



# Ill-Posedness and the Bias-Variance Tradeoff in Residual Stress Measurement Inverse Solutions

M. Beghini<sup>1</sup> · T. Grossi<sup>1</sup> · M.B. Prime<sup>2</sup> · C. Santus<sup>1</sup>

Received: 17 August 2022 / Accepted: 8 November 2022 / Published online: 19 December 2022  
© The Author(s) 2022

## Abstract

**Background** Relaxation methods determine residual stresses by measuring the deformations produced by incremental removal of a subdomain of the specimen. Measured strains at any given increment, determined by the cumulative effect of the relieved stresses, appear as an integral equation, which must be inverted to obtain residual stresses. In practice, stress distributions are discretized by a finite-dimensional basis, to transform the integral equations into a linear system of equations, which is often ill-conditioned.

**Objective** This article demonstrates that the problem is actually ill-posed and comes with an inherent bias-variance tradeoff.

**Methods** The hole drilling method is used as an example application, and the practical effects of ill-posedness are illustrated.

**Results** Traditional regularization of the solution by limiting the resolution of the discretization reduces solution variance (noise) at the expense of increased bias and often results in the ultimately harmful practice of taking fewer data points. A careful analysis including the alternate Tikhonov regularization approach shows that the highest number of measurements should always be taken to reduce the variance for a given regularization scheme. Unfortunately, the variability of a regularized solution cannot be used to build a valid confidence interval, since an unknown bias term is always present in the true overall error.

**Conclusions** The mathematical theory of ill-posed problems provides tools to manage the bias-variance tradeoff on a reasonable statistical basis, especially when the statistical properties of measurement errors are known. In the long run, physical arguments that provide constraints on the true solution would be of utmost importance, as they could regularize the problem without introducing an otherwise unknown bias. Constraining the minimum length scale to some physically meaningful value is one promising possibility.

**Keywords** Residual Stress · Relaxation Methods · Bias-variance Tradeoff · Ill-posedness · Inverse Problems · Hole Drilling Method

## Introduction

Many of the most established techniques to measure stress fields in a component consist of measuring the strain fields and exploiting the material constitutive relations to get the corresponding stresses [1]. However, many common strain-measuring instruments can only measure *relative* strains with respect to an initial condition. For example, strain gauges

measure deformations that are referenced to the moment they were glued; also non-contact techniques such as Digital Image Correlation (DIC) [2–6] or Electronic Speckle Pattern Interferometry (ESPI) [7, 8] measure *changes* in the displacement fields, by comparing an initial and a final state. In these cases, an often implicit assumption is that the reference state is *unstressed*. Then, the *relative* change of strain fields allows to compute the *absolute* stress fields.

Residual stresses are defined as non-null stress fields in components whose boundary surfaces are traction-free and when the component is not subject to body loads such as gravity [9, 10]. In practical terms, no external load is provided, yet the material has a self-equilibrated stress field. This happens due to the presence of inelastic incompatible strains (usually referred to as *inherent strains* [11, 12] or

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

<sup>1</sup> Dipartimento di Ingegneria Civile e Industriale, Università di Pisa, Largo Lucio Lazzarino, Pisa 56122, Italy

<sup>2</sup> Los Alamos National Laboratory, Los Alamos NM 87545, USA



*eigenstrains* [13–15]), which would lead to penetrations or gaps between material points connected in the undeformed component. Within strength limits, the component integrity is enforced by additional elastic strain fields (hence stress fields) that lead to compatible total strains.

When unknown residual stresses are to be measured, an *unstressed* state is not available anymore, as the target stresses are (by definition) those corresponding to a completely unloaded condition. This lack of an unstressed reference state defines the fundamental challenge to measuring residual stresses as compared to applied stresses.

