RESEARCH PAPER

Towards a Reliable Uncertainty Quantification in Residual Stress Measurements with Relaxation Methods: Finding Average Residual Stresses is a Well-Posed Problem

M. Beghini[1] · T. Grossi[1]

Abstract

Background In a previous work, the problem of identifying residual stresses through relaxation methods was demonstrated to be mathematically ill-posed. In practice, it means that the solution process is affected by a bias-variance tradeoff, where some theoretically uncomputable bias has to be introduced in order to obtain a solution with a manageable signal-to-noise ratio.

Objective As a consequence, an important question arises: how can the solution uncertainty be quantified if a part of it is inaccessible? Additional physical knowledge could—in theory—provide a characterization of bias, but this process is practically impossible with presently available techniques.

Methods A brief review of biases in established methods is provided, showing that ruling them out would require a piece of knowledge that is never available in practice. Then, the concept of average stresses over a distance is introduced, and it is shown that finding them generates a well-posed problem. A numerical example illustrates the theoretical discussion

Results Since finding average stresses is a well-posed problem, the bias-variance tradeoff disappears. The uncertainties of the results can be estimated with the usual methods, and exact confidence intervals can be obtained.

Conclusions On a broader scope, we argue that residual stresses and relaxation methods expose the limits of the concept of point-wise stress values, which instead works almost flawlessly when a natural unstressed state can be assumed, as in classical continuum mechanics (for instance, in the theory of elasticity). As a consequence, we are forced to focus on the effects of stress rather than on its point-wise evaluation.

Keywords Residual stress · Uncertainty quantification · Bias-variance tradeoff · Ill-posedness · Inverse problems

Introduction

On Ill-Posedness

Relaxation methods identify residual stresses from measurements of the relaxed strains produced by cuts in a given specimen. In a previous work [1], it was shown that they are *ill-posed* problems. This property should not be confounded with that of being *ill-conditioned*. Many ill-posed problems are also ill-conditioned in practice, but the reverse is not implied. Ill-posed problems are often ill-conditioned precisely *because* of their ill-posedness. As a consequence,

it is not a mistake to refer to relaxation methods as ill-conditioned problems—they certainly are—though it would leave out an important piece of the story.

Avoiding its rigorous definition (the interested reader can find the details in [1]), ill-posedness means absence of continuity in the solving operator, that is, arbitrarily small changes in the inputs can lead to enormous changes in the outputs. In most practical cases, ill-posedness arises in the inversion of an integral operator. The latter typically works as a low-pass filter in the spatial domain, so its inversion behaves as a high-frequency amplifier with a gain factor that diverges as the input gets more oscillatory. Since each measurement sample has some kind of independent error, inputs always include some white-like noise that spans the critical frequencies of the amplifier. As a result, the amplification factor through the inverse operator becomes high enough to completely dominate the solution.

✉ T. Grossi
   tommaso.grossi@phd.unipi.it

1   Dipartimento di Ingegneria Civile e Industriale, Università di Pisa, Largo Lucio Lazzarino, Pisa 56122, Italy

The key to understanding the peculiar features of ill-posedness is that those big output changes can be produced by *arbitrarily* small input changes; in other words, no input accuracy can limit the maximum error that may affect the results. The intuitive explanation of continuity is that a function $f(x)$ is continuous "when its graph can be drawn without ever lifting the pen". When a function is continuous, any desired uncertainty on the output value $y = f(x)$ can be achieved by making sure that the uncertainty on the input value x is small enough. Again, the pen metaphor is instructive: if one cannot lift the pen, then an infinitesimal advance in x must produce only an infinitesimal advance in y. If the function is not continuous—for instance, if it has a jump at a point—even an infinitesimal change in the input can produce a finite change in the output. The concept of "error sensitivity" fails; roughly speaking, it is an infinite value. Stretching the concept into the multidimensional setting, if a operator is continuous, a perturbed output is *sufficiently* close to the ideal one when the input is close *enough* to its true value. When continuity is lost, that extremely beneficial property does not hold anymore.

To escape the absence of continuity and obtain a meaningful solution, oscillatory noise has to be filtered out. Discretization—something that is actually required to numerically handle the problem —is a filter itself, as it bounds the number of degrees of freedom (DOFs) in the solution. Unless it is backed up by physical knowledge, discretization introduces a bias in the result; this bias is generally *uncomputable*. The rather trivial exception where the solution is known in advance is of little practical help.

As shown in [1]:

Proposition 1 *Every practical solution of an ill-posed problem comes with an inherent bias-variance tradeoff, controlled by the regularization level.*

Note that the lack of continuity does not prevent the possibility of obtaining a meaningful solution of the relaxation methods. As a matter of fact, the map between relaxed strains and residual stresses is definitely well-understood, owing to the great works of the last decades [2–15]. If anything, the process can be numerically cumbersome due to the number of FEM analyses needed and/or experimentally challenging due to the small strains that have to be measured, but a solution is seldom prevented by theoretical shortcomings—at least with the established methods. Nonetheless, continuity is a fundamental property for the evaluation of the uncertainties that come with the solution.

The biggest concern pivots on uncertainty quantification. Every experimental practitioner knows that inputs uncertainties can be propagated through the given model to obtain an estimate of the solution uncertainty. Regardless of the specific method used (mostly an explicit propagation of a covariance matrix or a Monte Carlo technique), a *finite* quantity is always reported for the obtained residual stresses, as every numerical implementation is implicitly regularized. Then, it is tempting to try to minimize that value, as this would apparently signal a more accurate solution. The optimization of the uncertainty of results has been targeted in many previous works, not only by means of improvements of input data, but also (and quite specifically) through changes in the mathematical post-processing of the experimental measurements.

Unsurprisingly, the uncertainty *seems to be* optimized when the solution DOFs are kept low, an operation which includes low-pass filtering of input and/or output data—that is, limiting DOFs in the frequency domain.

The problem lies in the fact that the uncertainty quantification is inevitably applied to a biased linear operator that is not the original ill-posed one, as the latter actually yields solutions that are completely dominated by an infinite-variance noise. When the bias introduced by the discretization is not considered, the obtained uncertainty actually becomes an observation of the solution variability, leaving out a potentially huge and dominant part of the total uncertainty. Any regularization scheme (including a naive discretization) *implicitly* assumes that the solution has some very specific properties, which may not be present in the true solution. This bias can be arbitrarily high and *cannot* be rigorously estimated without additional hypotheses (some exceptions are discussed later). Section "Uncertain Uncertainties" will expose some potentially dangerous and often counterintuitive consequences of this fact.

Considering the physical aspects of the problem, bias can be arbitrarily high only in theory, as residual stresses are obviously limited by the material strength properties—but, again, that is additional information with respect to the mathematical problem. Anyway, bias can eventually be high enough to completely change the results of a practical measurement.

For the sake of completeness, bias also affects well-posed problems. For instance, the numerical implementation of a problem whose mathematical model is a definite integral (usually a continuous operator) is subject to a bias that depends on the specific quadrature scheme and is bounded by the properties of the input and its derivatives *between* the known (namely, measured) integration points, something that is obviously missing and that requires additional knowledge. The important difference with respect to ill-posed problems is that, as long as it is computationally feasible, in that case the discretization scheme can be refined with a rational increase in DOFs at a level where bias is clearly negligible without compromising on variance. On the contrary, when facing ill-posed problems the best strategy is not to merely increase the DOFs, but to look for a balance. As a consequence, bias could even end up being higher in magnitude than the observed variance.

Motivation

This paper tries to answer the following questions:

1. If uncertainty is potentially infinite in the general case, what does one actually obtain when an uncertainty quantification technique is applied?
2. What are the hypotheses to which the validity of the uncertainties obtained through most common strategies is conditioned?
3. Is there a minimal set of assumptions that allow one to obtain more reliable uncertainty bounds for a given practical problem?
4. How should results and their uncertainty bounds be effectively communicated?

Regarding point 3, we will demonstrate that by giving up the pursuit of point-wise stress values—elusive even from a theoretical standpoint—a well-posed problem is obtained, to which conventional uncertainty quantification techniques can be rigorously applied without introducing any dangerous bias. It is here anticipated that no refined mathematical machinery can overcome an inherent and substantial limit of the underlying model. On the contrary, many advanced techniques turn out to be quite dangerous when applied to ill-posed problems, as their complexity hinders the clarity of their assumptions.

The discussion of the following sections is related to inverse problems arising from residual stress measurements with relaxation methods, whose general mathematical form is as follows:

$$\varepsilon(h) = \int_0^h A(h, z)\, \sigma(z)\, dz \tag{1}$$

where h denotes the cut length and z is used as a spatial coordinate. $A(h, z)$ has been called the *influence function*, *calibration function* or *kernel* of the problem and depends on constitutive properties of the material, on the geometry of the specimen and on that of the cut. Note that the notations $A(h, z)$ or $A(z, h)$ are equivalent in practice.

It must be added that the same considerations hold for all problems described by equation (1), including but not limited to cases where differentiation of a measured quantity is needed. For example, equation (1) includes the formulation of the residual stress identification problem in terms of *eigenstrains* [16].

This matter is anything but mathematical sophistry. Ill-posed problems are rooted in functional analysis, and many of their defining properties are quite difficult to grasp without some background in the subject. However, their practical applicability is mostly an engineering matter. Simply speaking, math demonstrates that the problem is not continuous

and that the total uncertainty cannot be estimated, *unless something else is known or assumed*. It all comes down to know something that mathematics alone does not know, and that is an application-specific piece of knowledge.

That is why, in terms of pure solution accuracy, experienced residual stress practitioners are probably already near-optimal, precisely because (even without a rigorous way to frame the problem) they use their personal engineering judgment to handle the bias-variance tradeoff and choose a rational (if not optimal) solution for their application. On the other hand, we think that there is room for improving the quantification of uncertainties: we are often quite puzzled by reproducibility tests where 95% confidence intervals *consistently don't match* among different experiments.

Probably the main point of this work goes against chasing elusive targets that can be proved to be uncomputable in the general case, so that research efforts can be effectively directed to things that can actually be improved in the present state of the art. Just to mention one of them, the huge quantity of input data provided by full-field methods is very promising [17–20], yet the framework used to process them is still rather tied to the usual strain gauge formalism; some interesting ideas have been discussed in [21, 22].

The paper is organized as follows:

- In Section "Continuum Bodies and Point-Wise Values", the discussion is introduced by analyzing the basic problem of measuring density throughout a solid; this fundamental and scalar property is unrelated to residual stresses, but the discussion serves to highlight general considerations.
- In Section "Preliminaries", the well-established instruments that allow one to obtain the solution of the residual stress inverse problem are briefly recalled. Additionally, a numerical reference problem is presented, to be paired with the theoretical discussions.
- In Section "Uncertain Uncertainties", the hidden biases of the solution process are discussed by using well-established techniques as examples. The main focus is not on the discretization error caused by finite DOFs, but rather on the other deceptive biases that impair a reliable uncertainty quantification. A fundamental question is raised: what is the point of reporting an uncertainty value that is, in turn, highly uncertain?
- In Section "A Constructive Proposal", the findings of the previous Section are used to constructively build a proposal that is aimed at minimizing the amount of *uncomputable* biases, thereby yielding a more reliable uncertainty quantification.

As stated in [1], whose this paper is the continuation, this discussion is not by any means a collection of arguments *against* the use of relaxation methods. Other ones are way

more susceptible to the material structure at the micro-scale, so a comprehensive evaluation of their uncertainty is possibly even more difficult than the present case. Again, the aim of this paper is to provide guidance on how to best squeeze out *reliable* information from experimental results.

Continuum Bodies and Point-Wise Values

The Density of a Mysterious Solid

There exists a much more graspable physical analogue to the ill-posedness of relaxation methods (see Fig. 1). Assume that a company has realized a solid of given shape in a newly-engineered graded material where the density is made variable across the body. To verify the accuracy of the manufacturing process, the solid is given to an external lab, asking for a characterization of the mass density field—namely, the evaluation of the mass density as a function of the position in the solid, defined point-by-point. Nothing else is told to the laboratory, aiming at a truly blind verification.

First of all, laboratory technicians measure mass and volume of the whole specimen. By dividing those two quantities, the *average* density of the whole body is obtained. If the specimen material was known to be homogeneous, the experiment would stop here. Since no assumption can be made, more measurements are needed. The specimen is carefully cut into small sub-components, say, of approximately a hundredth of the initial size, whose masses, volumes, and original locations in the specimen are recorded. This process yields the average density of each piece. Altogether, the target density function is known through a map of its averaged values on some subsets of its domain.

The laboratory head may question whether the achieved resolution is enough for the customer's needs and demands to increase the number of parts. Each sub-component is further divided into smaller sub-sub-components, whose average density is measured through mass and volume measurements. The process starts becoming quite challenging: the absolute precision of the laboratory instruments turns out to be quite poor in relative terms for these very tiny pieces. Moreover, the cut width stops being negligible. Anyway, by taking the uncertainties of the laboratory instruments into account, one can propagate them into uncertainty values of the obtained averaged densities. The smaller the parts, the more those densities are affected by high relative uncertainties, so at some point one must stop cutting, for example when the uncertainties on average densities exceed a certain threshold.

Regardless of the number of cut iterations, what is obtained is always a map of average densities over some given volumes. Here comes the main problem: are the *point-wise* values of the target density function constrained to fall within the obtained uncertainties? Without further assumptions (as in the present case), absolutely not. Indeed, a counterexample is trivial: each average density could arise from a few small but massive particles embedded in an almost vanishing matrix. There is no upper bound to the absolute error that could be made: as long as the massive particles are sufficiently small, their point-like density could be arbitrarily high. Ironically, that is *actually* the case.

At the atomic scale, matter is essentially a collection of tiny supermassive particles—nuclei—whose average density is in the order of 10^{17} kg/m^3, embedded in low-density regions, which are filled only by electronic orbitals. At higher length scales, those two densities average out into
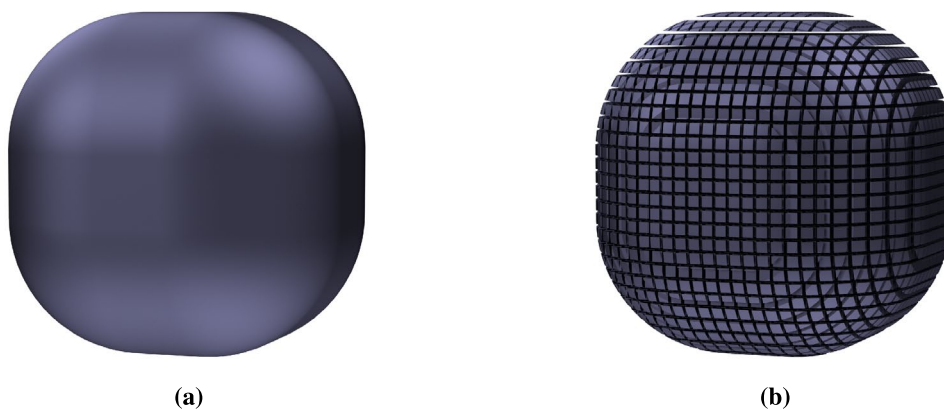

(a)   **(b)**

Fig. 1 (**a**) A mysterious solid whose unknown variable density is to be mapped throughout its volume. (**b**) To this aim, the solid is divided into smaller pieces, whose masses, volumes, and original locations in the specimen are recorded. The cutting scheme is arbitrary, but it profoundly affects the physical nature of the obtained result. By dividing the mass of a tiny piece by its volume, its *average* density is obtained. Unless some additional knowledge is available for assuming that a point-by-point evaluation of density is negligibly varying inside each small piece, *nothing* can be further inferred

the commonly known and used material density, which is fine for many practical applications.

