + (\bar{\epsilon}_k^h - \epsilon_k^h)] \, .$$

We find

$$\begin{aligned} \int_\Omega \sigma_k^h \cdot \epsilon_k^h \leq \|f_k\|_{(H^1_\Gamma)'} \|u_k - U_k\|_{H^1_\Gamma} + \|\bar{\sigma}_k^h\| \|U_k\|_{H^1_\Gamma} \\ + C(\|\sigma_k^h\| + \|\epsilon_k^h\|) \, (\|\bar{\sigma}_k^h - \sigma_k^h\| + \|\bar{\epsilon}_k^h - \epsilon_k^h\|) + \|\bar{\sigma}_k^h - \sigma_k^h\| \, \|\bar{\epsilon}_k^h - \epsilon_k^h\| \, . \end{aligned}$$

We can now exploit the pointwise estimate $|\sigma|^2 + |\epsilon|^2 \leq C(\epsilon \cdot \sigma + C_g(\eta))$ for $(\epsilon, \sigma) \in \mathcal{D}_{*,\mathrm{loc}}^M(\eta; x)$. Together with the boundedness of differences $\|\bar{\sigma}_k^h - \sigma_k^h\|$ and $\|\bar{\epsilon}_k^h - \epsilon_k^h\|$, which follows from (3.6), we obtain

$$\int_\Omega |\sigma_k^h|^2 + |\epsilon_k^h|^2 \leq C \left(C + \int_\Omega C_g(\eta_{k-1}^h) + \|u_k\|_{H^1_\Gamma} + \|\sigma_k^h\| + \|\epsilon_k^h\| \right) .$$

Upon an application of the Cauchy-Schwarz inequality, this provides the uniform boundedness of the solution sequence. We note that the uniform linear growth condition on H_h yields the bound for η_k^h in $L^2(\Omega)$. This also provides, by the quadratic growth of C_g, the boundedness of $\int_\Omega C_g(\eta_{k-1}^h)$. □

4 Data convergence

The next aim is to analyze data convergence. We think of a situation where a limiting data set $\mathcal{D}$ is approximated by data sets $\mathcal{D}_h$. Following the above reasoning, the limit analysis is performed with reduced data sets $\mathcal{D}_{*,h}^M$. Our aim is to show that the approximate solutions of scheme (3.1), introduced in Lemma 3.2, converge to a solution of the scheme for the limiting data set.

Assumption 4.1. *We assume that the sequence $\mathcal{D}_{*,h}^M$ of reduced data sets approximates the limiting reduced data set $\mathcal{D}_*^M$ finely and uniformly. Here, fine approximation is defined by*

$$\forall (q, \eta) \in \mathcal{D}_*^M \quad \exists \text{ sequence } (q_h, \eta_h) \in \mathcal{D}_{*,h}^M : \quad (q_h, \eta_h) \to (q, \eta) \, . \tag{4.1}$$

Uniform approximation is defined by

$$\forall \text{ sequence } (q_h, \eta_h) \in \mathcal{D}_{*,h}^M : \quad \mathrm{dist}((q_h, \eta_h), \mathcal{D}_*^M) \to 0 \, . \tag{4.2}$$

We recall that the distances are $L^2(\Omega)$-distances.

We can expect that the following holds: Let $\mathcal{D}_{h,\mathrm{loc}}$ be a sequence of evolution subsets as in (3.4) and assume that this sequence approximates finely and uniformly the limit set $\mathcal{D}_{\mathrm{loc}}$. Let H be a fixed propagator. Then Assumption 4.1 is satisfied for the reduced data sets as defined by (1.10). A further analysis of such implications is part of another project.

Theorem 4.2 (Data convergence of solutions)**.** *Let Assumption 2.1 hold for the limit data set $\mathcal{D}_*^M$ with a monotonicity coefficient $\gamma > 0$. Let the sets $\mathcal{D}_{*,h}^M$ converge to the limit data set as in Assumption 4.1. We assume a uniform growth condition on the sets $\mathcal{D}_{*,h}^M$. Finally, let the propagator H and the family of propagators H_h be continuous with uniform linear growth and with uniform convergence $H_h \to H$.*

Let $(\epsilon, \sigma)(.) \in Z$ be the solution to the limit problem as found in Lemma 3.1. Let $(\bar\epsilon^h, \bar\sigma^h), (\epsilon^h, \sigma^h) \in Z$ be sequences of quasi-minimizers as in Lemma 3.2. Then there holds

$$\bar\epsilon^h \to \epsilon \quad \text{and} \quad \epsilon^h \to \epsilon \quad \text{in} \quad \ell^2(\mathbb{N}_K, L^2(\Omega))\,. \tag{4.3}$$

Proof. Step 0. Since we are treating only finitely many time instances $k \le K$, it is sufficient to show the following: The convergences $\bar\epsilon_l^h \to \epsilon_l$ and $\epsilon_l^h \to \epsilon_l$ for every $l \le k-1$ imply $\bar\epsilon_k^h \to \epsilon_k$ and $\epsilon_k^h \to \epsilon_k$.

We note that our convergence assumption for the previous time steps together with our assumptions on $H_h \to H$ implies also $\tilde\eta_{k-1}^h \to \eta_{k-1}$. Indeed, a simple induction argument shows that $\tilde\eta_l^h = H_h(\bar\epsilon_l^h, \tilde\eta_{l-1}^h) \to H(\epsilon_l, \eta_{l-1})$ as $h \to 0$ for every $l \le k-1$.

Let us conclude our preparations of the proof with the observation that the growth assumptions on $\mathcal{D}_{*,h}^M$ together with the linear growth assumption on H_h imply the boundedness of the approximate solution sequence by Lemma 3.2.

Step 1. By construction, $\bar q_k^h \in \mathcal{E}_*^k$ and $(q_k^h, \eta_{k-1}^h) \in \mathcal{D}_{*,h}^M$ satisfy (3.6). We claim that, as $h \to 0$, the right hand side of (3.6) actually converges to 0 and, hence,

$$\mathrm{dist}^2(\bar q_k^h, q_k^h) \to 0 \quad \text{and} \quad \mathrm{dist}^2(\eta_{k-1}^h, \tilde\eta_{k-1}^h) \to 0\,. \tag{4.4}$$

A consequence of (4.4) is $\eta_{k-1}^h \to \eta_{k-1}$.