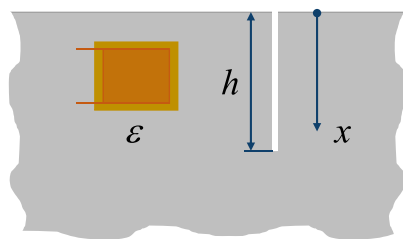
X-Rays diffraction [16] and Neutron Diffraction [17] techniques use *in situ* measurements of the lattice spacing to infer elastic strains by comparison to an unstressed reference value, whose availability is not trivial. Indirect strain measurements, (e.g. Magnetic [18] or Ultrasonic Methods [19]) have been studied too, although they suffer from the same issue.

Another common strategy is to induce a *partially relaxed* state by removing or disconnecting a subdomain of the component [20]: by doing this, new traction-free surfaces are created, some compatibility conditions are removed, and the strain fields are redistributed to a new configuration. Then, the original stress fields can be identified from the strain changes following material removal, which can be captured with any *relative* strain measuring instrument. These methods fall under the name of *relaxation methods*.

For any relaxation method, the measured strains at a given point are determined by a *cumulative* effect of all the stresses that have been relaxed on a new free surface, due to Bueckner's principle [21]. Under a continuum mechanical model, this relation appears as integral equations [22–25]. For example, in the simple 2D case of Fig. 1, it holds that:

$$\varepsilon(h) = \int_0^h A(h, x) \sigma(x) dx \quad (1)$$

where  $h$  denotes the cut length and  $x$  is used as a spatial coordinate. Equation (1) states that the measured deformations can



**Fig. 1** Typical example of the experimental setup for a residual stress measurement through relaxation methods, in the simple case of a 2D problem. When a cut of length  $h$  is incrementally introduced in the specimen, a strain gauge (or any other measuring instrument) registers a deformation  $\varepsilon$ . Its magnitude depends (through equation (1)) on the residual stresses that have been relaxed by the introduction of the cut

be obtained through a *weighted infinite sum* of the relieved stresses  $\sigma(x)$  on the cut surface, each of which contributes proportionally to  $A(h, x)$ . The latter is also known as the *influence function* or *kernel function* of the problem. See Fig. 1 for a representative example of a relaxation experiment. To obtain  $\sigma(x)$ , the relation between strains and stresses in equation (1) must be inverted. In general, a closed-form inversion is not possible.

Since the first pioneering studies [26–28], it was clear that numerical inversion of integral equations like equation (1) presented some significant difficulties, due to an often unmanageable ill-conditioning of the discretized equations. All subsequent research had to deal with this fact and proposed methods to quantify and/or decrease the solution uncertainty [29–32]. To the authors' best knowledge, all the proposed strategies involve limiting the degrees of freedom (DOFs) available for the stress solutions to limit sensitivity to noise and/or filtering the data or the stress profile to limit noise itself. How to choose the *optimal* DOFs amount or filter design is then a crucial matter.

A statement of contribution can now be formulated. In this paper, it will be shown that:

- The inversion of the integral equations arising from relaxation methods is an *ill-posed* problem, which is a very specific mathematical property and must be distinguished from a general *ill-conditioning*. A very active branch of applied mathematics is specifically devoted to the study of this class of problems: the *Theory of ill-posed problems*, whose origins trace back to the works of Tikhonov in the past century [33–35].
- Ill-posed problems require additional assumptions to obtain a physically meaningful solution. This process falls under the name of *regularization*.
- Any numerical implementation of the original functional problem is itself a form of regularization; it comes with *implicit* hypotheses and associated *biases*. An unregularized solution is both meaningless (having unbounded confidence intervals) and impossible to achieve in a practical setting.
- From a practical point of view, having a biased solution means that the distribution of solutions generated by imperfect data inputs is not centered around the *true* solution.
- Uncertainty quantification technique based on the analysis of input errors can only quantify the *variability* of the solution corresponding to the variance of problem inputs but it leaves out the *bias* introduced by regularization. Unless that bias can be bounded by physical considerations, it is potentially infinite. It can be reduced only at the expense of an increase of the solution variance, a fact known in the mathematical literature as *bias-variance tradeoff* [36].
- In contrast to solution variance, the solution bias is always inaccessible to the user, except for the remarkable case

where the solution is *already* known through other means. This situation is known as *ground truth* [37] in the mathematical and computer science literature and it is of fundamental importance for a quantitative performance evaluation of various regularization techniques.