The mathematical structure of this problem is akin to the one of relaxation methods. From a given density function $\rho(\mathbf{x})$, the mass m of a solid Ω can be obtained as:

$$m(\Omega) = \iiint_\Omega \rho(\mathbf{x})\,dV \tag{2}$$

which is the simplest form of equation (1), obtained by setting a $3D$ unit function as kernel. One measures $m(\Omega)$, characterizes Ω, and tries to identify $\rho(\mathbf{x})$. In other words, one wants to find a function through the results of some integrals over appropriate domains. Needless to say, this problem can be shown to be mathematically ill-posed.

What is literally ill-posed—even from a conceptual point of view—is the question itself. What is actually a point-wise mass density? Electrons do not occupy fixed positions in space, while thermal energy is stored in the vibrational motion of atoms in the crystal lattice, so strictly speaking density should be also a function of time. The point here is that density is almost always desired in its *continuum mechanics* definition, rather than in its strictly point-like value, and the two are substantially different. In fact, the first assumes that there exists an averaging volume that is sufficiently small for the averaged quantity to be considered practically point-wise, but sufficiently large to contain enough atoms so that density fluctuations due to individual motions can be neglected. The resulting averaged density becomes a function of spatial position only. The mass densities of materials, used in common engineering practice, fall within this definition, which—despite being openly artificial and missing the physical reality—serves as a valuable practical quantity. In fact, one can effectively use it to determine the mass of solids of any given shape through equation (2), provided that they are larger than the averaging volume; all while forgetting the discrete nature of matter and exploiting the advantages of calculus. In other words, we are trading a bit of physical rigor for invaluable practical convenience.

When trying to measure density in a mechanical component, finding a suitable averaging volume is not a particularly difficult task. For example, a 1 μm^3 cube of iron contains about 10^{11} atoms, while being definitely small for most applications in mechanical engineering. If the laboratory was able to provide the company with a characterization of the mass density averaged on 1 μm^3 cubes, it is quite safe to say that all its practical needs would be satisfied. Note that this piece of information is just a reasonable bet, but it is not implied in any way by the properties of the problem. It is an assumption, not a deduction.

To sum up, the laboratory faces two issues when solving a problem like this one:

- A *theoretical* issue. The solution of the problem equation looks for a point-wise quantity that actually does not exist or has significant conceptual shortcomings. Nonetheless, in many practical cases the point-wise values are not even needed. Density always presents itself inside integrals that yield measurable quantities (e.g., masses or moments of inertia). We are interested in predicting the results of those integrals, not in the mass density itself. Hence, why bother about the values and the (infinitely large) uncertainties of the point-wise density? All we need is an averaging volume that is small enough not to influence the results of integrals of the mass density, such as a refined discretization scheme.

- A *technological/metrological* issue. For a fixed absolute error of the measuring instruments, smaller averaging volumes imply higher relative errors in the results. Clearly, results have a practical relevance if their uncertainties fall below a certain threshold, so the accuracy of the available instruments set some limitations on the minimum achievable averaging volume, which may be significantly larger than the one that would overcome the previous issue.

Eventually, the problem is solved if three conditions are met:

1. there exists an averaging volume that is large enough to exclude the influence of matter discreteness;
2. that volume is small enough to capture any spatially localized effect of the measured quantity that may affect the result being evaluated—usually an integral value such as mass;
3. the accuracy of the available instruments allows one to obtain those integral values with reasonable uncertainties.

When at least one of these conditions is not met, the problem is *unsolvable*. Note that solvability is a user-dependent property and not an absolute one: the same unsolvable mass density measurement of a given material might be solvable (using the same instruments) for someone that will produce bigger parts out of that material, because the negligible influence of density variations at lower scales would allow a bigger averaging volume to be chosen. In our example, a 1 μm^3 averaging volume would miserably fail the needs of the customer if the solid is made out of a nano-architected material whose density is intentionally designed to vary at the nm scale.

Ill-posedness is often *caused* by the *theoretical* issue and *appears* through the *technological* issue, as a coupling between the requirements set by conditions 2 and 3. In well-posed problems, the two requirements can be solved almost independently. In those cases, discretization schemes only marginally affect the sensitivity of results to the measurement uncertainties; measuring function values at more points with a given absolute precision does not decrease the overall precision in knowing the whole function. In ill-posed

problems, the two requirements are instead coupled and define the *bias-variance tradeoff* [1].

Back to Residual Stresses

Let us apply the obtained results to residual stresses and relaxation methods. Again, what is a point-wise value of stress? If the whole problem of relaxation methods was reformulated in terms of inter-atomic forces, it would be well-posed—it has an extremely high number of degrees of freedom, though not an infinite one. Clearly, this approach is completely out of reach and probably more trouble than it is worth, so inter-atomic forces are averaged into the continuum mechanics formalism, which only works at some intermediate length scale.

By some manner, we are spoiled by the well-posedness of classical elasticity. In the ideal world of continuous homogeneous elastic solids, when suitable boundary conditions are applied, a unique solution in terms of *point-wise* stress fields is guaranteed and that solution is stable with respect to input variations. The boundary problem of classical elasticity is well-posed because of (not solely, of course) a rather strong mathematical assumption: a compatible *natural state* exists, in which the point-wise stress and strain fields are null throughout the body with no forces or displacements applied. When that holds, finding point-wise values of stresses and strains from suitable boundary conditions is a well-posed problem. Moreover, because of Saint Venant's principle, boundary loads tend to produce regular stress and strain fields far from their application point, so that we consider regularity as a natural property of the stress field. Unfortunately, the above assumption must be dropped when considering residual stresses, because often it is the opposite of what actually generates them. Indeed, the latter are produced by incompatible permanent strain fields that aren't necessarily regular (in terms of continuity or differentiability), and there are myriads of them that correspond to almost identical boundary conditions for the whole body. Even a simple prismatic bar under a single axial load could present an extremely intricate solution that satisfies compatibility and equilibrium, if the *initial* strain fields are not forced to be null.

Nonetheless, only *integral* effects of point-wise stresses can be observed and measured. We can measure forces, displacements, deformations of a given gauge length, etc., all of which depend on surface or volume integrals of the point-wise stress fields. Many failure criteria are defined—arguably for convenience—on point-wise values of stresses but are already known to cause some problems in relevant practical cases such as sharp notches, and most workarounds rely on switching to related integral quantities such as strain energy release rates [23], the Line Method in the Theory of Critical Distances [24], the Averaged

Strain Energy Density criterion [25], etc. If it is not a bit of a stretch, we claim that stress in a continuum is a useful *bridge* quantity between other measurable quantities that depend on stress integrals over some finite spatial domain.

This is the same condition discussed in the mass density experiment, where it is necessary to give up on the evanescent concept of point-wise values of some intensive quantity and to settle for an averaged version that is refined enough to compute the desired integrals. Nonetheless, the theoretical and technological issues are arguably much more challenging than in the density problem, as the stress is a tensor and the averaging size that satisfies both requirements is likely to be confined in a very narrow interval.

Note that the very nature of relaxation methods forces some features of the averaging process. For instance, in the hole-drilling method, residual stresses are averaged over cylindrical surfaces with *fixed* diameter; for the slitting method, residual stresses are averaged over planar surfaces of *given* length; and so on. Except for the contour method, whose averaging surface could be freely chosen, in principle, other methods allow one to change the averaging size only over a single dimension. Regarding other directions, the bias introduced by the averaging size is simply inevitable and not measurable, requiring careful evaluation by the stress analyst. For instance, if stress fields exhibit fine patterns on the plane where the hole is drilled, nothing in the experiment can account for it, nor can it estimate the error we introduce by assuming a constant averaged value on the plane. It is important to remember that the assumption of constancy in the averaging dimensions imposed by the measurement process itself must be supported by other information and cannot be confirmed retrospectively.

As long as we want to avoid the effects of crystal grain discreteness and orientation, we are forced to average point-wise stresses over a significant number of grains. Then, at the very least, the averaging length is bounded below by the grain size. For a typical structural steel, that would be around $10\,\mu$m [26, 27]. It is worth noting that this limit changes drastically if the measurement is capable of analyzing residual stresses within a grain, while accurately modeling its specific anisotropy in space—that is the case with FIB-DIC measurements [28]; at that point, the lower limit is likely to become the breakdown of the continuum body assumption.

The averaging length is also limited from above by the requirements set by integrals that must be computed to evaluate some material performances of interest, and the problem becomes extremely application-specific: the following discussion should be raised before every residual stress measurement.

By trying to provide a typical example, it is reasonable to assume that residual stresses are often characterized because of their significance for structural integrity. The characteristic length used by common fatigue or fracture criteria is

generally in the order of 100 μm for structural steels [24]. Then, the average residual stress fields over that length become the actual target of the measurement process.

Eventually, there remains the technological/metrological issue. The current state of the art is definitely not too far from the theoretical target—at least in this proposed example. A spatial resolution of 100 μm is not a tough request for a hole-drilling measurement at the present state of the art; for a slitting measurement it is much more of a challenge, although an impressive recent work proves that it is possible [29]. However, we are not simply interested in the average residual stresses over that length, since the whole point of this paper is to dig into their uncertainties. This fact sets additional requirements.

As shown in Section "Uncertain Uncertainties", if one uses the Integral Method and a 100 μm step, the actual averaged solution over 100 μm steps is not achieved; instead, a biased result is obtained. The same applies to Tikhonov regularization. We should not rely on any of the methods that introduce *uncomputable* biases—in the sense that the true solution is needed to compute them. When solving ill-posed problems, obtaining results is deceptively much easier than computing their uncertainties.

A brief side point: as repeatedly remarked in [1], if some physical knowledge suggests that the solution has a specific form depending on few parameters, no bias is introduced at all when the problem is solved with respect to those parameters, so this whole discussion becomes pointless.

Preliminaries

A Reference Example

As done in [1], the discussion is aided by a reference example of an ideal and known but realistic residual stress distribution, hypothetically produced by a shot peening treatment on a thick steel component ($E = 206$ GPa, $v = 0.3$). The chosen residual stress distribution is inspired by the classic textbook of Schulze on shot peening [30]. Expressing the in-depth coordinate z in mm and $\sigma(z)$ in MPa, the following expression was assumed (plotted in Fig. 2):

$$\sigma(z) = \begin{cases} -500 \cos\left[\frac{\pi}{4}(12z - 1)\right] & z \leq 0.25 \\ 0 & z > 0.25 \end{cases} \quad (3)$$

For simplicity, having assumed a sufficiently thick specimen, tensile residual stresses for $z > 0.25$ mm are neglected. In accordance with the ASTM E837 procedure [31] the relaxed strains measured by a standard type-A strain rosette with diameter $D = 5.13$ mm were numerically calculated by assuming a hole with diameter $D_0 = 2.05$ mm. The corresponding influence function was taken from [32].
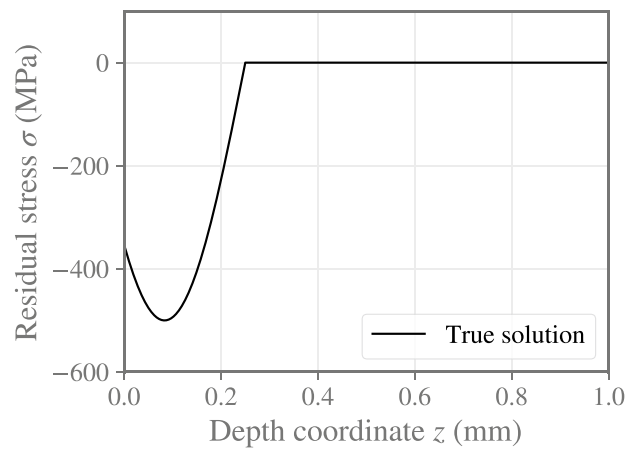
Fig. 2 Equi-biaxial residual stress distribution used as an ideal example to be paired with theoretical discussions

To simulate the typical measurement in state-of-the-art conditions, 100 strain samples were generated, corresponding to a constant drilling step of 10 μm. When measurement errors in relaxed strains had to be considered, a Gaussian noise having a standard deviation of $\varsigma_\varepsilon = 1$ $\mu\varepsilon$ was superimposed on the calculated relaxed strain. A Monte Carlo simulation with 10^4 trials were employed to evaluate the corresponding distributions of results.

Besides its illustrative functions, a toy problem like this one has a much more fundamental purpose. The problem with residual stress measurements is that each method is kind of specialized in measuring residual stresses at a given length scale and for a range of depths from the specimen surface—see the useful chart in the first chapter of [14]. A reference laboratory measurement where most uncertainties are minimized to a negligible level (compared to practical applications) is arguably still missing. As a consequence, uncertainty quantification techniques are almost unverifiable: how can we be sure that a confidence interval evaluation strategy works, if true values cannot be obtained with a significantly higher accuracy? If two techniques yield different values, which one is wrong?

On the other hand, any proposed uncertainty quantification strategy *must* work *at least* in a controlled numerical setting, and here is where the toy problem comes in handy. It cannot prove that an uncertainty quantification works—the real world can always show errors that were not included in our computations—but it can effectively disprove it.

Degrees of Freedom cannot be Infinite

Ill-posedness manifests itself in a very subtle way. In practice, only a *finite* collection of relaxed strain readings is available; as a consequence, at most, the same number of

DOFs can be assumed to represent the residual stress distributions. When the DOFs are made finite, the integral equation in equation (1) becomes a linear system and the operator is reduced to a suitable matrix.

A matrix is a representation of a linear finite-dimensional continuous operator: it sends infinitesimal inputs to infinitesimal outputs. Besides rigorous proofs, it can be directly verified by mere calculations. More interestingly, an input uncertainty cannot be amplified by the linear operator more than the maximum singular value of the matrix [33]. Did this fact overcome ill-posedness? Unfortunately no, even though the last fact is true.