In order to prove our claim about the right hand side of (3.6), we use the solution of the limit problem. We recall that $(q_k, \eta_{k-1}) \in \mathcal{D}_*^M$ with $q_k \in \mathcal{E}_*$ is constructed by solving the monotone problem $F(v_k) = f_k$ and setting $u_k = U_k + v_k$, $\epsilon_k = \nabla^s u_k$ and $\sigma_k = S_k(\epsilon_k)$, where the map S_k is determined by η_{k-1}. Because of the fine convergence of the data set $\mathcal{D}_{*,h}^M$ to $\mathcal{D}_*^M$ according to Assumption 4.1, there exists $(\check q_k^h, \check\eta_{k-1}^h) \in \mathcal{D}_{*,h}^M$ with $\mathrm{dist}((q_k, \eta_{k-1}), (\check q_k^h, \check\eta_{k-1}^h)) \to 0$. Since the infimum is not larger than one of the values, there holds

$$\inf_{\bar q \in \mathcal{E}_*} \mathrm{dist}^2((\bar q_k, \eta_{k-1}), \mathcal{D}_{*,h}^M) \le \mathrm{dist}^2((q_k, \eta_{k-1}), (\check q_k^h, \check\eta_{k-1}^h)) \to 0\,. \tag{4.5}$$

Because of $\tilde\eta_{k-1}^h \to \eta_{k-1}$, the claim follows.

Since we obtain relation (4.4), it is now sufficient to show only the first convergence in (4.3).

Step 2. Since both (ϵ_k, σ_k) and $(\bar{\epsilon}_k^h, \bar{\sigma}_k^h)$ are in $\mathcal{E}_*$, there exist potentials u_k and $\bar{u}_k^h$ and there holds $0 = -\nabla \cdot \sigma_k + \nabla \cdot \bar{\sigma}_k^h$. Multiplication of the latter equation with $u_k - \bar{u}_k^h$ provides

$$0 = \int_\Omega (\sigma_k - \bar{\sigma}_k^h) \cdot (\epsilon_k - \bar{\epsilon}_k^h) \, .$$

We now exploit the fact that $(q_k^h, \eta_{k-1}^h) \in \mathcal{D}_{*,h}^M$. The data sets $\mathcal{D}_{*,h}^M$ converge uniformly to $\mathcal{D}_*^M$. This implies that we can find $(\hat{q}_k^h, \hat{\eta}_{k-1}^h) \in \mathcal{D}_*^M$ such that the distance satisfies $\mathrm{dist}((q_k^h, \eta_{k-1}^h), (\hat{q}_k^h, \hat{\eta}_{k-1}^h)) \to 0$.

We have already obtained $\eta_{k-1}^h \to \eta_{k-1}$ and hence also $\hat{\eta}_{k-1}^h \to \eta_{k-1}$. We now define $\check{\epsilon}_k^h := \hat{\epsilon}_k^h$ and define $\check{\sigma}_k^h$ as the corresponding stress, more precisely: $\check{q}_k^h = (\check{\epsilon}_k^h, \check{\sigma}_k^h)$ satisfies $(\check{q}_k^h, \eta_{k-1}) \in \mathcal{D}_*^M$. The continuity assumption on the limiting data set, Item 3 in Assumption 2.1, allows to conclude $\mathrm{dist}(\check{q}_k^h, \hat{q}_k^h) \to 0$. This allows to calculate, for each k,

$$\begin{aligned}
0 &= \int_\Omega (\sigma_k - \bar{\sigma}_k^h) \cdot (\epsilon_k - \bar{\epsilon}_k^h) \\
&= \int_\Omega (\sigma_k - \check{\sigma}_k^h) \cdot (\epsilon_k - \bar{\epsilon}_k^h) + (\check{\sigma}_k^h - \bar{\sigma}_k^h) \cdot (\epsilon_k - \bar{\epsilon}_k^h) \\
&= \int_\Omega (\sigma_k - \check{\sigma}_k^h) \cdot (\epsilon_k - \check{\epsilon}_k^h) + (\sigma_k - \check{\sigma}_k^h) \cdot (\check{\epsilon}_k^h - \bar{\epsilon}_k^h) + (\check{\sigma}_k^h - \bar{\sigma}_k^h) \cdot (\epsilon_k - \bar{\epsilon}_k^h) \\
&= \int_\Omega (\sigma_k - \check{\sigma}_k^h) \cdot (\epsilon_k - \check{\epsilon}_k^h) + o(1) \left(\|\epsilon_k - \bar{\epsilon}_k^h\| + \|\sigma_k - \check{\sigma}_k^h\| \right) \, ,
\end{aligned}$$

where, in the last step, we exploited (4.4) and the constructions of $\hat{q}_k^h$, and $\check{q}_k^h$: All the pairs q_k^h, $\bar{q}_k^h$, $\hat{q}_k^h$, and $\check{q}_k^h$, are close to each other.

We now exploit that (ϵ_k, σ_k) and $(\check{\epsilon}_k^h, \check{\sigma}_k^h)$ satisfy the same strictly monotone relation, namely

$$(\epsilon_k, \sigma_k), (\check{\epsilon}_k^h, \check{\sigma}_k^h) \in \mathcal{D}_*^M(\eta_{k-1}) \, .$$

The uniform monotonicity of $\mathcal{D}_*^M(\eta_{k-1})$ provides

$$\gamma \|\epsilon_k - \check{\epsilon}_k^h\|^2 \leq \int_\Omega (\sigma_k - \check{\sigma}_k^h) \cdot (\epsilon_k - \check{\epsilon}_k^h) \leq o(1) \left(\|\epsilon_k - \bar{\epsilon}_k^h\| + \|\sigma_k - \check{\sigma}_k^h\| \right) \, .$$

Boundedness of all solution sequences provides

$$\gamma \|\epsilon_k - \check{\epsilon}_k^h\|^2 \leq o(1) \, .$$

By construction of $\check{q}_k^h$ we also have $\|\epsilon_k - \bar{\epsilon}_k^h\| \to 0$ and thus the result. □

5 Numerical experiments

Our numerical experiments have the aim of investigating the potential of data driven schemes for time-dependent problems. We use the setting as described above. This means, in particular, that, given a data set of evolutions of (ϵ, σ)-pairs, we enrich the data set with an η-variable and determine the values of this variable with a

propagator H that is chosen by the solver; we recall that the model is unknown to us, hence the "correct" propagator is unknown. The choice of the propagator H in the solver will be one of the most important parameters in our experiments.

In all data driven experiments, we generate the data set with one of the standard pde-models, using either elasticity or plasticity. The plasticity model defines the "correct" propagator H_{model}. The process of generating data leaves us some freedom. On the one hand, we are free how to choose ϵ-matrices (in elasticity) or time-dependent ϵ-paths (in plasticity). With the choice of the ϵ-entries the model yields σ-values, but we are free to add some stochastic perturbation to the latter.