- Many strategies that have been proposed to overcome the typical ill-conditioning of relaxation methods can be unified into the same mathematical framework.

In addition, since Tikhonov regularization and the Morozov discrepancy principle have received increasing attention from the residual stress community in the last twenty years [25, 31, 32, 38–41], a practical explanation of their effect on the inverse problem solution is provided. It is shown that Tikhonov regularization is not an *additional* noise filter applied to the discrete solution to smooth out results but rather an *alternative* regularization technique.

In the “[Theory of Ill-posed Problems](#)” section, some basic concepts of inverse problem theory will be introduced, through the conceptually simple problem of differentiating a measured function. In fact, differentiation (namely, the inverse problem of integration) shares many mathematical properties with the inversion of integral equations at the core of relaxation methods.

In “[The Hole Drilling Method](#)” section, the ill-posedness of relaxation methods will be investigated through a guiding example concerning the hole-drilling method [22, 23, 42]. Its practical consequences will be analyzed in detail, as many intuitive concepts that can be straightforwardly applied to a well-posed problem show some pitfalls when the problem is ill-posed.

The current state of the art of relaxation methods is well-established both in academia and industry, as proven by several studies that validated the consistency of their results and witnessed by the formulation of standard procedures [43]. However, in the authors’ opinion, there is still quite a gap between relaxation methods and the almost century-old mathematical theory of ill-posed problems, precisely concerning this class of problems. This article starts to bridge this gap, formalizing many concepts that are currently implicit or left to the sound discretion of the operator, hopefully leading in the future to improving solution accuracy and uncertainty estimation and a better understanding of the fundamental limits of inverse solutions.

## Theory of Ill-posed Problems

### Basic Concepts

When a physical behavior is mathematically modelled, the set of equations is required to be consistent with both the general laws and the experimental observations [44] at least in

the domain of interest. Consistency requires that predictions resulting from the mathematical model have some basic properties: 1) existence, 2) uniqueness and 3) continuity. Existence guarantees that a solution can be obtained for any admissible set of input data. In a deterministic framework, the solution is also expected to be unique. When a model is continuous, results depart “without jumps” from the exact output as measurements become more inaccurate. Correspondingly, as input errors tend to zero (i.e., by technological advances), so do output errors.

If a mathematical model satisfies these properties, its predictions can always be unambiguously compared with imperfect experimental outcomes, hence (at least in theory) it allows a verification of its consistency. In this case, the model is called well-posed in the sense of Hadamard [45] (“*bien posé*” in his original paper). A problem is considered ill-posed if it is not well-posed, hence if any of the above conditions is not met. The previous properties are desirable but they should not be regarded as necessary to produce a useful model of a physical process.

Quite commonly, a well-posed mathematical model is also needed in a reversed form, in which some outputs are available and some inputs are missing and they need to be evaluated through the model. Missing inputs are quantities that characterize the system of interest but cannot be accessed directly. A typical example is the density distribution measured by a CT scan, or the reconstruction of an image that has been blurred; another common case arises when the current state of a system is given but a past state has to be reconstructed. Well-posedness can be substantially impaired if the known and unknown quantities are inverted.

The original well-posed problem is usually called *direct problem* or *forward operator* and the reversed relationship between inputs and outputs is called *inverse problem*. “Forward” and “Inverse” do not necessarily point to a causal link and to its logical inversion, but, if it is the case, it is easy to identify the direct problem. Usually, the direct problem is the one that was studied first and is often well-posed. Inverse problems led to development of mathematical techniques aimed to overcome their typical ill-posedness, known as the *Theory of ill-posed problems* [46–49].

## The Simplest Ill-posed Problem

### Problem statement

Suppose that a 100 kg object is thrown out of an airplane at cruising altitude  $Z_0$ . The object is equipped with a GPS sensor that can measure the current altitude at any given sample rate. No other sensor is available.

