Big Data in Experimental Mechanics and Model Order Reduction: Today's Challenges and Tomorrow's Opportunities — Source link

Jan Neggers, Olivier Allix, Olivier Allix, François Hild ...+1 more authors

Institutions: Université Paris-Saclay, Institut Universitaire de France

Topics: Big data and Data flow diagram

Related papers:

- Estimation of Elastoplastic Parameters via Weighted FEMU and Integrated-DIC

- Comparison of Local and Global Approaches to Digital Image Correlation

- Digital Volume Correlation: Review of Progress and Challenges

- Digital volume correlation : Three-dimensional strain mapping using X-ray tomography

- Toward 4D mechanical correlation

Big Data in Experimental Mechanics and Model Order Reduction: Today's Challenges and Tomorrow's Opportunities

Jan Neggers, Olivier Allix, François Hild, Stéphane Roux

Big Data in Experimental Mechanics and Model Order Reduction: Today's Challenges and Tomorrow's Opportunities

Jan Neggers · Olivier Allix · François Hild · Stéphane Roux

Abstract Since the turn of the century experimental solid mechanics has undergone major changes with the generalized use of images. The number of acquired data has literally exploded and one of today's challenges is related to the saturation of mining procedures through such big data sets. With respect to Digital Image/Volume Correlation one of tomorrow's pathways is to better control and master this data flow with procedures that are optimized for extracting the sought information with minimum uncertainties and maximum robustness. In this paper emphasis is put on various hierarchical identification procedures. Based on such structures *a posteriori* model/data reductions are performed in order to ease and make the exploitation of the experimental information far more efficient. Some possibilities related to other Model Order Reduction techniques like the Proper Generalized Decomposition are discussed and new opportunities are sketched.

Keywords Data and model reduction · Full-field measurements · Identification · Images

1 Introduction

This paper is focused on experimental (solid) mechanics, which is related to the process of testing materials (or structures) by applying, say, a time sequence of mechanical loads through one or several actuators. The goal of the test is to extract an intrinsic mechanical characterization, such as one constitutive law (or several ones for multiphase materials). It may also be reduced to a simple subset of constitutive parameters such as modulus or strength

J. Neggers, O. Allix, F. Hild, S. Roux
LMT, ENS Paris-Saclay / CNRS / Université Paris-Saclay
E-mail: {jan.neggers;olivier.allix;francois.hild;stephane.roux}@ens-paris-saclay.fr

O. Allix
Institut Universitaire de France

including their statistical variability, or deal with slow time changes such as encountered, say, in cyclic fatigue. Up until recently, say 20 years ago, mechanical tests made use of local measurements (*e.g.,* provided by displacement transducer, strain gage, extensometer). This situation has drastically changed and today a huge and ever increasing number of data is routinely collected during an experiment. This is mainly due to the use of various imaging systems that give access to Mpixel-images and Gvoxel-volumes. For instance, nowadays experiments on flat samples use, typically, 1000 16-bit images of 1 to 100-Mpixel definition for the measurement of 2D displacement fields (*i.e.,* 2-200 Gbyte data sets). For more complex shapes, more than one imaging device is considered. For $N_p > 1$ cameras, stereocorrelation [1] or multiview formulations [2] are utilized to measure 3D surface displacements (*i.e.,* $2N_p$-$200N_p$ Gbyte data). With the emergence of 3D imaging systems (*e.g.,* X-ray tomography [3,4] and laminography [5,6], magnetic resonance imaging [7], optical coherence tomography [8]) an additional increase of data is observed. In tomography or laminography *each* reconstructed volume is obtained from a set of 6000×12-Mpixel radiographs (*i.e.,* 144 Gbytes per considered step for 16-bit digitizations).

Most of the full-field measurement techniques used in experimental mechanics are based on those various imaging devices [9]. They are not only available in very specialized facilities (*e.g.,* tomography on synchrotron beamlines) or institutions (*e.g.,* ultra high-speed cameras at defense labs) but also in academia and industry. This democratization of imaging tools has led to easier access to such "big data" at "reasonable" cost. To illustrate this trend, Figure 1 shows the cumulated number of image data (*i.e.,* pixels and voxels) used in peer-reviewed publications of the lab of the authors. The dashed line corresponds to an exponential interpolation of the image data and the solid line to Moore's law. This plot shows that in a matter of 15 years, the cumulated yearly image data sets have grown from 5×10^7 to 10^{12}, *i.e.,* faster than standard Moore's law (*i.e.,* the amount of data is doubled over a period of 18 months) that would have predicted (*i.e.,* 5×10^{10}). In the reported results, there are three regimes for which standard Moore's law applies. However, there are two jumps that correspond to the first use of tomographic data in 2007, and generalized analyses with tomographic acquisitions starting in 2010. These results show that with the development of 3D imaging the amount of available data will drastically increase. Once all these images have been acquired there is also a growing effort required to analyze the data sets to extract the sought information. More than often the duration of the experiment *per se* is very short in comparison with its design and its post-analysis.
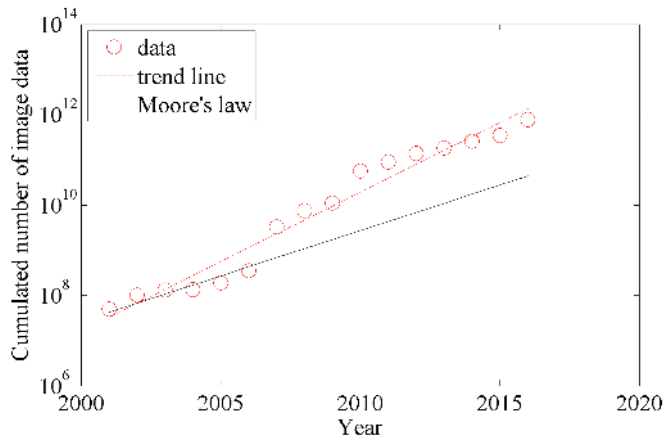
Fig. 1 Cumulated number of image data used in publications of LMT from 2001 to 2016. The dashed line corresponds to Moore's law and the solid line to exponential interpolation

The benefit of having access to such big data sets makes experimental mechanics a data provider that can reach the same level as that available in computational mechanics. This is very important for simulation-based engineering sciences, which call for validation procedures [10]. Meeting this challenge requires experiments and simulations to be compared as thoroughly as possible to probe the numerical models. It calls for full-field comparisons instead of point-wise differences. This observation opens the way for full (in)validations of numerical models of materials and structures [9].

These approaches are based on computed *spatiotemporal* fields that will eventually be compared with their experimental counterparts once the experiment has been prepared and performed or even during the experiment itself [11]. This remark shows that online control of experiments still requires developments on the measurement and computational side to achieve real-time hybrid interactions. However, the feasibility of tests controlled via displacement *fields* has been shown in the recent past [11–14]. It enters the framework of dynamic data-driven application systems [15] whose emergence is one critical aspects of simulation-based engineering sciences [10].

The aim of the present paper is to discuss the opportunities and challenges induced by the fact that big data sets are now available in experimental mechanics, and how they may change the dialog between experiments and computations. Section 2 introduces the principle to measure displacement fields via Digital Image Correlation, which has become one of the most popular measurement technique in solid mechanics [16, 17]. The next step is to extract material parameters from the measured fields. This mechanics-based data mining procedure consists of reducing the available data into meaningful parameters to be used in computations. The process of data reduction may be coupled with model reduction techniques (Section 4). Such approaches are illustrated on a typical example in Section 5. Last, some advanced procedures are discussed in

Section 6 to pave the way for the future of experimental mechanics in relation with the acquisition of big data sets.

The emergence of simulation-based engineering sciences calls, among many challenges [10], for uncertainty quantifications to achieve reliable predictions and probe (*i.e.,* validate or invalidate) material and numerical models. One of the sources of uncertainties is related to data acquisition (*e.g.,* standard imaging devices such as cameras are prone to acquisition noise). Even though the modeling of such uncertainties is possible and more importantly desirable, it will not be discussed herein at the acquisition step. Conversely, its propagation when dealing with data reduction and/or data mining will be illustrated throughout the paper. It will be shown that outputs of data processing steps are not only the quantities of interest but also their corresponding covariance matrix. The latter requires consistent metric to be introduced.

2 Full-field measurements in experimental solid mechanics

The most salient characteristics of imaging systems are (i) a *massive* assembly of (ii) rather *poor* sensors (pixelwise detectors). "Poor" refers to the fact that each elementary detector (*i.e.,* pixel) delivers an elementary measurement of light intensity, *i.e.,* gray levels encoded over few bits, typically 8 to 16, which suffers from noise that (more than the gray level depth itself) limits the reliable part of the information. "Massive" has already been discussed in the previous section. The idea here is that the huge number of pixels will allow for high spatial resolutions. What is said for optical camera also holds for electron microscopy, although the *scanning* character of acquisition introduces specific signatures. Noise in particular displays correlations that are inherited from the scanning procedure [18]. Similarly, atomic force microscopy (AFM) is also a scanning measurement device, with even stronger noise as resulting from the mechanical probe itself and its piezo-controller [19].

The specific challenge of utilizing images, in particular for experimental mechanics but not only, is to address the above two characteristics, the *massive amount of data* and their *noise*. The very reason why such a challenge can be overcome is that images will be used to measure quantities (*e.g.,* kinematic fields or material parameters) that require *less* or *much less* data. The mathematical treatment is thus to be dictated by the noise properties, using a suited probabilistic treatment (*e.g.,* maximizing the likelihood) and redundancy is to be exploited to either validate or invalidate the desired model, and in the latter case, provide clues to resolve discrepancies.

2.1 DIC principles

Digital Image Correlation (DIC) aims at measuring the motion that allows for the registration of two or more images. Images are here represented by gray levels f at each pixel or voxel at position $\boldsymbol{x}$ and time t, $f(\boldsymbol{x}, t)$. In order

to allow for sub-pixel/voxel measurement interpolation with cubic splines (or higher order) are assumed to hold. The most common assumption expresses that the gray level is a signature of the observed volume or surface element that is simply (passively) transported with the medium without any further modification. Hence it is expected that in the absence of noise the so-called residual

$$\tau(\boldsymbol{x}, t) = f(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, t), t) - f(\boldsymbol{x}, t_0) \tag{1}$$

should be identically 0. The condition $\tau = 0$ is the consequence of gray level (or brightness) conservation. It is to be emphasized that more elaborate descriptions of how images are modeled to evolve with time may be proposed, in particular, when images where the gray level has a different meaning. If it stands for a physical elevation as for topographic images encountered with AFM, then τ may represent the out of plane displacement, say $u_z(\boldsymbol{x})$ [20, 21]. If f is a digital level of an infrared camera, its value is expected to be equal to the product of the emissivity by the local sample temperature, and hence after a measurement of the emissivity field, τ can be expressed as a temperature field [22].

It will be considered in the following that a mere brightness conservation holds. However, in addition to the assumption of the underlying physical model for image formation, as earlier mentioned, images are prone to noise, and hence at least $\tau(\boldsymbol{x}, t)$ is equal to the difference in the noise $\eta(\boldsymbol{x}, t)$ affecting $f(\boldsymbol{x}, t)$ and $f(\boldsymbol{x}, t_0)$, or $\tau(\boldsymbol{x}, t) = \eta(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, t), t) - \eta(\boldsymbol{x}, t_0)$. For a Gaussian white noise characterized by a variance σ^2, $\langle \eta(\boldsymbol{x}, t) \eta(\boldsymbol{x}', t') \rangle = \sigma^2 \delta(\boldsymbol{x} - \boldsymbol{x}') \delta(t - t')$, where $\langle \bullet \rangle$ designates the mathematical expectation of $\bullet$ for different realizations of the noise, and δ the Dirac distribution. The noise difference can be gathered into an effective noise affecting only images at times later than t_0 and whose variance is doubled $2\sigma^2$.

For such white and Gaussian noise, the probability of observing $\tau = \eta$ at a pixel $\boldsymbol{x}$ and time t, is written as $p(\boldsymbol{x}, t) \propto \exp(-\tau^2/4\sigma^2)$ when not specifying an unimportant multiplicative constant. Hence, for the entire picture, the absence of correlation implies that this probability read

$$\begin{aligned} P &\propto \prod_{\boldsymbol{x}, t} \exp(-\tau(\boldsymbol{x}, t)^2/4\sigma^2) \\ &\propto \exp\left(-\frac{1}{4\sigma^2} \sum_{\boldsymbol{x}, t} \tau(\boldsymbol{x}, t)^2\right) \end{aligned} \tag{2}$$

Hence the registration consists in finding the displacement field $\boldsymbol{u}(\boldsymbol{x}, t)$ such that it maximizes the likelihood of τ, or equivalently such that it minimizes the argument of the exponential, or the L2-norm of τ. The following cost function is thus introduced

$$\chi^2[\boldsymbol{u}(\boldsymbol{x}, t)] = \frac{1}{2\sigma^2 N_x N_t} \sum_{\boldsymbol{x}, t} \tau(\boldsymbol{x}, t)^2 \tag{3}$$

where a normalization with respect to the number of pixels N_x in the region of interest and the number N_t of considered images (in addition to the reference one), makes this quantity intensive, and its comparison with 1 is to be interpreted as whether or not the residual reduces to noise ($\chi \approx 1$) or not ($\chi \gg 1$).

This minimization problem is *strongly nonlinear* because of the fact that the unknown appears in the argument of a function $f(\bullet, t)$ that characterizes the studied sample and whose regularity is out of control. However, if an initial displacement field is known and sufficiently close to the sought solution, then a Newton procedure provides a convenient route to the solution. Let us note that this minimization problem is global in time and space. Choosing global interpolation functions, as will be done later, will therefore provide a regularization of the problem both in space and time. If, for example, time is considered, a large series of images can be used to determine a given regular interpolation in time. This allows for a much better resolution in space. In the same manner a large spatial grid could allow for a very fine resolution in time.