In all cases, we use the standard data driven solver technique of [7] and [8], namely an iterated projection (once or in each time step): Loosely speaking, an approximate solution in $\mathcal{D}$ is projected onto $\mathcal{E}$, the result is projected back onto $\mathcal{D}$; a fixed point of this iteration is used as an approximate solution. We note that, by the analysis in [6], this approximate solution does typically not coincide with the correct solution, defined as the (ϵ, σ)-pair that minimizes a distance functional.

Reference solution

We consider the isothermal, isotropic, quasi-static elastoplastic model problem with Prandtl-Reuss flow rule and von Mises yield criterion and both kinematic and isotropic hardening. We use the isotropic stiffness tensor $\mathbb{C}e := 2\mu e + \lambda(\operatorname{tr} e)I$. The flow rule reads

$$(\dot{p}, \dot{a}, \dot{b}) \in \partial I_S(\sigma, \alpha, \beta) \tag{5.1}$$

where I_S is the support function of the elastic domain

$$S := \{(\sigma, \alpha, \beta) \ : \ |\operatorname{dev}(\sigma + \beta)| < \sigma_y(1 + \alpha_+)\}$$

and $\alpha = -h_{iso}a$, $\beta = -h_{kin}b$. The stress-strain relation is $\sigma(\epsilon) = \mathbb{C}(\epsilon - p)$.

Given a sequence length $T \in \mathbb{N}$ and a path $\epsilon_t \in \mathbb{R}^{n\times n}_{\text{s}}$, $1 \le t \le T$, we can calculate the corresponding stress path with a discretization of the flow rule (5.1). We set $p_0 = a_0 = b_0 = 0$ and iteratively compute (p_t, a_t, b_t) such that

$$\left(\frac{p_t - p_{t-1}}{\Delta t}, \frac{a_t - a_{t-1}}{\Delta t}, \frac{b_t - b_{t-1}}{\Delta t}\right) \in \partial I_S(\mathbb{C}(\epsilon_t - p_t), -h_{\text{iso}}a_t, -h_{\text{kin}}b_t)\,.$$

Note that this relation does not depend on the value of $\Delta t > 0$ and we might as well choose $\Delta t = 1$.

In our experiments we use the following material parameters:

$$\lambda = \frac{E\nu}{(1-2\nu)(1+\nu)}\,, \quad \mu = \frac{E}{2(1+\nu)}\,, \quad E = 210000\,, \quad \nu = 0.3\,,$$
$$\sigma_y = 500\,, \quad h_{\text{kin}} = 20000\,, \quad h_{\text{iso}} = 0.2\,.$$

Here and in the following we omit physical units, all numbers are in the S.I. system: Length in mm (millimeter), time in s (second), mass in t = 10^3kg (ton), force in N=t mm/s^2 (Newton), stress in MPa = N/mm^2 (Mega-Pascal), see Table 1.

Geometry

Our calculations are performed in space dimension $n = 2$. We consider an experiment in which a plate with a hole in the center is pulled on two opposite sides. The plate is quadratic with side length 200 (we recall that the unit is mm), the hole is a circle with radius 50. Using the symmetries of the domain, we can restrict the computation on one quadrant, $(x_1, x_2) \in (0, 100) \times (0, 100)$. The circle $B = B_{50}((100, 0))$ is removed from the domain. We prescribe the force with a force density parameter $f \in \mathbb{R}$. With the unit outer normal vector ν, the boundary conditions are:

$$\begin{aligned}
\sigma e_1 &= f e_1 \text{ on } \{x_1 = 0\}\,, \\
u_1 &= 0 \text{ and } e_2 \cdot \sigma e_1 = 0 \text{ on } \{x_1 = 100\}\,, \\
u_2 &= 0 \text{ and } e_1 \cdot \sigma e_2 = 0 \text{ on } \{x_2 = 0\}\,, \\
\sigma e_2 &= 0 \text{ on } \{x_2 = 100\}\,, \\
\sigma \nu &= 0 \text{ on } \partial\Omega \cap \partial B\,.
\end{aligned}$$

The finite element discretization uses 206 nodes and 348 triangular elements. For a given material model and a given geometry, a reference solution can be calculated in a finite element framework. Figure 2 indicates how reference solutions behave and, in particular, where plastic deformation takes place. Table 2 shows norms of the reference solution. Figure 1 shows the deformation u of the solution at time instance $t = 4$.

Quantity	Unit
Length	mm
Force	N=$(10^3$kg mm$)$/s^2
Mass	10^3kg
Time	s
Stress	MPa (N/mm^2)

Table 1: Units for the different quantities.

We fix a loading path for all our experiments: The force f is increased in 4 steps from 0 to 1000, then we unload the specimen and decrease the load in 2 steps to 500. This loading path is chosen in order to generate plastic deformations and in order to see memory effects. Given the plasticity material model and the geometry, a reference solution can be calculated in a finite element framework. Results are shown in Figure 2 and Table 2.

Data generation

Loosely speaking, we generate paths $(\epsilon_t)_{t \le T}$ by setting $\epsilon_0 = 0$ and choosing random increments $\epsilon_t - \epsilon_{t-1}$. The more detailed description is as follows: We first generate the reference solution from the model and evaluate the statistics of the strain entries. We then generate strain paths which match these statistics. In this step, we evaluate the (matrix component-wise) mean values and standard deviations of the strain

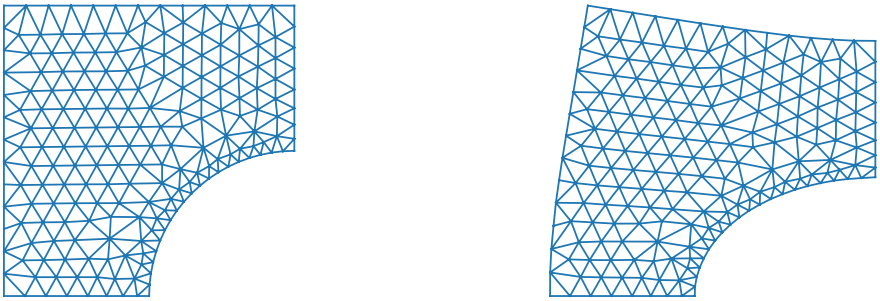

Figure 1: Left: Plot of the finite element grid. Right: All grid-points are moved according to the solution u at time $t = 4$; we multiplied all deformations with the scale factor 5 in order to have a clearly visible deformation.