When the object touches the ground, GPS data are collected, and the problem is to retrieve the vertical speed profile during the fall. No additional information is available



about the presence of engines, wings or parachutes that may alter the object trajectory.

In their functional form, GPS altitude  $Z(t)$  and vertical speed  $v(t)$  are related by the simple kinematic equation:

$$Z(t) = Z_0 + \int_0^t v(\tau) d\tau \tag{2}$$

Note that equation (2) can be rearranged as a particular case of equation (1):

$$Z(t) - Z_0 = \int_0^t A(t, \tau)v(\tau) d\tau \tag{3}$$

where the influence function  $A(t, \tau)$  is identically set to 1. Equation (2) allows one to obtain  $Z(t)$  from any given speed profile  $v(t)$ . Actually, the inverse relation between those two quantities is needed. In this case, the inverse operator corresponding to equation (2) can be explicitly obtained in a closed form:

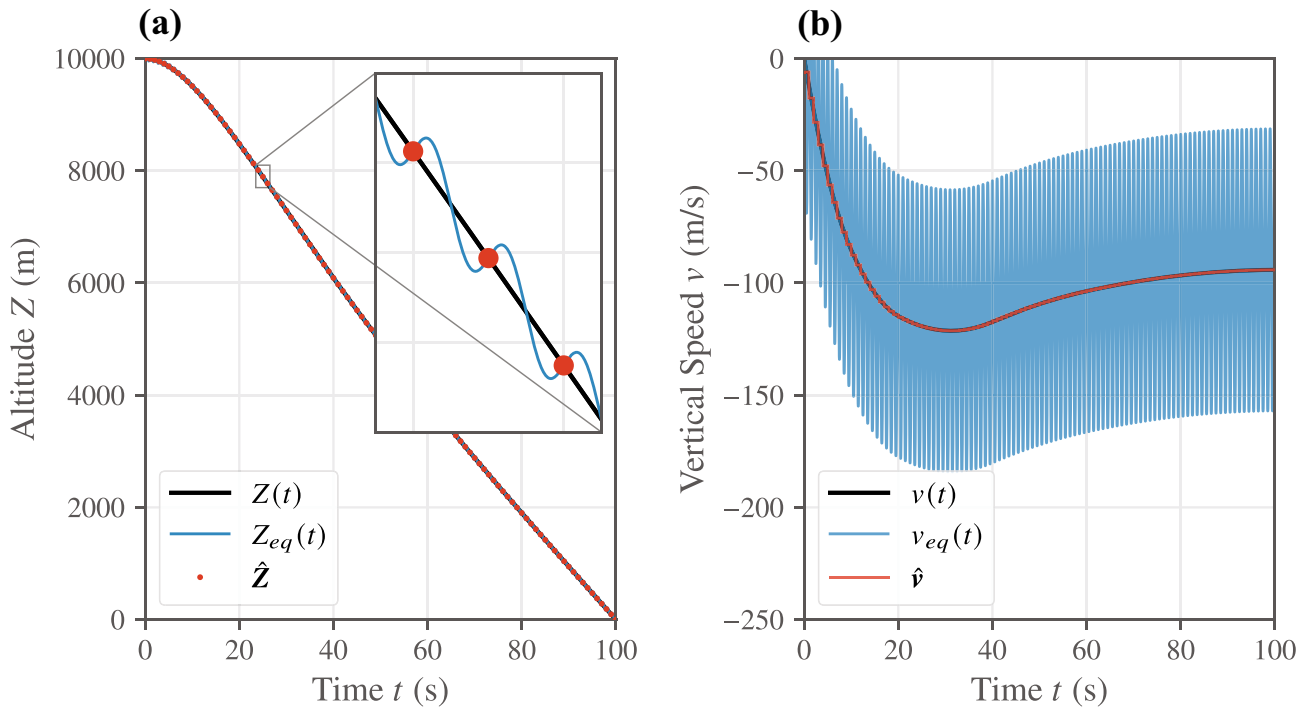
$$v(t) = \frac{dZ(t)}{dt} \tag{4}$$

If  $Z(t)$  was known in a differentiable form and in absence of measurement errors, equation (4) would solve the problem. A solution example is shown in Fig. 2.