A modified Newton iteration consists in correcting the displacement field by $\boldsymbol{\delta u}$, such that

$$[\boldsymbol{H}]\{\boldsymbol{\delta u}\} = \{\boldsymbol{j}\} \tag{4}$$

where the matrix $[\boldsymbol{H}]$ and second member vector $\{\boldsymbol{j}\}$ read

$$\boldsymbol{H}(\boldsymbol{x}, t) = \frac{1}{2\sigma^2} \boldsymbol{\nabla} f(\boldsymbol{x}, t_0) \otimes \boldsymbol{\nabla} f(\boldsymbol{x}, t_0)$$

$$\boldsymbol{j}(\boldsymbol{x}, t) = \frac{1}{2\sigma^2} \boldsymbol{\nabla} f(\boldsymbol{x}, t_0)\, \tau(\boldsymbol{x}, t) \tag{5}$$

One should note that when $\boldsymbol{u} = \boldsymbol{u}^*$ is the sought solution, $\boldsymbol{H}(\boldsymbol{x})$ corresponds to the Hessian of χ^2 with respect to $\boldsymbol{u}$. This explains the efficiency of using $[\boldsymbol{H}]$, which will therefore be referred to as the Hessian in the sequel. Once $\{\boldsymbol{\delta u}\}$ is computed, $\{\boldsymbol{u}\}$ is updated to $\{\boldsymbol{u} + \boldsymbol{\delta u}\}$, and the residual τ is refreshed. In the above writing, the second member is a column-vector of size DN_xN_t where D is the space dimensionality, N_x the number of pixels in the ROI, and N_t the number of instants of time. The Hessian is a square matrix of size $DN_xN_t \times DN_xN_t$.

As written, such problem is ill-posed and cannot be solved because it attempts at finding a vector (*i.e.*, $\boldsymbol{u}(\boldsymbol{x}, t)$) per scalar quantity (*i.e.*, $f(\boldsymbol{x}, t)$) at every point in space and time. To make the problem well-posed many strategies can be adopted. A natural one is to restrict the displacement field to a kinematic subspace. The choice of this set is critical, and will be discussed repeatedly in the following. Let us assume that the displacement fields are specialized to a subspace generated by a set of N_aN_t vector fields $\boldsymbol{P}_i(\boldsymbol{x}, t)$ such that

$$\boldsymbol{u}(\boldsymbol{x}, t) = \sum_i a_i \boldsymbol{P}_i(\boldsymbol{x}, t) \tag{6}$$

The above Newton scheme can be used in order to determine the unknown amplitudes gathered in a column-vector $\{\boldsymbol{a}\}$ of size N_aN_t, with an iterative

correction $\{\boldsymbol{\delta a}\}$,

$$[\boldsymbol{H_a}]\{\boldsymbol{\delta a}\} = \{\boldsymbol{j_a}\} \qquad (7)$$

Minimizing the cost function (3), a Galerkin procedure is used where the Hessian $[\boldsymbol{H_a}]$ and second member $\{\boldsymbol{j_a}\}$ read

$$\begin{aligned} [\boldsymbol{H_a}] &= [\boldsymbol{P}]^\top [\boldsymbol{H}][\boldsymbol{P}] \\ \{\boldsymbol{j_a}\} &= [\boldsymbol{P}]^\top \{\boldsymbol{j}\} \end{aligned} \qquad (8)$$

where $[\boldsymbol{P}]$ is a rectangular matrix of size $DN_x N_t \times N_a N_t$. It is quite natural at this stage to preserve the tensorial nature of the representation in orientation, space and time directions. That is, all components i_1 of the displacement field are treated in a similar way using an identical scalar shape function in space and scalar shape function in time. The latter can also be decomposed as the product of spatial shape functions $\phi_{i_2}(\boldsymbol{x})$ and a temporal one $\psi_{i_3}(t)$, so that now the index i refers to the triplet (i_1, i_2, i_3), and

$$\boldsymbol{P}_i(\boldsymbol{x}, t) = \boldsymbol{e}_{i_1} \phi_{i_2}(\boldsymbol{x})\psi_{i_3}(t) \qquad (9)$$

where $\boldsymbol{e}_{i_1}$ denotes a unit vector pointing along direction i_1.

This formulation of the problem is similar to the way it is tackled from the numerical modeling standpoint. In the latter case, it is standard to use a finite-element mesh for $\phi_{i_2}(\boldsymbol{x})$ so that i_2 refers to the nodal degrees of freedom. If the image contrast is sufficiently fine-scaled with respect to the element size, then the resulting Hessian $[\boldsymbol{H_a}]$ may be strictly positive and hence the problem may be solvable. This condition also implies that the mesh cannot be arbitrarily fine, as along with fineness comes first ill-conditioning, and second ill-posedness (*i.e.,* rank-deficient Hessian) as for the pixel-wise case.

After convergence, the residual field $\tau(\boldsymbol{x}, t)$ provides a good indication of the quality of the obtained solution. As earlier mentioned, it is ideally expected to consist of white noise, with an expected dimensionless variance of 1. When such a condition is not met, it is often easy to be in a position to interpret the cause. This may call for an enriched model (such as the inclusion of brightness or contrast corrections, when the residual has a non zero mean, or looks like the reference image [23]). It may also reveal the presence of discontinuity (*e.g.,* a crack) that has to be included in the description of the displacement field [24], or simply an unreliable part of the observed surface that has to be masked or ignored.

An important aspect of the analysis is that, from the noise η affecting the images, the uncertainty of the measured displacement amplitudes $\{\boldsymbol{a}\}$ can be computed analytically within the present framework [25] as

$$\{\boldsymbol{\delta a}\} = [\boldsymbol{H_a}]^{-1}\{\boldsymbol{j_a}\} \qquad (10)$$

with $\{\boldsymbol{j_a}\} = [\boldsymbol{P}]^\top \{\boldsymbol{\nabla} f \eta\}$. By construction, the tangent problem is linear and hence the displacement measurement is unbiased. Its covariance matrix $[\boldsymbol{C_a}]$ reads

$$[\boldsymbol{C_a}] = \langle \{\boldsymbol{\delta a}\}\{\boldsymbol{\delta a}\}^\top \rangle = [\boldsymbol{H_a}]^{-1} \qquad (11)$$

Therefore the above procedure delivers not only the estimated displacement field but also its associated uncertainty. This is an essential aspect of the DIC methodology, since displacements are almost never the ultimate quantity of interest, but rather intermediate data that are to be further processed to estimate the sought mechanical properties. Without the attached uncertainty, there is no way to weigh the measured information as it should.

Before closing this section, let us introduce a formal representation of the DIC process that will be used all along the paper in different contexts. A procedure, such as DIC, is schematically pictured in Figure 2. It is fed with entries $\boldsymbol{X}$ that are images for DIC, together with a norm that is issued from the noise affecting images. As above argued, the simple case of a Gaussian white noise corresponds to a uniform L2-norm to be used when registering images. The procedure itself can be seen as a black box, where algorithmic aspects are here irrelevant. However, it contains a *model*, here a relationship between images (*e.g.,* brightness conservation) that is parameterized by quantities generically denoted as $\boldsymbol{Y}$ in Figure 2, which are the nodal displacements $\{\boldsymbol{a}\}$ in the case of DIC. The model is formulated in a variational way because noise forbids a strict obedience to the model, and there lies the legitimacy of introducing a "measure" of the violation of the model assumptions that is fitted to the suspected cause of this violation noise. The *procedure* provides an estimate of the model (output) parameters, $\boldsymbol{Y} = \boldsymbol{Y}^*$, but also uncertainties attached to them, which may be encoded in an output metric. The procedure conveys the input noise affecting $\boldsymbol{X}$ into one affecting $\boldsymbol{Y}$, and this is essential for a proper usage of the output data. Alternatively, one may say that the output of the procedure is an "equivalent" variation problem, $\chi^2(\boldsymbol{Y}) \approx \chi^2_{osc}(\boldsymbol{Y}) = \|\boldsymbol{Y} - \boldsymbol{Y}^*\|^2$, where for DIC the output metric is $\|\{\boldsymbol{a} - \boldsymbol{a}^*\}\|^2 = \{\boldsymbol{a} - \boldsymbol{a}^*\}^\top [\boldsymbol{H}_a]\{\boldsymbol{a} - \boldsymbol{a}^*\}$. Additionally, the procedure provides also residuals that allow the (in)validity of the model to be assessed.
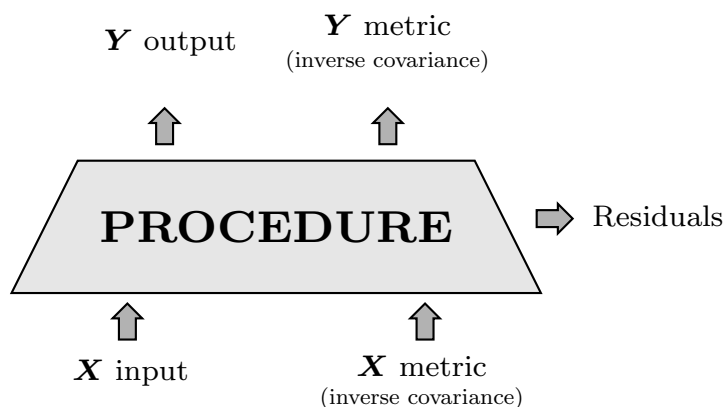
Fig. 2 Schematic representation of a procedure "distilling" input data $\boldsymbol{X}$ together with a metric describing in a statistical sense the noise that may affect $\boldsymbol{X}$, to provide measured $\boldsymbol{Y}$ data, together with the processed noise characteristics to be used for a suited metric on the output

Let us stress that the model itself may be highly nonlinear, and such is the case for DIC as the unknowns are arguments of an "image texture", that is a highly irregular function apart from the needed ability to interpolate gray levels at sub-pixel accuracy. Yet, at convergence, the problem has been substituted by its osculatory quadratic minimization approximation, χ^2_{osc}, constructed from the Hessian of the original problem, which is nothing but the inverse of the covariance of the processed noise affecting the output. Because of the intrinsic nonlinearities of the problem, one can trust the osculating problem only in a close neighborhood to the solution. Yet, if this condition is met, the osculating problem delivers a very robust but equivalent formulation, where robustness comes from the quadratic nature of the output.

Last, the number of data input in the procedure, N_X, is typically much larger than the number of output data, $N_Y \ll N_X$, schematically shown as a shorter output level than the base input in Figure 2. This difference is the redundancy that allows for extracting fewer meaningful data from many noisy data.

3 Constitutive parameter identification

3.1 FEMU

Identification may have a broad meaning. In the present case, it is proposed to embrace constitutive laws (CL), or elements of the constitutive law such as criteria (for the onset of yield, damage, cracking) but also shape (S), boundary conditions (BC), or any subset of parameters extracted from these three categories. All of these are formally represented by a column-vector

$\{\boldsymbol{p}\} = \{p_1, ..., p_{N_p}\}^\top$. The idea is to introduce as many parameters as possibly unknown — or at least subject to corrections — quantities. Calibrating a Young's modulus suggests that the latter is not known precisely, and this is natural in the context of experimental mechanics. Including shape parameters as unknowns may be relevant when for instance accounting for an unknown crack extension or an offset from a nominal CAD model. Last, especially when an instability is present, boundary conditions may differ from what was initially designed.

One of the most versatile identification technique corresponds to Finite Element Model Updating (FEMU) [9]. The underlying philosophy is to start from an initial approximation to the solution $\{\boldsymbol{p}_0\}$, and to iteratively correct it based on the comparison between computed and measured displacements. The cost function is thus defined to be

$$\chi^2_{\text{FEMU-a}}(\{\boldsymbol{p}\}) = \|\{\boldsymbol{a}^{\text{comp}}(\{\boldsymbol{p}\})\} - \{\boldsymbol{a}^{\text{meas}}\}\|^2 \tag{12}$$

At this stage, many norms can be used [9]. A Bayesian framework indicates that the best suited norm [26, 27] for Gaussian noise corrupting the displacement field is that given by the inverse covariance matrix of $\{\boldsymbol{u}\}$. With the chosen FE discretization, the inverse covariance is simply given by the reduced Hessian $[\boldsymbol{H_a}]$, and hence, for any i, the equation to solve is

$$\sum_{k,l} S_{ki} (H_a)_{kl} \left(a_l^{\text{comp}} + \sum_j S_{lj} \delta p_j - a_l^{\text{meas}} \right) = 0 \tag{13}$$

where

$$S_{ij} \equiv \frac{\partial a_i^{\text{comp}}}{\partial p_j}(\{\boldsymbol{p}\}) \tag{14}$$

are the sensitivity fields [26, 28] with respect to the parameter p_j expressed within the nodal description. The above linear system (13) is rewritten as

$$[\boldsymbol{H_p}]\{\boldsymbol{\delta p}\} = \{\boldsymbol{j_p}\} \tag{15}$$

with

$$\begin{aligned} [\boldsymbol{H_p}] &= [\boldsymbol{S}]^\top [\boldsymbol{H_a}][\boldsymbol{S}] \\ \{\boldsymbol{j_p}\} &= [\boldsymbol{S}]^\top [\boldsymbol{H_a}]\{\boldsymbol{j_a}\} \end{aligned} \tag{16}$$

Because the problem has been linearized up to convergence, the Gaussian character of image noise is transported to the displacement level where it remains Gaussian. However, even if the image noise were white, the DIC procedure transforms the noise on the image pixel into noise affecting the nodal displacement with nontrivial correlations. This is the very reason why the Hessian should come into play in the definition of the norm.