What one observes is a bounded sensitivity of the finite-DOFs residual stress solution with respect to the input error. Unfortunately, the solution may have arbitrarily high frequency contents outside those specific DOFs that were considered in the calculations. Just to give the idea, the residual stress solution could present extremely high oscillations inside a given discretization step, as long as the integral in equation (1) yields the same relaxed strain values. In mathematical terms, discretization is itself a regularization technique. That is why it is often important to choose a basis for the residual stress functions which suitably represent a given experimental case—but that is engineering knowledge, not maths.

One may give up on the possibility to attain the true point-wise residual stress function and settle for its best approximation (in a least squares sense) in the chosen finite-dimensional space. For example, when the well-known Integral Method [5–7] is used, the results become the averaged residual stress along the given calculation intervals. Surprisingly, the following proposition holds, as shown in Section "Uncertain Uncertainties":

Proposition 2 *In general, a least squares approximation in the space of relaxed strains* **is not** *a least squares approximation in the space of residual stresses.*

As a consequence, not only the obtained least squared solution could be arbitrarily far from the true one, *it is not even the best one* among the members of the chosen discretization scheme. In other words, the solution is also biased with respect to the element that best represents the true residual stress fields in terms of least-squares distance.

This last sentence is counter-intuitive and deserves an additional ideal reference example. Let us consider a thick specimen with an equi-biaxial residual stress distribution, assumed to be known with infinite accuracy, and a hole-drilling measurement performed in—say—ten equal steps. The relaxed strains are measured with no errors, too. The setup is so precise that the experiment could be repeated in any point of the specimen and the same relaxed strains are obtained. When the inverse problem is solved by the Integral Method, the same ten residual stress values are obtained for any hole. Given

the accuracy of the setup, it is reasonable to expect that the obtained residual stress values are the average of the real stress distribution along each calculation step: in general, *they are not*. From a practical point of view, the solution obtained with the Integral Method is not an unbiased estimator of the *average* residual stresses in each interval. It is not at all a specific limit of the Integral Method; it is an inherent limit of equation (1).

Nonetheless, every bias vanishes as the discretization scheme is pushed to a high number of degrees of freedom. As that happens, the discretized operator better resembles the original (discontinuous) operator, and this fact manifests itself as an increase of the solution sensitivity to input errors towards unmanageable values, until the solution is practically dominated by noise. A trade-off between bias and variance becomes necessary; when it is not explicitly chosen, a hidden bias is inevitably being introduced.

Practical Solution

To solve equation (1), the residual stress is assumed to belong to the span of a n-dimensional basis $\boldsymbol{\beta} = \left[\beta_1(z), \beta_2(z) \ldots \beta_n(z) \right]$, while relaxed strains are sampled at a finite number of hole depths $\mathbf{h} = \left[h_1, h_2 \ldots h_m \right]$. Linearity is exploited to evaluate the effect of every component of the residual stress basis to each relaxed strain measurement:

$$A_{ij} = \int_0^{h_i} A(h, z)\, \beta_j(z)\, dz \tag{4}$$

so that, for a given residual stress distribution $\sigma(z) = \sum_{j=1}^n s_j \beta_j(z)$, the following relation holds:

$$
\begin{aligned}
\varepsilon(h_i) &= \int_0^{h_i} A(h, z)\, \sigma(z)\, dz \\
&= \int_0^{h_i} A(h, z) \sum_{j=1}^n s_j \beta_j(z) \\
&= \sum_{j=1}^n s_j \int_0^{h_i} A(h, z)\, \beta_j(z) \\
&= \sum_{j=1}^n A_{ij} s_j
\end{aligned}
\tag{5}
$$

By defining an array of measured strain samples $\mathbf{e} = \left[\varepsilon(h_1), \varepsilon(h_2) \ldots \varepsilon(h_m) \right]$ and another array of residual stress components with respect to the chosen basis $\mathbf{s} = \left[s_1, s_2 \ldots s_n \right]$, the usual linear system is obtained:

$$\mathbf{As} = \mathbf{e} \tag{6}$$

Depending on the specific choice of the basis $\boldsymbol{\beta}$ and on the solution strategy of equation (6), several approaches have been proposed to solve this inverse problem. The hidden biases introduced by those procedures are discussed in the following section.

Uncertain Uncertainties

Integral Method

The celebrated Integral Method [5–7] assumes that the residual stress distribution is a piecewise constant function (also known as *staircase* function). In mathematical terms, the interval $[0, h_{max}]$ is divided into a finite number of subintervals $H_j \triangleq [h_{j-1}, h_j]$, and their related indicator functions $\chi_{H_j}(z)$ are used as a basis. Recall that $\chi_{H_j}(z)$ is defined as:

$$\chi_{H_j}(z) \triangleq \begin{cases} 1 & \text{if } z \in H_j \\ 0, & \text{if } z \notin H_j \end{cases} \tag{7}$$

The coordinates s_j of $\sigma(z)$ with respect to that basis are the piecewise constant values of stresses in the chosen subintervals. The solution is obtained as:

$$s_j = \frac{1}{h_j - h_{j-1}} \int_{H_j} \sigma(z)\, dz \tag{8}$$

Apart from the rather unreasonable case of knowing that $\sigma(z)$ *actually is* a staircase function, there exists a representation error $\sigma^\perp(z)$ such that:

$$\sigma(z) = \sum_{j=1}^{n} s_j \chi_{H_j}(z) + \sigma^\perp(z) \tag{9}$$

By plugging equation (9) into equation (1), for a given relaxed strain sample e_i one gets:

$$\begin{aligned} e_i &= \int_0^{h_i} A(h, z)\, \sigma(z)\, dz \\ &= \int_0^{h_i} A(h, z) \sum_{j=1}^{n} s_j \chi_{H_j}(z) + \int_0^{h_i} A(h, z)\, \sigma^\perp(z)\, dz \\ &= \sum_{j=1}^{n} A_{ij} s_j + \int_0^{h_i} A(h, z)\, \sigma^\perp(z)\, dz \end{aligned} \tag{10}$$

There is no guarantee that the second term at right-hand side is null. Depending on the specific h_i, it could be either positive or negative, randomly perturbing the linear system.

Eventually, the coefficients s_j are found by inverting equation (6), but that is not the *true* model of physical reality, which is actually found in equation (10). A slightly wrong model is used both to find the solution *and* to compute its uncertainties. It is worth noting that this error is added to the representation error. Equation (9) already indicates that the best solution would anyway miss the component $\sigma^\perp(z)$. *Additionally*, the values of s_j are biased by a wrong inversion model. The correct one depends on $\sigma^\perp(z)$ and it is not accessible to the analyzer. This happens *even if the input data is unaffected by any kind of error*. In practice, even with perfect measurements, the results would not correspond to

the average values of $\sigma(z)$ over the subintervals H_j. The only way to reasonably avoid this bias is to include the highest number of DOFs, so that $\sigma^\perp(z)$ can be deemed as physically negligible (this point is articulated in Section "A Constructive Proposal"). See Fig. 3 for a numerical example.

Another bias impairs the inversion process. When the regularization level is controlled by the size of the subintervals H_j, the resulting DOFs are almost always less than the measurement points—that is, the matrix **A** is rectangular and has more rows than columns. The linear system in equation (6) becomes overdetermined, and **s** is evaluated in a least-squares sense.

Least-squares solutions of discrete inverse problems come with a fairly overlooked subtlety: the residuals are minimized in the only space where that operation is possible, which is the space of the measured relaxed strains. The least-squares solution $\mathbf{s}^\dagger$ is the one that best approximates **e** in terms of relaxed strains. In geometrical terms, it is the one such that $\mathbf{As}^\dagger$ is the orthogonal projection of **e** on the span of possible relaxed strains.

Unfortunately, the fact that $\mathbf{As}^\dagger$ is the best approximation of **e** does not imply that $\mathbf{s}^\dagger$ is the best approximation of the real $\sigma(z)$, as that is hardly ever the case. In formal terms, the operator does not preserve orthogonality so, in general, an orthogonal projection in the space of measured relaxed strains is not an orthogonal projection in the space of residual stresses. As a consequence, not only the solution
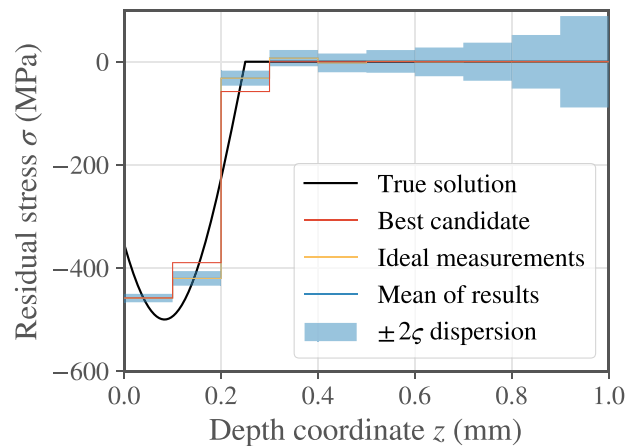
Fig. 3 Solutions obtained through the classic Integral Method. To achieve a reasonable output variance while avoiding additional regularization, the domain is divided into 10 subintervals, each having a 100 μm depth. The yellow solution corresponds to *ideal* errorless measurements and coincides with the expected value of experimental solutions (in blue), while the shading shows the corresponding $\pm 2\varsigma$ scatter band. Confidence intervals would have the same size, although they would be centered on a specific experimental solution. The actual average of the true solution inside each subinterval is shown in red. The red and the yellow curve *do not correspond*. In some subintervals, the actual average is even outside the $\pm 2\varsigma$ results band

is affected by discretization error, it is not even the best one among those spanned by the chosen basis too.

When a technique for uncertainty quantification is applied to the usual linear system in equation (6), it is commonly assumed that a correct model of the physical reality is available. Then, the uncertainties of its constituents can be included. Uncertainties on relaxed strains can be considered in **e**; all kinds of geometrical and constitutive uncertainties, in particular those on the cut depth measurements, can be included in **A**; but the fact that equation (6) is not the true model of the physical process *cannot* be rigorously considered, since its associated error—exposed in equation (10)—depends on the unknown true value of the solution.

Eventually, the statistical distribution of **s** can be completely characterized but it is not possible to know where it is positioned with respect to the values of $\sigma(z)$, neither to their averages over the calculation intervals, unless—as anticipated—some physical knowledge can rule out bias. Moreover, as a consequence of the bias-variance tradeoff, by varying the bias the variance of the solution actually changes and would give rise to different total uncertainties under the *same* estimation framework.

How can uncertainties be trusted, if they are themselves highly uncertain? A, say, 95% confidence interval on s_j is interpreted as a region in which the *ideal* values of s_j—namely, obtained with errorless input data— are included with a 95% probability. Nonetheless, the physical significance of being just an ideal result is rather questionable, as those ideal values are biased by an unknown quantity with respect to the actual average values of $\sigma(z)$ over the subintervals H_j, which are instead quantities tied to the physical reality. The 95% coverage holds only for those ideal values but *not* for the averages of $\sigma(z)$, let alone for the point-like values of $\sigma(z)$.

Power Series Method

Instead of indicator functions, a set of polynomials could be assumed as the basis $\boldsymbol{\beta}$ that spans the space of admissible solutions:

$$\sigma(z) = \sum_{j=1}^{n} s_j z^{j-1} \tag{11}$$

Equation (11) defines the so-called Power Series Method, which was also explored in Schajer's pioneering paper [5].

All the biases discussed in the previous section still hold, though possibly in an even more underhand manner. The solutions of the Power Series Method s_j are not estimators of spatially averaged residual stress values; instead, they are the coefficients of a polynomial—each having a different dimension and measurement unit. As for the integral method, the observed variance only concerns the

variability of the s_j's with respect to their ideal errorless values. Nonetheless, the latter have a rather limited physical meaning: the, say, fifth coefficient itself of a polynomial stress distribution has hardly ever a direct influence on the structural behavior.

As a consequence, the uncertainties of coefficients are frequently combined to obtain the point-like stress uncertainties, which are instead inherently biased by the fact that, in practice, a truncated polynomial expansion never represents exactly the true solution; in other words, there always exists a $\sigma^{\perp}(z)$ component as in equation (10). See Fig. 4 for a numerical example. In fact, no one can have a prior knowledge that a residual stress distribution *is* a polynomial function. There's only a remarkable exception, which holds whenever the residual stress is known to have at most a linear trend. If a slender beam-like component is only subjected to far-field incompatibilities, as in the classical case of a straight beam in a two-dimensional statically indeterminate structure, then Navier's formula holds, and residual stresses are described by a first-order polynomial. Note that in this favorable case the problem is well-posed and probably also sufficiently conditioned, so it is far form the "traps" of ill-posedness.

As a side note, it can be observed that a residual stress distribution could be considered better approximated by a
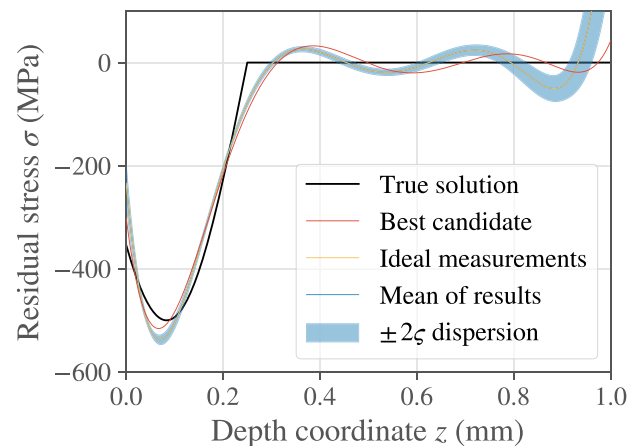
Fig. 4 Solutions obtained through the Power Series Method and a Legendre polynomial basis. To achieve a reasonable output variance without additional regularization, sixth-order polynomials are employed. The yellow solution corresponds to *ideal* errorless measurements and coincides with the expected value of experimental solutions (in blue), while the shading shows the corresponding $\pm 2\varsigma$ dispersion. Confidence intervals would have the same size, although they would be centered on a specific experimental solution. The candidate that best approximates the true solution in a least-squares sense is reported in red. The red and the yellow curve *do not correspond*. The red curve is frequently outside the $\pm 2\varsigma$ results dispersion. Since the physical meaning of the red curve is definitely questionable, one may be tempted to compare the obtained solution with the true one itself, but the situation is even worse

polynomial than by a staircase function, at least in components made by a homogeneous material. The discontinuities of a staircase distribution would require the process that generated residual stresses to produce an intricate distribution of eigenstrains, which is not realistic for common manufacturing processes. Therefore, the disadvantages of the Power Series Method could be slightly compensated by the fact that $\sigma^\perp(z)$ is likely smaller than the one found with the Integral Method for a given number of DOFs. Nonetheless, this fact breaks down as soon as the residual stresses show high gradients that are better modeled by a discontinuity than by a smooth function.

