MODEL-FREE DATA-DRIVEN METHODS IN MECHANICS: MATERIAL DATA IDENTIFICATION AND SOLVERS

LAURENT STAINIER, ADRIEN LEYGUE, AND MICHAEL ORTIZ

ABSTRACT. This paper presents an integrated model-free data-driven approach to solid mechanics, allowing to perform numerical simulations on structures on the basis of measures of displacement fields on representative samples, without postulating a specific constitutive model. A material data identification procedure, allowing to infer strain-stress pairs from displacement fields and boundary conditions, is used to build a material database from a set of mutiaxial tests on a non-conventional sample. This database is in turn used by a data-driven solver, based on an algorithm minimizing the distance between manifolds of compatible and balanced mechanical states and the given database, to predict the response of structures of the same material, with arbitrary geometry and boundary conditions. Examples illustrate this modelling cycle and demonstrate how the data-driven identification method allows importance sampling of the material state space, yielding faster convergence of simulation results with increasing database size, when compared to synthetic material databases with regular sampling patterns.

1. INTRODUCTION

Kirchdoerfer and Ortiz [7, 8, 9] have recently proposed a new class of problems in static and dynamic elasticity, referred to as *Data-Driven problems*, defined on the space of strain-stress field pairs, or phase space. The problems consist of minimizing the distance between a given material data set and the subspace of compatible strain fields and stress fields in equilibrium. Classical solutions are recovered in the case of elasticity and conditions for convergence of Data-Driven solutions corresponding to sequences of material data sets have been derived by Conti *et al.* [1]. Data-Driven elasticity effectively reformulates the classical initial-boundary-value problem of elasticity directly from material data, thus bypassing the empirical material modelling step altogether. By eschewing empirical models, material modelling empiricism, modelling error and uncertainty are eliminated entirely and no loss of experimental information is incurred.

Data-driven elasticity relies on the availability of a material data set (referred to as a material database in the following), which may be obtained in different ways. Building this database computationally, for example through micro-macro approaches such as FE^2 [4, 5] is expensive and might require efficient model order reduction and high dimensional interpolation techniques. From an experimental point of view on the other hand, it is far from trivial to be able to somehow measure strain and stress over a wide range of deformations. Most identification methods rely on the postulate of a constitutive model, for which optimal parameters are obtained that minimize the distance between experimental measures and numerical results, for example obtained by Finite Element [14]. Leygue *et al.* [10] recently proposed an alternative approach, completely avoiding the postulate of a specific constitutive model, exploiting the distance-minimization paradigm of Data-Driven elasticity, and hence coined as Data-Driven Identification.

This paper shows how the two approaches naturally combine to provide a consistent Data-Driven strategy, opening new perspectives on the computational design cycle of structural parts. Examples presented here demonstrate the added value of complex tests on technological samples

---

which, when combined with field measurement techniques (e.g. DIC, tomography), provide rich material databases and importance sampling of the material state space.

The paper is organized as follows. We start by recalling the essential elements of the distance minimization paradigm for Data-Driven Computational Mechanics (DDCM), in section 2. Section 3 discusses the solver's algorithm, showing how it is connected to traditional linear elasticity solvers, opening the path to interfaces with off-the-shelf computational engines (commercial or other). The paper then proceeds to recall the Data-Driven Identification (DDI) algorithm, in section 4, before presenting examples of the application of the two approaches (DDI+DDCM) combined in section 5. Material databases of various sizes are first identified from a biaxial loading test on a plate, fitted with several holes to generate heterogeneous mechanical fields. This material database is then used to solve two different boundary-value problems: a plate with a single hole (different geometry and loading than the sample used for identification) and a L-shaped beam. Results obtained by this DDI+DDCM approach are confronted to those obtained by DDCM with purely synthetic material databases, with uniform sampling of material state space, and to reference solutions obtained by classical FEM using a constitutive model. The paper closes with a discussion of the methodology, results and perspectives.

2. MATERIAL DATA DISTANCE MINIMISATION PARADIGM

Let us start by recalling the main lines of the Data-Driven Computational Mechanics (DDCM) approach for elasticity, as recently proposed by [7, 8]. For simplicity, we consider discrete, or discretized, systems consisting of N nodes and M material points. Such systems typically arise from a Finite Element (FE) discretization, but other methods can be included in the framework as well. The system undergoes displacements $\boldsymbol{u} = \{\boldsymbol{u}_a\}_{a=1}^N$, with $\boldsymbol{u}_a \in \mathbb{R}^{n_a}$ and n_a the dimension of the displacement at node a, under the action of applied forces $\boldsymbol{f} = \{\boldsymbol{f}_a\}_{a=1}^N$, with $\boldsymbol{f}_a \in \mathbb{R}^{n_a}$. The local state of the system is characterized by stress and strain pairs $\{(\boldsymbol{\epsilon}_e, \boldsymbol{\sigma}_e)\}_{e=1}^M$, with $\boldsymbol{\epsilon}_e, \boldsymbol{\sigma}_e \in \mathbb{R}^{m_e}$ and m_e the dimension of stress and strain at material point e. We regard $\boldsymbol{z}_e = (\boldsymbol{\epsilon}_e, \boldsymbol{\sigma}_e)$ as a point in a local phase space $Z_e = \mathbb{R}^{m_e} \times \mathbb{R}^{m_e}$ and $\boldsymbol{z} = \{(\boldsymbol{\epsilon}_e, \boldsymbol{\sigma}_e)\}_{e=1}^M$ as a point in the global phase space $Z = Z_1 \times \cdots \times Z_M$.

The internal state of the system is subject to the compatibility and equilibrium constraints of the general form

(1a)
$$\boldsymbol{\epsilon}_e = \mathbf{B}_e \boldsymbol{u}, \quad e = 1, \ldots, M,$$

(1b)
$$\sum_{e=1}^M w_e \mathbf{B}_e^T \boldsymbol{\sigma}_e = \boldsymbol{f},$$

where $\{w_e\}_{e=1}^M$ are elements of volume and $\mathbf{B}_e$ is a discrete strain operator for material point e. We note that constraints (1) are universal, or material-independent. They define a subspace, or constraint set,

(2)
$$E = \{\boldsymbol{z} \in Z : (1a) \text{ and } (1b)\},$$

consisting of all compatible and equilibrated internal states. In (2) and subsequently, the symbol : is used to mean 'given' or 'subject to' or 'conditioned to'. Within this subspace, the local state satisfies the power identity

(3)
$$\boldsymbol{f} \cdot \boldsymbol{u} = \sum_{e=1}^M w_e \, \boldsymbol{\sigma}_e \cdot \boldsymbol{\epsilon}_e.$$

In classical elasticity, the problem (1) is closed by appending local material laws, e. g., functions of the general form

$$\boldsymbol{\sigma}_e = \hat{\boldsymbol{\sigma}}_e(\boldsymbol{\epsilon}_e), \quad e = 1, \ldots, M, \tag{4}$$

where $\hat{\boldsymbol{\sigma}}_e : \mathbb{R}^{m_e} \to \mathbb{R}^{m_e}$. However, often material behavior is only known through a material data set D_e of points $\boldsymbol{z}_e = (\boldsymbol{\epsilon}_e, \boldsymbol{\sigma}_e) \in Z_e$ obtained experimentally or by some other means. Again, the conventional response to this situation is to deduce a material law $\hat{\boldsymbol{\sigma}}_e$ from the data set D_e by some appropriate means, thus reverting to the classical setting (4).

The Data-Driven reformulation of the classical problems of mechanics consists of formulating boundary-value problems directly in terms of the material data, thus entirely bypassing the material modeling step altogether [7]. A class of Data-Driven problems consists of finding the compatible and equilibrated internal state $\boldsymbol{z} \in E$ that minimizes the distance to the global material data set $D = D_1 \times \cdots \times D_M$. To this end, we metrize the local phase spaces Z_e by means of norms of the form

$$|\boldsymbol{z}_e|_e = \left(\mathbb{C}_e \boldsymbol{\epsilon}_e \cdot \boldsymbol{\epsilon}_e + \mathbb{C}_e^{-1} \boldsymbol{\sigma}_e \cdot \boldsymbol{\sigma}_e\right)^{1/2}, \tag{5}$$

for some symmetric and positive-definite matrices $\{\mathbb{C}_e\}_{e=1}^M$, with corresponding distance

$$d_e(\boldsymbol{z}_e, \boldsymbol{y}_e) = |\boldsymbol{z}_e - \boldsymbol{y}_e|_e, \tag{6}$$

for $\boldsymbol{y}_e, \boldsymbol{z}_e \in Z_e$. The local norms induce a metrization of the global phase Z by means of the global norm

$$|\boldsymbol{z}| = \left(\sum_{e=1}^M w_e |\boldsymbol{z}_e|_e^2\right)^{1/2}, \tag{7}$$

with associated distance

$$d(\boldsymbol{z}, \boldsymbol{y}) = |\boldsymbol{z} - \boldsymbol{y}|, \tag{8}$$

for $\boldsymbol{y}, \boldsymbol{z} \in Z$. The distance-minimizing Data-Driven problem is, then,

$$\min_{\boldsymbol{y} \in D} \min_{\boldsymbol{z} \in E} d(\boldsymbol{z}, \boldsymbol{y}) = \min_{\boldsymbol{z} \in E} \min_{\boldsymbol{y} \in D} d(\boldsymbol{z}, \boldsymbol{y}), \tag{9}$$

i. e., we wish to find the point $\boldsymbol{y} \in D$ in the material data set that is closest to the constraint set E of compatible and equilibrated internal states or, equivalently, we wish to find the compatible and equilibrated internal state $\boldsymbol{z} \in E$ that is closest to the material data set D.