The tangent problem is again linear. Consequently the parameter calibration is unbiased. Its covariance matrix $[\boldsymbol{C_p}]$ then becomes

$$[\boldsymbol{C_p}] = \langle \{\boldsymbol{\delta p}\}\{\boldsymbol{\delta p}\}^\top \rangle = [\boldsymbol{H_p}]^{-1} \tag{17}$$

and is the result from propagating the uncertainties from the pixel level up to the parameter level.

It is to be emphasized that even if the used finite element mesh is extremely fine, leading potentially to a singular Hessian, it can be used in the above norm. The kernel of the Hessian may lead to ill-posedness, but when the displacement field is sought in a subspace, which is generated by the sensitivity fields, the restriction of the Hessian to this subspace is likely to be strictly positive, and better conditioned as the subspace dimensionality decreases.

It is noteworthy that FEMU can be cast into the scheme of procedures shown in Figure 2. Here the input (resp. output) is $\{a\}$ (resp. $\{p\}$). The similar format of the input and output was chosen so that one can easily "plug" the two procedures, and perform them sequentially to achieve identification from images. It is also possible to merge the two steps into a single integrated scheme [29–33].

3.2 Integrated DIC

The fact that the Hessian of the DIC problem comes into play in the expression of the FEMU-a kernel shows that the above norm used in χ^2_{FEMU} is such that

$$\|\boldsymbol{\delta u}(\boldsymbol{x},t)\|^2 \propto \chi^2[\boldsymbol{u}^{meas}(\boldsymbol{x},t) + \boldsymbol{\delta u}(\boldsymbol{x},t)] \tag{18}$$

when $\boldsymbol{\delta u}$ is sufficiently small (*i.e.,* less than the image correlation length). More precisely the combination DIC+FEMU is osculatory to the above cost function. This can easily be shown from the corresponding Hessian

$$[\boldsymbol{H_p}] = [\boldsymbol{S}]^\top[\boldsymbol{H_a}][\boldsymbol{S}] = [\boldsymbol{S}]^\top[\boldsymbol{P}]^\top[\boldsymbol{H}][\boldsymbol{P}][\boldsymbol{S}] \tag{19}$$

Hence, this specific FEMU variant (*i.e.,* FEMU-a using $[\boldsymbol{H_a}]$ as a norm) is mathematically equivalent to solving for the DIC problem, with the help of the identification framework as a very well-suited regularization, rather than a convenient heuristics as usually brought by a Tikhonov regularization [34]. It implies that DIC and FEMU can be performed in a *single* step, called I-DIC (for Integrated-DIC). Within this framework, the mesh almost disappears as it is just an intermediate tool that allows for the interpolation of the displacement field at any pixel position, or equivalently, I-DIC, corresponds to choosing as a mesh the pixel representation, that is a pixel-wise FEMU.

Using the schematic convention of procedure as building bricks (Figure 2), the assembly of DIC and FEMU into I-DIC is shown in Figure 3. In the present case, since FEMU is based on the measured displacement fields associated with the chosen kinematic subspace (see Equation (6)), it is designated as FEMU-a hereafter (*i.e.,* to refer to the measured amplitudes $\{a\}$ used as input data and the corresponding covariance matrix $[\boldsymbol{H_a}]^{-1}$).
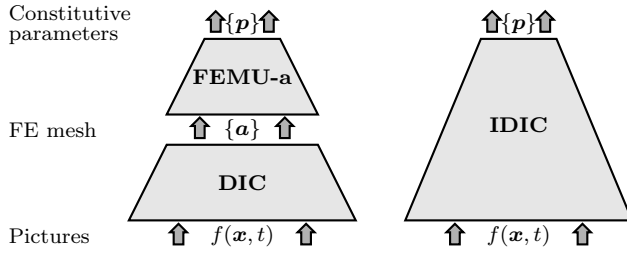
Fig. 3 DIC and FEMU-a can be assembled as two sequential operations (left) or merged into a single one (right)

It is noteworthy that for integrated-DIC, the pixel-wise and the (nodal) FEMU-a identification are mathematically equivalent, provided the auxiliary steps consisting in projecting the displacement fields onto the predefined basis do not introduce any significant model error (explicitly, the sensitivity fields $[S]$ should be well described by the finite element mesh), or that the latter is kept under control. The similarity of the different formulations is established when no model error occurs (*e.g.,* the mesh should not be too coarse, the identified model should not pull the DIC solution outside the domain where the Taylor expansion based on pixel-size displacement breaks down). Otherwise the parabolic expansion of the three comparison metrics are *identical* at the DIC solution [31], so that the variational problems are osculatory. However, the quadratic cost functions strengthen the robustness of the algorithm. Moreover, the nodal version may be computed for different mesh sizes. Coarser meshes lead to fewer unknowns, and hence less demanding computations, with potentially a better conditioning. Thus as long as the approximation error due to the mesh remains well below the uncertainty of DIC, a coarse mesh is beneficial [33]. Conversely, when the sample geometry becomes more complex and/or the picture contrast is poor, fine meshes can be used with integrated approaches. However, the very same meshes would not allow a standard DIC code to converge [29, 35].

3.3 Illustration with an elastoplastic test case

To illustrate IDIC, a test case is discussed in which the elastoplastic material parameters of J2-flow rule with freeform isotropic hardening are calibrated from a tensile experiment on an aluminum alloy (AA2219) sample. The latter is a dog-bone geometry with a hole to increase the sensitivity to the sought material parameters (Figure 4(a)). It is loaded cyclically, where the maximum displacement for the first cycles is incrementally increased (Figure 4(b)). A 4-Mpixel camera is used (with 16-bit digitization). The physical size of each pixel is 11.14μm. A telecentric lens $\times 0.25$ is selected to minimize the effects of out-of-plane motions.
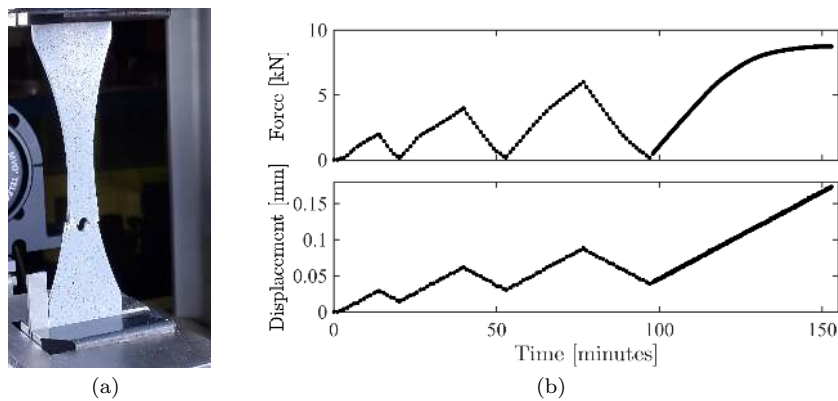
Fig. 4 (a) Tensile test on dog-bone sample with a hole. (b) Load and stroke of the analyzed test as measured at the upper grip of the tensile tester

To drive IDIC calculations, Dirichlet boundary conditions are measured via global DIC with a triangular mesh (Figure 5). In the present case, DIC and IDIC analyses are performed with the *same* surface mesh. This requirement is not mandatory [29, 35] and discretization convergence of the IDIC results is not discussed herein. It was shown on other cases [33] that rather coarse meshes allowed for very good estimates of the material parameters provided that the geometry of the sample is well captured [29, 35]. It is worth noting that even though 2D pictures are considered and the sample thickness is small (*i.e.,* equal to 2 mm), IDIC analyses are performed with 3D finite element simulations. The latter ones are performed within the large transformation framework. In elastoplasticity in particular, 3D simulations are needed even for sub-millimeter thin samples [35]. Consequently, the measured displacements are extruded along the thickness of the sample where they are prescribed (Figure 5).
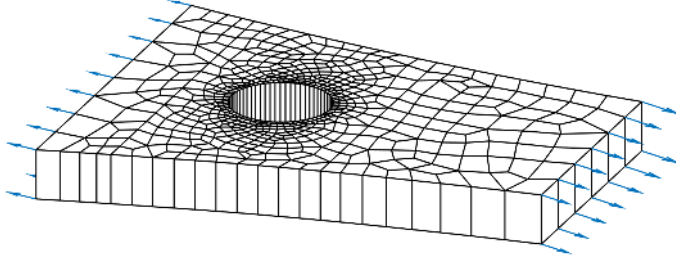
Fig. 5 Mesh used for IDIC analyses, which are based on measured and extruded Dirichlet boundary conditions (blue arrows)

Figure 6 shows the results of DIC analyses of the whole test. It should be noted that displacement fields are defined over space and time (or increment), and they are vector fields with in-plane components for the displacements u_x, u_y and the resultant force F is defined over time only. Figure 6 only shows two cross-sections for each displacement component, one in space for increment 637 (of 639) and one in increment at $x = 1400$ pixels (see vertical dashed lines in Figure 6). This type of representation will be used in the remainder of the paper. Figure 6(a) shows the displacement fields at the last increment. The corresponding in-plane logarithmic strain fields are shown in Figure 6(b). At the end of the test, four strained bands are observed. They all emanate from the central hole and have developed over the last 200 increments. It is also concluded that the probed longitudinal strain range is of the order of 10 %.

In the following, the gray level residuals $\tau(\boldsymbol{x}, t)$ are normalized with respect to the dynamic range Δf_0 of the picture in the reference configuration (*i.e.*, $\Delta f_0 = \max_{\text{ROI}} f(\boldsymbol{x}, t_0) - \max_{\text{ROI}} f(\boldsymbol{x}, t_0)$). Figure 6(a) indicates that the registration quality is generally very good since the residuals remain very close to the levels associated with acquisition noise (*i.e.*, of the order of 1.6 % of Δf_0). However, there are areas where they are a bit higher throughout the whole load history. This is in part due to dust on the objective lens. At the end of the test, higher residuals are observed in the strained regions. They are due to paint cracking and the fact that the FE mesh is not able to fully capture the localized character of these bands. The overall root mean square residual is equal to 1.9 % of the dynamic range, which is deemed very satisfactory
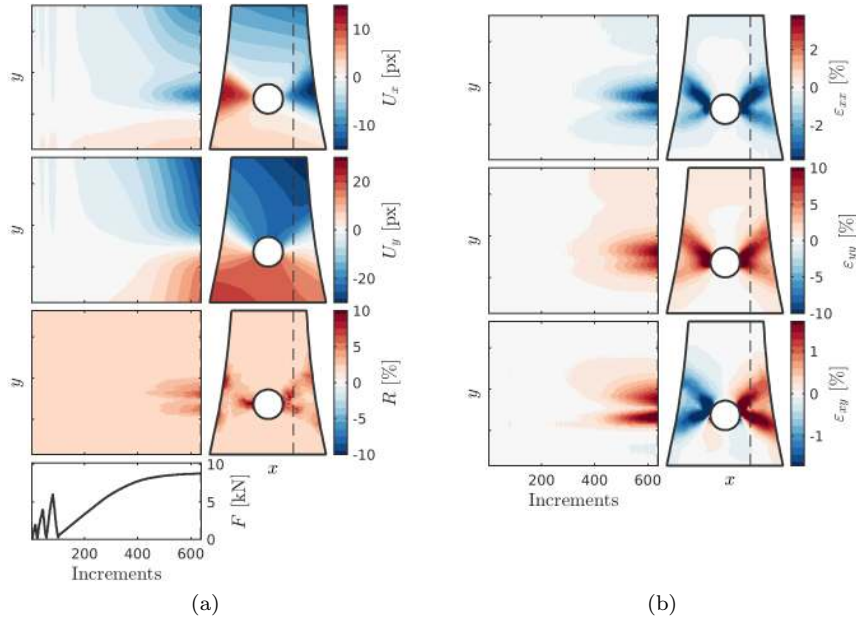
Fig. 6 Global DIC results for the analyzed test. (a) Displacement fields expressed in pixels and normalized gray level residual with respect to the dynamic range of the picture in the reference configuration. (b) Corresponding in-plane strain fields

Ten material parameters (*i.e.,* $\{\boldsymbol{p}\} = \{E, \nu, \sigma_y, \Delta\sigma_1 \ldots \Delta\sigma_7\}^\top$) are unknown. The first two are Young's modulus E and Poisson's ratio ν. The third parameter corresponds to the yield stress σ_y. Then seven additional stress increments $\Delta\sigma_i$ are considered. The corresponding strain levels are chosen in increasing amplitude to properly capture the initial yielding (*i.e.,* at $[0.1, 0.2, 0.5, 1.1, 2.8, 6.0, 9.5]$ % equivalent plastic strain). Since stress related parameters are sought, load measurements are accounted for in identification steps. In the present case, the resultant forces $\{\boldsymbol{F}^{\mathrm{comp}}\}$ induced by Dirichlet boundary conditions are compared with measured load data $\{\boldsymbol{F}^{\mathrm{meas}}\}$. The corresponding global equilibrium residual reads

$$\chi_{\mathrm{F}}^2(\{\boldsymbol{p}\}) = \|\{\boldsymbol{F}^{\mathrm{comp}}(\{\boldsymbol{p}\})\} - \{\boldsymbol{F}^{\mathrm{meas}}\}\|^2 \qquad (20)$$

To be consistent with the present framework, for Gaussian noise corrupting the load measurements is that given by the inverse covariance matrix of $\{\boldsymbol{F}^{\mathrm{meas}}\}$. In the present case, the load measurements are uncorrelated and their variance is denoted by σ_F^2. With the used load cell, $\sigma_F = 4$ N. Consequently, the global equilibrium is rewritten as

$$\chi_{\mathrm{F}}^2(\{\boldsymbol{p}\}) = \frac{1}{N_t \sigma_F^2} \left(\{\boldsymbol{F}^{\mathrm{comp}}\} - \{\boldsymbol{F}^{\mathrm{meas}}\}\right)^\top \left(\{\boldsymbol{F}^{\mathrm{comp}}\} - \{\boldsymbol{F}^{\mathrm{meas}}\}\right) \qquad (21)$$

There is an additional (static) Hessian $[\boldsymbol{H_F}]$

$$[\boldsymbol{H_F}] = \frac{1}{\sigma_F^2}[\boldsymbol{S_F}]^\top[\boldsymbol{S_F}] \tag{22}$$

that is formed by gathering all the load sensitivities

$$S_{F_{ij}} \equiv \frac{\partial F_i^{\mathrm{comp}}}{\partial p_j}(\{\boldsymbol{p}\}) \tag{23}$$

in a rectangular sensitivity matrix $[\boldsymbol{S_F}]$.