The application of the Power Series Method exposes the core problem of ill-posedness: how many polynomial terms should be included? If too few are employed, the solution is very robust to input noise but highly biased; if too many are included, the contrary holds. Again, bias is not observable, so making this choice by evaluating the solution variance would actually demand a constant function—namely, a zero-order polynomial—to be used.

To the authors' best knowledge, Prime and Hill were the first to acknowledge this effect in a now well-known paper [34], where they made clear the distinction between a "measurement uncertainty" and a "model uncertainty", that is, respectively, variance and bias. They also proposed a heuristic method to evaluate the latter, which, despite not having a proven general validity, certainly raised awareness on the topic (see the recent works by Olson et al. [35, 36], where they call it "regularization uncertainty"), in addition to providing good results in many practical cases.

Smit and Reid [37–39] proposed another heuristic: the best order is the one with the lowest uncertainty *among the ones that have converged*. They actually seized a very important effect in ill-posed problems, called *semiconvergence*, which works as follows. Call *n* the number of DOFs for a given discretization scheme. If the true solution and input noise have markedly distinct features with respect to the given DOFs (for instance, one is very smooth while the other is rough and oscillatory), then the solution is seemingly observed to converge with increasing *n* to a given result, before starting to diverge again for much higher values of *n*. Needless to say, that convergent plateau is a sweet spot for the bias-variance tradeoff, and the strategy of choosing the *least variant* solution among results that are similarly biased is reasonably optimal (it is indeed quite similar to the well-established *L-Curve* criterion [40] and to the *Quasi-Optimality* principle [41]). Two problems arise. The first is that optimal does not mean unbiased: bias is null only if the solution is *exactly* modeled by a polynomial of the chosen order, and the solution may have some features that are only picked by very high-order terms—yet are lost in the noise sensitivity associated with that number of DOFs. The second is that semiconvergence

may not even show up: especially if the residual stress distribution is not well modeled by a low-order polynomial, by increasing the maximum order one observes both a continuously changing solution and an ever increasing noise effect, without that useful "dead-zone".

In a recent article, Brítez et al. [42] explored the use of nonconsecutive polynomial orders with the Power Series Method and proposed another heuristic to pick the best solution. When testing their algorithm and comparing it with other literature proposals, they asked what we believe to be the *right* question. If one constructs a 95% confidence interval, then it should have a 95% frequentist coverage, that is, the obtained bounds should include the true values 95% of the times that the procedure is applied. How does that figure compare with the *actual* coverage of the intervals provided by the methods? In other words, what is the accuracy of the given uncertainty? Results are extremely interesting, as none of the tested methods consistently achieves the right coverage for all datasets. Note that all tested methods include some heuristic estimation of the model error (namely, bias); if the same test was carried out on an uncertainty quantification strategy that does not even include this term, coverage would be shockingly low.

There is a minor statistical flaw in testing the coverage of confidence intervals by counting the fraction of residual stress values of a *single* solution that fall within the computed point-like intervals, as done in [35, 36, 42]. In fact, the entries of the obtained solution are correlated in space, and they are more and more correlated as the regularization level is increased. Confidence intervals come with a probability of including the underlying true value upon *independent repetitions* of the same bounds estimation procedure, something that does not hold for the intervals computed at different depths (being them correlated), while it does for a repetition of the whole experiment.

There is indeed an instructive example. Assume that the solution values are perfectly correlated in space, as when error is due to a constant shift from the true solution. In this case, either all or none of the confidence intervals at different depths include the corresponding true solution values. Nonetheless, upon repetition of the experiment, each confidence interval at a given depth will anyway include the corresponding true value with the desired probability. Clearly, this effect is not so dramatic in residual stress measurements, but in this work we will stick to the rigorous definition of confidence intervals coverage.

Tikhonov Regularization

Tikhonov regularization improves the ill-conditioning of refined discretization schemes, but, in exchange, it introduces a bias in the solution, whose form is anything but trivial. Its rigorous characterization requires some knowledge of

functional analysis (see [43–45]), but an intuitive discussion can be provided by adopting a bit of linear algebra.

The coefficient matrix $\mathbf{A}$ can be written in terms of its Singular Value Decomposition (SVD) [33] as:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \qquad (12)$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{\Sigma}$ is a diagonal matrix having the so-called *singular values* λ_i on the diagonal (usually arranged in descending order). The SVD provides a useful interpretation of how a finite-dimensional linear operator (namely, a matrix) acts on a vector. First, it finds its coordinates with respect to an orthogonal reference system $\mathbf{V}$, then it *amplifies* the i-th component by the singular value λ_i, and finally it adjusts results through a rotation $\mathbf{U}$.

Hansen et al. [46] showed that low i's (namely, high $|\lambda_i|$'s) are related to smoothly varying components, while high i's (namely, low $|\lambda_i|$'s) are associated with rapidly oscillating ones, suggesting that the SVD of $\mathbf{A}$ somehow resembles a kind of Fourier transform in the spatial frequency domain. As a consequence, some terms and concepts in the following have been borrowed from signal processing, such as *low-pass filter* and *high-frequency amplifier* (see [1]).

The ill-posedness of equation (1) causes the singular values of $\mathbf{A}$ to accumulate at zero [45], that is, the more degrees of freedom are present, the more $\lambda_i \to 0$ for high i. In other words, the forward problem (from residual stresses to relaxed strains) is a *low-pass filter* with respect to the basis $\mathbf{V}$.

A useful feature of the SVD is that it allows the least-squares solution of $\mathbf{As} = \mathbf{e}$ to be obtained in a convenient form:

$$\mathbf{s}^\dagger \triangleq \arg\min_{\mathbf{s}}\|\mathbf{As} - \mathbf{e}\|^2 = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^T\mathbf{e} \triangleq \mathbf{A}^\dagger\mathbf{e} \qquad (13)$$

where $\mathbf{\Sigma}^\dagger$ is the diagonal matrix whose elements are given by the reciprocal of each singular value (null entries are left at zero). Its effect is the same: it projects on $\mathbf{U}$, amplifies by $1/\lambda_i$, and rotates by $\mathbf{V}$.

If for high i $\lambda_i \to 0$, then $1/\lambda_i \to \infty$, and that is precisely the curse of ill-posedness. Some components receive an extremely high amplification, and if errors spans those components—which is almost always the case—noise ends up dominating the final result. The inverse problem is a *high-frequency amplifier* with respect to the basis $\mathbf{U}$.

Behind its practical implementation as a *penalized least squares* problem, Tikhonov regularization actually applies a filter on the terms $1/\lambda_i$ [44]. Instead of amplifying by $\frac{1}{\lambda_i}$, classical Tikhonov regularization substitutes that term with by $\frac{1}{\lambda_i + \alpha}$, so that as $\lambda_i \to 0$ the maximum amplification factor does not grow to infinity anymore and for large enough λ_i the solution is negligibly affected by the value of α. The second-order Tikhonov regularization included

in the ASTM E837-20 standard [31] works similarly, although in a slightly more sophisticated mathematical context. This operation comes at the cost of a bias: by filtering high-order components, both noise and true content in that zone get distorted. This bias could be characterized, if some solution properties with respect to those components were known.

Nonetheless, a practical problem lies in the fact that $\mathbf{U}$ and $\mathbf{V}$ have hardly ever some physical meaning in relaxation methods. They are simply some orthogonal bases (in $\mathbb{R}^n$, unrelated with geometrical orthogonality in the real world of the experiment) that allow one to write $\mathbf{A}$ as a SVD decomposition. As a consequence, it is practically impossible to have prior knowledge on how the solution behaves in those bases, such as the sparsity of its representation, so the corresponding bias cannot be known.

Tikhonov regularization is frequently coupled with some criteria on how to *optimally* choose the regularization parameter α. One of the most used strategy is the Morozov discrepancy principle, which is also suggested in the ASTM E837-20 standard [31]. Its details are covered in [1]; here, we should delve into the true meaning of the above-mentioned optimality.

A formal discussion of the Morozov discrepancy principle can be found in Chapter 7 of [44] and is way outside the scope of this paper, but some concepts are useful in the engineering practice too. A choice strategy of a regularization parameter is optimal if the obtained solution is the closest to the true one, among all possible values of the α parameter. If this definition is assumed, the Morozov discrepancy principle is *not* optimal and, to our knowledge, a strategy that deterministically picks the optimal α parameter does not exist. In a statistical discrete setting, the Morozov discrepancy principle is said to be *convergent* and *order-optimal*. Convergent means that the expected value of the root mean square solution error tends to zero as the degrees of freedom of the problem grow to infinity. Order-optimal means that the chosen parameter approaches the optimal one up to a multiplicative constant, as the degrees of freedom grow to infinity; strictly speaking, in the case of the Morozov criterion it requires some technical assumptions that will be omitted here. Note that both these properties are focused on the *asymptotic* behavior with respect to increasing the degrees of freedom. Similar considerations hold for other "optimal" parameter choice strategies.

From a practical point of view, the last paragraph leads to two important consequences:

1. When Tikhonov regularization is used, the highest number of measurement points and the corresponding most refined discretization scheme should always be exploited. That experimental strategy ensures that the regularization parameter is chosen in the best possible

way. It is impossible to know how much the obtained solution is better than another one coming from a less refined discretization scheme, but this ensures that the most favorable conditions are chosen for the method to work. Giving up on experimental data (such as reducing the number of physical drilling steps) *makes no sense* other than being a pure time-saving measure.

2. Because of the previous point, for a given number of degrees of freedom there is no mathematical result that guarantees that the chosen parameter is close to the optimal one, even in a statistical sense. Depending on the problem and on the properties of the true solution, some methods can systematically over- or under-regularize the solution [47]. Sometimes the Morozov discrepancy principle yields a regularization parameter that is blatantly unreasonable, and there's nothing wrong with this outcome. In practice, parameter choice criteria are just sound and well-grounded heuristics.

Strictly speaking, the Morozov discrepancy principle is agnostic to the specific regularization technique applied. For instance, it could be used to automatically choose also the size of the discretization step in the Integral Method or the maximum degree of polynomials in the Power Series Method.

Tikhonov regularization is particularly convenient because it allows one to easily play with its (single) regularization parameter and explore the bias-variance tradeoff without having to recompute the entire discretization scheme, thus avoiding a computationally burdensome procedure. Then, by mixing personal judgment and the outputs obtained by various parameter selection criteria, the analyst can choose a suitable solution. Nonetheless, a practical quantification of uncertainty is still an open—possibly unsolvable—problem. When uncertainty is not that big of a concern (for instance, in image deblurring, when only a *nice* reconstruction is needed), the upsides of this approach are particularly advantageous, but when its results are used to make safety-critical decisions (as when considering the effect of residual stresses in a failure assessment), it should be handled with *extreme* care.

As a side note, recall that the beneficial mathematical properties of the Morozov discrepancy principle hold under the assumption of *deterministically known* variance of the input, which is instead a not-so-trivial piece of information in the real world of measurements. The ASTM E837-20 provides a technique that *estimates* the input variance, but the variance of the estimated variance itself can be a cause for concern. See Figs. 5 and 6 for a numerical example. The interested reader can find a statistical analysis in Appendix A.

(a)   **(b)**
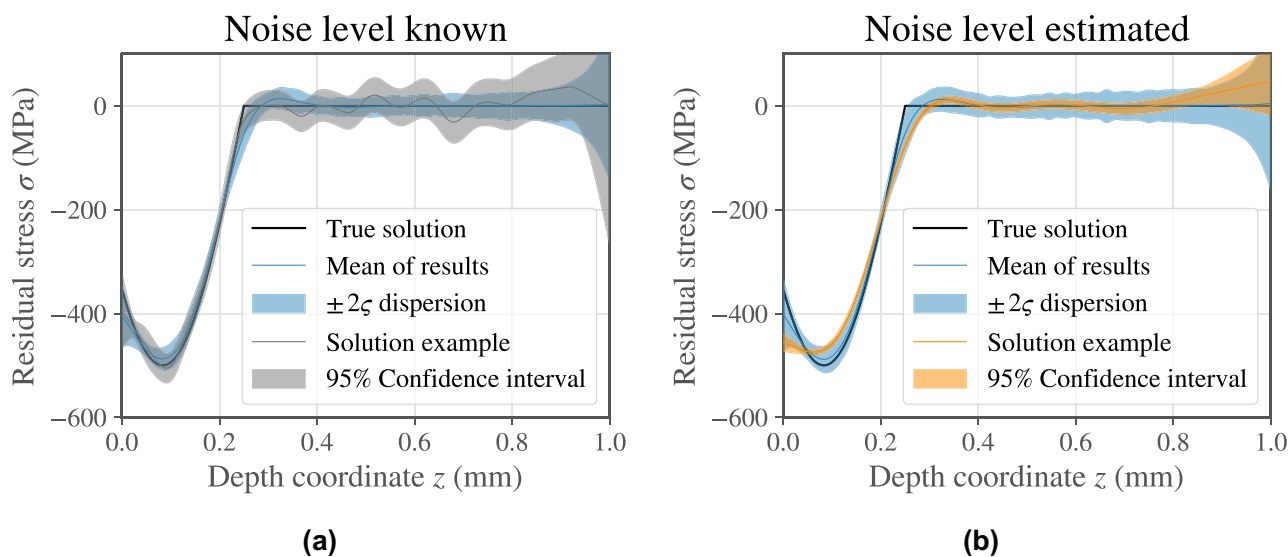
Fig. 5 Solutions obtained through Tikhonov regularization and the Morozov discrepancy principle. Since DOFs are not limited by this approach, the best possible solution coincides with the true one, and it would be achieved with perfect measurements. A single realization of a solution is also reported, together with its computed 95% confidence interval. (**a**) Noise level known a priori and equal to $\varsigma_\varepsilon = 1\,\mu\varepsilon$. (**b**) Noise level estimated through the procedure of ASTM E837-20 standard. Note that results are centered along a solution that is biased with respect to the true one, especially in areas where features vary on a small spatial scale. For example, the effect is rather pronounced at the surface. The fact that the noise level is not known a priori adds another layer of variability to the results, but it can have an apparently opposite effect on a single solution: when the noise level is overestimated, the *same* solution needs more regularization, and that *decreases* its perceived variability
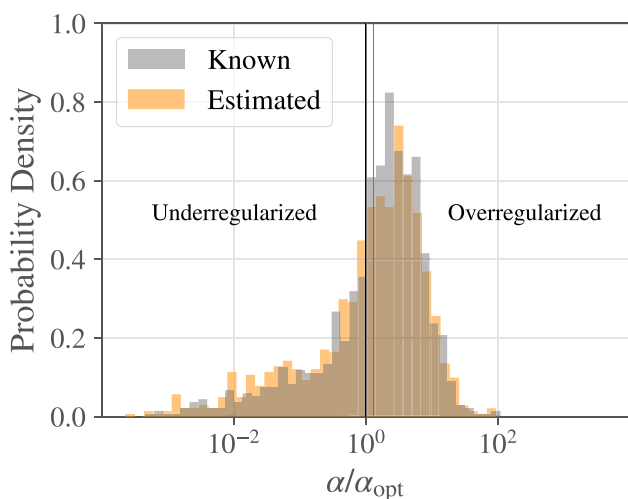
Fig. 6 Distribution of the ratio between the regularization parameter α chosen by the Morozov discrepancy principle and the optimal value corresponding to the minimum solution error in a least squares sense. The two cases of known vs. estimated noise level are reported. Note that the picked value of α is slightly overregularized on average, but prone to some cases of severe underregularization

Machine Learning Applications

The development of statistics and data science provided residual stress engineers with a plethora of tools to tackle the inverse problem of relaxation methods. Neural networks have been employed [48, 49], though mainly for the correction of non-linear effects on the measurements, such as plasticity. Anyway, the bias-variance tradeoff is a well-established concept in the context of neural networks [50], so the design and training of the network is generally planned accordingly.