**Ideal discrete solution**

Unfortunately, actual measurements are neither differentiable nor perfect. No matter how high the instrument sample rate is, a *finite* set of  $p$  measurements at timestamps  $\hat{t} = [t_1, t_2 \dots t_i \dots t_p]$  will be available, collected in the array  $\hat{Z} = [Z_1, Z_2 \dots Z_i \dots Z_p]$ . For now,  $\hat{Z}$  is still assumed to be an array of perfect measurements and timestamps  $t$  are supposed to be known with infinite accuracy. A fixed sample rate is also assumed, so that  $t_i - t_{i-1} = \Delta t$ . Equation (4) needs to be discretized too. A reasonable option is to assume that the *true* solution  $v(t)$  is piecewise constant between two timestamps. This hypothesis projects  $v(t)$  on a finite-dimensional subspace (piecewise constant functions on given intervals); therefore, its values can be represented in an array  $\hat{v} = [v_1, v_2 \dots v_i \dots v_p]$ , which will constitute an *ideal discrete* solution. Here “ideal” points to the fact that it comes from measurements that have no error, so it is the best solution available with the given discretization scheme. Each value can be obtained through equation (4):

$$v_i = \frac{Z_i - Z_{i-1}}{\Delta t} \tag{5}$$



**Fig. 2** Finding the vertical speed of a falling object from altitude measurements. Although the data are made-up, the reported curves are plausible. In practical settings,  $Z(t)$  is never known in an analytical closed form. On the contrary, a finite set of measurements is available, gathered in the array  $\hat{Z}$ . Discretization yields a piecewise constant solution, whose values are stored in the array  $\hat{v}$ . The difference between the *ideal discrete* solution  $\hat{v}$  and the *true* solution  $v(t)$  can be arbitrarily high. For example, define  $Z_{eq}(t)$  by superposing a sinusoidal perturbation of

period  $\Delta t$  to  $Z(t)$ . If the true object trajectory was actually  $Z_{eq}(t)$ , the sampling process would yield the same array  $\hat{Z}$ , but the corresponding  $v_{eq}(t)$  would be significantly far from the obtained discrete solution. No discontinuity is needed to obtain this effect:  $Z_{eq}(t)$  is still a  $C^\infty$  function, exactly like  $Z(t)$ . Similarly, even higher errors can be generated. Nevertheless, the *ideal discrete* solution  $\hat{v}$  converges to the *true* solution  $v(t)$  as the sample rate is increased, no matter how oscillatory the *true* solution is



assuming that the cruising altitude  $Z_0$  (where the object is released) is known as a boundary condition. Equation (5) solves the problem but introduces an error through the assumed form of the *true* solution  $v(t)$ . In fact, the actual solution to equation (4) is seldom a piecewise constant function, especially considering that a massive body cannot have jumps in its speed. This error could in principle be arbitrarily high:  $v(t)$  is free to show any behavior between two given timestamps, as long as its integral is kept constant (see Fig. 2). Nevertheless, that bias vanishes as the discretization scheme is refined, and the *ideal discrete* solution  $v(t)$  converges to the *true* solution  $v(t)$  of equation (4).

**Perturbed solution**

Measurements are always affected by errors, made up by a bias and a variability component. Hence, the array of true values  $\hat{Z}$  is actually never attainable. Instead, the GPS sensor yields a collection of inexact measurements  $\hat{Z}_m = \hat{Z} + \eta$ , where errors are denoted with the array  $\eta = [\eta_1, \eta_2 \dots \eta_i \dots \eta_p]$ . Timestamps  $\hat{t}$  are also never known exactly in a real-world case; this will be another source of error, often quite difficult to estimate. For the sake of simplicity, in this example it will be assumed that  $\hat{t}$  is known with negligible error. The interested reader will find a detailed treatise of this matter in [50–52].