	f	$\max \lvert u \rvert$	$\lVert u \rVert$	$\lVert \epsilon \rVert$	$\lVert \sigma \rVert$
1	250	0.43	0.27	0.93	0.89
2	500	1.06	0.68	2.30	1.79
3	750	1.91	1.23	4.16	2.68
4	1000	2.87	1.85	6.33	3.57
5	750	2.46	1.60	5.48	2.69
6	500	2.05	1.34	4.64	1.81

Table 2: Norms of the reference solution. For the domain Ω, we table shows $\left(|\Omega|^{-1} \int_\Omega \|\epsilon(x)\|_{\mathbb{C}}^2 \, dx\right)^{1/2}$ and $\left(|\Omega|^{-1} \int_\Omega \|\sigma(x)\|_{\mathbb{C}^{-1}}^2 \, dx\right)^{1/2}$, such that the entries are typical values of normalized strain and normalized stress. When we calculate errors in data driven experiments, we calculate with the same norm. When we report relative errors, we divide norms of differences by the corresponding values in this table.

increments across the elements of the finite element discretization. We then generate N strain paths (typically $N = 10^6$) by sampling strain increments from normal distributions with these statistics. This guarentees that the generated dataset covers the relevant portions of phase space.

In some experiments, we enrich the dataset with the data from the reference solution, i.e., for each element of the finite element discretization, we add the strain path of that element in the reference solution to the set of sampled paths.

Once a strain path is chosen, we generate the corresponding stress path according to the reference plasticity model.

When we include stochastic perturbations, we add Gaussian noise to the sigma values, the amount is regulated with the parameter $c_{\text{noise}} > 0$. More precisely, we generate symmetric matrices E with entries from independent standard normal distributions. We then add $c_{\text{noise}} \mathbb{C}^{1/2} E$ to the stress matrix. The effect of c_{noise} is shown in Figure 7.

Once that a path is generated, we "forget" the model variables p, a, b and only

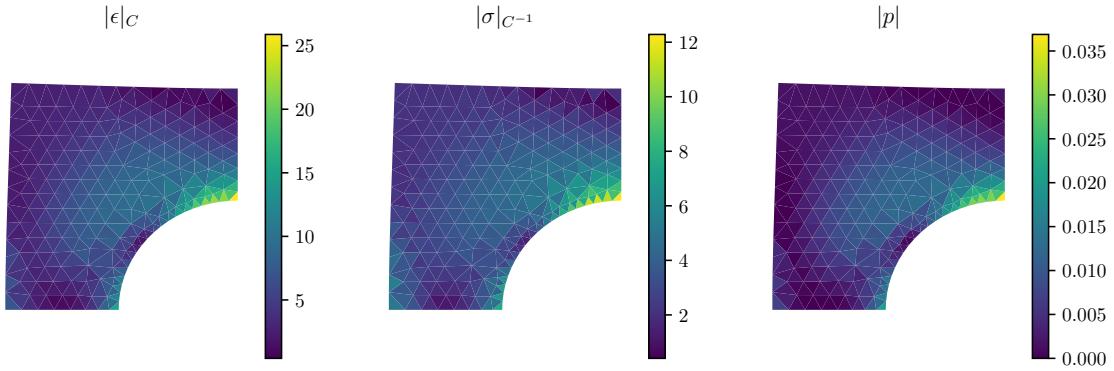

Figure 2: The finite element reference solution at time $t = 6$. We use the norms $\|\epsilon(x)\|^2_{\mathbb{C}} = \epsilon(x) : \mathbb{C}\epsilon(x)$ and $\|\sigma(x)\|^2_{\mathbb{C}^{-1}} = \sigma(x) : \mathbb{C}^{-1}\sigma(x)$ in order to have the two quantities of the same order. The color/grey level at point x indicates the norm of the solution components $\epsilon(x,t)$, $\sigma(x,t)$, and $p(x,t)$ for $t = 6$.

store the evolutions of (ϵ, σ). The idea is that only evolutions of (ϵ, σ) are measurable, no internal variables are accessible in experiments. The size of of the dataset depends on the number N of sampled paths and the length T of the individual paths. Regarding the dependence of solutions on N we refer to Figure 5.

Data driven solver and propagator functions

The data consists of strain-stress paths. The underlying idea of our approach is to enrich the data set with a history surrogate η in which the relevant history is stored. In a pre-processing step we therefore enrich the data set with some propagator H to generate triples $\{(\epsilon_t, \sigma_t, \eta_{t-1}) | t \leq T\}$. In this data set, we apply a minimization scheme in each time step.

For the minimization, we must fix a norm in $\mathbb{R}^{n\times n}_{\mathrm{s}} \times \mathbb{R}^{n\times n}_{\mathrm{s}} \times \mathbb{R}^p$ for triples (ϵ, σ, η). As noted above, this norm should induce an ℓ^2-type distance. We fix a stiffness tensor $\mathbb{C}$ and a constant vector $A \in \mathbb{R}^m_{>0}$ and define

$$\|(\epsilon, \sigma, \eta)\|^2 := \mathbb{C}\epsilon : \epsilon + \sigma : \mathbb{C}^{-1}\sigma + c_\eta |\eta|^2_A \,.$$

The weight vector A is important, since, in general, the order of magnitude of η is arbitrary. We therefore evaluate the statistics of the η-variable in our data set and then choose A such that $A_i\eta_i$ has a unit standard deviation for each component $i \leq m$. We introduced additionally a relative weight $c_\eta > 0$, which implements a tradeoff: A large value of c_η puts a large penalty on any mismatch of the history variable η so that only datapoints with an almost-correct value of η are used. These, however, will be scarce which may necessitate large errors in the $\epsilon - \sigma$ space. We investigated how the method depends on c_η, see Figure 6 for results. Loosely speaking, the choice $c_\eta = 1$ (equal weight for η-differences and stress-strain differences after normalization) is indeed a good choice.