We emphasize that the local material data sets can be graphs, point sets, 'fat sets' and ranges, or any other arbitrary set in phase space. Evidently, the classical problem is recovered if the local material data sets are chosen as

$$D_e = \{(\boldsymbol{\epsilon}_e, \hat{\boldsymbol{\sigma}}_e(\boldsymbol{\epsilon}_e))\}, \tag{10}$$

i. e., as *graphs* in Z_e defined by the material law (4). Thus, the Data-Driven reformulation (9) extends –and subsumes as special cases– the classical problems of mechanics.

3. Data-Driven simulation algorithm

Note that, for fixed $\boldsymbol{y} \in D$, the closest point projection $\boldsymbol{z} = P_E \boldsymbol{y}$ onto E follows by minimizing the quadratic function $d^2(\cdot, \boldsymbol{y})$ subject to the constraints (1). The compatibility constraint (1a) can be enforced directly by introducing a displacement field $\boldsymbol{u}$. The equilibrium constraint (1b) can then be enforced by means of Lagrange multipliers $\boldsymbol{\lambda}$ representing virtual displacements of the system.

With $\boldsymbol{y} \equiv \{(\boldsymbol{\epsilon}_e^*, \boldsymbol{\sigma}_e^*)\}_{e=1}^M$ given, e. g. from a previous iteration, the corresponding Euler-Lagrange equations are [7]

$$(11a) \qquad \left(\sum_{e=1}^M w_e \mathbf{B}_e^T \mathbb{C}_e \mathbf{B}_e\right) \boldsymbol{u} = \sum_{e=1}^M w_e \mathbf{B}_e^T \mathbb{C}_e \boldsymbol{\epsilon}_e^*,$$

$$(11b) \qquad \left(\sum_{e=1}^M w_e \mathbf{B}_e^T \mathbb{C}_e \mathbf{B}_e\right) \boldsymbol{\lambda} = \boldsymbol{f} - \sum_{e=1}^M w_e \mathbf{B}_e^T \boldsymbol{\sigma}_e^*,$$

which define two standard linear displacement problems. The closest point $\boldsymbol{z} = P_E \boldsymbol{y} \in E$ then follows as

$$(12a) \qquad \boldsymbol{\epsilon}_e = \mathbf{B}_e \boldsymbol{u}, \quad e = 1, \dots, M,$$

$$(12b) \qquad \boldsymbol{\sigma}_e = \boldsymbol{\sigma}_e^* + \mathbb{C}_e \mathbf{B}_e \boldsymbol{\lambda}, \quad e = 1, \dots, M.$$

A simple Data-Driven solver then consists of the fixed point iteration [7]

$$(13) \qquad \boldsymbol{z}_{j+1} = P_E P_D \boldsymbol{z}_j,$$

for $j = 0, 1, \dots$ and $\boldsymbol{z}_0 \in Z$ arbitrary, where P_D denotes the closest point projection of a point in Z onto D. Iteration (13) first finds the closest point $P_D \boldsymbol{z}_j$ to $\boldsymbol{z}_j$ on the material data set D and then projects the result back to the constraint set E. The iteration is repeated until $P_D \boldsymbol{z}_{j+1} = P_D \boldsymbol{z}_j$, i. e., until the data association to points in the material data set remains unchanged. Note that more sophisticated algorithms can be considered to solve this combinatorial optimization problem [6].

Equations (11) define two standard linear elasticity problems, which can be interpreted as follows. Equation (11a) states that displacement field should be compatible with material strains $\{\boldsymbol{\epsilon}_i^*\}$ in a weak sense, i. e. strains computed from $\boldsymbol{u}$ at material points associated to the same data point $\boldsymbol{\epsilon}_i^*$ should average to that value. In view of (11b) or (12b), Lagrange multipliers $\boldsymbol{\lambda}$ can be interpreted as a discrete displacement field corresponding to the mismatch between mechanical and material stresses, resp. $\boldsymbol{\sigma}_e$ and $\boldsymbol{\sigma}_e^*$. From a practical point of view, the above linear elasticity problems can be treated as classical problems, with arbitrary elastic properties $\mathbb{C}_e$ (possibly non-homogeneous), and subject to eigen-strain or eigen-stress fields. In particular, system (11a) corresponds to a linear elasticity problems where the physical kinematic boundary conditions are enforced, and where the only other loading consists in the eigen-strain field described by $\{\boldsymbol{\epsilon}_e^*\}$, i. e. where applied body forces and (non-zero) static boundary conditions of the physical problem are not enforced. On the other hand, system (11b) corresponds to a linear elasticity problem with homogeneous kinematic boundary conditions, where the physical loading is applied (body forces and static boundary conditions) together with a field or eigen-stresses (or residual stresses) described by $\{\boldsymbol{\sigma}_e^*\}$. Ideally, this eigen-stress field should be balanced with the applied loading, yielding null Lagrange multipliers. Finally note that the linearity of systems (11) to be solved is completely independent of the linearity or non-linearity of the material behavior described by D.

The convergence properties of the fixed-point solver (13) have been investigated in [7]. The Data-Driven paradigm has been extended to dynamics [9], finite kinematics [13] and objective functions other than phase-space distance can be found in [8]. The well-posedness of Data-Driven problems and properties of convergence with respect to the data set have been investigated in [1].

4. Material data identification

The DDCM paradigm exposed in section 2 relies critically on the availability of a material data set D. For three-dimensional elasticity, sufficient phase-space coverage (i. e. importance sampling) may require a very large number of data points, which may not be easily amenable by classical mechanical testing (uniaxial, biaxial or shear loadings). To address this challenge, a material data set can be

directly constructed, as proposed by [10], from a collection of displacement and (non homogeneous) strain fields, associated with a series of known boundary conditions, i. e. imposed displacements and/or (resultant) forces. These fields could for example be obtained by Digital Image Correlation (DIC). This Data-Driven Identification (DDI) method, summarized below, simultaneously identifies the stress component of the mechanical state for each loading condition and the full material states database, which is common to all loading conditions.

For each data item α (or snapshot), the following quantities are available:

- discrete (e. g. nodal) displacements $\boldsymbol{u}^\alpha = \{\boldsymbol{u}_a^\alpha\}_{a=1}^N$,
- the (discretized) geometry, encoded through matrices $\mathbf{B}_e^\alpha$, which can compute the mechanical strain $\boldsymbol{\epsilon}_e^\alpha = \mathbf{B}_e^\alpha \boldsymbol{u}^\alpha$,
- applied nodal forces $\boldsymbol{f}^\alpha$,
- prescribed nodal displacements.

The aim of the DDI technique is to compute a number N^* of material states $(\boldsymbol{\epsilon}_i^*, \boldsymbol{\sigma}_i^*)$ such that:

- for each snapshot α and material point e, it is possible to compute the mechanical state $\boldsymbol{\sigma}_e^\alpha$ which satisfies mechanical equilibrium,
- for each snapshot, a material state $(\boldsymbol{\epsilon}_{ie}^*, \boldsymbol{\sigma}_e^*)$ can be assigned to each material point e which is the closest to the mechanical state according to norm (5).