Recently, a technique called Gaussian Process Regression (GPR) was employed to solve this inverse problem [51, 52]. A thorough analysis of its mathematical properties would require a whole separate paper, so the interested reader is referred to an impressive online article [53] and to a thorough yet not too technical review [54]. Nonetheless, its basic principles are discussed hereafter.

In summary, Gaussian processes are the functional equivalent of multivariate random variables: they establish a probability distribution over possible functions. The functions are distributed around a *mean function*, and their point-like values are correlated in space through a *covariance kernel*. The more values that are far from each other are positively correlated, the more the function is expected to be smooth; in other words, a function can hardly show abrupt changes if each time the points "immediately after" are highly correlated with the current one. The mean function is often assumed to be identically null, while the covariance kernel is assumed to follow a given mathematical law, controlled by a finite set of

values called *hyperparameters*. A nice feature of GPR is that it provides a Bayesian framework to incorporate the observed measurements and obtain a posterior joint distribution of values, even outside of measurement points, from *prior* information about the Gaussian process. Then, one can propagate this distribution through any model of choice, obtaining the distributions of other quantities that depend on it.

When applied to residual stress measurements, a GPR application is straightforward: the relaxed strains (or the displacements, in the case of the Contour method) are assumed to be modeled by a Gaussian process; given the measured strains, a posterior distribution over the space of relaxed strains is obtained, which is propagated through the inverse operator to find the distribution of residual stresses. Being Gaussian, the mean of the posterior distribution is also the corresponding *maximum a posteriori* (MAP) estimator. Since the covariance kernel usually enforces smoothness, a sample from the posterior distribution typically appears as a filtered version of the input strains. Here is where regularization comes into play: highly oscillating inputs are filtered, so the inverse problem variance is tamed. The more the *prior* assumption of a covariance kernel incorporates smoothness assumptions, the more this will be reflected in the posterior distribution through Bayesian inference, and the problem will be regularized at the cost of bias—since the true solution may not be as smooth as expected.

If the input relaxed strains are *actually* a Gaussian process with that specific covariance kernel and hyperparameters, everything works, but, as usual, that piece of information cannot be available in a practical case. As a consequence, the procedure is carried out in two steps.

First, the hyperparameters of the assumed kernel are usually *inferred* from data via *maximum likelihood estimation* (MLE): the intrinsic smoothness level of data is chosen as the most statistically plausible given the observed measurements and the measurement error, if known. This rationale is the same of the Morozov discrepancy principle, although the math is a bit more sophisticated. Since there is measurement noise, it is more likely that the measured input strains be a slightly corrupted version of smooth true data, than them being very accurate observations of actually rough true data.

The regression phase is consequently carried out by assuming that the true data is a Gaussian process with that covariance kernel having the estimated hyperparameters.

Like the term says, the estimation of hyperparameters is done on a *maximum likelihood* basis, that is: among all possible models, the one under which the probability of obtaining the observed data is maximized (with respect to other models) is selected. Three concerns must be raised, as they actually hold for MLE in general, also when it is applied in other frameworks:

- MLE looks for the "best" model among the ones indexed by the hyperparameters, but it provides no measure of the absolute probability of it being actually the true one. Although typical covariance kernels are fairly general, data may depart from the implied assumptions—for instance, discontinuities would be quite an issue for many established kernels.
- MLE is likely to be statistically optimal in some sense, which means that the expected value of the results is probably closer to the true solution than with other choices of hyperparameters. However, again, optimal does not mean unbiased: it is the best with the given setup, but the bias introduced by the smoothness assumptions is active and affects the uncertainty quantification.
- MLE is a statistical estimator, not a deterministic truth, in the sense that the values of the hyperparameters are uncertain, and ill-posedness is very punishing with respect to that kind of uncertainty. The uncertainty of residual stress can be computed by assuming that the posterior data distribution is exact, but there are some less-probable-though-not-so-unlikely hyperparameters (corresponding to lower data smoothness) that correspond to much greater variances in the residual stress solution; inevitably, uncertainty quantification is, in turn, quite uncertain.

For the sake of completeness, it is possible to avoid MLE and evaluate the posterior data distribution by marginalizing over the whole domain of hyperparameters (for instance, see Chapter 5 in [55]), although the operation is overly complex for the scope of a residual stress measurement. Anyway, that would leave the choice of a kernel family as the only remaining bias, at the cost of a rather sizable variance.

A common argument in favor of GPR claims that by using a so-called *universal kernel*, every continuous function can be approximated with infinite precision by a suitable mean of a posterior distribution predicted by GPR (formally, the reproducing kernel Hilbert space of a universal kernel is dense in the space of continuous functions). In simple words, regardless of the given kernel and hyperparameters, *when fed with sufficient data* the mean of the posterior distribution predicted by GPR can converge to any given solution. However, this concerns just an asymptotic behavior with respect to the size of the dataset: actually, for a *given* dataset, kernels inject a hugely influential prior in the Bayesian inference and favor some solutions over others at the expense of bias. Knowing that with increasing data the GPR estimate would converge to the truth—possibly in contrast to prior smoothness assumptions—is of little practical help, as convergence may be extremely slow [56]. Like the Morozov discrepancy principle, mathematics just states that having the greatest amount of data puts us in the most favorable conditions for GPR to work, though we don't know how favorable they actually are.

Anyway, it is fair to say that GPR is likely to perform slightly better than the Tikhonov-Morozov approach, because of a combination of three factors: 1) since true relaxed strains are the output of a low-pass filter (equation (1)), a general smoothness assumption is reasonable; 2) GPR employs that prior assumption to build a MAP estimate of relaxed strains that is then propagated through an *unbiased* model (except for the bias induced by the discretization scheme, which should be kept as refined as possible); 3) GPR estimations are seemingly not so affected by lack of knowledge about the input noise. See Fig. 7 for a numerical example.

To conclude, it is important to recall that ill-posedness cannot be solved with mathematical/statistical tricks; if anything, maybe what is missing right now is a convenient way to formally incorporate the physical knowledge about the engineering problem into its equations, and that is where the authors think should be directed our efforts. The following section is in attempt at facing this problem.

A Constructive Proposal

Main Idea

After having analyzed the shortcomings of the available techniques, it is now time to apply the conclusions of Section "Continuum Bodies and Point-Wise Values" in order to formulate a proposal on how relaxed strain measurements should be processed to obtain a residual stress distribution. The proposal is based on the fact that ill-posedness comes from the quest for *point-wise* stress values, while most likely what is requested is some *integral* value of residual stresses along a given "problem related" length. Therefore, it is suggested to give up on finding point-wise residual stresses with their uncertainties and to directly look for properly averaged quantities instead.

The procedure is grounded on the existence of three characteristic lengths, which depend on the particular phenomenon to be engineeringly faced and on the specific residual stress measurement:

1. An averaging length l_σ. This is the length at which residual stresses manifest their effects, being it required for fatigue models of crack initiation, fracture mechanics approaches for crack growth, distortion analyses, studies of tribological properties, or any other application of choice. In absence of application-specific information, l_σ is a length at which we *accept* that results will be averaged. Everything that happens at a scale under l_σ is assumed not to affect the considered phenomenon, and it is deemed as not technically relevant.
2. A measuring step l_m. This is the maximum resolution that can be technologically achieved by the adopted

Noise level known
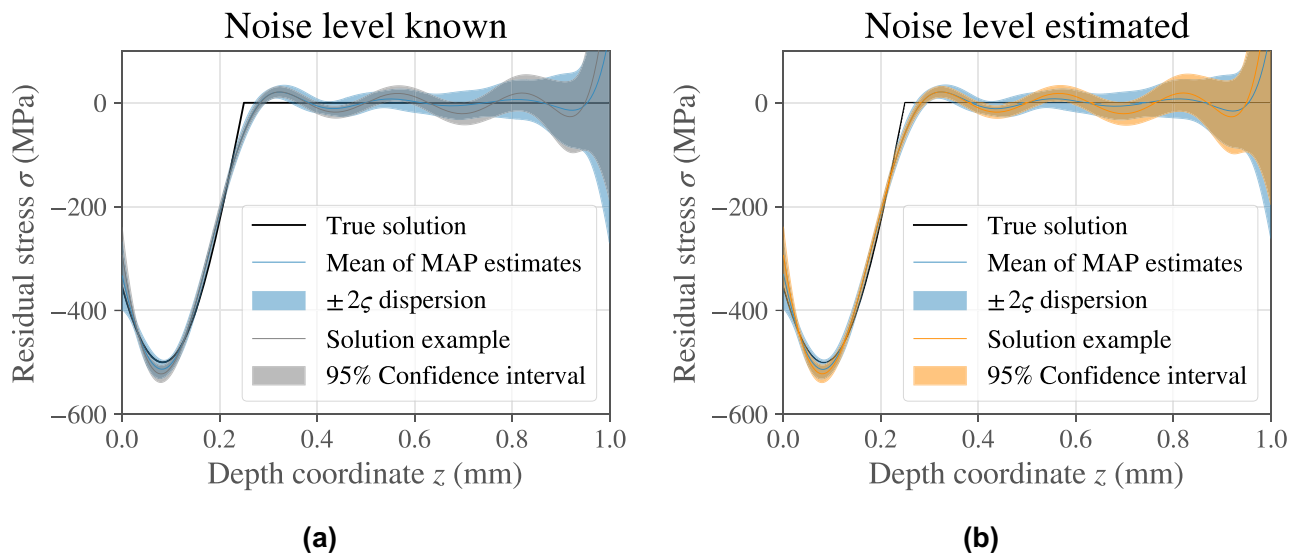

(a)

Noise level estimated

(b)

Fig. 7 Solutions obtained through Gaussian Process Regression applied to relaxed strain data. Since DOFs are not limited by this approach, the best possible solution coincides with the true one, and it would be achieved with perfect measurements. A single realization of a solution is also reported, together with its computed 95% confidence interval. (**a**) Noise level known a priori and equal to $\varsigma_\varepsilon = 1\,\mu\varepsilon$. (**b**) Noise level estimated by MLE as an additional hyperparameter. Note that results are centered along a solution that is biased with respect to the true one, especially in areas where features vary on a small spatial scale. For example, the effect is rather pronounced at the surface. The fact that the noise level is not known a priori has a seemingly negligible effect both on the dispersion of results and on a single realization of a solution. That is likely due to the fact that a higher estimated noise comes with a smoother kernel but at the same time it decreases the statistical power of observations, so the overall effect is less dramatic than with Tikhonov-Morozov approaches

relaxation technique, *regardless* of the variance that would come with that resolution. In simple terms, it is the minimum step of the cutting process that can be performed consistently and whose effect can be reasonably captured by the chosen relaxed strain measuring method.

3. A continuum length l_c. This is the length scale where the chosen model of continuum—under which the influence functions for residual stress evaluation are deducted—starts failing at representing the local material properties. Typically, when measuring macro-stresses in a metallic component, considering that the models are based on homogeneous isotropic linear elastic materials, l_c is in the order of the average grain size.

If $l_c > l_\sigma$ the mathematical framework used for the elaboration is inadequate, as the residual stress distribution would be significant at a length scale that is not modeled by a homogeneous isotropic material. Moreover, for the proposed procedure to work, l_m must be smaller than l_σ (ideally, one order of magnitude less).

The main result is the following. If l_m is approximately equal to or less than l_c, the procedure yields virtually exact confidence intervals on residual stresses averaged along l_σ. Otherwise, a technological/metrological limit of the process appears, which has to be reported as a result. To sum up, the following conditions are necessary:

$$l_m \le l_c \ll l_\sigma \tag{14}$$

When equation (14) holds, the procedure exploits the physical/technical knowledge that comes with l_c. If everything that happens below the l_c scale falls outside the chosen continuum model, at that length scale the solution cannot be biased—the corresponding true solution does not even exist at lower length scales. As a matter of fact, it is the most refined solution one can get that still satisfies the continuum hypothesis.

This fact is used to carry out an ill-posed inversion while avoiding any bias, at the cost of a likely huge variance. To minimize the uncertainty of the uncertainty, what can't be known has to be minimized; in this case, bias is the issue. Once that unbiased results are obtained, their variance is subsequently reduced by averaging them along l_σ.

The trick here is that if the solution to an ill-posed problem is stripped out of *uncomputable* components, from that point onward every uncertainty can be safely propagated to compute the uncertainties of derived quantities. Ill-posedness did not disappear: the results validity is still conditioned to the value (and to the existence) of l_c. On the other hand, it is arguably one of the weakest, less restrictive, and most general hypotheses that can help one tame ill-posedness.

Formalization

For the sake of simplicity, a one-dimensional setting is considered, such as for most relaxation methods. The case of contour method easily follows by extending line integrals to surface integrals, without significant conceptual effort.

If the averaging length l_σ has been chosen, an averaged residual stress σ_{av} at a point ζ such that $\frac{l_\sigma}{2} \leq \zeta \leq h_{max} - \frac{l_\sigma}{2}$ is defined as:

$$\sigma_{av}(\zeta) \triangleq \frac{1}{l_\sigma} \int_{\zeta - \frac{l_\sigma}{2}}^{\zeta + \frac{l_\sigma}{2}} \sigma(z)\, dz \tag{15}$$

Quite astonishingly, the following proposition holds [57]:

Proposition 3 *The problem of finding $\sigma_{av}(\zeta)$ from $\varepsilon(h)$ is well-posed.*

In Appendix B, a formal proof is proposed. In other words, small changes in $\varepsilon(h)$ cannot give rise to arbitrarily high changes in $\sigma_{av}(\zeta)$; the uncertainty of $\sigma_{av}(\zeta)$ is bounded by the uncertainty of $\varepsilon(h)$. It is worth noting that this proposal is not an original mathematical approach, as it falls under the broader class of *mollifier* methods for ill-posed problems [57–60]. However, it is technically useful if the key quantity l_σ actually exists and it is correctly chosen. Being well-posed, results are brought to convergence by successive refinements of the discretization scheme. As intuition suggests, the ill-conditioning of the problem is controlled by the value of l_σ; the greater l_σ is, the lower the result is sensitive to input errors, as noise gets filtered by the averaging integral in equation (15).