In the pre-processing step, the data set is enriched with η-entries; this requires some propagator function H. Since the "true" propagator of the model is not known

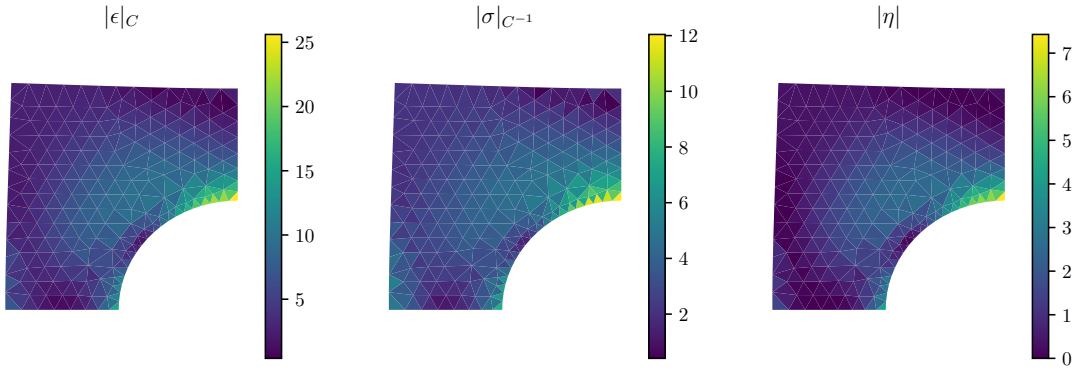

Figure 3: A data driven solution when the propagator of the data generating model is used in the data pre-processing. The experiment is run with $N = 10^6$ data points, no stochastic perturbation, and $c_\eta = 1$. The figure shows the solution components at $t = 6$. We see the qualitative agreement with the reference solution of Figure 2. Regarding the history surrogate we emphasize that η has now more components and no canonical norm; we use some ℓ^2-norm to show concentrations in this figure.

for real life data, we perform experiments for various choices of the propagator function H. Since our data are artificially generated, we can use the propagator function H of the model. We do this in most of our experiments to investigate the sources of errors. We investigave two other choices of propagators. One possibility is to use the propagator H of the model where the model parameters are altered. Another possibility is to use a neural network propagator H. Details on the latter are described in the next subsection. For results regarding the three different choices of the propagator see: Table 3 (propagator from the data generating model), Table 4 (propagator from a model with altered material parameters), Table 5 (neural network propagator).

	u error	u error (%)	ϵ error	ϵ error (%)	σ error	σ error (%)
1	0.00	0.1%	0.01	0.9%	0.01	1.1%
2	0.01	0.0%	0.02	1.0%	0.02	1.3%
3	0.03	0.2%	0.06	1.6%	0.06	2.2%
4	0.04	0.2%	0.13	2.0%	0.12	3.3%
5	0.08	0.3%	0.16	3.0%	0.19	7.0%
6	0.26	1.3%	0.22	4.6%	0.23	12.8%

Table 3: Data driven scheme with a data set that is pre-processed with the propagator of the data generating model. The numbers measure the errors that are shown in Figure 4. We use $N = 10^6$ paths to generate the data, no stochastic perturbation, an optimal initialization in each time step, and $c_\eta = 1$. The loading path is as in the reference solution. The absolute errors are norms of the differences to the reference solution, we use the natural norms that are imposed by $\mathbb{C}$. The relative errors are computed by dividing the absolute errors by the norm of the reference solution.

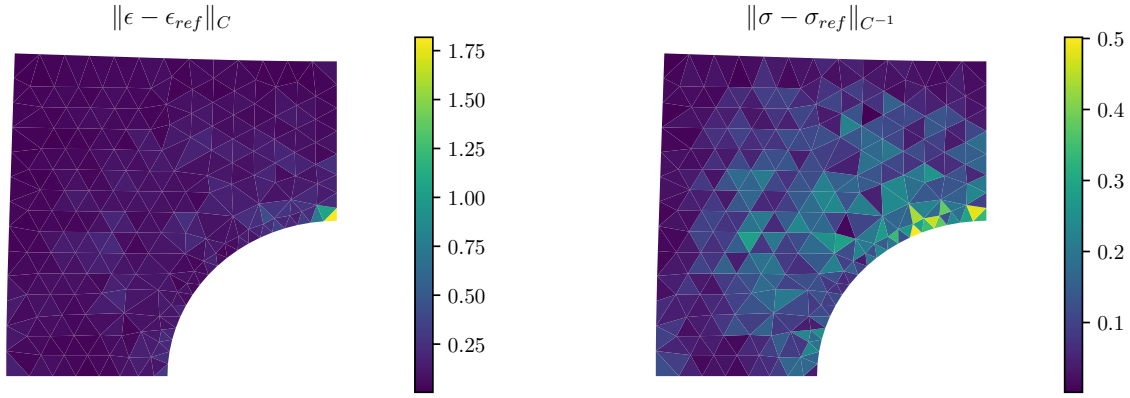

Figure 4: Differences of reference solution and data driven solution as shown and described in Figures 2 and 3. We see that the error in strain is concentrated in a corner where strains are large. Slightly different is the distribution of the stress error, which is more scattered. The corresponding norms of the error are reported in Table 3.

The neural network propagator

We use basic methods from deep learning to generate a propagator function H purely from data. In addition to H, which computes $\eta_t := H(\epsilon_t, \eta_{t-1}) = \eta_{t-1} + \hat{H}(\epsilon_t, \eta_{t-1})$, we also learn a stress funcion S with $\hat{\sigma}_t := S(\epsilon_t, \eta_t) \approx \sigma_t$. We optimize $\hat{H}$ and S by stochastic gradient descent with the objective to minimize the ℓ^2-loss $\|\hat{\sigma}_t - \sigma_t\|$. We optimize S in the space of linear functions and $\hat{H}$ in a space of feedforward neural networks. In our experiments we use 3 affine transformations for $\hat{H}$ with rectified linear units in between. The dimension of the two hidden layers is 64. The learned parameters are the matrices and vectors of the affine transformations (known as weights and biases in deep learning lingo).

The kind of neural network we implement is known as a recurrent neural network since the function H is repeatedly applied to its own output. In our experiments we use $m = 8$ as the dimension of the η-space.

It is worthwhile to ponder the relationship of S and H. In the end, we are only interested in H. But the quality of H cannot be evaluated in and of itself: It is of good quality when the hidden variables η which it produces are predictive of the $\epsilon - \sigma$ relationship. We thus learn H and S simulateneously and evaluate the σ-predictions. With S and H we have a full material model which could be used directly in computations without making any further use of the material data. However, in order to remain in the framework proposed by Kirchdörfer and Ortiz, we discard S an only use H. The advantage of this method is that it evades the danger of wrong extrapolations in the learned function S by always referencing back to the actual material data.