It results the following minimization problem

$$(14) \qquad \min_{\boldsymbol{\sigma}_e^\alpha, \boldsymbol{\epsilon}_i^*, \boldsymbol{\sigma}_i^*, ie^\alpha} \sum_\alpha d(\{(\boldsymbol{\epsilon}_e^\alpha, \boldsymbol{\sigma}_e^\alpha)\}, \{(\boldsymbol{\epsilon}_{ie^\alpha}^*, \boldsymbol{\sigma}_{ie^\alpha}^*)\}),$$

subject to constraints

$$(15) \qquad \sum_{e=1}^M w_e^\alpha \mathbf{B}_e^{\alpha T} \boldsymbol{\sigma}_e^\alpha = \boldsymbol{f}^\alpha \quad \forall \alpha.$$

In (14), ie^α denotes a mapping between material points (e) and data points (i) for snapshot α. For an arbitrary state mapping ie^α, the equilibrium constraints (15) are enforced by means of Lagrange multipliers $\boldsymbol{\eta}^\alpha$, yielding the following stationarity equations [10]:

$$(16a) \qquad \left(\sum_{e=1}^M w_e^\alpha \mathbf{B}_e^{\alpha T} \mathbb{C}_e \mathbf{B}_e^\alpha\right) \boldsymbol{\eta}^\alpha - \sum_{e=1}^M w_e^\alpha \mathbf{B}_e^{\alpha T} \boldsymbol{\sigma}_{ie^\alpha(e)}^* = \boldsymbol{f}^\alpha \quad \forall \alpha,$$

$$(16b) \qquad \sum_\alpha \sum_{e:ie^\alpha(e)=i} w_e^\alpha \mathbf{B}_e^\alpha \boldsymbol{\eta}^\alpha = 0 \quad \forall i \in [1:N^*].$$

This linear system of equations is solved to simultaneously determine $\{\boldsymbol{\sigma}_i^*\}_{i=1}^{N^*}$ and $\{\boldsymbol{\sigma}_e^\alpha\}_{e=1}^M$ ($\forall \alpha$) (through $\boldsymbol{\eta}^\alpha$, see [10]). The following fixed-point iterations algorithm is then used to compute the material data set, mechanical stresses, and the state mapping:

(1) simultaneously initialize $\{\boldsymbol{\epsilon}_i^*\}$ and $\{ie^\alpha\}$ by a *k-means* algorithm on $\{\boldsymbol{\epsilon}_e^\alpha\}$,
(2) simultaneously compute $\{\boldsymbol{\sigma}_i^*\}$ and $\{\boldsymbol{\eta}^\alpha\}$ from (16),
(3) update the value of $\{\boldsymbol{\sigma}_e^\alpha\}$ as

$$(17) \qquad \boldsymbol{\sigma}_e^\alpha = \boldsymbol{\sigma}_{ie^\alpha}^* + \mathbf{B}_{ie^\alpha}^\alpha \boldsymbol{\eta}^\alpha,$$

(4) compute new state mappings $\{ie^\alpha\}$ with:

$$(18) \qquad \{ie^\alpha\} = \arg\min_{ie^\alpha} \sum_\alpha d(\{(\boldsymbol{\epsilon}_e^\alpha, \boldsymbol{\sigma}_e^\alpha)\}, \{(\boldsymbol{\epsilon}_{ie^\alpha}^*, \boldsymbol{\sigma}_{ie^\alpha}^*)\}),$$

(5) update $\{\boldsymbol{\epsilon}_i^*\}$ from

$$(19) \qquad \sum_\alpha \sum_{e:ie^\alpha(e)=i} w_e^\alpha \mathbb{C}_e(\boldsymbol{\epsilon}_e^\alpha - \boldsymbol{\epsilon}_{ie^\alpha}^*) = 0,$$

(6) iterate steps 2–5 until convergence of $\{ie^\alpha\}$.

Note that the most numerically expensive part are steps 2 and 4, which involve the solution of a large linear system and a database search, respectively.

5. APPLICATION EXAMPLES

In the following, we present a complete cycle of material data identification from strain fields obtained from a set of mechanical tests on a representative sample, followed by data-driven simulation of a different mechanical problem (involving the same material, of course) using the data identified in the first step.

5.1. **Data identification.** For data identification, we use displacement fields and reaction force curves obtained from (displacement controlled) tests on the sample illustrated in Fig. 1. In the present case, these tests were performed numerically, on a linear isotropic elastic material ($E = 217.5$ GPa, $\nu = 0.3$, plane strain), but such displacement fields could come from DIC measures as well. Numerically generated results will be free of experimental noise, but in the following we will focus more on questions of importance sampling than on the effect of noisy data.
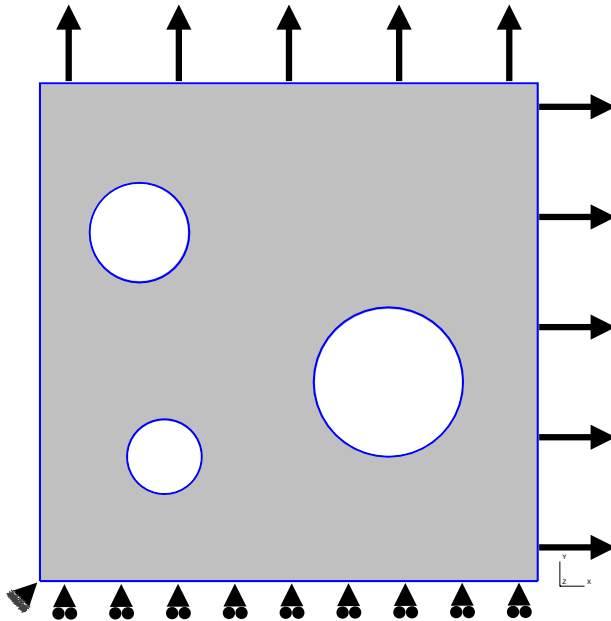

FIGURE 1. Biaxial test sample used for Data Identification

Figure 2 provides resultant loads in both directions, as a function of time. This information will be used, together with snapshots of displacement fields, to construct the database. Two examples of displacement snapshots (at times $t = 1$ and $t = 2$, corresponding to peak loads) are illustrated at Fig. 3.

Three different databases have been constructed using the DDI algorithm. They will be referred to as DDI-DB1, DDI-DB2, and DDI-DB3, and contain $10\,000$, $25\,000$, and $100\,000$ data points
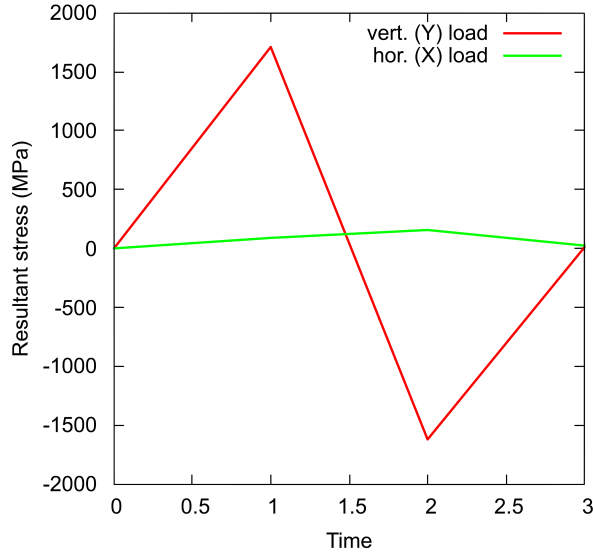
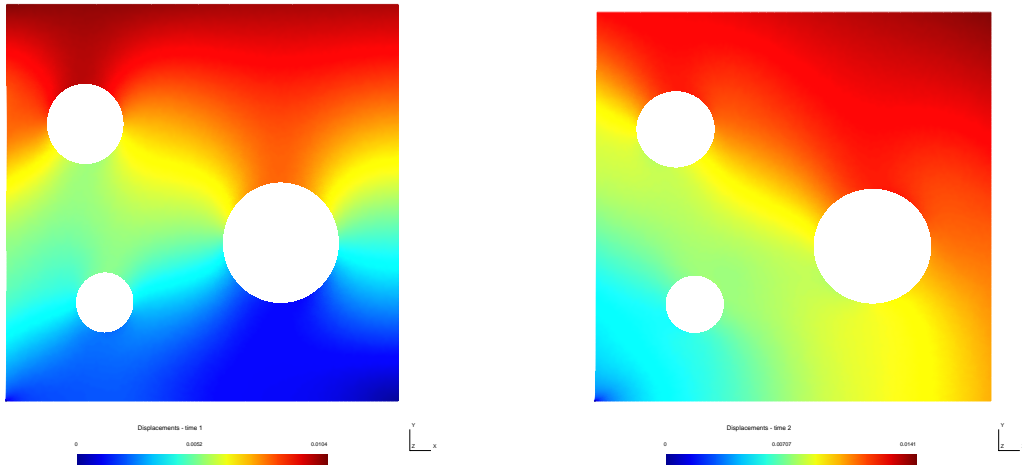FIGURE 2. Biaxial loading of sample used for Data Identification

FIGURE 3. Representative snapshots of displacement fields used for Data Identification

(i.e. strain-stress pairs), respectively. The typical ratio in number of measured mechanical states (M times number of load cases) to number of identified material states (N^*) which ensures good performance of the DDI algorithm is about 200. The number of measured mechanical states can be increased by considering more snapshots and/or higher resolution displacement fields, in order to be able to evaluate gradients at more points in space. Figure 4 shows the set of strain-stress pairs contained in DDI-DB2, with a comparison to the reference linear elastic behavior. We see that the identification process leads to significant variations around this reference behavior, the level of this "noise" being related to the heterogeneity of the input fields.