For IDIC the total residual is formed as the weighted sum of the dimensionless gray level and load residuals [36]

$$\chi_{\mathrm{tot}}^2 = \frac{N_x}{1+N_x}\chi^2 + \frac{1}{1+N_x}\chi_{\mathrm{F}}^2 \tag{24}$$

For the other identification methods, similar weighings are used.

Figure 7 shows the ten corresponding sensitivity fields. They are computed using finite differences with a perturbation factor ϵ of 1 %. Within figure 7 the sensitivity fields are shown as multiplied with their respective parameter value in order to bring them all to the same units and ease comparison. This makes them effectively relative sensitivity fields. They can be read as, the correction to the simulated displacement fields for a change in parameter of amplitude 1. Additionally, the correction to the simulated reaction force is also shown. Note that the force and displacement contributions are connected and cannot change individually. These figures can be used to evaluate any identification problems. In practice, two concepts are important. The first is that the sensitivity fields have amplitudes that are within the measurable range. For this experiment the displacement uncertainty as about 0.0144 px, and all sensitivity fields express stronger displacement amplitudes. However, this would be for a parameter change of one, which is typically larger than the desired identification uncertainty. Consequently, the contributions from the force are important and significantly enhance the sensitivity. The second important concept is that they are unique. If any linear combination of sensitivity fields can express another field, than this will result in conditioning issues of the identification Hessian. The chosen set of parameters each have their support relatively local in time. That is to say, the parameters are activated sequentially. Therefore, no uniqueness problems are expected.

(a) Young's modulus E (b) Poisson's ν (c) Yield stress σ_y

(d) Hardening stress $\Delta\sigma_1$ (e) Hardening stress $\Delta\sigma_2$ (f) Hardening stress $\Delta\sigma_3$

(g) Hardening stress $\Delta\sigma_4$ (h) Hardening stress $\Delta\sigma_5$ (i) Hardening stress $\Delta\sigma_6$

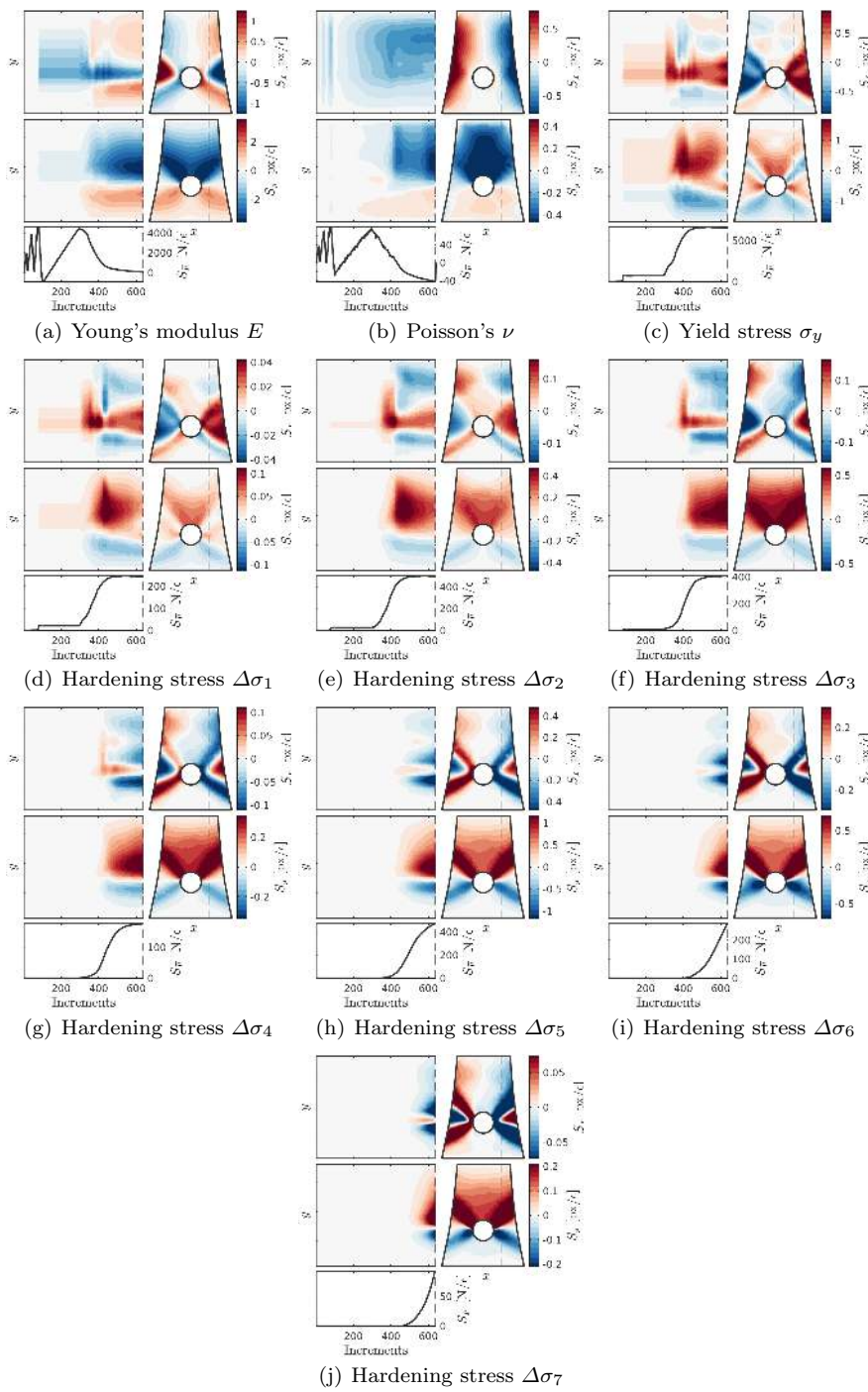(j) Hardening stress $\Delta\sigma_7$

Fig. 7 Sensitivity fields for all ten material parameters for the final set of material parameters. The plotted fields are $[\boldsymbol{S}_i]p_i$, this multiplication of p_i is done to bring them all to the same units.

When these sensitivity fields are utilized for calibration purposes, the identification residual is successfully minimized and the optimal parameters are reported in Table 1. The calibrated elastic and plastic properties are consistent with literature and industrial data on this type of alloy.

Table 1 Identified parameters for the four different identification cases and the initial guess

	E GPa	ν -	σ_y MPa	$\Delta\sigma_1$ MPa	$\Delta\sigma_2$ MPa	$\Delta\sigma_3$ MPa	$\Delta\sigma_4$ MPa	$\Delta\sigma_5$ MPa	$\Delta\sigma_6$ MPa	$\Delta\sigma_7$ MPa
$\{p_0\}$	43.2	0.168	225	9.2	16.9	13.7	7.8	16.0	15.1	6.5
IDIC	75.3	0.319	376	14.2	28.8	23.6	10.3	28.6	24.9	9.9
FEMU-a	75.3	0.319	376	14.1	30.0	22.2	10.6	28.5	25.3	12.3
FEMU-b	76.0	0.310	373	16.1	31.9	19.4	11.4	29.1	25.3	14.0
FEMU-b (secant)	76.0	0.310	373	16.0	31.9	19.4	11.4	29.1	25.4	14.0

Figure 8 shows different residuals for the initial step of the IDIC procedure and at convergence. Three different types of residuals are reported for IDIC analyses. The global residual that is minimized is a weighted sum of gray level and equilibrium residuals based on a Bayesian foundation [36]. Since DIC analyses are also performed, a residual on DIC and IDIC displacements is also constructed.

First, displacement residuals, which are the displacement differences between DIC and IDIC analyses, are reported. Since measured (via DIC) displacements are prescribed on the top and bottom boundaries of IDIC analyses, their difference vanishes. Further, until increment 300, the displacement differences remain very small (and close to the measurement uncertainty whose standard deviation is 0.0144 pixel for the chosen discretization. In this part of the experiment, the sample is essentially loaded in the elastic regime and then small scale yielding occurs. In the second part of the experiment, larger residuals are observed. This is an indication of model error. Table 2 shows that the root mean square displacement difference R_u is about ten times higher than the measurement uncertainty. First, the mesh may not be fine enough. Second, the material model, even though based on a freeform hardening model, is not able to fully capture all the kinematic details of the experiment. Third, measured boundary conditions have been extruded (Figure 5), which is not necessarily totally consistent with the experiment [35]. Last, the displacement residuals do not decrease very much when the first iteration results are compared with those at convergence. This is due to the fact that measured boundary conditions are considered herein.

Gray level residuals are also reported. This quantity is one way of characterizing the overall quality of the IDIC registration. It also gradually degrades as the increment number increases and a larger zone of the sample yields. When comparing the residuals at the beginning and the end of the minimization process, their level only slightly decreases, in line with the fact that the displacement fields did not vary much. Interestingly, they are also very close

to DIC residuals (Figure 6(a)). Table 2 shows that the root mean square gray level residual R_f normalized by the dynamic range Δf_0. If it were only related to acquisition noise, its level is equal to 1.62 %.
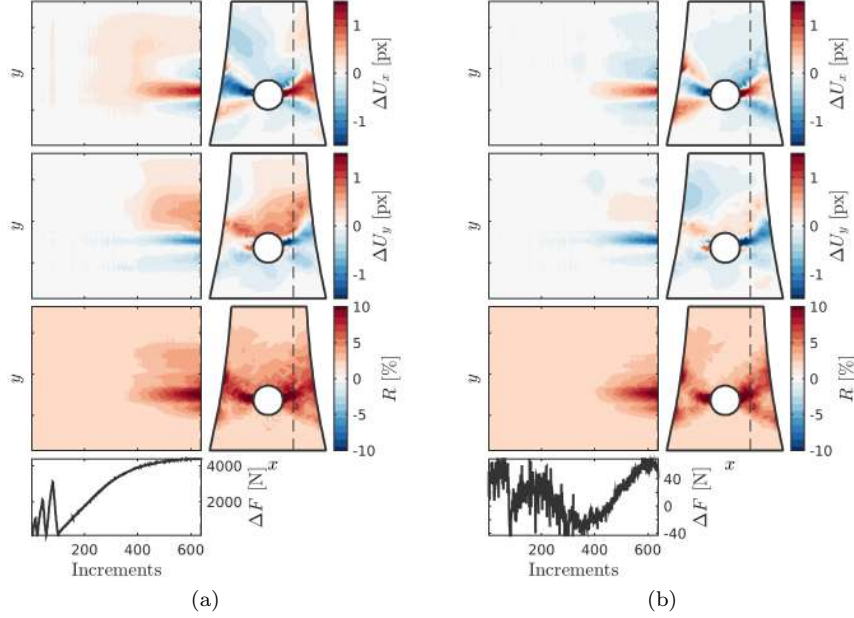

Fig. 8 IDIC residuals for the analyzed test. Displacement residuals expressed in pixels, normalized gray level residuals and load residuals for the first (a) and converged (b) solution

Finally the load residuals show that there is a clear decrease (by a factor of 100 for the root mean square difference, see Table 2) when the initial guess is compared with the final result. In the present case, this quantity is the most sensitive. This is due to the fact that the initial guess of the parameter set was about 50 % off from the converged solution, which results in stress levels that are significantly lower (Figure 9). The final RMS residual is still higher (by a factor of 10) in comparison with load uncertainty. This again indicates that model errors remain at the end of the analysis.

Table 2 Global residuals for the four different identification cases and the initial guess

	R_u [px]	R_F [N]	R_f [%]
DIC	-	-	1.90
$\{p_0\}$	0.399	3203	2.50
IDIC	0.344	35	2.15
FEMU-a	0.340	33	2.15
FEMU-b	0.338	40	2.16
FEMU-b (secant)	0.338	41	2.16

Even though the analysis of the residuals indicates that no perfect match was achieved, the levels of the reported residuals remain low (*i.e.,* only ten times the acquisition noise) so that for the analyzed strain range, it is believed that the calibrated freeform model (Figure 9) is a rather good approximation of the actual behavior of the studied aluminum alloy.
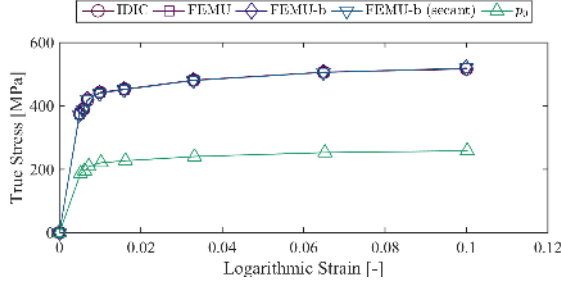

Fig. 9 Stress-strain curve of the studied material when different calibration techniques are used, and for the initial set of material parameters $\{\boldsymbol{p}_0\}$

Last, the original kinematic sensitivity fields, when expressed at the pixel level, contain approximately 11.9×10^9 values (see Table 3) when 639 increments are considered. When written at the nodes, the number of values is reduced to 11.7×10^6, and only 10 constitutive parameters are to be determined in the end.