In other words, with respect to the original ill-posed problem of finding point-like values of residual stresses, l_σ acts as a meaningful regularization parameter, and it introduces a bias with respect to the *true* solution of the original problem—which is actually deemed as purely theoretical. In fact, there is no interest anymore in the point-wise values of residual stresses, as the regularized (average) solution is *itself* the desired quantity, with a precise and useful physical meaning. By contrast, a truncated polynomial expansion or a specific α in Tikhonov regularization *do not have a clear physical meaning*, so they are interpreted as *biased surrogates* of the true point-wise residual stresses that have to be integrated later, to apply a given engineering model.

Step 1 – Minimizing the Unknown

To evaluate the integral in equation (15), the values of $\sigma(z)$ must be known at a more refined length scale than l_σ, and here is where l_m comes into play. In practical terms, as many measurement steps are performed, then $\sigma(z)$ can be found by inverting the problem through the classical Integral Method

with a square coefficient matrix, *without any other form of regularization or statistical averaging*. By doing this, the bias-variance tradeoff gets totally unbalanced towards variance, which most likely will dominate over true information in the obtained solution.

Note that bias has not been strictly zeroed, since discretization—no matter how refined—always includes some regularization. The good part is that the existence of l_c guarantees that bias is practically null. As a matter of fact, it is unreasonable to think that stresses obtained under a continuum body assumption have meaningful content on a length scale where the continuum assumption itself fails. As argued in Section "Back to Residual Stresses", these length scales are now technologically feasible (or close to) in many relaxation measurement techniques.

Even if one was unsure about l_c, there is another argument in favor of this strategy: it is the one that involves the least possible amount of bias in uncertainty quantifications, unless additional physical knowledge is available. In other words, even in cases where the assumptions about l_c are not verified exactly, it is probably the best strategy in terms of accuracy of the uncertainty quantification.

Step 2 – Reducing the variance

An *unmanageably noisy* but *unbiased* solution has been obtained; it will be called the *unbiased solution* in the following. The unbiased solution is still not a final result, as the residual stresses at this refined length scale are not useful; instead, an averaged value on the length l_σ is requested according to equation (15).

Even if the target is always to find averaged values of residual stresses, it is worth noting the subtle but significant difference with respect to the classic Integral Method. In the latter, $\sigma(z)$ is *assumed* to be constant on intervals of given length. Here, an unbiased solution is *first* obtained with minimal assumptions, *then* its data points are aggregated. The great advantage lies in the fact that the uncertainty of the unbiased solution can safely be propagated through equation (15) to find the uncertainty of the averaged solution. See Fig. 8 for a numerical example.

As a result, equation (15) significantly *reduces* the variance of the averaged solution, with respect to the unbiased one. As the latter is likely to present many anti-correlated samples—ill-posedness gives rise to oscillatory solutions—it is important to propagate uncertainties rigorously, to exploit the advantages of averaging anti-correlated quantities.

If measurement errors are independent and have variance ς_ε^2, their covariance matrix is $\mathbf{\Sigma_e} = \varsigma_\varepsilon^2 \mathbf{I}$. When equation (6) is applied to solve the inverse problem, $\mathbf{\Sigma_e}$ is propagated to the covariance matrix $\mathbf{\Sigma_s}$ of the unbiased solution $\mathbf{s}$:

$$\mathbf{\Sigma_s} = \mathbf{A}^{-1}\, \mathbf{\Sigma_e}\, \mathbf{A}^{-T} \tag{16}$$

Fig. 8 (**a**) Solutions obtained through the proposed procedure. A $l_\sigma = 100\,\mu m$ is assumed and clearly reported. The problem is linear and well-posed. The mean of the solutions coincides with the averaged true solution (which is also achieved by ideal measurements). The domain appears shrunk by the averaging process, but the first sample of each solution is actually the average of the first $100\,\mu m$ from the surface. Since nothing has been assumed concerning smoothness, solutions are quite rough; in return, their confidence intervals are exact. (**b**) Coverage of 95% confidence intervals over repetitions of the experiment, compared among the proposed method, Tikhonov-Morozov, and GPR. Only the first consistently achieves a 95% coverage, as it is theoretically guaranteed to do so. This is possible due to the fact that a physically meaningful averaged solution is being estimated, not the point-wise true solution. Other strategies fail in areas that are more prone to the effects of bias. The true solution is reported out-of-scale in the background, for reference

Note that $\boldsymbol{\Sigma_s}$ is, in general, a full matrix—that is, it is not diagonal like $\boldsymbol{\Sigma_e}$ anymore.

Equation (15) must be discretized with a suitable numerical integration scheme. Since the discretization scheme of $\sigma(z)$ is forced by the Integral Method solution $\mathbf{s}$, basically one just needs to apply a rectangular rule of numerical integration to equation (15) by averaging groups of stress intervals together. For example, say that a $l_\sigma = 50\,\mu m$ has to be obtained from a $l_m = 10\,\mu m$; then, a rolling average of 5 consecutive intervals is applied to the unbiased solution:

$$\mathbf{s_{av}} \triangleq \boldsymbol{\Theta}\,\mathbf{s} \tag{17}$$

where:

$$\boldsymbol{\Theta} \triangleq \frac{1}{5}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & & & \\ & 1 & 1 & 1 & 1 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{18}$$

Eventually, $\boldsymbol{\Sigma_s}$ can be propagated through equation (17), obtaining the final covariance matrix of $\mathbf{s_{av}}$:

$$\boldsymbol{\Sigma_{av}} = \boldsymbol{\Theta}\,\boldsymbol{\Sigma_s}\,\boldsymbol{\Theta}^T = \boldsymbol{\Theta}\,\mathbf{A}^{-1}\,\boldsymbol{\Sigma_e}\,\mathbf{A}^{-T}\,\boldsymbol{\Theta}^T \tag{19}$$

Equation (19) allows one to obtain the covariance matrix of the averaged residual stresses from the input covariance matrix. Two points should be raised:

- The input covariance matrix $\boldsymbol{\Sigma_e}$ could be unknown. This issue is ubiquitous in uncertainty quantifications, and an accurate characterization of the measurement process is a responsibility of the laboratory. The input uncertainty could be estimated from data, but—for instance, as shown in Appendix A—the estimation has an uncertainty itself. In the worst case, equation (19) explicitly gives the exact distribution of results conditioned to the input distribution, and that is already a useful piece of information to interpret the measurements.

- In many practical cases, $\boldsymbol{\Sigma_e}$ is not the only error source. For example, uncertainties on measured cut depths, elastic constants, actual geometry, etc. also play a role and affect results through non-linear effects. Consequently, a Monte Carlo simulation method is usually employed, which yields correct uncertainty bounds, since no bias is introduced during the inversion. As the problem is now well-posed, there is no contraindication to using standard uncertainty quantification techniques. In this case, equation (19) only evaluates the output uncertainty that is caused by strain/displacement uncertainties, which in fact are the main limiting factor due to the ill-posedness of the problem.

Presenting the Results

The bias-variance tradeoff can produce some remarkable spurious phenomena when it comes to presenting the residual stress results.

Since the solution uncertainty *seems to be* low with high regularization levels—even though in that case only the variance is low—the user is often tempted to favor overregularized solutions as they seems more presentable *and* more accurate. That is why often a polynomial series expansion looks more convenient, as it achieves good results with just a few DOFs.

Say that a round-robin test is conducted, where several laboratories are asked to evaluate residual stresses in a specimen (for instance, see two notable recent works [61, 62]). Even if their measurement equipment and operators' skills are comparable, an overregularized solution reported by one of the labs is inevitably presented as more accurate. That solution is smoother *and* apparently has smaller confidence intervals.

As discussed in [1], ill-posedness produces an even more deceptive effect. Assume that all laboratories diligently apply the same parameter selection criterion such as the Morozov discrepancy principle but, as usual, some laboratories have less accurate measuring instruments. To compensate for higher noise, the "optimal" solution picked by the selection criterion must undergo a higher level of regularization—at the cost of higher biases. Here is the paradox: the ones with worse input data will report *smoother* results, by simply following the "statistically right" way of processing data.

Those effects are quite confusing for an analyst who has to interpret the results and use them in an engineering application. Hence, we advocate for a way of presenting the results that allows for an intuitive and rational comparison of the results.

Through the proposed method, it all comes down to clearly disclose the value of the averaging length l_σ and to consider results that do not share that value as non-comparable quantities. As a matter of fact:

- The problem of finding $\sigma_{av}(\zeta)$ from $\varepsilon(h)$ is well-posed (Proposition 3), so there is no bias-variance tradeoff that can spoil the uncertainty quantification process. The results are *convergent* with the discretization scheme; as long as the latter is sufficiently refined, results do not depend on it.
- How refined? That is where l_c becomes effective. By disclosing l_m and l_c (or, possibly, some related and measurable physical quantity such as the average grain size), one can find whether the results uncertainty can be safely considered bias-free or not. When not verified, this fact might lead to a slight underestimation of the total uncertainty. Actually, if compared with the severe uncertainty underestimation that comes with classical methods, in most practical cases, results will be satisfactory even when $l_m > l_c$, as long as l_m is sufficiently refined to evaluate equation (15).
- Sometimes, a given measurement process would not be accurate enough to find $\sigma_{av}(\zeta)$ with an acceptable sig-

nal-to-noise ratio. In those cases, l_σ can be increased to reduce the solution variance, but this operation will yield *another* physical quantity, whose results and uncertainties should not be compared with residual stresses averaged over smaller lengths. The technological challenge becomes achieving the same l_σ with higher accuracy or a smaller l_σ with the same accuracy.

- The value of l_σ becomes a convenient way for the spatial resolution to be driven by performance requirements and not by the measurement itself or by the sensibility of the residual stress engineer. In an ideal case, one has a given strength criterion in mind and asks for a residual stress measurement that is coherent with it.

Conclusion

This paper is a development of the discussion introduced in [1] about some counterintuitive mathematical properties of residual stress measurements with relaxation methods, mainly stemming from ill-posedness. The previous work was concerned with how results are obtained through regularization and shed light on the bias-variance tradeoff, whereas this paper is focused on the corresponding quantification of uncertainties.

The inherent limits of the existing methods used to quantify the uncertainties have been explored. In particular, it is shown that most common strategies only quantify the *sensitivity* of results to input uncertainties, that is, the relation between variances of input and output distributions. Few notable exceptions attempted at taking bias into account, by referring to it as "model error" or "regularization error".

To foster more reliable uncertainty quantifications in residual stress evaluations, a different approach to the elaboration of measurements was proposed. After having shown that achieving point-wise values of residual stresses is questionable and not useful, it was suggested to switch to *average* residual stresses over a specified length. That averaging length may be chosen on the basis of the specific application (such as the critical distance in fatigue failure criteria) or it can be arbitrarily agreed among laboratories involved in a round robin of residual stress measurements. It was demonstrated that the problem of finding average residual stresses is well-posed, hence immune to the bias-variance tradeoff and to most of the counterintuitive problems that come with ill-posedness.

It is hoped that this paper will raise awareness on the risks of carrying out uncertainty quantifications in ill-posed frameworks, and that the proposed strategy will stand the test of the residual stress community. As ill-posedness is actually ubiquitous—from simple numerical differentiations to complicated non-linear inverse problems—similar considerations are likely to hold in many other problems in the field of experimental mechanics, which will hopefully be studied soon.

Appendix A: The Morozov Discrepancy Principle in ASTM E837-20

In its finite-dimensional numerical implementation, the Morozov discrepancy principle assumes that the Euclidean distance between the measured relaxed strains vector and the *ideal* one is known, at least in a statistical sense. If the independent variance of each measurement sample is known *a priori* (for example from field experience or an instrument certificate) and is the only error source, the expected squared discrepancy is easy to calculate.

When the measurement error is not known at all, the mathematical literature provided ill-posed problems with regularization parameter selection criteria that are tailored to function even in the absence of such information, including *Generalized Cross-Validation* (GCV) [63], the *L-Curve* criterion [40], *Iterative Predictive Risk Optimization* (I-PRO) [47] or the *Quasi-Optimality* principle [41].

Schajer and Prime [13, 64] took a different approach, by complementing the method with a statistical estimation of the measurement error, performed directly on the measured strain data. This procedure is currently included in the ASTM E837-20 [31] hole-drilling standard and is analyzed in the following paragraphs, because of its significance to relaxation methods.