In the DDI framework, the sampling of strain space is dictated by the mechanics of the chosen test. As a way of illustration, the set of strains corresponding to database DDI-DB2 is shown in Fig. 5. One can observe that the loading is not purely radial, and also that the density of points is not uniform. The bulk of material points are concentrated around moderate values, with a few
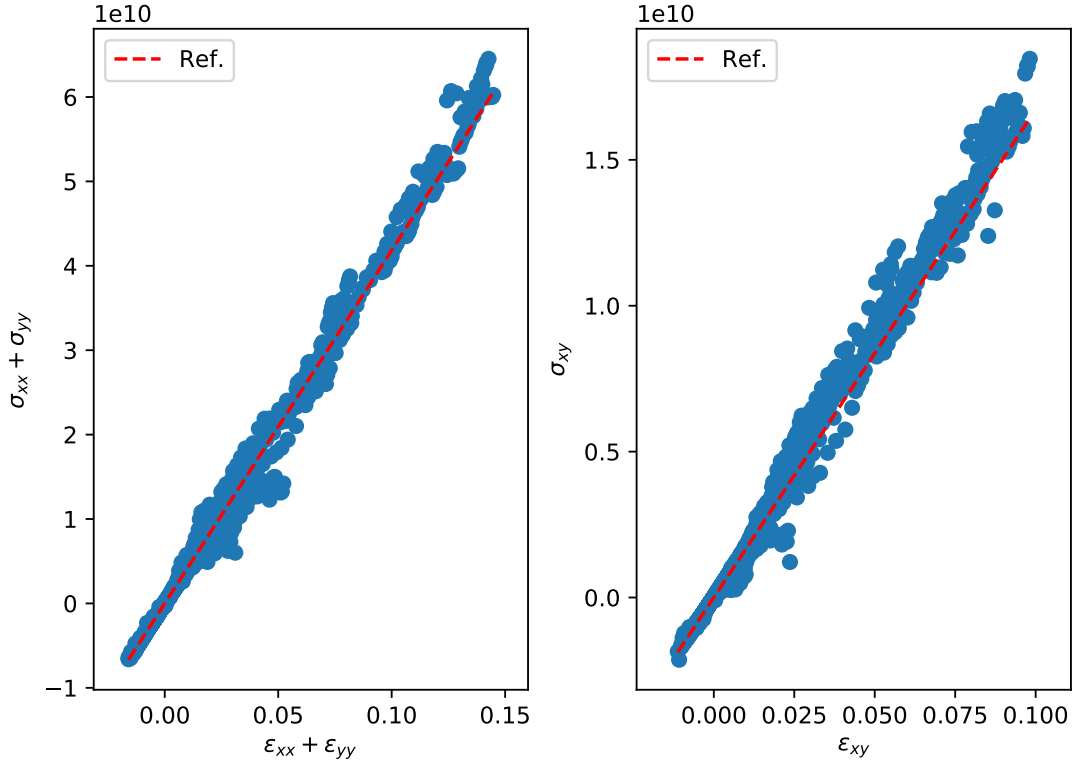
7

FIGURE 4. Set of srain-stress states in material database DDI-DB2

points, corresponding to the strain concentration in the lower left corner, exploring larger strain values.

5.2. **Data-driven simulation.** All following DDCM results were obtained with a uniform metric $\mathbb{C}_e$, corresponding to an isotropic Hookean material with $E = 100$ GPa and $\nu = 0.35$.

5.2.1. *Plate with a hole.* As a first example, we consider a thick plate (plane strain assumption) of width 12.8 R and height 20 R, where R is the radius of a circular hole located in the middle of the plate. By symmetry, only one quarter of the plate is modelled. The plate is subjected to a compression load in the vertical direction (average longitudinal strain of -0.4%). A reference solution is obtained by using a classical FE simulation, using the same elastic properties as those used to generate the input images for the DDI method above (i.e. $E = 217.5$ GPa, $\nu = 0.3$, plane strain), on a mesh of 492 bilinear quadrangular elements (450 nodes). The total elastic strain energy associated to the reference FE solution is 737 J. This number will constitute a useful reference when discussing the distances to and from material databases in the following.

This problem was simulated using the DDCM method described in section 3, using the different material databases identified by DDI (see above), and the same mesh as for the FE reference solution. For comparison, we also used three purely synthetic databases, labelled REG-DB1, REG-DB2, and REG-DB3, obtained by applying the constitutive equations to uniformly sampled strain tensors: we considered samplings of the $\{\epsilon_{xx}, \epsilon_{xy}, \epsilon_{yy}\}$ space of $30 \times 30 \times 30$, $50 \times 50 \times 50$, and $100 \times 100 \times 100$ points, respectively, in the range specified in Table 1. Note that these ranges were chosen in function of the strain distribution computed in the FE solution. On the contrary, the range of strains in the databases generated by DDI are not directly controlled. For reference,
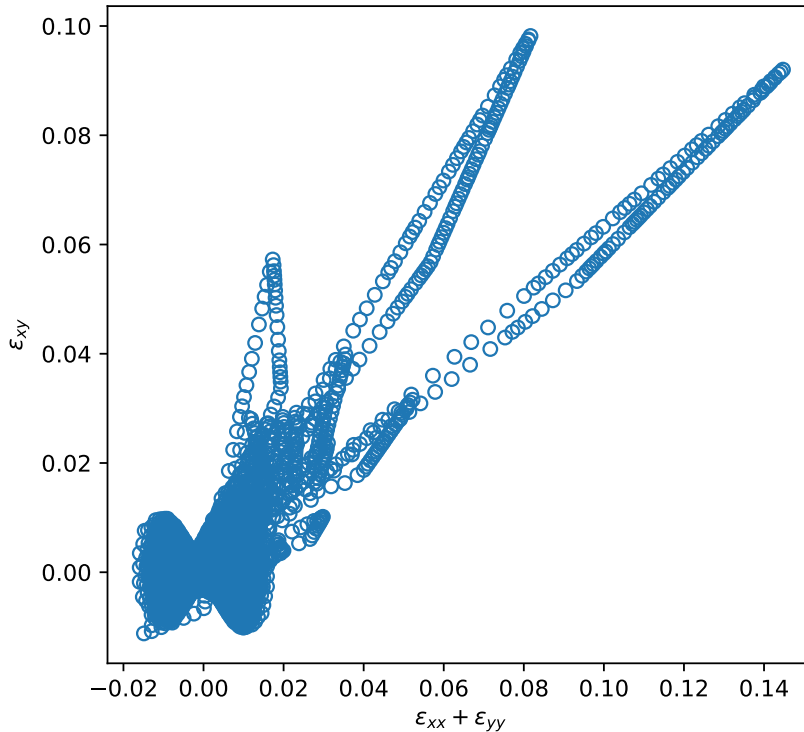
FIGURE 5. Sampling of strain space for material database DDI-DB2

Figures 6 and 7 show how the strain space is spanned by the databases (DDI-DB2 and REG-DB1), FEM and DDCM solutions.

TABLE 1. Range of strains in synthetic material databases

Bounds	ϵ_{xx}	ϵ_{xy}	ϵ_{yy}
min.	-0.002	-0.002	-0.015
max.	0.005	0.005	0.0025

Figure 8 shows a comparison of the stress field obtained by DDCM with respect to our reference solution, obtained by FEM. We observe an overall good agreement, with the DDCM solution correctly picking the stress concentration at the side of the hole. The distribution of absolute and relative errors (in the metric of $\mathbb{C}_e$) is shown in Fig. 9. For the database illustrated here (DDI-DB2), the error on the stress concentration amplitude is of the order of 10%. This error can be reduced by enriching the database and/or tuning the metric tensor $\mathbb{C}_e$. Note that with the values which were (arbitrarily) chosen here, the error on strain fields, with respect to the FE solution, is much lower (of the order of 1%, and less in most of the plate, for the case illustrated in Fig. 8) than the error on stress fields.