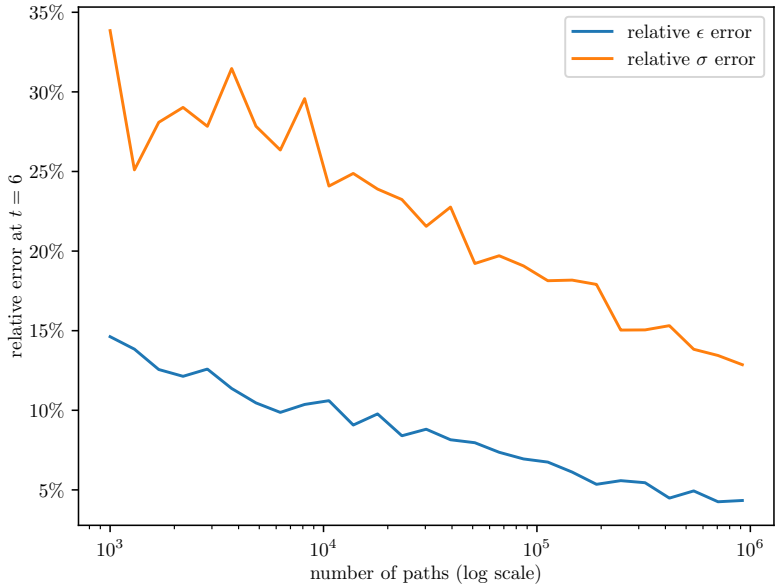

Figure 5: Dependence on the number N of paths in the data set. We use here the propagator of the data generating model and the parameter $c_\eta = 1$. No stochastic perturbation is included and we use an initialization with the reference solution in each time step.

Discussion of the numerical results

Let us summarize our findings. In an idealistic data driven solver (initialization of the solver with the reference solution, using the correct model for pre-processing the data, no stochastic perturbation), typical errors of u are about 1%, errors of ϵ are about 5%, errors of σ are about 13%. These errors are not very small.

A positive observation is that the errors are not very sensitive to the choice of the model for the data pre-processing: With the wrong model and with the neural network, the errors are of similar order, see Tables 3–5.

Also stochastic perturbations do not have a dramatic impact. With a factor $c_{\text{noise}} = 0.2$, we perturb the data by about 20% of their values, which we consider a quite large perturbation. Nevertheless, the errors are still of the same order, see Figure 7.

The dependence on the size of the data set is shown in Figure 5. We observe a logarithmic dependence. We also observed that the errors are not due to a non-optimal coefficient c_η, see Figure 6.

The above observations do not clearify the source of the error, since it does neither seem to be the model, nor the noise, nor the amount of data. To investigate the source of the error, we generated Table 6. One way of "helping" the minimization algorithm is to initialize the iteration with the reference solution. We emphasize that, if the algorithm were finding optimal solutions, this initialization would not change anything in the results. We observe that the error in ϵ is about a third when a good initialization is performed. This means that, indeed, the algorithm by itself

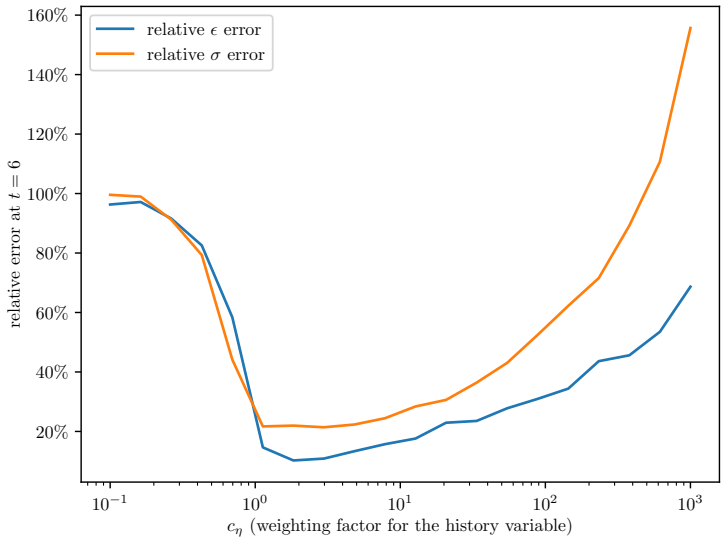

Figure 6: Many solutions of the data driven time dependent scheme, each solution is calculated on the basis of the same data set, which is generated with $N = 100000$ paths. We do not include stochastic perturbation and use the propagator of the data generating model. Plotted is the relative error of strains and stresses in the last time step in the usual norm. We vary the η weighting factor c_η from 10^{-1} to 10^3. The least errors are optained for c_η between 1 and 3.

does not find good solutions.

Another test of the quality of the solver is to include the data of the reference solution in the data set. Of course, when this is done *and* the iteration is initialized with the reference data, then the reference solution is found — this explains the vanishing errors in the right bottom. Interesting is the top row: When only the reference data are included in the data set, then errors produced by the minimization algorithm are not decreased. This is another hint that the minimization algorithm of iterated projections is indeed performing quite poorly.

Our conclusion is that the overall method is very promising. Storing a history surrogate and thus compressing the data set seems to be an adequate idea in the data driven analysis of evolutionary systems. Currently, the weakness of the method seems to be the solver of the minimization problem.

Acknowledgements.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the SPP “Variational Methods for Predicting Complex Phenomena in Engineering Structures and Materials” under grants BA 4195/3-1 and SCHW 639/12-1.

	u error	u error (%)	ϵ error	ϵ error (%)	σ error	σ error (%)
1	0.01	0.1%	0.01	1.0%	0.01	1.1%
2	0.01	0.1%	0.03	1.3%	0.03	1.5%
3	0.05	0.3%	0.07	1.7%	0.07	2.5%
4	0.08	0.3%	0.13	2.1%	0.12	3.5%
5	0.23	1.0%	0.16	2.9%	0.19	7.2%
6	0.36	1.8%	0.22	4.7%	0.24	13.3%

Table 4: Same calculations as in Table 3, but here the data are pre-processed with an altered model. We use the propagator H that is implied by the plasticity model with the parameters $\sigma_y = 200$, $h_{\text{kin}} = 5000$, and $h_{\text{iso}} = 0$. We observe that, choosing wrong material parameters in the data processing, we do not increase errors significantly.

	u error	u error (%)	ϵ error	ϵ error (%)	σ error	σ error (%)
1	0.01	0.3%	0.02	1.9%	0.01	1.5%
2	0.03	0.3%	0.04	1.7%	0.04	2.3%
3	0.10	0.5%	0.08	1.9%	0.08	2.9%
4	0.08	0.3%	0.14	2.2%	0.13	3.7%
5	0.26	1.1%	0.19	3.4%	0.21	7.8%
6	0.31	1.6%	0.23	5.0%	0.27	14.9%

Table 5: Same calculations as in Tables 3 and 4, but here the data are pre-processed with a neural network. We observe comparable orders of the error.