The array  $\eta$  is not deterministically known, otherwise it could be used to obtain the perfect measurements  $\hat{Z}$ . At most, some statistical properties of  $\eta$  may be available. The individual entries of  $\eta$  could be correlated with each other to some degree (as it happens with systematic errors), but they always show an independent and totally random component. In other words, every measurement sample is affected by an error, a part of which is independent from the error on other samples. For example, electrical noise and the machine representation error usually belong to this category. For the sake of simplicity, in the following it is assumed that the GPS sensor yields unbiased measurements and only independent errors are considered. Therefore, every  $\eta_i$  is assumed to be a random variable with zero mean and variance  $\zeta^2$ , independent of the measured value  $Z_i$ . When  $\hat{Z}_m$  is plugged into equation (5), a *perturbed* version of  $\hat{v}$  is obtained, denoted as  $\hat{v}_\delta$ :

$$\begin{aligned}
 v_{\delta,i} &= \frac{Z_{m,i} - Z_{m,i-1}}{\Delta t} \\
 &= \frac{Z_i - Z_{i-1}}{\Delta t} + \frac{\eta_i - \eta_{i-1}}{\Delta t} \\
 &= v_i + \frac{\eta_i - \eta_{i-1}}{\Delta t}
 \end{aligned}
 \tag{6}$$

The *true* solution  $v(t)$  was projected onto a finite-dimensional subspace and represented with the array  $\hat{v}$ , an *ideal discrete* solution. The latter was perturbed by measurement noise, so

a *perturbed* solution  $\hat{v}_\delta$  was achieved. This distinction between solutions is summarized in Fig. 3.

**Ill-posedness**

Each entry of  $\hat{v}_\delta$  deviates from the *ideal discrete* solution  $\hat{v}$ . Their difference is another random variable, whose mean and variance can be obtained as:

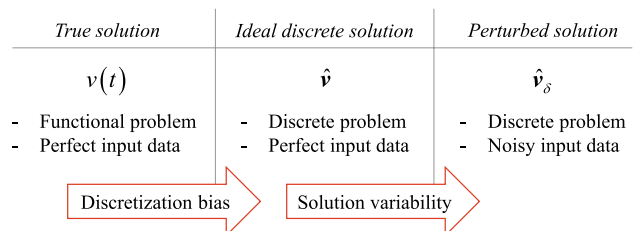
$$\mu(v_{\delta,i} - v_i) = \frac{\mu(\eta_i - \eta_{i-1})}{\Delta t} = 0
 \tag{7}$$

$$\text{Var}(v_{\delta,i} - v_i) = \frac{\text{Var}(\eta_i - \eta_{i-1})}{\Delta t} = \frac{2\zeta^2}{\Delta t}
 \tag{8}$$

Note from equation (7) that the *perturbed* solution  $\hat{v}_\delta$  is an unbiased estimator of the *ideal discrete* solution  $\hat{v}$ , as  $\mu(v_{\delta,i} - v_i)$  is null: if the experiment could be repeated arbitrarily many times in the exact same conditions, the point-wise average of the *perturbed* solution  $\hat{v}_\delta$  would converge to the *ideal discrete* solution  $\hat{v}$ , due to the central limit theorem. On the other hand, equation (8) provides information on how far  $\hat{v}$  could be from the obtained  $\hat{v}_\delta$ .

Even if  $\hat{v}_\delta$  is a fully characterized unbiased estimator of  $\hat{v}$ , the real solution error shall be evaluated with respect to the *true* solution  $v(t)$ . Unfortunately, the estimation properties of  $\hat{v}_\delta$  gets impaired when the discretization scheme is refined to diminish its bias. In fact, from equations (7) and (8),  $\mu(v_{\delta,i} - v_i)$  remains always null while  $\text{Var}(v_{\delta,i} - v_i)$  diverges as  $\Delta t \rightarrow 0$ . The key argument here is error independence, so the measurement variance  $\zeta^2$  is not affected by the sample rate.