For a more quantitative comparison, we can look at distances between the various solution sets, in the metric of $\mathbb{C}_e$. Figure 9 shows the distribution of the local distance between DDCM and FEM solutions, while Table 2 lists global measures (i.e. integrated on the whole domain). From the first row in Table 2, we see that DDI-generated databases are providing a better representation of, i.e. are closer to, the FE reference solution at a given database size. Note that the mechanical
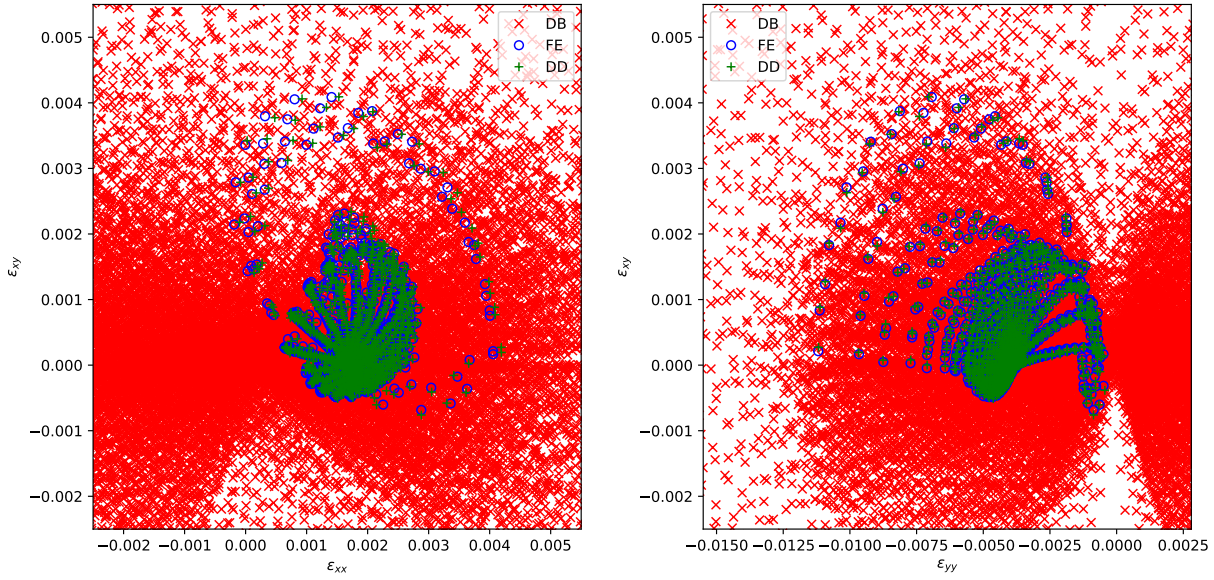
FIGURE 6. Sampling of strain space for material database DDI-DB2, associated DDCM, and FEM solutions
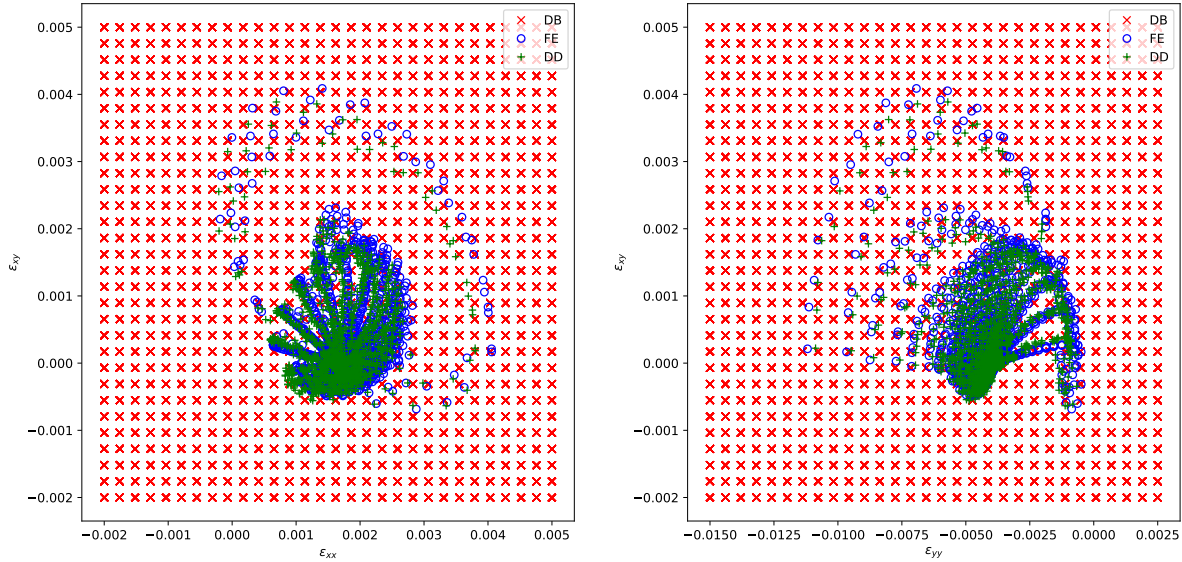

FIGURE 7. Sampling of strain space for material database REG-DB1, associated DDCM, and FEM solutions

boundary-value problem considered here is different from those used to identify the databases. From the third row in Table 2, we also see that a DDI-generated database of 25,000 points (DDI-DB2) allows to obtain a DDCM solution of similar quality as a 1,000,000 point database using uniform sampling (REG-DB3), providing a strong argument in favour of using the former. For reference, recall that the total elastic strain energy computed in the reference FE solution is 737 J. Figure 10 shows the convergence of the minimal distance attained by the DDCM algorithm, as

Stresses (DD)

2.08e+08   1.42e+09   2.63e+09

Stresses (FE)

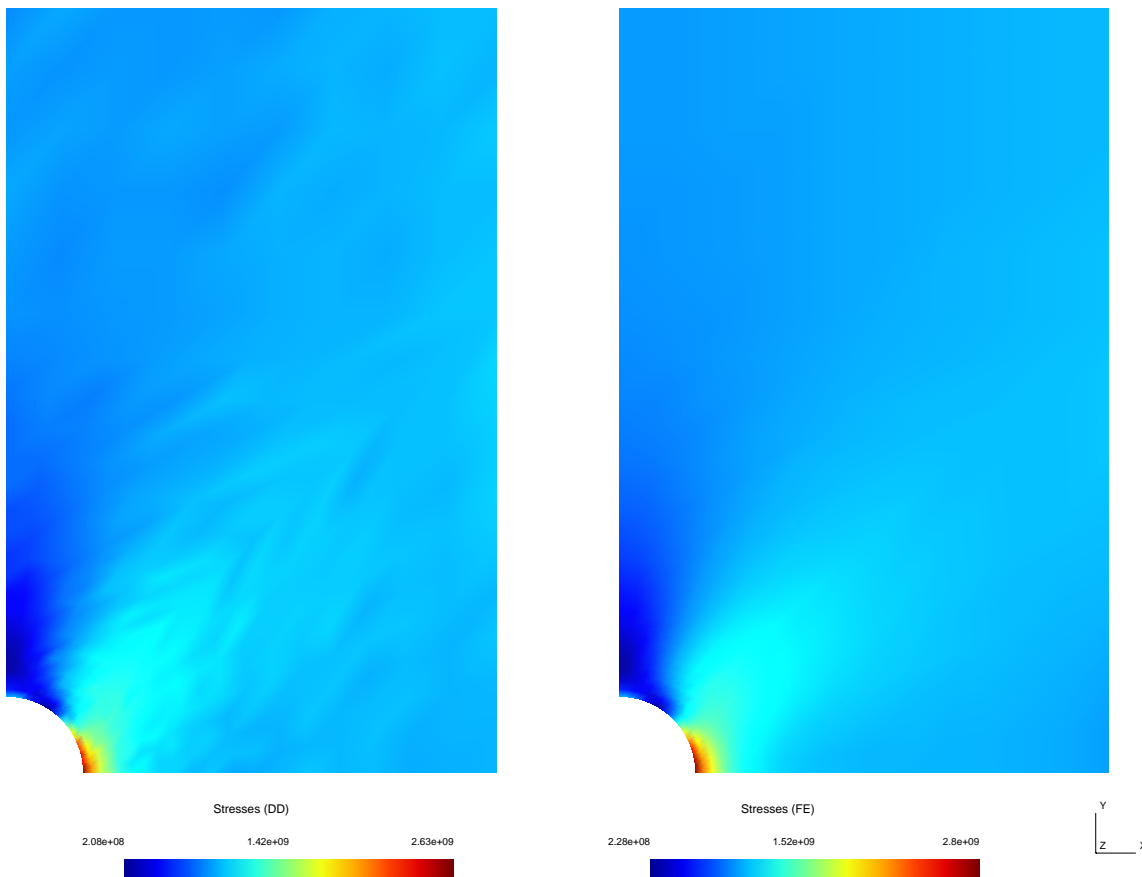2.28e+08   1.52e+09   2.8e+09

FIGURE 8. Comparison of stress fields obtained by DDCM and FEM on the case of a plate with a hole (DDCM results obtained with DDI-DB2)

a function of the database size, and of the error with respect to the FE reference solution. Both curves clearly show that DDI databases significantly outperform REG ones, not only at equivalent size but overall. Both sets of databases appear to have similar asymptotic convergence rates.