The relaxed strain measurements are assumed to vary smoothly with the hole depth h, so that a second-order Taylor series expansion produces negligible errors at every point. For now, it is also assumed that the p depth steps h_i are uniform with interval Δh. Moreover, the *ideal* strain distribution $\varepsilon(h)$ is assumed to be perturbed at every sample by additive, zero-mean Gaussian noise $\mathcal{N}(0, \varsigma_\varepsilon^2)$ having an unknown variance ς_ε^2. At a given hole depth h_i, the Taylor series of the measured strain distribution $\varepsilon_\mathrm{m}(h_i)$ can be expanded as:

$$
\begin{aligned}
\varepsilon_\mathrm{m}(h_i) &= \varepsilon(h_i) + \mathcal{N}(0, \varsigma_\varepsilon^2) + O(\Delta h^3) \\
\varepsilon_\mathrm{m}(h_{i+1}) &= \varepsilon(h_i) + \Delta h\, \varepsilon'(h_i) + \frac{\Delta h^2}{2}\varepsilon''(h_i) \\
&\quad + \mathcal{N}(0, \varsigma_\varepsilon^2) + O(\Delta h^3) \\
\varepsilon_\mathrm{m}(h_{i+2}) &= \varepsilon(h_i) + 2\Delta h\, \varepsilon'(h_i) + 2\Delta h^2 \varepsilon''(h_i) \\
&\quad + \mathcal{N}(0, \varsigma_\varepsilon^2) + O(\Delta h^3) \\
\varepsilon_\mathrm{m}(h_{i+3}) &= \varepsilon(h_i) + 3\Delta h\, \varepsilon'(h_i) + \frac{9\Delta h^2}{2}\varepsilon''(h_i) \\
&\quad + \mathcal{N}(0, \varsigma_\varepsilon^2) + O(\Delta h^3)
\end{aligned}
\tag{A.1}
$$

Terms having order three or higher are denoted with $O(\Delta h^3)$. In order to neglect them, it is assumed that the *ideal* strain distribution is *sufficiently smooth* with respect to the depth increments scheme. Equivalently, a parabola is assumed to describe four consecutive *ideal* samples with negligible error. The fit error on the measured samples $\varepsilon_\mathrm{m}(h_i)$ will only be attributed to the statistical error $\mathcal{N}(0, \varsigma_\varepsilon^2)$. *The effectiveness of the procedure is conditioned on this assumption.*

Equation (A.1) can be linearly combined to simplify terms which depend on $\varepsilon(h)$ and its derivatives:

$$
\begin{aligned}
&\varepsilon_\mathrm{m}(h_i) - 3\varepsilon_\mathrm{m}(h_{i+1}) + 3\varepsilon_\mathrm{m}(h_{i+2}) - \varepsilon_\mathrm{m}(h_{i+3}) \\
&\triangleq \xi(h_i) = \mathcal{N}\big(0, \big(1^2 + 3^2 + 3^2 + 1^2\big)\varsigma_\varepsilon^2\big) + O(\Delta h^3) \\
&\quad \approx \mathcal{N}(0, 20\varsigma_\varepsilon^2)
\end{aligned}
\tag{A.2}
$$

Since the left-hand side in equation (A.2) behaves as a random variable with a Gaussian distribution, that same evaluation can be repeated along the entire relaxed strain vector by increasing i from 1 to $p - 3$. Hence, $p - 3$ realizations of that random variable can be drawn. Exploiting the properties of a χ^2 distribution, an unbiased estimator of ς_ε^2 can be built. By denoting the expected value with its usual symbol $\mathbb{E}(\cdot)$, it follows:

$$
\mathbb{E}\left(\sum_{i=1}^{p-3} \frac{\xi(h_i)^2}{20(p-3)}\right) = \frac{\varsigma_\varepsilon^2}{p-3}\mathbb{E}\big(\chi^2(p-3)\big) = \varsigma_\varepsilon^2
\tag{A.3}
$$

since $\mathbb{E}(\chi^2(n)) = n$. The left-hand side of equation (A.3) is an unbiased estimator of ς_ε^2. If the hole steps are not uniform, Schajer [13] provided an analogous set of coefficients to apply in equation (A.2), to cancel out Taylor expansion terms and to obtain the random component as a result. The procedure follows similarly.

An important remark on the estimator in equation (A.2) is that its variance is, in turn, not null and often not negligible. It can be shown that:

$$
\begin{aligned}
&\mathrm{Var}\left(\sum_{i=1}^{p-3} \frac{\xi(h_i)^2}{20(p-3)}\right) = \\
&= \left(\frac{\varsigma_\varepsilon^2}{p-3}\right)^2 \mathrm{Var}\big(\chi^2(p-3)\big) = \frac{2\varsigma_\varepsilon^4}{p-3}
\end{aligned}
\tag{A.4}
$$

since $\mathrm{Var}(\chi^2(n)) = 2n$. Note that for, say, $p = 100$ measurements, the standard deviation of this estimation is about 15% of the expected value; moreover, it increases with decreasing p. Recall that an inaccurate variance input in the Morozov discrepancy principle adds another layer of uncertainty to the parameter choice. If a Monte Carlo method is used to explore the solution variance, an uncertain input variance will be given as input to an operator whose regularization level depends on the input variance itself, so the physical significance of the output variance is, again, highly questionable (and bias is yet to be added to the overall uncertainty). As argued in [1], this discussion also points out that a high number of measurement samples is always beneficial, as it improves the sharpness of the input noise estimation.

Note that the $p - 3$ realizations are not strictly independent, because they share some measurement samples. Consequently, the assumption of a χ^2 distribution is questionable, though in many practical cases the associated error is

negligible compared to the estimation variance. A detailed mathematical analysis can be found in [65]; this procedure is a particular case of a more general class of *a posteriori* noise estimation algorithms.

Appendix B: Well-Posedness of the Proposed Strategy

A proof of well-posedness of the proposed mollifier method involves some concepts of functional analysis, but it is reported for the sake of rigor and completeness.

Equation (1) is a Volterra integral equation of the first kind, with kernel $A(h, z)$ defined on a triangular domain $0 \le z \le h \le h_{\max}$. With physical arguments, the kernel $A(h, z)$ can be easily shown to be bounded on its domain. It defines an integral operator $K : L^2(0, h_{\max}) \rightarrow L^2(0, h_{\max})$ that takes an element $\sigma(z)$ of the functional space of residual stresses and generates the corresponding element $\varepsilon(h)$ of the relaxed strains function space.

Its adjoint operator satisfies:

$$(K^* f)(z) = \int_z^{h_{\max}} A(h, z) f(h) \, dh \qquad (A.5)$$

An averaging operator $\Theta : L^2(0, h_{\max}) \rightarrow L^2(l_\sigma/2, h_{\max} - l_\sigma/2)$ is defined by choosing a suitable *mollifier* function $\omega(\zeta, z) : [l_\sigma/2, h_{\max} - l_\sigma/2] \times [0, h_{\max}] \rightarrow \mathbb{R}$ and writing:

$$(\Theta \sigma)(\zeta) \triangleq \int_0^{h_{\max}} \omega(\zeta, z) \, \sigma(z) \, dz \qquad (A.6)$$

For Θ to be actually an averaging operator, it must hold that:

$$\int_0^{h_{\max}} \omega(\zeta, z) \, dz = 1 \quad \forall z \in [l_\sigma/2, h_{\max} - l_\sigma/2] \qquad (A.7)$$

The classical version of the proof assumes that for each ζ, it is: $\omega(\zeta, z) \in R(K^*)$. In that case, for each $\zeta \in [l_\sigma/2, h_{\max} - l_\sigma/2]$ there exists a member $\beta(h, \zeta)$ of $L^2(0, h_{\max})$ such that:

$$\omega(\zeta, z) = (K^* \beta)(z) = \int_z^{h_{\max}} A(h, z) \, \beta(h, \zeta) \, dh \qquad (A.8)$$

By substituting equation (A.8) into equation (A.6), one obtains:

$$\begin{aligned}(\Theta \sigma)(\zeta) &= \int_0^{h_{\max}} \left(\int_z^{h_{\max}} A(h, z) \, \beta(h, \zeta) \, dh \right) \sigma(z) \, dz \\ &= \int_0^{h_{\max}} \int_z^{h_{\max}} A(h, z) \, \beta(h, \zeta) \, \sigma(z) \, dh \, dz\end{aligned} \qquad (A.9)$$

Due to Fubini's theorem, the order of integration can be switched over the triangular domain, obtaining:

$$\begin{aligned}(\Theta \sigma)(\zeta) &= \int_0^{h_{\max}} \int_0^{h} A(h, z) \, \beta(h, \zeta) \, \sigma(z) \, dh \, dz \\ &= \int_0^{h_{\max}} \beta(h, \zeta) \left(\int_0^{h} A(h, z) \, \sigma(z) \, dz \right) dh \\ &= \int_0^{h_{\max}} \beta(h, \zeta) \, \varepsilon(h) \, dh \triangleq (B\varepsilon)(\zeta)\end{aligned} \qquad (A.10)$$

where the same result could have been obtained by exploiting the properties of the adjoint operator.

Since by hypothesis $\beta(h, \zeta) \in L^2(0, h_{\max})$ for every $\zeta \in [l_\sigma/2, h_{\max} - l_\sigma/2]$, $B : L^2(0, h_{\max}) \rightarrow L^2(l_\sigma/2, h_{\max} - l_\sigma/2)$ is a Hilbert-Schmidt integral operator; in particular, it is bounded and *continuous*, therefore the problem of finding $(B\varepsilon)(\zeta)$ from $\varepsilon(h)$ is well-posed.

Equation (15) specifies that the chosen mollifier $\omega(\zeta, z)$ is a sliding rectangle function (referred to with the symbol Π), so that:

$$\begin{aligned}\sigma_{av}(z) = \Theta(\sigma) &= \int_0^{h_{\max}} \Pi\left(\frac{z - \zeta}{l_\sigma} \right) \sigma(\zeta) \, d\zeta \\ &= \frac{1}{l_\sigma} \int_{z - \frac{l_\sigma}{2}}^{z + \frac{l_\sigma}{2}} \sigma(\zeta) \, d\zeta\end{aligned} \qquad (A.11)$$

Unfortunately, for every $\zeta \in (l_\sigma/2, h_{\max} - l_\sigma/2)$ the mollifier $\Pi\left(\frac{z - \zeta}{l_\sigma} \right)$ is discontinuous at $z = \zeta \pm l_\sigma/2$, so it cannot belong to the range of the adjoint operator $(K^* f)(z)$. On the other hand, $\overline{R(K^*)} = N(K)^\perp = L^2(0, h_{\max})$, since K has a trivial nullspace; this classical result can be found, for instance, in Theorem 1.4.1 of [66]. Strictly speaking, the latter requires that $A(h, h) \neq 0$ on $0 \le h \le h_{\max}$. As a matter of fact, the maximum depth $h_{\max}$ is specifically limited in real-world applications in order to avoid a zone where a depth increment would produce a negligible strain regardless of the residual stresses therein (for instance, see [67]), so the condition is generally fulfilled.

Since $\Pi\left(\frac{z - \zeta}{l_\sigma} \right) \in \overline{R(K^*)} \setminus R(K^*)$, for each ζ there exists a sequence of elements $\gamma_k(\zeta, z) \in R(K^*)$ that converges to $\Pi\left(\frac{z - \zeta}{l_\sigma} \right)$. It can be easily constructed as follows. Being the kernel $A(h, z)$ continuous on its domain for obvious physical reasons, the corresponding integral operator (defined in equation (1)) is compact, so it admits a singular value expansion $(\lambda_i, u_i(h), v_i(z))$. Its singular values λ_i are countably infinite and accumulate at 0; when arranged in decreasing order, $\lim_{i \rightarrow \infty} \lambda_i = 0$. This fact is the central source of ill-posedness. As done in the main text, note that singular values are indicated

with the symbol λ_i rather than the more common s_i, to avoid confusion with discretized stress components.

Being compact by Schauder's theorem [68], the adjoint K^* also admits a singular value expansion $(\lambda_i, v_i(z), u_i(h))$; note the inverted position of $u_i(h)$ and $v_i(z)$.

Now, for each ζ define:

$$\gamma_k(\zeta, z) \triangleq \sum_{i=1}^{k} \left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle v_i(z) \tag{A.12}$$

For $k \to \infty$ we have that $\gamma_k(\zeta, z) \to \Pi\left(\frac{z-\zeta}{l_\sigma}\right)$, as equation (A.12) becomes a decomposition of $\Pi\left(\frac{z-\zeta}{l_\sigma}\right)$ along the orthonormal system defined by $v_i(z)$.

Each $\gamma_k(\zeta, z)$ has a unique preimage under K^* for every ζ, which can be obtained as:

$$\beta_k(h, \zeta) = \sum_{i=1}^{k} \frac{\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle}{\lambda_i} u_i(h) \tag{A.13}$$

In turn, each $\beta_k(h, \zeta)$ defines—as in the previous case—an averaging operator Θ_k such that:

$$\begin{aligned}
(\Theta_k \sigma)(\zeta) &= \int_0^{h_{max}} \gamma_k(\zeta, z)\, \sigma(z)\, dz \\
&= \int_0^{h_{max}} \left(\int_z^{h_{max}} A(h, z)\, \beta_k(h, \zeta)\, dh \right) \sigma(z)\, dz \\
&= \int_0^{h_{max}} \beta_k(h, \zeta)\, \varepsilon(h)\, dh \\
&= \int_0^{h_{max}} \sum_{i=1}^{k} \frac{\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle}{\lambda_i} u_i(h)\, \varepsilon(h)\, dh \\
&= \sum_{i=1}^{k} \frac{\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle}{\lambda_i} \langle \varepsilon(h), u_i(h) \rangle
\end{aligned} \tag{A.14}$$

Clearly, $\Theta_k \to \Theta$ as $k \to \infty$, which is a compact operator, hence bounded and continuous. Nonetheless, continuity with respect to $\varepsilon(h)$ is yet to be proven.

As shown by Lamm in [69], since $A(h, h) \neq 0$ on $0 \leq h \leq h_{max}$ the operator defined by equation (1) is *one-smoothing*, thus $\lambda_i \sim i^{-1}$ as $i \to \infty$ (see for example [70] or the simpler discussion in the appendices of [71]). Although—to our best knowledge—general results are not available, $\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle$ can as well be easily shown to asymptotically decay as i^{-1} when $i \to \infty$, by explicit numerical computation of its terms. Some additional arguments show that the contrary is very unlikely.

Due to Parseval's identity, $\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle$ must asymptotically behave as $i^{-\alpha}$ with $\alpha > 1/2$, since $\Pi\left(\frac{z-\zeta}{l_\sigma}\right)$ is in $L^2(0, h_{max})$. As shown by Hansen et al. in [46], the terms $v_i(z)$ of a compact

operator behave very similarly to the Fourier series functions in $L^2(0, h_{max})$. On the other hand, having defined $\mu = \frac{l_\sigma}{h_{max}}$, a rectangle function has a well-known Fourier series expansion:

$$\Pi\left(\frac{z}{l_\sigma}\right) = \sum_{n=-\infty}^{\infty} \mu\, \mathrm{sinc}(n\pi\mu) \exp j \frac{2\pi n}{h_{max}} z \tag{A.15}$$

Then, its Fourier coefficients are $\lesssim n^{-1}$. Eventually, $\left\langle \Pi\left(\frac{z-\zeta}{l_\sigma}\right), v_i(z) \right\rangle \sim i^{-1}$ is a rather weak condition. As a matter of fact, it can be numerically shown that it is true for all the typical operators K involved in relaxation methods.

Coming back to equation (A.14), this means that, as $k \to \infty$, for each value of ζ we have that:

$$(\Theta_k \sigma)(\zeta) \sim \sum_{i=1}^{k} \langle \varepsilon(h), u_i(h) \rangle \tag{A.16}$$

Consequently, in L^2 norms:

$$\left\| (\Theta_k \sigma)(\zeta) \right\| \sim \|\varepsilon(h)\| \tag{A.17}$$

Therefore, $\Theta \circ K^\dagger : \overline{R(K)} \setminus R(K) \to L^2(l_\sigma/2, h_{max} - l_\sigma/2)$ is bounded and continuous, so it can be continuously extended to $\overline{R(K)} = L^2(0, h_{max})$, proving the well-posedness of the proposed scheme.

Equation (A.14) effectively provides an explicit way of evaluating $\Theta \circ K^\dagger$ as a series expansion without having to obtain $\sigma(z)$ as an intermediate result, since $K^\dagger$ is only defined on $\overline{R(K)} \setminus R(K)$. However, in a discrete practical setting this problem does not arise at all, since discretization of $K^\dagger$ yields a continuous operator defined on the whole finite-dimensional strains domain, so $\Theta \circ K^\dagger$ can be evaluated in two passages, as done in the main text.