Table 3 The size of objects required within the different proposed algorithms

	Number of values	% of zeros
Images	640	-
Image pixels	4.2M	-
ROI pixels N_x	1.8M	-
Increments N_t	639	-
Nodes	630	-
Nodal DOF N_a	1.9k	-
Modes N_b	100	-
Parameters N_p	10	-
$[\boldsymbol{P}][\boldsymbol{S}]$	11.7G	0
$[\boldsymbol{P}]$	3.5G	99.8
$[\boldsymbol{S}]$	12.1M	49.8
$[\boldsymbol{U}]$	639k	9.1
$[\boldsymbol{V}]$	189k	0
$[\boldsymbol{H}_a]$	3.6M	99.5
$[\boldsymbol{H}_b]$	10.0k	0
$[\boldsymbol{H}_p]$	100	0

4 Data and Model Reduction

In computational mechanics, model reduction techniques are used to solve problems, which are out of reach of direct numerical simulations even using parallel computing. Such problems are encountered for example in case of parametric models, the cost of which would increases exponentially with the number of parameters in the formulation otherwise. Those methods, unlike meta-modeling techniques, do not intend to simplify the model but to provide a framework for approximating the solution of the mechanical problem associated with a rather complex physical model. In this context model-order reductions usually involve two phases, namely, the building of the parameterized reduced-order model, usually carried out off-line, and an on-line phase, which consists in particularizing the solution. Among those techniques that have been quite extensively used over the last decade one can list the proper orthogonal decomposition (POD) method [37] (and the closely related Principal Component Analysis (PCA), singular-value decomposition (SVD) or Karhunen-Loeve (KL) techniques), the reduced-basis (RB) method [38] and the proper generalized decomposition (PGD) method [39, 40]. Some approaches (*e.g.,* POD, RB) rely on a learning phase, for which the solution of the problem is partially computed for given instants or values of parameters (snapshots) in order to extract the reduced-order basis that can be used online to find the approximation of another similar problem.

In contrast, PGD techniques do not rely on such learning stage as they directly build separated-variable approximations of the solution fields using a greedy algorithm, without any prior information on this solution. Such techniques allow for a strong reduction of the dimension of the space in which the solution is sought. However, in case of nonlinear problems, evaluations over the initial high dimensional space are needed and hence the net benefit of such techniques may be limited. Therefore, and depending on the type of techniques, a second reduction procedure aiming at reducing the evaluation step over a lower dimensional space were developed as the Empirical Integration Method for the RB approach [41–43], the hyper-reduction method [44, 45] or data compression method for PGD [46].

Another way of classifying these model reduction techniques is to consider the metric with which the reduced basis is chosen. For POD and its variants, the Euclidian L2-norm is implicitly chosen (and the orthogonality in its acronym indeed refers to the plain scalar product). Although not classically presented this way, it is straightforward to formally extend POD to more general quadratic norms than L2, with an arbitrary metric tensor. The multiplication of all fields from the library by the square root of the metric tensor leads to new fields that can be reduced using classical POD, and reverting back to the original space solves the problem. This observation may be useful because of the emphasis that should be given in utilizing a suited norm tailored to the noise (or uncertainty).

For PGD [39] and RB [38], the used metric is based on a variational formulation of the problem to solve. When model reduction aims at solving a

specific problem, it appears as satisfactory to evaluate residual errors using the very same formulation as that used for handling the problem directly. In the context of solid mechanics, PGD may use a constitutive law error [47] having the dimension of an energy, hence endowed with a concrete physical meaning.

4.1 The Multi-query LATIN method

The computation of the Hessian of the FEMU-a problem could be performed efficiently in a simple manner exploiting a nice property of the time-space PGD approach using the LATIN algorithm [48]. Let us first recall the basic ingredients of this algorithm, presented here in the context of nonlinear constitutive models [47,49] before considering its use for computing sensitivity fields.

The reference problem is quasi-static and isothermal for a structure defined over the time-space domain $\mathcal{I} \times \Omega$, with $\mathcal{I} = [t_0, T]$ the interval of study, assuming small perturbations. The structure is subjected to prescribed body forces $\boldsymbol{f}_d$ on Ω, to tractions $\boldsymbol{F}_d$ over a part $\partial_F \Omega$ of the boundary, and to prescribed displacements $\boldsymbol{u}_d$ over the complementary part $\partial_u \Omega$. The state of the structure is defined by the set of fields $\mathbf{s} = (\boldsymbol{\sigma}, \boldsymbol{\epsilon}_i, \{\boldsymbol{\alpha}\}, \{\boldsymbol{\beta}\})$, in which

- $\boldsymbol{\sigma}$ designates the Cauchy stress tensor;
- $\boldsymbol{\epsilon}_i$ the inelastic part of the strain tensor $\boldsymbol{\epsilon}$ corresponding to displacement field $\boldsymbol{u}$, with the classical additive decomposition into an elastic part $\boldsymbol{\epsilon}_e = \mathcal{H}^{-1}\boldsymbol{\sigma}$ and an inelastic part $\boldsymbol{\epsilon}_i = \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_e$;
- $\{\boldsymbol{\alpha}\}$ N_α internal variables;
- $\{\boldsymbol{\beta}\}$ the set of N_α variables conjugate to $\{\boldsymbol{\alpha}\}$;
- a normal formulation is adopted, which means that by a change of variable, if needed, the state equations are linear, *i.e.*, $\{\boldsymbol{\beta}\} = [\boldsymbol{L}]\{\boldsymbol{\alpha}\}$.

All these quantities are defined over $\mathcal{I} \times \Omega$ and assumed to be sufficiently regular. The nonlinearity appears for growth laws (25) formally defined by the monotone operator $\mathbf{B}$ such that

$$(\dot{\boldsymbol{\epsilon}_i}, \{\dot{\boldsymbol{\alpha}}\}) = \mathbf{B}(\boldsymbol{\sigma}, \{\boldsymbol{\beta}\}) \tag{25}$$

Let $\boldsymbol{\Gamma}$ designate the space of fields $\hat{\mathbf{s}}$ satisfying the growth laws (25). It is noteworthy that these nonlinear equations are local and can be processed in parallel. Let us also introduce the space $\mathbf{A_d}$ of *admissible fields* $\mathbf{s} = (\boldsymbol{\sigma}, \boldsymbol{\epsilon}_i, \{\boldsymbol{\alpha}\}, [\boldsymbol{L}]\{\boldsymbol{\alpha}\})$ such that the stress tensor $\boldsymbol{\sigma}$ and the displacement field $\boldsymbol{u}$ are respectively statically and kinematically admissible. This definition involves only linear operators with long-range couplings.

The solution to the previous nonlinear problem is then defined by the field $\mathbf{s_{ex}}$ that belongs both to $\mathbf{A_d}$ and $\boldsymbol{\Gamma}$, as sketched in Figure 10. The LATIN algorithm consists in solving the problem iteratively, projecting alternatively on these two spaces, as shown with the search directions $\mathbf{E}^+$ and $\mathbf{E}^-$. Usually $\mathbf{E}^-$ is defined as the tangent to $\boldsymbol{\Gamma}$ taken at the iterate $\hat{\mathbf{s}}_{n+\frac{1}{2}}$ belonging to $\boldsymbol{\Gamma}$, while $\mathbf{E}^+$ solves Equation (25).
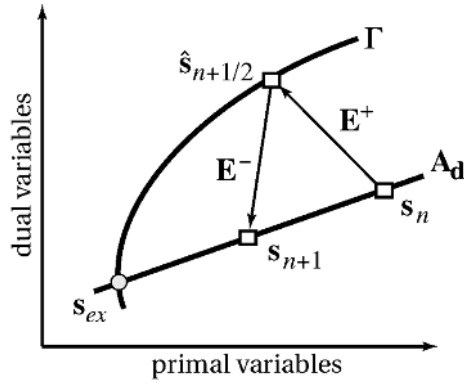
Fig. 10 Graphical sketch of the LATIN algorithm (after [49])

One aspect of the LATIN framework is that a reduced-order basis $\{\boldsymbol{u}_l(\boldsymbol{x},t)\}_{l=1}^{m'}$ is constructed. The approximation of the displacement field $\boldsymbol{u}(\boldsymbol{x},t)$ over the space-time domain $\mathcal{I} \times \Omega$ is iteratively computed as

$$\boldsymbol{u}(\boldsymbol{x},t) \approx \sum_{l=1}^{m'} \gamma_l(t)\,\boldsymbol{\Gamma}_l(\boldsymbol{x}) \tag{26}$$

More details about the LATIN algorithm can be found in Ref. [49].

In the context of this paper, the solution to the linearized FEMU-a problem (see Equation (13)), for small variations of the parameters, is a domain where PGD-like algorithms are expected to reveal very efficient. The LATIN method, and more specifically the so-called multi-query-PGD algorithm [48], allows for the initialization over the entire space and time domain of the fields to be computed, for each new set of parameters, as shown schematically in Figure 11.
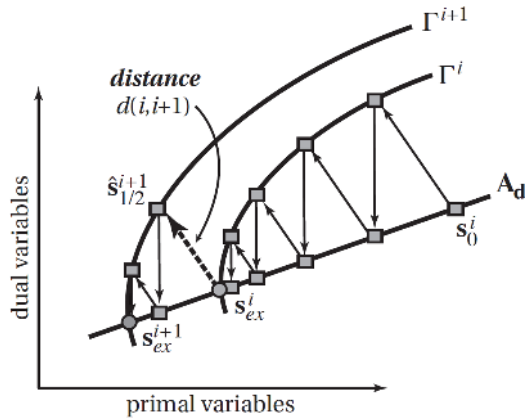

Fig. 11 Graphical sketch of the multi-query-PGD algorithm (after [48])

The first step updates the time functions $\{\lambda_l(t)\}_{l=1}^{m'}$ leading to low computational costs. When the sought parameters belong to a small region in the parameter space, this step is often sufficient. If not, very few additional spatial modes $\boldsymbol{\Gamma}_l(\boldsymbol{x})$ (in general, one or two) have to be added to precisely compute the solution for the new set of parameters. With this technique, the Hessian with respect to the parameters could be computed by looking for the solution for a small variation of each parameter, leading to a cheap computation of the sensitivity fields. In Ref. [50], more involved routes are followed allowing for the derivative of the time and space functions with respect to the parameters. Two techniques were tested, namely, a direct closed-form differentiation [51], and the adjoint state method [52]. These methods are efficient but require quite heavy code developments (*i.e.,* they are very intrusive).

4.2 PGD parameterized problem

In the FEMU context, the problem to be solved corresponds to the minimization of the cost function (21) over the parametric space used to describe the candidate constitutive law. As such a "brute force" parametric PGD [40,53] could be used by seeking the displacement field over the whole parametric space. The solution $\boldsymbol{u}$, which depends on a set of parameters of dimension N_p $\{p\} \in \boldsymbol{\mathcal{P}}$, is sought as a separated variables approximation of $\boldsymbol{u}$ (where separation is proposed for the constitutive parameters only)

$$\boldsymbol{u}(\boldsymbol{x}, t; \{p\}) \approx \sum_{k=1}^{m} \lambda_k(\{p\}) \, \boldsymbol{\Lambda}_k(\boldsymbol{x}, t) \tag{27}$$

with

$$\lambda_k(\{p\}) = \prod_{j=1}^{N_p} \lambda_k^j(p_j) \tag{28}$$

where m is the number of "modes" used in the approximation, λ_k^j scalar functions, and $\boldsymbol{\Lambda}_k(\boldsymbol{x}, t)$ the spatio-temporal modes of the displacement field. It is worth noting that this form does not separate space and time dimensions.

The PGD method provides a representation with separated variables requiring the use of greedy algorithm [39] updating at each iteration the scalar functions separately. What remains to be carried out is to look for the set of parameters that minimizes the FEMU-a error. One advantage of the method is that it allows for the determination of the best candidate in case of multiple local minima and does not require any gradient calculation. It is to be noted though that this method has mainly been applied to linear problems (for parameters such as Young's moduli, Poisson's ratios, diffusion coefficients) where affine decompositions of the operator allows for quite efficient and robust algorithms.

Fewer works have considered the solution of time-dependent *and* nonlinear parameterized problems. The LATIN method [47], which iteratively seeks solutions over the whole space-time domain, has been mostly used. It allows for

the possibility of adding constitutive parameters as extra coordinates [54]. Although very appealing, the efficiency of this approach for nonlinear operators is still an open issue.

5 Data processing and reduction applied to mechanical identification

5.1 Data reduction for displacement fields in DIC and DVC

The above section is a strong motivation for using model reduction or data reduction techniques in the introduced context of DIC and further for identification.

Volumetric registrations are usually concerned with very large data sets. Consequently, the number of measured kinematic degrees of freedom are generally numerous. It has been proposed to account for displacement fields in a separated form for different space dimensions. Such a specific form for displacement fields can be introduced in the formalism of DIC/DVC, retaining only the dominant mode repeatedly until the residual level has been exhausted [55]. This way of handling the problem is precisely in the spirit of PGD, where i) the different modes are introduced in a greedy approach (and not for reducing a set of prior fields), and ii) the separated form is inserted directly in the variational formulation of here DIC/DVC. However, because most of the computation cost is related to the correction by the measured displacement field of the deformed volume the CPU time gain is not spectacular in comparison with standard implementations.

It is worth noting that such strategies have also been proposed for 2D DIC registrations [56]. In this case, convergence is faster and equivalent uncertainty levels are observed as in standard Q4-DIC. However, when the displacement field contains discontinuities that are not aligned with the axes chosen for the separation, the procedure is detrimental.

The way DIC (and DVC) registrations were presented herein is via spatiotemporal formulations (3). These approaches are particularly well-suited when dealing with parameter calibration [31–33] and more generally when comparing measurements with simulations. They can also be followed for measurement purposes [57,33], exploiting space-time separation. However, it is worth noting that virtually all DIC (and DVC) codes only solve "incrementally" the minimization of χ^2 (*i.e.,* for each time step t *independently*). The additional cost of adding the time axis can be made minimal when using space/time decompositions [57], or by driving the FE code with all the measured boundary conditions for each image acquisition [31–33].