References

[1] P. Carrara, L. De Lorenzis, L. Stainier, and M. Ortiz. Data-driven fracture mechanics. *Comput. Methods Appl. Mech. Engrg.*, 372:113390, 27, 2020.

[2] S. Conti, S. Müller, and M. Ortiz. Data-driven finite elasticity. *Arch. Ration. Mech. Anal.*, 237(1):1–33, 2020.

[3] S. Conti, S. Müller, and M. Ortiz. Data-driven problems in elasticity. *Arch. Ration. Mech. Anal.*, 229:79–123, 2018.

[4] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, and M. Ortiz. Model-free data-driven inelasticity. *Comput. Methods Appl. Mech. Engrg.*, 350:81–99, 2019.

[5] R. Eggersmann, L. Stainier, M. Ortiz, and S. Reese. Model-free data-driven computational mechanics enhanced by tensor voting. *Comput. Methods Appl. Mech. Engrg.*, 373:113499, 18, 2021.

[6] Y. Kanno. Mixed-integer programming formulation of a data-driven solver in computational elasticity. *Optimization Letters*, 13(7):1505–1514, 2019.

[7] T. Kirchdoerfer and M. Ortiz. Data-driven computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, pages 81–101, 2016.

	without reference data	with reference data
without perfect initialization	14.1% / 16.2%	12.9% / 15.7%
with perfect initialization	4.6% / 12.8%	0.0% / 0.0%

Table 6: Performance of the minimization scheme. Entries of the table are of the form: relative ϵ-error / relative σ-error. We measure the quality of solutions when we switch on methods to facilitate the minimization scheme. Two methods are investigated: (i) include the data of the reference solution into the set of raw data, generated from random entries. (ii) initialize the minimization search with the reference solution. Values in the right bottom are 0.0%; the true values are of the order 10^{-14} and 10^{-7}. The errors in the bottom left are those of the last line in Table 3.

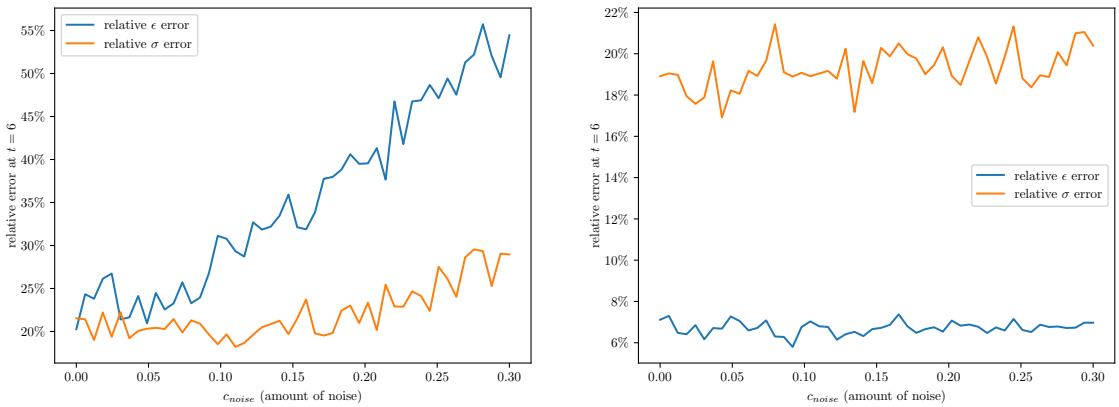

Figure 7: Dependence on the noise factor c_{noise}. We use $N = 100000$ paths and the propagator of the data generating model, $c_\eta = 1$ and no stochastic perturbation of the data. Currently: No perfect initialization. Plotted is the relative error of strains and stresses in the last time step (in the usual norms) against the noise factor c_{noise}.

[8] T. Kirchdoerfer and M. Ortiz. Data driven computing with noisy material data sets. *Comput. Methods Appl. Mech. Engrg.*, pages 622–641, 2017.

[9] T. Kirchdoerfer and M. Ortiz. Data-driven computing in dynamics. *Internat. J. Numer. Methods Engrg.*, 113(11):1697–1710, 2018.

[10] A. Leygue, M. Coret, J. Réthoré, L. Stainier, and E. Verron. Data-based derivation of material response. *Computer Methods in Applied Mechanics and Engineering*, 331:184–196, 2018.

[11] L. T. K. Nguyen and M.-A. Keip. A data-driven approach to nonlinear elasticity. *Computers & Structures*, 194:97–115, 2018.

[12] L. T. K. Nguyen, M. Rambausek, and M.-A. Keip. Variational framework for distance-minimizing method in data-driven computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, 365:112898, 39, 2020.

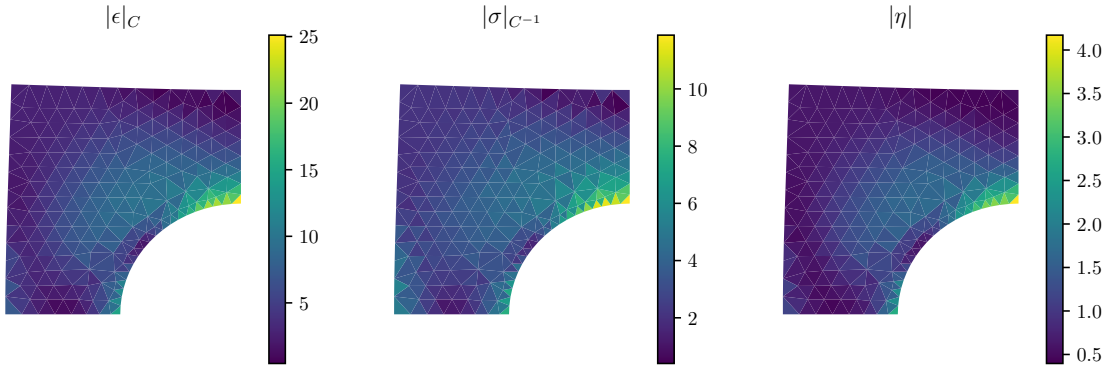

Figure 8: The data driven solution for a neural network propagator that was trained with the data.