TABLE 2. Distance between solutions and material database (distances in energy [J]) – plate example

Distance	DDI-DB1	DDI-DB2	DDI-DB3	REG-DB1	REG-DB2	REG-DB3
FEM-DB	1.30	0.70	0.28	5.86	1.87	0.45
DDCM-DB	4.01	2.82	1.15	22.91	8.38	2.05
DDCM-FEM	16.88	11.70	4.93	96.20	35.47	8.66

5.2.2. *L-beam.* As a second example, we consider a L-shaped beam (plane strain assumption) with the following geometrical characteristics: total width $W = 0.6\,H$, vertical branch of width $w = 0.2\,H$, horizontal branch of height $h = 0.3\,H$, total height $H = 1$m. The fillet has a curvature radius $r = 0.02\,H$. A hole of radius $R = 0.075H$ is positioned with its center at $(0.2\,H, 0.5\,H)$. The base is fixed so as to prevent rigid body modes, and a horizontal displacement of $0.002\,H$ is imposed at
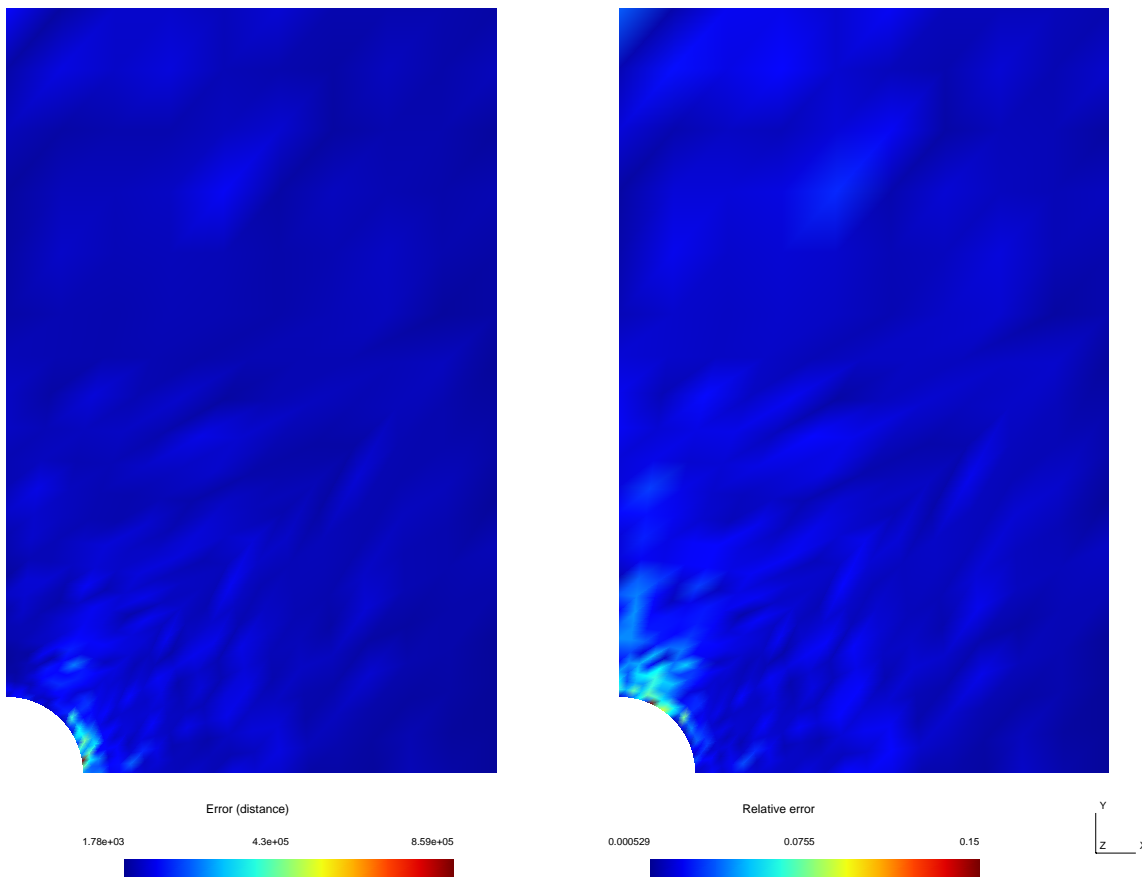
FIGURE 9. Distance between DDCM and FEM solutions on the case of a plate with a hole (DDCM results obtained with DDI-DB2). The relative error is with respect to the local energy density $\boldsymbol{\epsilon} \cdot \boldsymbol{\sigma}$.

the top. A reference solution is obtained by using a classical FE simulation, using the same elastic properties as those used to generate the input images for the DDI method above (i.e. $E = 217.5$ GPa, $\nu = 0.3$, plane strain), on a mesh of 847 quadratic triangular elements (1823 nodes). The total elastic strain energy associated to the reference FE solution is 1510 J. This number will constitute a useful reference when discussing the distances to and from material databases in the following.

This problem was simulated using the DDCM method described in section 3, using the different material databases identified by DDI (see above), and the same mesh as for the FE reference solution. As in the first example, we also used the three purely synthetic databases, labelled REG-DB1, REG-DB2, and REG-DB3, for comparison with DDI. The geometry and loading of this test case differ more significantly than in the previous one from those of the sample used for identification, which should provide a more straining test for the DDI databases.

Figure 11 shows a comparison of the stress field obtained by DDCM with respect to our reference solution, obtained by FEM. We observe an overall good agreement, with the DDCM solution correctly picking the stress concentration at the fillet. Figure 12 shows the distribution of absolute and relative errors (in the metric of $\mathbb{C}_e$). The maximal absolute error is located at the fillet, but the maximal relative error actually occurs at the top corners, where the loading (imposed displacement) is applied.
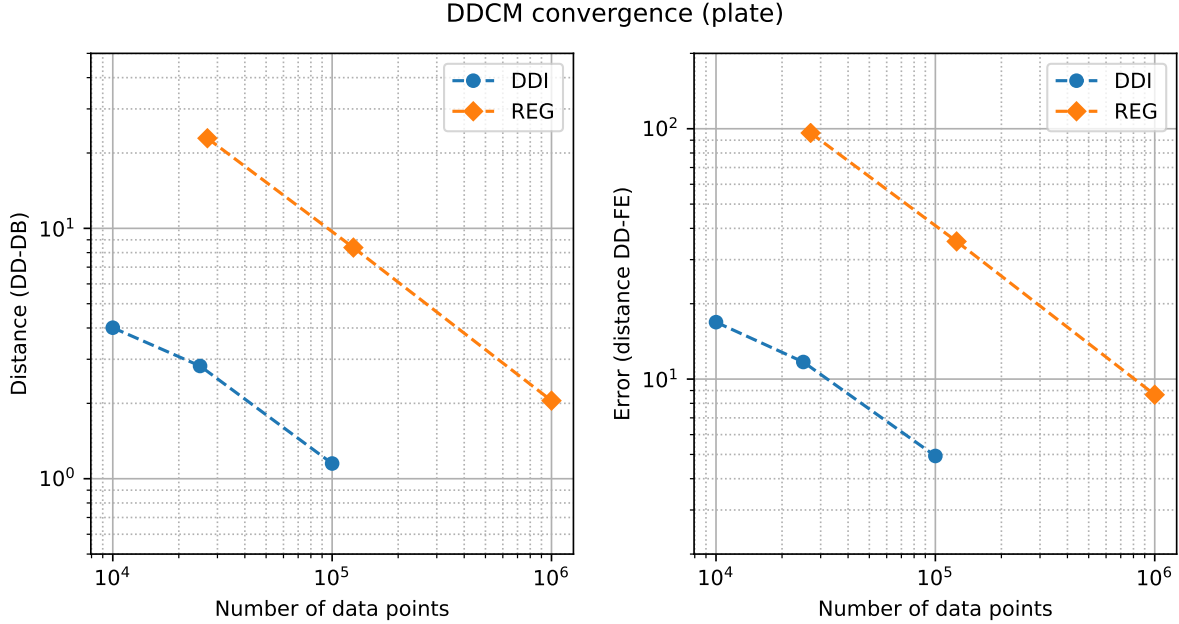
12

DDCM convergence (plate)

FIGURE 10. Convergence of the DDCM method (plate example). Left plot shows convergence of the distance between the mechanical states and the database. Right plot shows the convergence of the error with respect to the reference FE solution.

For a more quantitative comparison, we can look again at distances between the various solution sets, in the metric of $\mathbb{C}_e$. Figure 12 shows the distribution of the local distance between DDCM and FEM solutions, while Table 3 lists global measures (i.e. integrated on the whole domain). From the first row in Table 3, we see that, like in previous example, DDI-generated databases are providing a better representation of, i.e. are closer to, the FE reference solution at a given database size. From the third row in Table 3, we also see that a DDI-generated database of 25,000 points (DDI-DB2) allows to obtain a DDCM solution of better quality than a 1,000,000 point database using uniform sampling (REG-DB3), providing once more a strong argument in favour of using the former. Figure 13 shows the convergence of the minimal distance attained by the DDCM algorithm, as a function of the database size, and of the error with respect to the FE reference solution. Both curves clearly show that DDI databases significantly outperform REG ones, even with a limited number of data points.