In the following, model reduction or reduced basis will be mostly driven by the identification goal, rather than from the DIC algorithm.

5.2 Data pyramid

Figure 3 summarizes the different steps that are performed to process the acquired pictures in order to calibrate material parameters. As above shown, FEMU-a is based on sensitivity fields computed over the mesh with the nodal displacements $\{a\}$. When including DIC within identification as in I-DIC, these sensitivity fields require evaluations at the pixel level for each acquisition time, thereby inducing matrices of considerable size to store.

Alternatively, a sequential treatment consists of first measuring displacement fields (*e.g.,* via DIC or DVC). In the present case, it is assumed that finite-element based registration procedures are chosen. Such techniques give access to nodal displacements for the chosen spatial discretization over all the considered acquisition time steps. At this stage, the pictures are no longer needed since the displacement fields have been measured. To extract the sought material parameters, the measured fields and additional data (*e.g.,* load measurements) are to be compared with, say, finite element simulations.

Conversely, the Hessian matrix is usually of modest size (*i.e.,* with only $N_p \times N_p$ components, where N_p is the number of unknown material parameters). This contrast in size between the sensitivity fields and the Hessian is the consequence of the integrated DIC procedure where the maximum data set is applied to directly obtain the minimum number of unknowns. I-DIC has the advantage of allowing for an arbitrarily fine mesh, as dictated for instance by quality criteria used in the FE simulation, at the expense of large data sets [29,33]. In contrast, whenever a relatively coarse mesh is sufficient to faithfully describe the displacement field, it is more efficient to split the transformation into two parts, namely, first from the pixel space to the FE nodal space, then from the nodal space to the parameter space, and as earlier mentioned the two routes are mathematically equivalent provided no spurious minima has been retained in the first DIC step.

These sensitivity fields can be further reduced by computing the modes associated with a reduction procedure (*e.g.,* POD [58]). Then the material parameters associated with a chosen constitutive model are calibrated. Last, if additional failure criteria are sought, the simulations need to be further processed to analyze, say, various equivalent stresses or strains or any other internal variable that is not directly accessible in experiments [35].

The model reduction strategy (*e.g.,* POD [59]) introduces a new subspace, which is intermediate between the FE nodal space and the material parameter space as shown in Figure 12. In total, four discrete spaces have been introduced herein, namely, i) the pixel space or data acquisition space, ii) the FE mesh or computational space, iii) the dominant eigen modes of POD space, and iv) the material parameter space. With this additional subspace, FEMU is no longer performed at the level of pixels (*i.e.,* IDIC) or measured displacement fields (*i.e.,* FEMU-a) but at the level of the modes described by the amplitudes collectively gathered in vector $\{b\}$. Such approach will be referred to as FEMU-b.
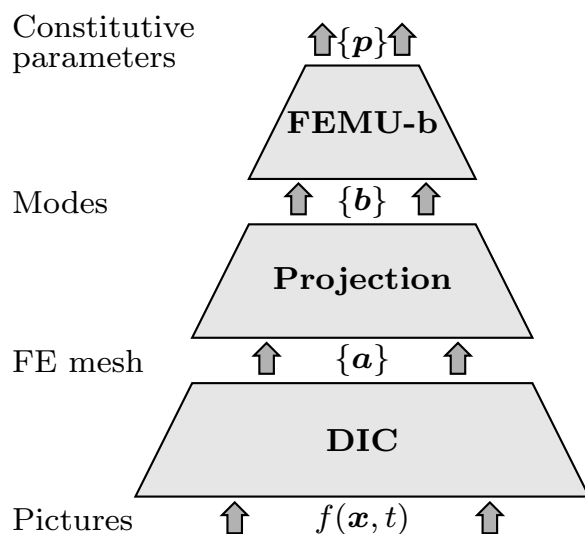
Fig. 12 Data pyramid in image-based identification. As compared to Figure 3, an intermediate level, "modes", has been included. FEMU is now performed by using the degrees of freedom associated with the modes instead of the degrees of freedom associated with the FE discretization. It is therefore referred to as FEMU-b

Assuming that the constitutive model is chosen and the experiment is performed, the first and last spaces are chosen. In any projection between two spaces, information is potentially lost. However, the proposed intermediate spaces are chosen to optimize the efficiency while maximizing the quality of the retained information. The identification methods being iterative, it is always possible to use more singular modes or a finer mesh when getting closer to convergence. The goal of model reduction then is to expedite the iterative process and gain in terms of storage of information [59]. This type of decomposition may also be used when performing 4D mechanical correlation [33] for which the raw data sets are even bigger. Once the constitutive parameters have been obtained, all detailed fields are computed, and residuals are generated to check the consistency of the entire chain (Figure 13).
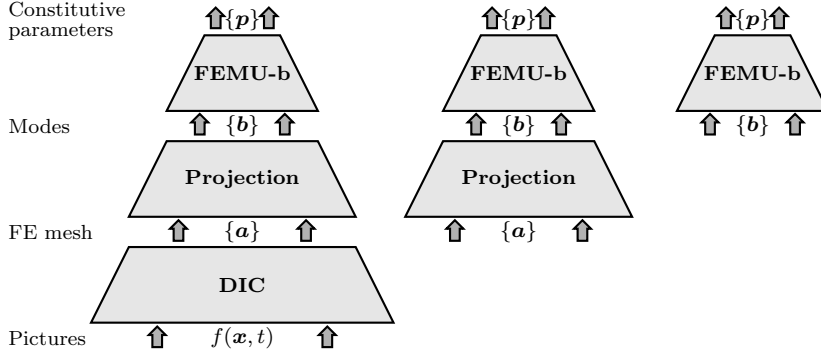
Fig. 13 Model reduction in image-based identification. Moving to the right leads to smaller size problems. Only the intrinsic complexities of the mechanical model are kept, while the data volume has been shrunk to a minimum. Once the constitutive parameters have been obtained, all detailed fields are computed, and residuals generated to check the consistency of the entire chain

5.3 Application of POD to an identification study

The sensitivity fields express the kinematics required to identify the parameters $\{p\}$. Therefore, they are ideal candidates to construct the proposed intermediate space $\{b\}$. In the following, SVD is applied to the sensitivity matrix $[S]$, serving the dual purpose of reducing the order of this significantly sized matrix (Table 3) and simultaneously constructing a basis with which to describe the residual displacement more efficiently. The SVD is applied to a rearranged version of the sensitivity matrix $[S^*]$ that has its rows and columns arranged such that the rows contain all values for each time increment and each parameter. Conversely, the columns are arranged to contain all nodal degrees of freedom. As a result, $[S^*]$ is of size $N_t N_p \times N_a$. Additionally, the square root of the DIC Hessian $[H_a]$ is applied to properly weigh the contribution of each node in the singular value decomposition

$$
\begin{aligned}
[S^*][\sqrt{H_a}] &= \sum_{j=1}^{N_b} [U_j(t,\{p\})][V_j(\{a\})][\sqrt{H_a}] \\
[U_j] &= \Sigma_j [Q_j]
\end{aligned}
\tag{29}
$$

in which the sensitivity fields are decomposed into a series of (nodal) spatial fields $[V_j]$. The singular values Σ_j and the left singular vectors Q_j are combined into temporal functions $[U_j]$, which depend on the sought parameters $\{p\}$. Each combination of $[V_j]$ and $[U_j]$ expresses a principal mode of the sensitivity fields, ordered by significance by their singular values. The approximation quality depends on the number of modes N_b applied in this factorization. The goal is to reduce the weight of $[S]$ and thus only a limited number of modes (*i.e.,* $N_b = 100$) will be kept.

The displacement field correction is written within the reduced basis as

$$\{\boldsymbol{\delta a}\} = \sum_{j=1}^{N_b} b_j \{\boldsymbol{V_j}\} \tag{30}$$

and, as before, the Hessian expressed in modes reads

$$\begin{aligned} [\boldsymbol{H_b}] &= [\boldsymbol{V}]^\top [\boldsymbol{H_a}][\boldsymbol{V}] \\ \{\boldsymbol{j_b}\} &= [\boldsymbol{V}]^\top \{\boldsymbol{j_a}]\} \end{aligned} \tag{31}$$

so that the Hessian for FEMU-b becomes

$$\begin{aligned} [\boldsymbol{H_p}] &= [\boldsymbol{U}]^\top [\boldsymbol{H_b}][\boldsymbol{U}] \\ \{\boldsymbol{j_p}\} &= [\boldsymbol{U}]^\top \{\boldsymbol{j_b}]\} \end{aligned} \tag{32}$$

This SVD is now applied to the test case of Section 3.3. It is recalled that the original sensitivity fields, at the pixel level, contain 11.7×10^9 values (Table 3). When split into FE shape function matrix $[\boldsymbol{P}]$ and nodal sensitivity field $[\boldsymbol{S}]$ this size is reduced to 3.5G and 12.1M values, of which the former is only 0.2% full and can thus be stored efficiently in sparse form. When 100 singular modes are considered in the SVD decomposition the required data footprint of $[\boldsymbol{S}]$ is reduced to 639k and 189k for $[\boldsymbol{U}]$ and $[\boldsymbol{V}]$, respectively. Figure 14 shows the singular values and the L2 error made when truncating. Notably, a truncation to 100 modes results in an approximation error of less than 1 %.
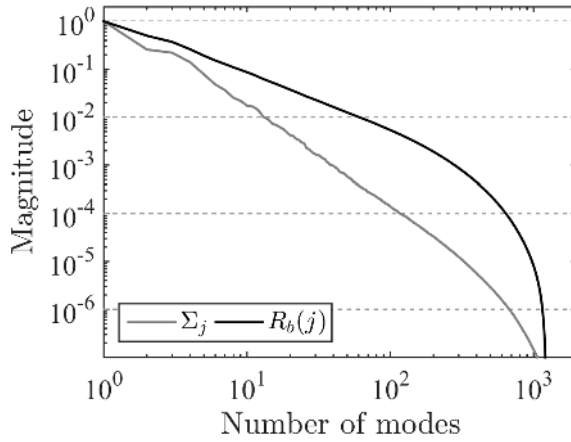

Fig. 14 Spectrum of singular values and L2 truncation error as a function of the number of modes R_b

Figure 3 illustrated two ways of extracting material parameters, and Figure 12 exemplified model reduction in the previous framework. In the following,

three approaches will be compared for the test case of Section 3.3. First, IDIC (Figure 3(right)), which is the most direct way linking sets of pictures with material parameters. Second, weighted FEMU-a (Figure 3(left)), which requires a first DIC analysis and then processes the measured fields to extract the material parameters. Last, weighted FEMU-b that uses the reduced basis instead of the measurement basis (Figure 12). For this last approach, two routes are followed based on the observations of sensitivity fields (see Figures 7). One approach will update the Hessians needed in FEMU-b for each iteration, and the other one, as in DIC, will not update the Hessian.

As in Section 3.3, three different residuals will be reported (*i.e.*, on displacements, pictures and resultant forces, see Table 2). For IDIC, the weighted sum of registration and load residuals is minimized. They reach very low values at the end of the 11 iterations. Figure 15 shows the changes of three different residuals normalized between 0 and 1 corresponding to the smallest and largest occurring residual as reported in Table 2 per residual type (*i.e.*, displacement, force and image). Since all identification cases start with the same initial guess $\{p_0\}$, their initial residuals are all equal. All identification cases reduce their residuals with approximately equal rates. For IDIC, the image residual is one of the optimization target and R_f is the lowest of the four identification routes. For the other cases the residual displacements in nodal form or reduced form are the optimization target. Ergo, for the FEMU-a and FEMU-b cases, the displacement residuals R_u are slightly better when compared to IDIC. It is to be noted that these displacement residuals are comparison with the DIC measurement of displacement fields, which is not necessarily a golden standard. Last the load residuals very similar for IDIC and FEMU-a. They degrade a bit when FEMU-b is considered.

Figure 15 also shows a fourth case, "FEMU-b (secant)" that differs only from the FEMU-b case by not updating the sensitivity fields when the update in the parameters is small (*i.e.*, $\{\delta p\} < 10$ %). Additionally, there is no need to update $[V]$ when the sensitivity fields are fixed. This has the apparent effect that the secant method requires less iterations. However, it is emphasized that the obtained solutions are not strictly equal, which clouds any comparisons of the number of required iterations.
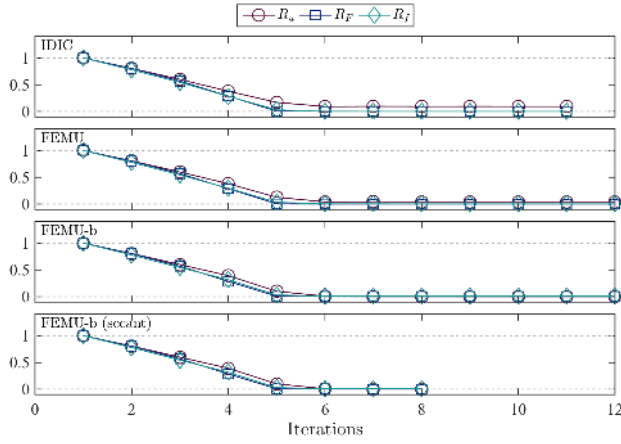
Fig. 15 Changes of the global residuals as functions of iteration number and type of identification method, all values are normalized between the smallest and largest values reported in Table 2

Figure 16 shows the changes of the material parameters for all the studied identification methods. In the present case, the reference corresponds to the converged set of IDIC (Table 1). Because a trust region algorithm was used (*i.e.,* the parameters were not allowed to vary by more than 10 % of their current estimate per iteration) the changes have two regimes. The first one associated with the trust region up to the fourth iteration, and a second one up to converge to a solution for which each parameter variation is less than 0.05 % of the current estimate. Even though the initial guess was not close to the converged solution (Figure 9), only 11 iterations are needed and similar amounts for the other methods.