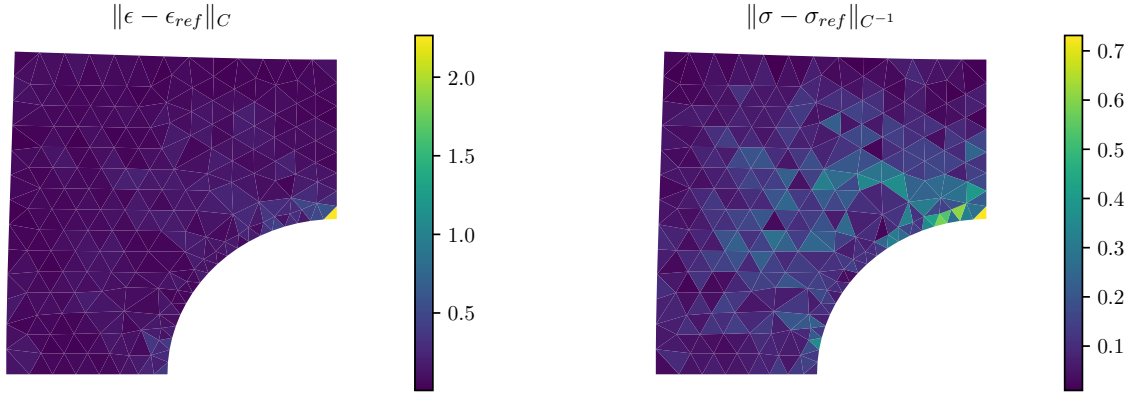

Figure 9: Differences of reference solution and data driven solution generated with a neural network, as shown and described in Figures 2 and 8. The corresponding norms of the error are reported in Table 5.

[13] M. Röger and B. Schweizer. Relaxation analysis in a data driven problem with a single outlier. *Calc. Var. Partial Differential Equations*, 59(4):Paper No. 119, 22, 2020.

[14] L. Stainier, A. Leygue, and M. Ortiz. Model-free data-driven methods in mechanics: material data identification and solvers. *Comput. Mech.*, 64(2):381–393, 2019.

Preprints ab 2015/08

2021-02 **Klaas Poelstra, Thorsten Bartel, Ben Schweizer**
A data driven framework for evolutionary problems in solid mechanics

2021-01 **Mathias Schäffner, Ben Schweizer, Yohanes Tjandrawidjaja**
Domain truncation methods for the wave equation in a homogenization limit

2020-05 **Koondanibha Mitra, Ben Schweizer, Andreas Rätz**
Travelling wave solutions for gravity fingering in porous media flows

2020-04 **Sergio Andraus, Kilian Hermann and Michael Voit**
Limit theorems and soft edge of freezing random matrix models via dual orthogonal polynomials

2020-03 **Michael Voit and Jeannette H.C. Woerner**
Limit theorems for Bessel and Dunkl processes of large dimensions and free convolutions

2020-02 **Patrizia Donato, Agnes Lamacz, Ben Schweizer**
Sound absorption by perforated walls along boundaries

2020-01 **Christoph Schumacher**
Habilitationsschrift: Concentration inequalities in random Schrödinger operators

2019-06 **Matthias Röger and Ben Schweizer**
Relaxation analysis in a data driven problem with a single outlier

2019-05 **Klaas Hendrik Poelstra, Ben Schweizer and Maik Urban**
The geometric average of curl-free fields in periodic geometries

2019-04 **Agnes Lamacz and Ben Schweizer**
Representation of solutions to wave equations with profile functions

2019-03 **Ben Schweizer**
Existence results for the Helmholtz equation in periodic wave-guides with energy methods

2019-02 **Ben Schweizer and Maik Urban**
On a limiting absorption principle for sesquilinear forms with an application to the Helmholtz equation in a waveguide

2019-01 **Michael Voit and Jeannette H.C. Woerner**
Functional central limit theorems for multivariate Bessel processes in the freezing regime

2018-08 **Ben Schweizer**
Effective Helmholtz problem in a domain with a Neumann sieve perforation

2018-07 **Sergio Andraus and Michael Voit**
Limit theorems for multivariate Bessel processes in the freezing regime

2018-06 **Michael Voit**
Central limit theorems for multivariate Bessel processes in the freezing regime

2018-05 **Elena El Behi-Gornostaeva, Koondanibha Mitra and Ben Schweizer**
Traveling wave solutions for the Richards equation with hysteresis

2018-04 **Mario Ohlberger, Ben Schweizer, Maik Urban and Barbara Verfürth**
Mathematical analysis of transmission properties of electromagnetic meta-materials

2018-03 **Margit Rösler and Michael Voit**
Beta distributions and Sonine integrals for Bessel functions on symmetric cones

2018-02 **Michael Voit**
Continuous Association Schemes and Hypergroups

2018-01 **Merdan Artykov and Michael Voit**
Some central limit theorems for random walks associated with hypergeometric functions of type BC

2017-05 **Ben Schweizer and Florian Theil**
Lattice dynamics on large time scales and dispersive effective equations

2017-04 **Frank Klinker and Christoph Reineke**
A note on the regularity of matrices with uniform polynomial entries

2017-03 **Tomáš Dohnal and Ben Schweizer**
A Bloch wave numerical scheme for scattering problems in periodic wave-guides

2017-02 **Matthias Röger and Ben Schweizer**
Strain gradient visco-plasticity with dislocation densities contributing to the energy

2017-01 **Ben Schweizer and Maik Urban**
Effective Maxwell´s equations in general periodic microstructures

2016-05 **Robert Lipton and Ben Schweizer**
Effective Maxwell´s equations for perfectly conducting split ring resonators

2016-04 **Ben Schweizer**
Resonance meets homogenization - Construction of meta-materials with astonishing properties

2016-03 **Ben Schweizer**
On Friedrichs inequality, Helmholtz decomposition, vector potentials, and the div-curl lemma

2016-02 **Michael Voit**
Generalized commutative association schemes, hypergroups, and positive product formulas

2016-01 **Agnes Lamacz and Ben Schweizer**
Effective acoustic properties of a meta-material consisting of small Helmholtz resonators

2015-13 **Christian Eggert, Ralf Gäer, Frank Klinker**
The general treatment of non-symmetric, non-balanced star circuits: On the geometrization of problems in electrical metrology

2015-12 **Daniel Kobe and Jeannette H.C. Woerner**
Oscillating Ornstein-Uhlenbeck processes and modelling electricity prices

2015-11 **Sven Glaser**
A distributional limit theorem for the realized power variation of linear fractional stable motions

2015-10 **Herold Dehling, Brice Franke and Jeannette H.C. Woerner**
Estimating drift parameters in a fractional Ornstein Uhlenbeck process with periodic mean

2015-09 **Harald Garcke, Johannes Kampmann, Andreas Rätz and Matthias Röger**
A coupled surface-Cahn-Hilliard bulk-diffusion system modeling lipid raft formation in cell membrans

2015-08 **Agnes Lamacz and Ben Schweizer**
Outgoing wave conditions in photonic crystals and transmission properties at interfaces