Note that the choice of $\Pi\left(\frac{z-\zeta}{l_\sigma}\right)$ was arbitrary, and other mollifiers (for instance, see [59]) could be picked for this problem—some of them would not require this second part of the proof, due to their smoothness. Nonetheless, a rectangle function was chosen because of its simplicity, as it represents a plain unweighted average along a given distance.

Declarations

References

1. Beghini M, Grossi T, Prime MB, Santus C (2023) Ill-Posedness and the Bias-Variance Tradeoff in Residual Stress Measurement Inverse Solutions. Exp Mech 63(3):495–516. https://doi.org/10.1007/s11340-022-00928-5

2. Sachs G, Espey G (1941) The measurement of residual stresses in metal. Iron age 148(63–71):148

3. Treuting RG, Read WT (1951) A Mechanical Determination of Biaxial Residual Stress in Sheet Materials. J Appl Phys 22(2):130–134. https://doi.org/10.1063/1.1699913

4. Beaney EM (1976) Accurate measurement of residual stress on any steel using the centre hole method. Strain 12(3):99–106. https://doi.org/10.1111/j.1475-1305.1976.tb00194.x

5. Schajer GS (1981) Application of Finite Element Calculations to Residual Stress Measurements. J Eng Mater Technol 103(2):157–163. https://doi.org/10.1115/1.3224988

6. Schajer GS (1988) Measurement of Non-Uniform Residual Stresses Using the Hole-Drilling Method. Part I Stress Calculation Procedures. J Eng Mater Technol 110(4):338–343. https://doi.org/10.1115/1.3226059

7. Schajer GS (1988) Measurement of Non-Uniform Residual Stresses Using the Hole-Drilling Method. Part II Practical Application of the Integral Method. J Eng Mater Technol 110(4):344–349. https://doi.org/10.1115/1.3226060

8. Cheng W, Finnie I (1993) A Comparison of the Strains Due to Edge Cracks and Cuts of Finite Width With Applications to Residual Stress Measurement. J Eng Mater Technol 115(2):220–226. https://doi.org/10.1115/1.2904210

9. Liu X, Schajer GS (1997) More Reliable Calculations for Layer-Removal Residual Stress Measurements. In: Proceedings of SEM Spring Conference on Experimental Mechanics, p 255–256

10. Prime MB (1999) Residual Stress Measurement by Successive Extension of a Slot: The Crack Compliance Method. Appl Mech Rev 52(2):75–96. https://doi.org/10.1115/1.3098926

11. Prime MB (2001) Cross-Sectional Mapping of Residual Stresses by Measuring the Surface Contour After a Cut. J Eng Mater Technol 123(2):162–168. https://doi.org/10.1115/1.1345526

12. Schajer GS, Prime MB (2007) Residual stress solution extrapolation for the slitting method using equilibrium constraints. J Eng Mater Technol 129(2):227–232. https://doi.org/10.1115/1.2400281. ISBN: 0094-4289

13. Schajer GS (2007) Hole-Drilling Residual Stress Profiling With Automated Smoothing. J Eng Mater Technol 129(3):440–445. https://doi.org/10.1115/1.2744416

14. Schajer GS (2013) Practical Residual Stress Measurement Methods. John Wiley & Sons

15. Schajer GS (2022) Hole Eccentricity Correction for Hole-Drilling Residual Stress Measurements. Exp Mech. https://doi.org/10.1007/s11340-022-00881-3

16. Jun T-S, Korsunsky AM (2010) Evaluation of residual stresses and strains using the Eigenstrain Reconstruction Method. Int J Solids Struct 47(13):1678–1686. https://doi.org/10.1016/j.ijsolstr.2010.03.002

17. Schajer GS, Winiarski B, Withers PJ (2013) Hole-Drilling Residual Stress Measurement with Artifact Correction Using Full-Field DIC. Exp Mech 53(2):255–265. https://doi.org/10.1007/s11340-012-9626-0

18. Baldi A (2014) Residual Stress Measurement Using Hole Drilling and Integrated Digital Image Correlation Techniques. Exp Mech 54(3):379–391. https://doi.org/10.1007/s11340-013-9814-6

19. Harrington JS, Schajer GS (2017) Measurement of Structural Stresses by Hole-Drilling and DIC. Exp Mech 57(4):559–567. https://doi.org/10.1007/s11340-016-0247-x

20. Hagara M, Trebuňa F, Pástor M, Huňady R, Lengvarský P (2019) Analysis of the aspects of residual stresses quantification performed by 3D DIC combined with standardized hole-drilling method. Measurement 137:238–256. https://doi.org/10.1016/j.measurement.2019.01.028

21. Schajer GS, Steinzig M (2005) Full-field calculation of hole drilling residual stresses from electronic speckle pattern interferometry data. Exp Mech 45(6):526–532. https://doi.org/10.1007/BF02427906

22. Schajer GS (2021) Optical Hole-Drilling Residual Stress Calculations Using Strain Gauge Formalism. Exp Mech 61(9):1369–1380. https://doi.org/10.1007/s11340-021-00740-7

23. Anderson TL (2017) Fracture Mechanics: Fundamentals and Applications. CRC press

24. Taylor D (2007) The Theory of Critical Distances: a New Perspective in Fracture Mechanics. Elsevier, Amsterdam, Boston

25. Berto F, Lazzarin P (2014) Recent developments in brittle and quasi-brittle failure assessment of engineering materials by means of local approaches. Mater Sci Eng R Rep 75:1–48. https://doi.org/10.1016/j.mser.2013.11.001

26. Sha W (2013) Steels: From Materials Science to Structural Engineering. Springer, London. https://doi.org/10.1007/978-1-4471-4872-2. https://link.springer.com/10.1007/978-1-4471-4872-2

27. Callister WD (2019) Materials Science and Engineering. Wiley, New York. OCLC: 1083469263

28. Salvati E, Korsunsky AM (2018) A simplified FEM eigenstrain residual stress reconstruction for surface treatments in arbitrary 3D geometries. Int J Mech Sci 138–139:457–466. https://doi.org/10.1016/j.ijmecsci.2018.02.016

29. Strantza M, Vrancken B, Prime MB, Truman CE, Rombouts M, Brown DW, Guillaume P, Van Hemelrijck D (2019) Directional and oscillating residual stress on the mesoscale in additively manufactured Ti-6Al-4V. Acta Mater 168:299–308. https://doi.org/10.1016/j.actamat.2019.01.050

30. Schulze V (2005) Modern Mechanical Surface Treatment: States, Stability, Effects, 1st edn. Wiley. https://doi.org/10.1002/3527607811

31. American Society for Testing and Materials (2020) Test Method for Determining Residual Stresses by the Hole-Drilling Strain-Gage Method. West Conshohocken, PA. https://doi.org/10.1520/E0837-20

32. Beghini M, Bertini L (2000) Analytical expressions of the influence functions for accuracy and versatility improvement in the hole-drilling method. J Strain Anal Eng Des 35(2):125–135. https://doi.org/10.1243/0309324001514071

33. Van Loan CF (1976) Generalizing the Singular Value Decomposition. SIAM J Numer Anal 13(1):76–83. https://doi.org/10.1137/0713009

34. Prime MB, Hill MR (2006) Uncertainty, Model Error, and Order Selection for Series-Expanded, Residual-Stress Inverse Solutions. J Eng Mater Technol 128(2):175–185. https://doi.org/10.1115/1.2172278

35. Olson MD, DeWald AT, Hill MR (2020) An Uncertainty Estimator for Slitting Method Residual Stress Measurements Including the Influence of Regularization. Exp Mech 60(1):65–79. https://doi.org/10.1007/s11340-019-00535-x

36. Olson MD, DeWald AT, Hill MR (2021) Precision of Hole-Drilling Residual Stress Depth Profile Measurements and an Updated Uncertainty Estimator. Exp Mech 61(3):549–564. https://doi.org/10.1007/s11340-020-00679-1

37. Smit TC, Reid RG (2018) Residual Stress Measurement in Composite Laminates Using Incremental Hole-Drilling with Power

Series. Exp Mech 58(8):1221–1235. https://doi.org/10.1007/s11340-018-0403-6

38. Smit TC, Reid RG (2020) Use of Power Series Expansion for Residual Stress Determination by the Incremental Hole-Drilling Technique. Exp Mech 60(9):1301–1314. https://doi.org/10.1007/s11340-020-00642-0

39. Smit TC, Nobre JP, Reid RG, Wu T, Niendorf T, Marais D, Venter AM (2022) Assessment and Validation of Incremental Hole-Drilling Calculation Methods for Residual Stress Determination in Fiber-Metal Laminates. Exp Mech. https://doi.org/10.1007/s11340-022-00848-4

40. Hansen PC (1999) The L-curve and its use in the numerical treatment of inverse problems

41. Glasko VB, Kriksin YA (1984) On the quasi-optimality principle for ill-posed problems in Hilbert space. USSR Comput Math Math Phys 24(6):1–7

42. Brítez DL, Prime MB, Werda S, Laheurte R, Darnis P, Cahuc O (2023) Uncertainty reduction in residual stress measurements by an optimised inverse solution using nonconsecutive polynomials. Strain 59(1). https://doi.org/10.1111/str.12430

43. Tikhonov AN, Goncharsky AV, Stepanov VV, Yagola AG (1995) Numerical Methods for the Solution of Ill-Posed Problems. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-8480-7. http://link.springer.com/10.1007/978-94-015-8480-7

44. Vogel CR (2002) Computational Methods for Inverse Problems. SIAM

45. Hansen PC (2010) Discrete Inverse Problems: Insight and Algorithms. SIAM

46. Hansen PC, Kilmer ME, Kjeldsen RH (2006) Exploiting Residual Information in the Parameter Choice for Discrete Ill-Posed Problems. BIT Numer Math 46(1):41–59. https://doi.org/10.1007/s10543-006-0042-7

47. Benvenuto F, Jin B (2020) A parameter choice rule for Tikhonov regularization based on predictive risk. Inverse Prob 36(6):065004. https://doi.org/10.1088/1361-6420/ab6d58

48. Chupakhin S, Kashaev N, Klusemann B, Huber N (2017) Artificial neural network for correction of effects of plasticity in equibiaxial residual stress profiles measured by hole drilling. J Strain Anal Eng Des 52(3):137–151. https://doi.org/10.1177/0309324717696400

49. Halabuk D, Návrat T (2022) Universal Procedure for Correction of Plasticity Effect in Hole-Drilling Uniform Residual Stress Measurement. Exp Mech. https://doi.org/10.1007/s11340-022-00869-z

50. Geman S, Bienenstock E, Doursat R (1992) Neural Networks and the Bias/Variance Dilemma. Neural Comput 4(1):1–58. https://doi.org/10.1162/neco.1992.4.1.1

51. Tognan A, Laurenti L, Salvati E (2022) Contour Method with Uncertainty Quantification: A Robust and Optimised Framework via Gaussian Process Regression. Exp Mech. https://doi.org/10.1007/s11340-022-00842-w

52. Barile C, Carone S, Casavola C, Pappalettera G (2023) Implementation of Gaussian Process Regression to strain data in residual stress measurements by hole drilling. Measurement 211:112590. https://doi.org/10.1016/j.measurement.2023.112590

53. Görtler J, Kehlbeck R, Deussen O (2019) A Visual Exploration of Gaussian Processes. Distill 4(4):17. https://doi.org/10.23915/distill.00017

54. Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G (2021) Gaussian Process Regression for Materials and Molecules. Chem Rev 121(16):10073–10141. https://doi.org/10.1021/acs.chemrev.1c00022

55. Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. Adaptive computation and machine learning. MIT Press, Cambridge, Mass. OCLC: ocm61285753

56. Sollich P (2001) Gaussian process regression with mismatched models. Adv Neural Inform Process Syst 14

57. Louis AK, Maass P (1990) A mollifier method for linear operator equations of the first kind. Inverse Prob 6(3):427–440. https://doi.org/10.1088/0266-5611/6/3/011

58. Murio DA (1993) The Mollification Method and the Numerical Solution of Ill-posed Problems. Wiley, New York

59. Hegland M, Anderssen RS (1998) A mollification framework for improperly posed problems. Numer Math 78(4):549–575. https://doi.org/10.1007/s002110050325

60. Jonas P, Louis AK (2001) A Sobolev space analysis of linear regularization methods for ill-posed problems. J Inverse Ill-Posed Prob 9(1). https://doi.org/10.1515/jiip.2001.9.1.59

61. D'Elia CR (2020) Interlaboratory Reproducibility of Contour Method Data Analysis and Residual Stress Calculation. Exp Mech 13

62. D'Elia CR, Carlone P, Dyer JW, Lévesque JB, De Oliveira JA, Prime MB, Roy MJ, Spradlin TJ, Stilwell R, Tucci F, Vasileiou AN, Watanable BT, Hill MR (2022) Interlaboratory Reproducibility of Contour Method Data in a High Strength Aluminum Alloy. Exp Mech 62(8):1319–1331. https://doi.org/10.1007/s11340-022-00849-3

63. Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21(2):215–223

64. Schajer GS, Prime MB (2006) Use of Inverse Solutions for Residual Stress Measurements. J Eng Mater Technol 128(3):375. https://doi.org/10.1115/1.2204952

65. Czesla S, Molle T, Schmitt JHMM (2018) A posteriori noise estimation in variable data sets: With applications to spectra and light curves. Astron Astrophys 609:39. https://doi.org/10.1051/0004-6361/201730618

66. Brunner H (2017) Volterra Integral Equations: An Introduction to Theory and Applications, 1st edn. Cambridge University Press. https://doi.org/10.1017/9781316162491

67. Schajer GS (2020) Compact Calibration Data for Hole-Drilling Residual Stress Measurements in Finite-Thickness Specimens. Exp Mech 60(5):665–678. https://doi.org/10.1007/s11340-020-00587-4

68. Schauder J (1930) Über lineare, vollstetige Funktionaloperationen. Studia Math 2:183–196. ISBN: 0039-3223 Publisher: Instytut Matematyczny Polskiej Akademii Nauk

69. Lamm PK (2000) A Survey of Regularization Methods for First-Kind Volterra Equations. In: Colton D, Engl HW, Louis AK, McLaughlin JR, Rundell W (eds) Surveys on Solution Methods for Inverse Problems. Springer, Vienna, pp 53–82

70. Tuan VK, Gorenflo R (1994) Asymptotics of singular values of fractional integral operators. Inverse Prob 10(4):949–955. https://doi.org/10.1088/0266-5611/10/4/013

71. Kirsch A (2021) An Introduction to the Mathematical Theory of Inverse Problems. Applied Mathematical Sciences, vol. 120. Springer, Cham. https://doi.org/10.1007/978-3-030-63343-1. http://link.springer.com/10.1007/978-3-030-63343-1