Some of the material parameters related to isotropic hardening (*i.e.,* $\Delta\sigma_3$ and $\Delta\sigma_7$) are different for the four identification results (Table 1). However, their effect of the calibrated stress-strain remains very small (Figure 9). In particular, FEMU-b leads to more significant differences, yet they do not degrade too much the stress/strain response of the material. The obtained solutions can be used to re-initialize the DIC measurement. This is currently not implemented for FEMU-a and FEMU-b cases. Consequently, the presence of any measurement artifacts in the DIC measurements remains throughout the identification. This improvement is trivial but was excluded to reduce the complexity of the discussion. Further, the fact that the Hessian is updated or not does not significantly change the parameter levels at convergence. Consequently, the secant method may be preferred as only one FE calculation is be needed for each iteration instead of $N_p + 1$ when the Hessians needs to be updated.
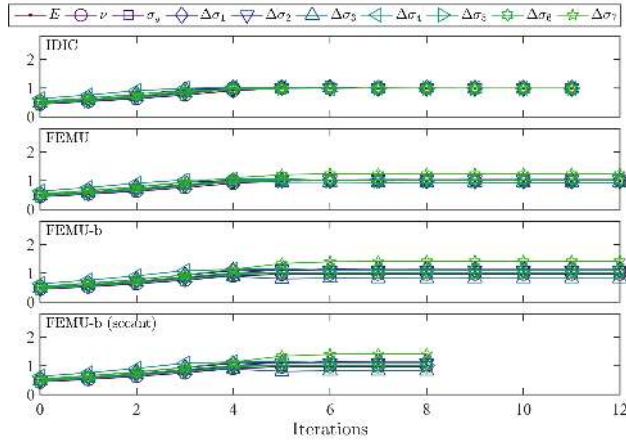
Fig. 16 Changes of the material parameters as functions of iteration number and type of identification method

Last, it is worth noting that the orthogonal modes can also be used as tools to identify where a high spatial or temporal discretization quality is required. Therefore, these modes can be applied to guide auto-meshing routines or time-stepping routines to create FE models that only have a refined mesh where needed for identification purposes.

6 Change of Paradigm?

This last part aims to discuss some new opportunities that can be envisioned with the current state of the art in experimental and computational mechanics.

6.1 High-throughput experimental mechanics

One of the recent trends in biology and chemistry is high-throughput screening [60,61] (a.k.a. combinatorial chemistry, or combinatorial biology [62,63]), meaning that many different products can be tested simultaneously using *e.g.,* lab-on-chips, micro-fluidics devices, robot automated tests. It can be argued that experimental mechanics also, although driven by different opportunities, enters into the era of high-throughput identification.

The new lever for experimental mechanics is not only the availability of new micro-technology, but rather the fact that full-field measurements are now accessible. It means that a single "non-simple" test can be analyzed. What is meant by non-simple may be either a complicated geometry of the sample, or a complex loading history (*e.g.,* multiaxial loading where the different actuators may allow for non radial loading paths), or both. Geometry and loading allow to probe in space and time a multiplicity of stress/strain paths, and hence they may offer a wealth of information from which complex constitutive laws can

be calibrated. It is to be appreciated how such an observation is precisely opposite to the traditional way of designing mechanical tests [64]. Classically, the strategy is rather one test/one measured parameter (or very few), and a wealth of creativity was spent to isolate each parameter with the least coupling to others, to approach a "pure" measurement as much as possible. Additionally, the exploitation of tests was often done as if they were perfect and followed precisely the theoretical design. This called for extreme care, but an obvious limitation is the presence of instabilities that do break down symmetries giving an unbreakable ceiling to the ambition of even the best experimentalists. Full-field measurements offer the opportunity to by-pass those obstacles and limitations. Imperfect tests may be tolerable if the imperfections are measured and accounted for in the modeling [65].

Allowing for complex geometries and loadings naturally opens a wide avenue to transport the methodology to industrial parts, to assemblies of different parts and to service conditions (provided that image acquisition remains accessible). Again, this is a major evolution of experimental mechanics, being thereby endowed with much more credit in decision making.

However, in order for these opportunities to flourish, it is important that they may rely on a sophisticated modeling including relevant nonlinearities, most generally in 3D. 3D may mean in the entire volume as can be offered by tomography, although this remains today an exception. 3D also means surface measurement from stereo-vision where multiple cameras bring their 2D projections that can be combined to generate the observed surface in 3D, usually after a calibration step [1]. The same procedure following a speckle pattern or other markings on the surface provides a measurement of 3D surface motions (in which case, the procedure that exploits DIC is called stereocorrelation or 3D-DIC).

Moreover, tight connections are to be established between the geometry of the model and the actual shape. To this aim, specific strategies based on CAD models have been proposed for stereocorrelation [66, 67]. This allows all projected views (for each camera) to be related to a *unique* 3D model that becomes the reference, in contrast to classical 3D-DIC procedures [2]. Going beyond shape or displacement measurements, so as to include a mechanical modeling necessitates a full 4D (3D plus time) computation. Along the lines that were above detailed in Section 2 for 2D-DIC (or DVC), the entire procedure can be extended to stereocorrelation.

This constitutes a major revolution for experimental mechanics, which ironically is very demanding on the computational side. Now that the methodology is well established with solid foundations and some case studies have been published to illustrate the different aspects, the last barrier to a more systematic deployment of such approaches is the computational cost, and the associated management of the involved "big data." Model Reduction is expected to be very instrumental in addressing this last challenge.

6.2 "Wide scope" experimental mechanics

Images may be more or less prone to misinterpretation. Optical distortions, change in illumination, spurious reflections may require additional corrections between the raw picture acquisition and the image to be used as input to DIC [1, 23] or DVC [68]. The above listed examples are rather straightforward to deal with. Some other cases may be more demanding. In the above discussed stereo-vision, a projection model (in addition to distortion corrections) has to be applied to transform the 3D object into the captured view by any camera, and as mentioned in the introduction, tomographic images are *computed* from radiographs.

The alternative option that is preferred here, and for which the expression "wide scope" is proposed, is to include the processing of the raw experimental data into the modeling chain. This may be seen as an additional ground floor on which rests the entire pyramid shown in Figure 3. For stereo-vision, it consists in adding the projection operator as well as distortion corrections for each camera [69], while in the calibration stage, the speckle pattern on the sample surface has been transported onto the 3D reference object from the observed 2D views in each camera capture (the CAD model gives a good starting information on the sample shape, but generally does not know about the surface speckle [2]). For X-CT, similarly, the generation of radiographs from the 3D object (also called projection) is a direct problem that can be included [70, 71].

However, it may be observed that adding more floors to the pyramid requires the addition of even more data, a trend that may appear to be opposite to what was advocated earlier. This is only partly true. In the first run, it is advantageous to remove entire levels in the pyramid, provided the result is accompanied by its metric. Yet many of these problems are nonlinear and maybe imperfect convergence has been achieved. Thus it is very important after the top of the pyramid has been reached to redistribute the local solution using the transformation matrices constructed first for data reduction. This downflow gives access to residuals, either at a coarse, medium or fine scale. These data now incorporate information coming from the entire pyramid, and thus are expected to be much more reliable (unless a significant model error has been made). For instance displacement fields incorporate a constitutive law, but also balance equation, and boundary conditions. Thus these fields have the potential to correct data that were initially poorly converged. Similarly artifacts and instrumental ill-calibration from the lower levels of the pyramid may be accessed. Ideally, convergence should be declared only when all levels of the pyramid have fulfilled their own satisfaction criteria. The above strategy of model reduction is thus not to be considered as being done once for all and never questioned. On the contrary, it is designed to make convergence easier and faster because aggregated data are used with no prejudice to the underlying physics. However, the essential merit of the pyramid is to have a "holistic" view of an experiment allowing for information flow upward and downward.

6.3 "Data pruning" in experimental mechanics

As big data are distilled into key data, a lot is lost from level to level of the pyramid (Figures 3 and 12). However, "loss" is to be considered with circumspection. Part comes from redundancy and is needed to reduce the noise level so that an accurate estimation of quantities of interest can be extracted from the noisy data. Redundancy is also extremely useful to detect model errors, as the residual fields (*i.e.,* the unexplained part of the data [72]) may be sometimes easily interpreted. For instance, unanticipated cracks can often be revealed in DIC residuals (see *e.g.,* [73]). In such cases, it gives a handle on how to correct or enrich the model. Residuals thus appear to be very rich and informative.

Yet, the ratio between initial and processed data is so large that one may wish to exploit part of this redundancy to acquire less data. *Frugality* may be seen as the ultimate way of fighting big data. However, the potential abundance of data prompts the question of how to choose the "best" data to be kept. At convergence of all degrees of the pyramid, the (correction of the) quantities of interest are given by linear extractors $\boldsymbol{e}^\top$ applied to the initial data $f(\boldsymbol{x}, t)$. For instance the correction δp_i with respect to a nearby reference value p_i^0 can be read as

$$\delta p_i = \sum_{(\boldsymbol{x},t)} e(\boldsymbol{x}, t) f(\boldsymbol{x}, t) \tag{33}$$

The initial data can be seen as a very long (big data) vector gathering all pixel values at all times, and hence the extractor, which is conjugate to it is a similar size object. One iteration is thus a very trivial operation, *i.e.,* a scalar product of two vectors. However, their size is such that it may take time. One natural strategy for pruning is to discard all those pixels having a small weight in the corresponding extractor. This is however not very operational as the discarded pixels are smeared in the entire data set and even keeping 10 % of the original data would presumably cost about the same time and memory as keeping the entire set.

As one climbs up the pyramid, the temporal or spatial correlation in the modes or in the sensitivity fields may prompt for a more efficient selection of data. When considering the question of the sensitivity determining the amplitude of a specific mode or sensitivity field, one may ask the question of the marginal benefit/cost of adding/removing a specific time frame, or a particular region in space (ideally product of indicator functions along each dimension for better efficiency). Because the pyramidal structure can be read as a tree-like structure, pruning (that is cutting a branch), is all the more important as it takes place higher in the pyramid. In some sense, the POD approach can be read as a pruning process where subdominant modes are simply ignored. The efficiency in terms of data handling is obvious. Within the pyramid, one may envision a domain decomposition mesh where pruning will invite one to discard one or several domains. As one gets down to the bottom level, the question may turn to neglect some frames of the picture series. The tremendous amount of data suggests that indeed a very large benefit in terms of data

handling can be obtained. The limit is that uncertainty and accuracy will degrade as more and more data are discarded, but selection based on sensitivity (how much does that particular datum matter?) and on measurement reliability (how uncertain is that particular measurement?), that is the downward and upward flow in the pyramid certainly provides the wisest way to select important data.

In the same spirit, images can very classically be stored in a pyramidal multi-scale format, either redundant or not, using Laplacian pyramids, or any other more sophisticated wavelet representation. Considering only the coarsest part of images, or in other words, pruning out the high frequencies, generally allows for a robust and fast (because of a reduced data volume) process. This is classically done in DIC in order to capture large scale displacements [74]. The price to pay is that uncertainty is large. Hence, a standard procedure consists in progressively restoring finer scales details of images to achieve a more accurate determination of the displacements. Because the multi-scale organization of images is a pyramidal structure different from the one discussed earlier for instance in Figure 13, the two may be combined to endow each extractor with such a multi-scale representation.

Another attractive route is to exploit the residuals in order to focus on the unconverged part of the computation. Very often, in DIC, it appears that the displacement converges fast to the solution in some region of space, whereas other regions are slower and stay far from the actual solution for many iterations. This is usually due to the nonlinearities of the DIC that appear when displacements are far off the solution (and possibly pinned to spurious secondary minima of χ). However, within a global approach it is natural to solve for the entire system even when the bad convergence is limited to a small domain. What is said for space also holds true for time, where some frames may resist the DIC analysis whereas others before or after are more easily accessible. Thus, based on residuals, one may design a local indicator, that allows either to focus only on the part that deserves attention for a better convergence, when moving downward in the pyramid, or alternatively, when progressing upward, that is when the results are used, to give more or less credit to the processed data. The latter aspect plays a similar role as the metric accounting for uncertainty, but rather than being the ideal uncertainty that is achieved at convergence, it would also contain the uncertainty due to ill-convergence as indicated by the local residuals.

6.4 The issue of the constitutive law representation

Up to now, the constitutive law was assumed to fit within a parametric representation. However, quite often, the very choice of such a form is inherited from tradition, or empirical expressions that do not necessarily offer the flexibility or the well-behaved laws that one would enjoy. All these possibilities are associated to a somehow conventional description of constitutive laws. As a result *model errors* may hinder a proper identification. The above discus-

sion suggests that one may also question the very framework within which the constitutive model is expressed. In this spirit, the above identification example was parameterized so that a large freedom was left to the shape of the hardening law (Figure 9).

However, one could be even more ambitious, namely, opening for much more versatility in the description of the material behavior. Considering the success of big data science and machine learning techniques a new route was recently proposed in the emerging context of data-driven computations [75]. This route aims at avoiding as much as possible the use of empirical material models or constitutive equations [76,77]. The idea is to directly make use of data sets consisting of a "large" collection of strain-stress pairs. In Ref. [76] the data-driven nonlinear elastic identification consists in the minimization in the data set space of a distance function between admissible strain and stress fields, defined as the sum of a somehow arbitrary reference strain and complementary energy densities. One nice feature of the approach is to avoid any curve fitting and to use the data as given. In Ref. [77] a machine learning strategy based on a nonlinear dimensionality reduction techniques is designed to extract a constitutive manifold. An appealing aspect of the method is the possibility to extract latent parameters that somehow may correspond to internal variables. An implicit constitutive law results from this analysis.