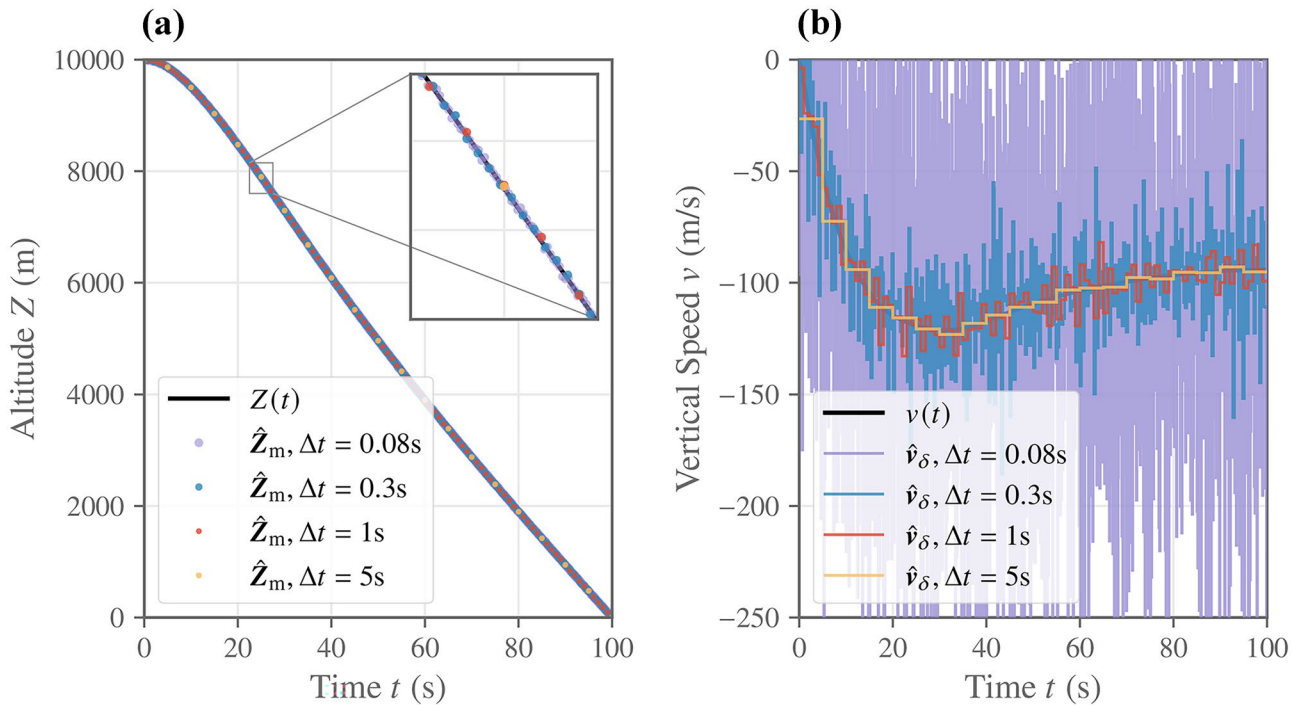
As an attempt to reduce the discretization bias by refining the numerical scheme, the only available *perturbed* solution  $\hat{v}_\delta$  becomes an increasingly imprecise approximation of the *ideal discrete* solution  $\hat{v}$ . This fact is shown in Fig. 4. In practical terms, the differentiation of a function amplifies



**Fig. 3** Terminology of the inverse problem solutions. The *true* functional solution is assumed to belong to the span of a finite-dimensional basis, so the actual solution space is *projected* to a finite-dimensional space. In general, this introduces a bias. Then, the coordinates with respect to that basis allow to *represent* the solution through an array; namely, an element of  $\mathbb{R}^p$ . The latter constitutes an *ideal discrete* solution, corresponding to ideally perfect measurements. In presence of measurement errors, only a *perturbed* solution is available, whose variability depends on the variance of input errors







**Fig. 4** Solution examples corresponding to altitude measurements having a standard deviation of 5 m. Low sampling rates yield a precise but inaccurate solution; however, if the sample rate is increased, the per-

turbed solution  $\hat{v}_\delta$  becomes more and more variable until it is completely dominated by random errors

high-frequency noise, and the latter increases with sample rate, until  $\hat{v}_\delta$  becomes substantially dominated by noise.