TABLE 3. Distance between solutions and material database (distances in energy [J]) – beam example

Distance	DDI-DB1	DDI-DB2	DDI-DB3	REG-DB1	REG-DB2	REG-DB3
FEM-DB	329.68	174.18	91.18	3434.35	1014.91	303.48
DDCM-DB	300.72	195.52	115.15	3976.76	998.25	693.04
DDCM-FEM	1195.73	617.84	433.73	11616.54	3779.12	2522.43

5.2.3. *Material symmetries.* If it is *a priori* known that the material behavior is isotropic, associated symmetries can be exploited and the intrinsic dimension of the material database can be reduced to that of invariants (3 for strains and 3 for stresses in 3D) [7]. As a consequence, the distance used
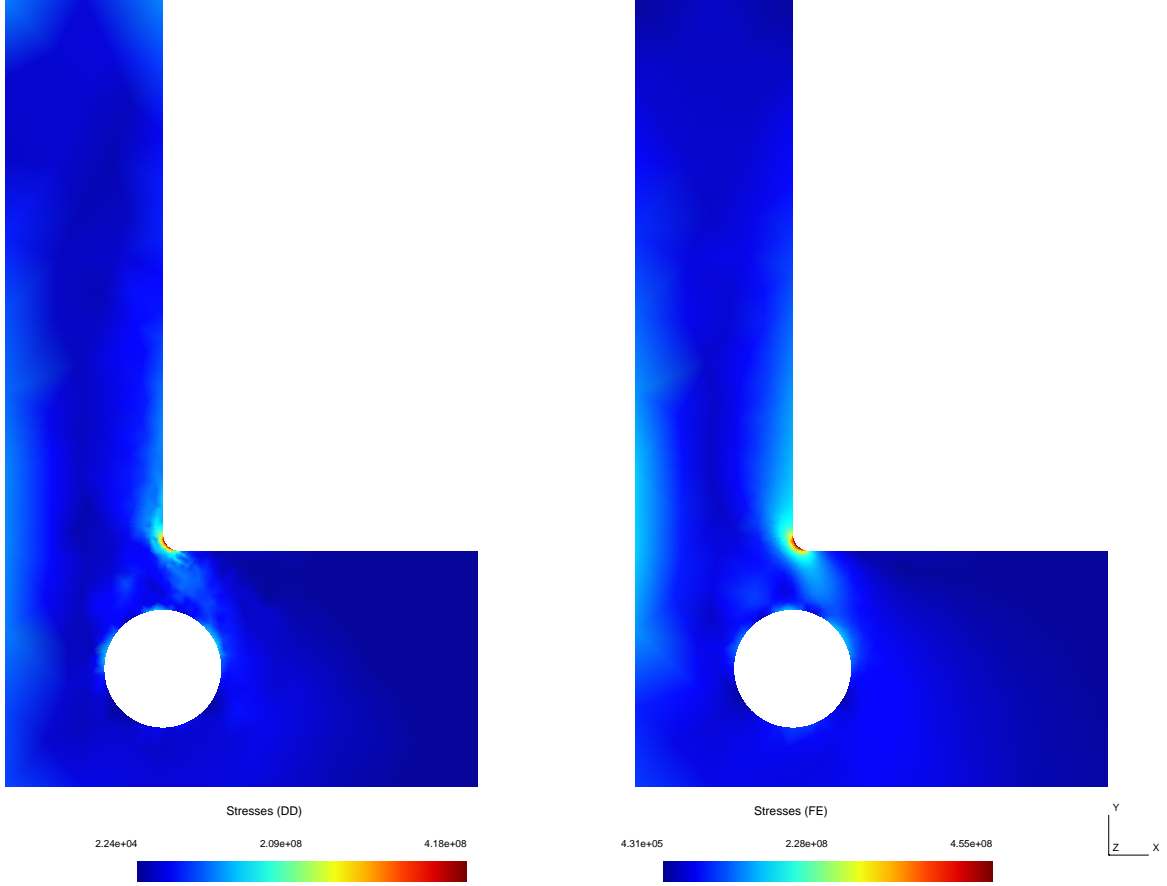
FIGURE 11. Comparison of stress fields obtained by DDCM and FEM on the case of a L-beam (DDCM results obtained with DDI-DB2)

in algorithm (13) can be modified to

$$(20) \qquad \hat{d}(\boldsymbol{z}, \boldsymbol{y}) = \min_{\{\boldsymbol{R}_e\}} \Big(\sum_{e=1}^{M} w_e |\boldsymbol{z}_e - \hat{\boldsymbol{y}}_e|_e^2 \Big)^{\frac{1}{2}}$$

where $\hat{\boldsymbol{y}}_e = \{\hat{\boldsymbol{\epsilon}}_e^*, \hat{\boldsymbol{\sigma}}_e^*\} = \{\boldsymbol{R}_e^T \boldsymbol{\epsilon}_e^* \boldsymbol{R}_e, \boldsymbol{R}_e^T \boldsymbol{\sigma}_e^* \boldsymbol{R}_e\}$, with $\boldsymbol{R}_e \in SO(3)$ a rotation tensor associated to material point e. When $\boldsymbol{\epsilon}_e$ and $\boldsymbol{\sigma}_e$ share the same eigenvectors, the rotation tensor can easily be derived analytically, but this is generally not the case, since they are updated independently from (12a) and (12b). The minimization can then be performed numerically (this operation is purely local for each material point), using a parametrization of the rotation tensor. In the 2D case considered in the above examples, such parametrization reduces to a single angle, and the minimization problem can be solved analytically.

Results obtained using this isotropic distance are shown in Table 4, for the case of the plate and DDI databases, and Table 5, for the case of the L-beam and DDI databases. When compared to values obtained with the standard distance (Tab. 2 and 3), we see a reduction of the minimum distance, but the error with respect to the reference FE solution does not systematically decrease (e.g. in the case of the plate with DDI-DB1). Moreover, this improvement in the minimum distance typically comes at the cost of more iterations in the DDCM algorithm. The computational cost *vs.* precision trade-off is thus not always beneficial. Here, this approach appears more relevant for
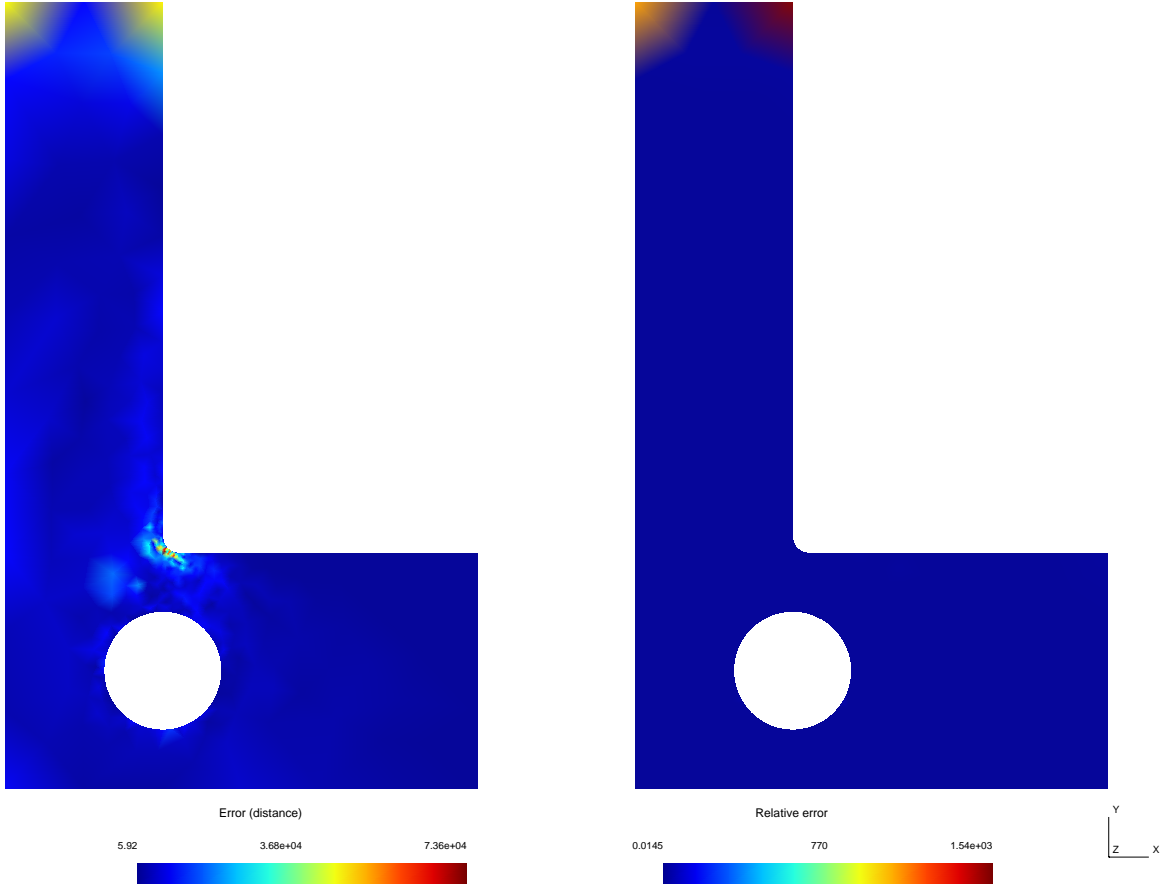
FIGURE 12. Distance between DDCM and FEM solutions on the case of a L-beam (DDCM results obtained with DDI-DB2). The relative error is with respect to the local energy density $\boldsymbol{\epsilon} \cdot \boldsymbol{\sigma}$.

the beam case, where the loading differs more significantly from that of the identification sample than in the case of the plate. Accounting for full $SO(3)$ orbits of material points in the database then enrich the latter, leading to improved results of the DDCM algorithm. Yet, this requires the *a priori* assumption of isotropy of the material.