Let us note that the above two papers do not address the issue related to data generation (and, in particular, the difficulty of measuring local stresses independently from strains). However, the tackled problem — even if reformulated to focus on more experimentally accessible quantities — is challenging and the proposed framework attractive. In this context, image-based techniques and especially DIC offers great opportunities.

7 Conclusion

The emergence and development of full-field measurement techniques has transformed experimental solid mechanics, where huge and ever increasing amounts of data are routinely collected during experiments. The treatment of this experimental information has led to specific and elaborate numerical developments aiming at integrating measurements and simulations into a unique framework. Integrated DIC (or weighted FEMU) techniques, as described herein, merging of DIC (or DVC) and numerical tools traditionally used in computational mechanics (*e.g.,* Finite Element Methods) constitute emblematic examples.

What has also changed in the domain of experimental mechanics is the fact that the tests are designed more carefully because of the high cost of each experiment. The design of the test concerns for instance the sample shape that may become more complex because heterogeneous displacement and strain fields may allow for better calibration of material models [64]. Such type of study can be undertaken thanks to the propagation of uncertainties from image acquisition to the sought material parameters as illustrated herein.

The paper addressed the question of model/data reduction in order to ease and make the exploitation of the experimental information far more efficient. Emphasis was put on the transfer at the different possible levels of reduction (*i.e.,* from pixel, to FE displacement fields, to extracted modes based on sensitivity analyses, and finally to the sought parameters) keeping the basic information coming from the images and its "value" or uncertainty (via the Hessian) along all the reduction steps. This reduction framework is used to propagate the linearization of the problem to be solved across all the scales, thereby significantly reducing the data footprint associated with the determination of the sought parameters.

It is expected that much more could be performed regarding time and memory savings while keeping all the relevant information including uncertainties. The paper discussed some possibly more efficient reduction techniques related to the use of, for instance, modified POD with the metric associated with the covariance matrix for the mode selection or PGD. (The latter in particular will be the subject of future work as a very promising and efficient solution generator.)

Some other largely open issues were discussed in Section 6. Subsection 6.2 was devoted to the question of achieving higher resolution by the treatment of experimental artifacts through "ab-initio" simulations of the acquisition process and its regularization by the input of all reliable information associated with it. Subsection 6.3 discussed the possibility of alleviating the *big data* problem by the selection of a subset of raw data without loosing the information regarding the targeted *key data* and their uncertainties. All these venues also invite the computational mechanics community, as already advocated [76,77], to reconsider the issue of the constitutive law representation (Section 6.4). The dream would be to identify constitutive equations with the least constraints, avoiding all possible empiricism, but incorporating known symmetries and well-established thermodynamic principles [78,79].

On behalf of all authors, the corresponding author states that there is no conflict of interest

References

1. M. Sutton, J. Orteu, H. Schreier, *Image correlation for shape, motion and deformation measurements: Basic Concepts, Theory and Applications* (Springer, New York, NY (USA), 2009)
2. J.E. Dufour, F. Hild, S. Roux, J. Strain Analysis **50**(7), 470 (2015)
3. J. Baruchel, J. Buffière, E. Maire, P. Merle, G. Peix (eds.), *X-Ray Tomography in Material Sciences* (Hermès Science, Paris (France), 2000)
4. E. Maire, P.J. Withers, Int. Mat. Rev. **59**(1), 1 (2014)
5. L. Helfen, T. Baumbach, P. Mikulk, D. Kiel, P. Pernot, P. Cloetens, J. Baruchel, Appl. Phys. Lett. **86**(7), 071915 (2005)

6. L. Helfen, A. Myagotin, A. Rack, P. Pernot, P. Mikulk, M. Di Michiel, T. Baumbach, Phys. Stat. Sol. (a) **204**, 2760 (2007)
7. A. Benoit, S. Guérard, B. Gillet, G. Guillot, F. Hild, D. Mitton, J. Périé, S. Roux, J. Biomech. **42**, 2381 (2009)
8. D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, J. Fujimoto, Science **254**(5035), 1178 (1991)
9. M. Grédiac, F. Hild (eds.), *Full-Field Measurements and Identification in Solid Mechanics* (ISTE / Wiley, London (UK), 2012)
10. J. Oden, T. Belytschko, J. Fish, T. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, S. Yip, Simulation-based engineering sciences. Final report, NFS (www.nsf.gov/pubs/reports/sbes_final_report.pdf) (2006)
11. A. Carpiuc, Innovative tests for characterizing mixed-mode fracture of concrete: from pre-defined to interactive and hybrid tests. Phd thesis (2015)
12. X. Fayolle, S. Calloch, F. Hild, Exp. Tech. **31**(3), 57 (2007)
13. E. Durif, J. Réthoré, A. Combescure, M. Fregonese, P. Chaudet, Exp. Mech. **52**, 1021 (2012)
14. X. Fayolle, F. Hild, Exp. Mech. (accepted) (2013)
15. F. Darema, *Dynamic Data Driven Applications Systems: A New Paradigm for Application Simulations and Measurements* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), pp. 662–669
16. M. Sutton, Appl. Mech. Rev. **65**(AMR-13-1009), 050802 (2013)
17. M. Sutton, F. Hild, Exp. Mech. **55**(1), 1 (2015)
18. M.A. Sutton, N. Li, D. Joy, A.P. Reynolds, X. Li, Exp. Mech. **47**(6), 775 (2007)
19. H. Teyssedre, S. Roux, G. Régnier, A. Tracz, J. Strain Analysis **46**(5), 361 (2011)
20. K. Han, M. Ciccotti, S. Roux, EuroPhys. Lett. **89**(6), 66003 (2010)
21. J. Neggers, J. Hoefnagels, F. Hild, S. Roux, M. Geers, Exp. Mech. **54**(5), 717 (2014)
22. A. Maynadier, M. Poncelet, K. Lavernhe-Taillard, S. Roux, Exp. Mech. **52**(3), 241 (2011)
23. F. Hild, S. Roux, *Digital Image Correlation* (Wiley-VCH, Weinheim (Germany), 2012), pp. 183–228
24. J. Réthoré, S. Roux, F. Hild, C. R. Mécanique **335**, 131 (2007)
25. F. Hild, S. Roux, Exp. Mech. **52**(9), 1503 (2012)
26. A. Tarantola, *Inverse Problems Theory. Methods for Data Fitting and Model Parameter Estimation* (Elsevier Applied Science, Southampton (UK), 1987)
27. J. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, vol. 160 (Springer Science & Business Media, 2006)
28. J. Mottershead, M. Link, M. Friswell, Mech. Syst. & Signal Proc. **25**(7), 2275 (2011)
29. H. Leclerc, J. Périé, S. Roux, F. Hild, *Integrated digital image correlation for the identification of mechanical properties* (Springer, Berlin (Germany), 2009), vol. LNCS 5496, pp. 161–171
30. J. Réthoré, Int. J. Num. Meth. Eng. **84**(6), 631 (2010)
31. F. Mathieu, H. Leclerc, F. Hild, S. Roux, Exp. Mech. **55**(1), 105 (2015)
32. J. Neggers, J. Hoefnagels, M. Geers, F. Hild, S. Roux, Int. J. Num. Meth. Eng. **203**(3), 157 (2015)
33. F. Hild, A. Bouterf, L. Chamoin, F. Mathieu, J. Neggers, F. Pled, Z. Tomičević, S. Roux, Adv. Mech. Simul. Eng. Sci. (submitted for publication) (2016)
34. A. Tikhonov, V. Arsenin, *Solutions of ill-posed problems* (J. Wiley, New York (USA), 1977)
35. D. Lindner, F. Mathieu, F. Hild, O. Allix, C. Ha Minh, O. Paulien-Camy, J. Appl. Mech. **82**(7), 071014 (2015)
36. M. Bertin, F. Hild, S. Roux, F. Mathieu, H. Leclerc, P. Aimedieu, J. Strain Anal. **51**(2), 118 (2016)
37. A. Chatterjee, Current Science **78**(7), 808 (2000)
38. Y. Maday, E.M. Ronquist., SIAM Journal on Scientific Computing **26(1)**, 240 (2004)
39. P. Ladevèze, in *Separated Representations and PGD-Based Model Reduction* (Springer, 2014), pp. 91–152
40. F. Chinesta, A. Ammar, E. Cueto, Archives of Computational Methods In Engineering **17**(4), 327 (2010)

41. M. Grepl, Y. Maday, N. Nguyen, A. Patera, Modélisation mathématique et analyse numérique **41**(3), 575 (2007)
42. M. Barrault, Y. Maday, N. Nguyen, A. Patera, Comptes Rendus Académie des Sciences. Paris. **Ser. I**(339), 667 (2004)
43. S. Chaturentabut, D. Sorensen, Society for Industrial and Applied Mathematics **32**(5), 2737 (2010)
44. D. Ryckelynck, International Journal for Numerical Methods in Engineering **77**(1), 75 (2009)
45. C. Farhat, P. Avery, T. Chapman, J. Cortial, International Journal for Numerical Methods in Engineering **98**(9), 625 (2014). DOI 10.1002/nme.4668
46. D. Néron, P. Ladevèze, in *11th Biennial Conference On Engineering Systems Design And Analysis*, ed. by ASME (ESDA 2012, 2012)
47. P. Ladevèze, *Nonlinear Computational Structural Mechanics - new approaches and non-incremental methods of calculation*. Mechanical Engineering Series (Springer New York, 1999)
48. D. Néron, P.A. Boucard, N. Relun, International Journal for Numerical Methods in Engineering **103**(4), 275 (2015)
49. N. Relun, D. Néron, P.A. Boucard, Computational Mechanics **51**, 83 (2013)
50. O. Allix, P. Vidal, Computer Methods in Applied Mechanics and Engineering **191**(1), 2727 (2002)
51. R. Mahnken, E. Stein, Int. J. of Plasticity **12**(4), 451 (1996)
52. A. Constantinescu, N. Tardieu, Inverse Problems in Engineering **9**, 19 (2001)
53. E. Nadal, F. Chinesta, P. Diez, F. Fuenmayor, F. Deniac, Comput. Methods Appl. Mech. Engrg **296**, 113 (2015)
54. M. Vitse, D. Néron, P.A. Boucard, Computational Mechanics **54**(6), 1529 (2014)
55. L. Gomes Perini, J.C. Passieux, J.N. Périé, Strain **50**(4), 355 (2014)
56. J.C. Passieux, J.N. Périé, Int. J. Num. Meth. Eng. **92**(6), 531 (2012)
57. G. Besnard, H. Leclerc, S. Roux, F. Hild, J. Strain Analysis **47**(4), 214 (2012)
58. J. Neggers, F. Mathieu, S. Roux, F. Hild, in *Photomechanics* (2015)
59. J. Neggers, F. Mathieu, S. Roux, F. Hild, *Reducing Full-Field Identification Cost by Using Quasi-Newton Methods* (Springer International Publishing, 2017), pp. 135–140
60. J. Kononen, L. Bubendorf, A. Kallionimeni, M. Bärlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter, O.P. Kallionimeni, Nature medicine **4**(7), 844 (1998)
61. S.A. Sundberg, Current opinion in biotechnology **11**(1), 47 (2000)
62. P.T. Corbett, J. Leclaire, L. Vial, K.R. West, J.L. Wietor, J.K. Sanders, S. Otto, Chemical reviews **106**(9), 3652 (2006)
63. M. Sarikaya, C. Tamerler, A.K.Y. Jen, K. Schulten, F. Baneyx, Nature materials **2**(9), 577 (2003)
64. M.B.R. Bertin, F. Hild, S. Roux, Strain **52**(4), 307 (2016)
65. O. Allix, P. Feissel, H. Nguyen, Eng. Comput. **22**(5-6), 487 (2005)
66. B. Beaubier, J. Dufour, F. Hild, S. Roux, S. Lavernhe-Taillard, K. Lavernhe-Taillard, Exp. Mech. **54**(3), 329 (2014)
67. J.E. Dufour, B. Beaubier, F. Hild, S. Roux, Exp. Mech. **55**(9), 1657 (2015)
68. N. Limodin, J. Réthoré, J. Adrien, J. Buffière, F. Hild, S. Roux, Exp. Mech. **51**(6), 959 (2011)
69. J. Dufour, F. Hild, S. Roux, Opt. Lasers Eng. **56** (2014)
70. H. Leclerc, S. Roux, F. Hild, Exp. Mech. **55**(1), 275 (2015)
71. T. Taillandier-Thomas, S. Roux, F. Hild, Physical Review Letters **117**(2), 025501 (2016)
72. J. Neggers, F. Mathieu, F. Hild, S. Roux, N. Swiergiel, Int. J. Solids Struct. (in press) (2017)
73. F. Hild, A. Bouterf, S. Roux, Int. J. Fract. **191**(1), 77 (2015)
74. F. Hild, B. Raka, M. Baudequin, S. Roux, F. Cantelaube, Appl. Optics **IP 41**(32), 6815 (2002)
75. B. Peherstorfer, K. Willcox, Comput Methods Appl Mech Eng **291**, 21 (2015)
76. T. Kirchdoerfer, M. Ortiz, Comput. Methods Appl. Mech. Engrg **304**, 81 (2015)
77. R. Ibanez, E. Abisset-Chavanne, J.V. Aguado, D. Gonzalez, E. Cueto, F. Chinesta, Arch Computat Methods Eng (2017)
78. P. Germain, Q. Nguyen, P. Suquet, ASME J. Appl. Mech. **50**, 1010 (1983)
79. J. Lemaitre, J. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge (UK), 1990)