TABLE 4. Minimal isotropic distance reached by DDCM and standard distance between DDCM and FEM solutions (distances in energy [J]) – plate example

Distance	DDI-DB1	DDI-DB2	DDI-DB3
DDCM-DB	2.53	1.45	0.44
DDCM-FEM	33.13	10.43	3.10

6. DISCUSSION AND PERSPECTIVES

In this paper, we have shown how the DDI algorithm and DDCM solver naturally combine to constitute a computational design toolbox eliminating epistemic uncertainty linked to traditional constitutive models in mechanics. Not only does DDI provide a way to infer stresses directly from
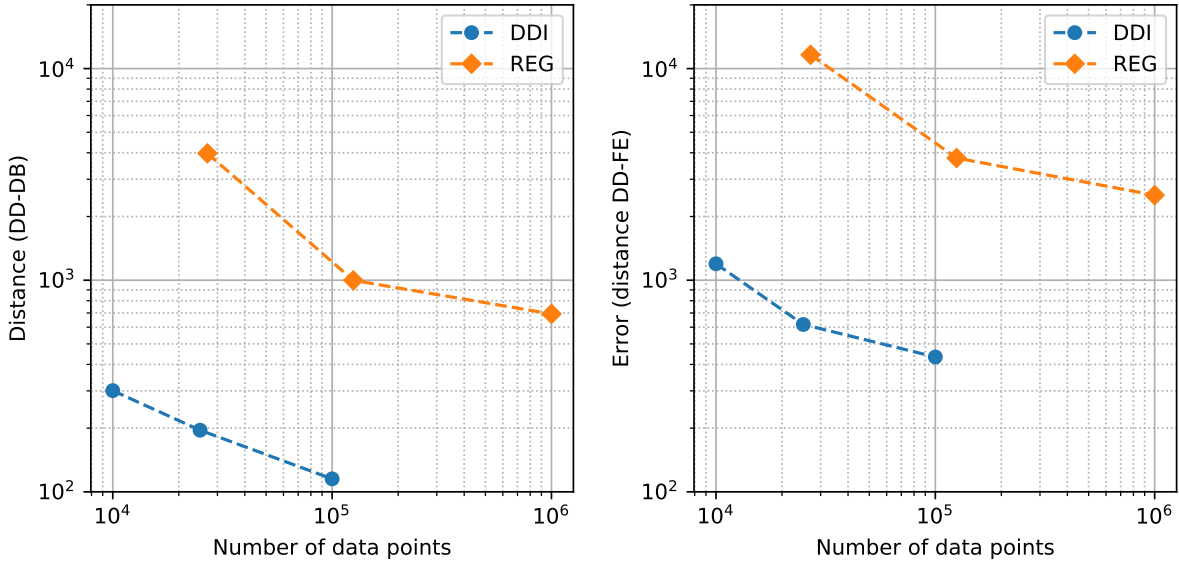
15

FIGURE 13. Convergence of the DDCM method (beam example). Left plot shows convergence of the distance between the mechanical states and the database. Right plot shows the convergence of the error with respect to the reference FE solution.

TABLE 5. Minimal isotropic distance reached by DDCM and standard distance between DDCM and FEM solutions (distances in energy [J]) – beam example

Distance	DDI-DB1	DDI-DB2	DDI-DB3
DDCM-DB	89.26	45.26	18.10
DDCM-FEM	420.71	200.24	68.91

experimental field measures, where classical model updating techniques must postulate a specific constitutive model, it also generates databases of material states respecting some importance sampling. This latter feature proved to yield a greatly improved efficiency for DDCM in the examples considered.

DDCM also proved its robustness in the above examples, by its capacity to perform simulations on various geometries and loadings using a given database. Qualitatively, the method requires only limited data to predict the major features of strain and stress fields (e.g. location of maximal values). Quantitatively, the precision obtained depends directly on the quality of the database (number and sampling of data points), but one should note that the DDCM comes with its own error estimate. Indeed, the minimal distance which can be attained between mechanical states (i.e. those verifying compatibility and balance constraints) and material states (i.e. those in the database) provides a direct indicator of the adequacy of the given database for the considered mechanical problem. If this distance cannot be sufficiently decreased by the algorithm, it can be considered as an indication that more points are needed in the database.

Linear systems occurring in DDCM algorithm correspond to standard elasticity problems with pre-strain/pre-stress loading. This opens a perspective to implementing this step through third-party FE software computing engines. These softwares would need to allow for individual specification of pre-strain / pre-stress at each integration point. If this feature seems available in

the open-source software Code_Aster [2] for example, it may require going through user-defined elements in most commercial softwares.

In practice, most of the computational cost resides in database searches. For linear system solutions, a single factorization is typically necessary, even in the case of a multi-step simulation with non-linear material behavior, unless one wants to adapt the metric at each step, which is of course possible. Thus, although the above examples were treated using a basic search algorithm, computational efficiency could significantly be improved by using data structures and search algorithms adapted to manipulating and navigating very large data sets (e.g. [12]).

It is also important to note that the whole DDI+DDCM framework remains valid, without any modifications, in the presence of material non-linearities. Indeed, those will be entirely contained within the database, and the linearity of the systems of equations solved in DDI and DDCM is unaffected by material non-linearities. Those will thus have no effect on computational performance either. For material behaviors presenting irreversibilities and/or history dependence, first results have been established, both in DDI [11] and DDCM [3].

Finally, the Data-Driven paradigm, and in particular the DDI+DDCM framework, opens promising perspectives in terms of public repositories of material databases. Such web-based platforms would allow and encourage capitatlization and sharing of large amounts of material data, experimental and/or synthetic (e.g. multiscale computations), which are typically thrown away after constitutive model identification in the current dominant model-based paradigm. Of course, this would probably require a preliminary stage of standardization of database formats (including metadata) and curation methods by the community.

References

[1] Conti, S., Müller, S., Ortiz, M.: Data-driven problems in elasticity. Archive for Rational Mechanics and Analysis **229**, 79–123 (2018)

[2] EDF: Code_aster. `https://www.code-aster.org/spip.php?rubrique2`

[3] Eggersmann, R., Kirchdoerfer, T., Reese, S., Stainier, L., Ortiz, M.: Model-free data-driven inelasticity. Computer Methods in Applied Mechanics and Engineering **to appear** (2019)

[4] Feyel, F.: Multiscale FE 2 elastoviscoplastic analysis of composite structures. Computational Materials Science **16**(1-4), 344–354 (1999)

[5] Feyel, F.: A multilevel finite element method (FE2) to describe the response of highly non-linear structures using generalized continua. Computer Methods in Applied Mechanics and Engineering **192**(28-30), 3233–3244 (2003)

[6] Kanno, Y.: Mixed-integer programming formulation of a data-driven solver in computational elasticity. Optimization Letters (2019). DOI https://doi.org/10.1007/s11590-019-01409-w

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

[8] Kirchdoerfer, T., Ortiz, M.: Data driven computing with noisy material data sets. Computer Methods in Applied Mechanics and Engineering **326**, 622–641 (2017)

[9] Kirchdoerfer, T., Ortiz, M.: Data driven computing in dynamics. International Journal for Numerical Methods in Engieering **113**(11), 1697–1710 (2018)

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

[11] Leygue, A., Seghir, R., Réthoré, J., Coret, M., Verron, E., Stainier, L.: Non-parametric material state field extraction from full field measurements. Computational Mechanics **this issue** (2019)

[12] Muja, M., Lowe, D.G.: Scalable nearest neighbor algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence **36** (2014)

[13] Nguyen, L.T.K., Keip, M.A.: A data-driven approach to nonlinear elasticity. Computers and Structures **194**, 97115 (2018)

[14] Réthoré, J.: A fully integrated noise robust strategy for the identification of constitutive laws from digital images. International Journal for Numerical Methods in Engineering **84**, 631–660 (2010)

Research Institute in Civil and Mechanical Engineering (GeM - UMR 6183 CNRS/ECN/UN), Ecole Centrale Nantes, F-44321 Nantes, France
  E-mail address: `Laurent.Stainier@ec-nantes.fr`

Research Institute in Civil and Mechanical Engineering (GeM - UMR 6183 CNRS/ECN/UN), Ecole Centrale Nantes, F-44321 Nantes, France
  E-mail address: `Adrien.Leygue@ec-nantes.fr`

Division of Engineering and Applied Sciences, California Institute of Technology, Pasadena, CA 91125, USA
  E-mail address: `ortiz@aero.caltech.edu`