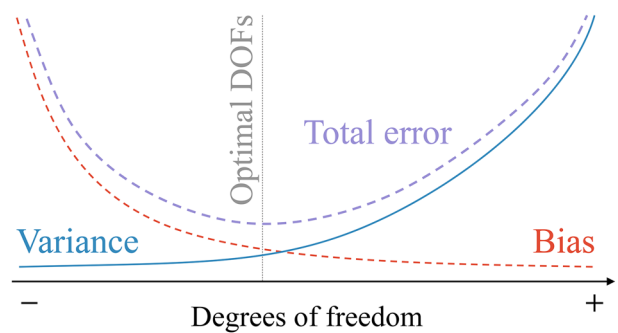
In the original functional setting of equation (4), where no bias has been introduced yet, even an infinitesimal perturbation of  $Z(t)$  can lead to huge output errors, so there is a sudden jump between the *true solution* and the outputs of imperfect input data. In fact, it can be shown that differentiation (namely, the inverse problem of integration) is an ill-posed problem in the sense of Hadamard, because it violates the continuity condition (see Part One, Sect. II in [53]).

As a result, a refinement of the discretization scheme is not always beneficial to the solution. In the inverse problem literature, this phenomenon is known as *bias-variance tradeoff*. A reduction in the former yields an increase in the latter, and vice-versa. A graphical representation is shown in Fig. 5. Every analytical or numerical estimation of the solution sensitivity to errors in inputs will only quantify the variance of  $\hat{v}_\delta$ , leaving out the unknown and potentially boundless bias given by the chosen discretization scheme. No mathematical caveat can overcome this curse, which is typical of ill-posed problems.

The same is not true for the integration of a function: a refined discretization scheme leaves out only high-frequency content, which has negligible effects on the obtained solution, and noise also gets attenuated by the low-pass properties of the integration, so both bias and variance can be reduced indefinitely.

### Regularization

In most cases, some general properties are known about the solution. This additional information is independent of the mathematical equations alone and comes almost always from physical arguments and context-based evaluations.



**Fig. 5** Illustration of the bias-variance tradeoff. Dashed quantities are unknown in a practical setting. By increasing DOFs to reduce the bias of the discrete problem, a corresponding increase in variance is obtained, and vice-versa. The total error (sum of bias and variance) shows a minimum, which corresponds to an *optimal* number of DOFs. Since the bias actually depends on the unknown *true* solution, in practice the optimal point cannot be determined, unless some properties of the *true* solution are known *a priori*



For instance, it is certainly difficult to believe that the most refined solution in Fig. 4 is the true vertical speed profile of the falling object, even if it is actually the only solution to equation (5) corresponding to the given data. The involved accelerations would require technologically unfeasible thrust forces and inertial loads that would probably pulverize the object, no matter the strength of its material.

As an imaginary example, after a careful inspection of the fallen object, it may be noted that no moveable aerodynamic devices or engines are present. Thus, besides gravitational force and aerodynamic drag (which together induce a vertical acceleration between  $-g$  and  $\approx 0$  on a falling object), the only variable forces acting during the fall are due to the turbulent flow around the object. The magnitude of those forces might be estimated as no higher than the object’s weight, thanks to a specific CFD analysis. Consequently, according to Newton’s second law, the vertical acceleration can be constrained to lie between  $-2g$  and  $+g$ . Since the unique exact solution to equation (5) has already been found and reported in Fig. 4, these new constraints violate equation (5). That is coherent with the fact that true measurements  $Z_i$  are inaccessible and only imperfect  $Z_{m,i}$  are available, so the *ideal discrete* solution  $\hat{v}$  itself will not satisfy equation (5) when  $\hat{Z}_m$  is used. Therefore, when the new constraints are imposed equation (5) can be solved in a least-squares sense:

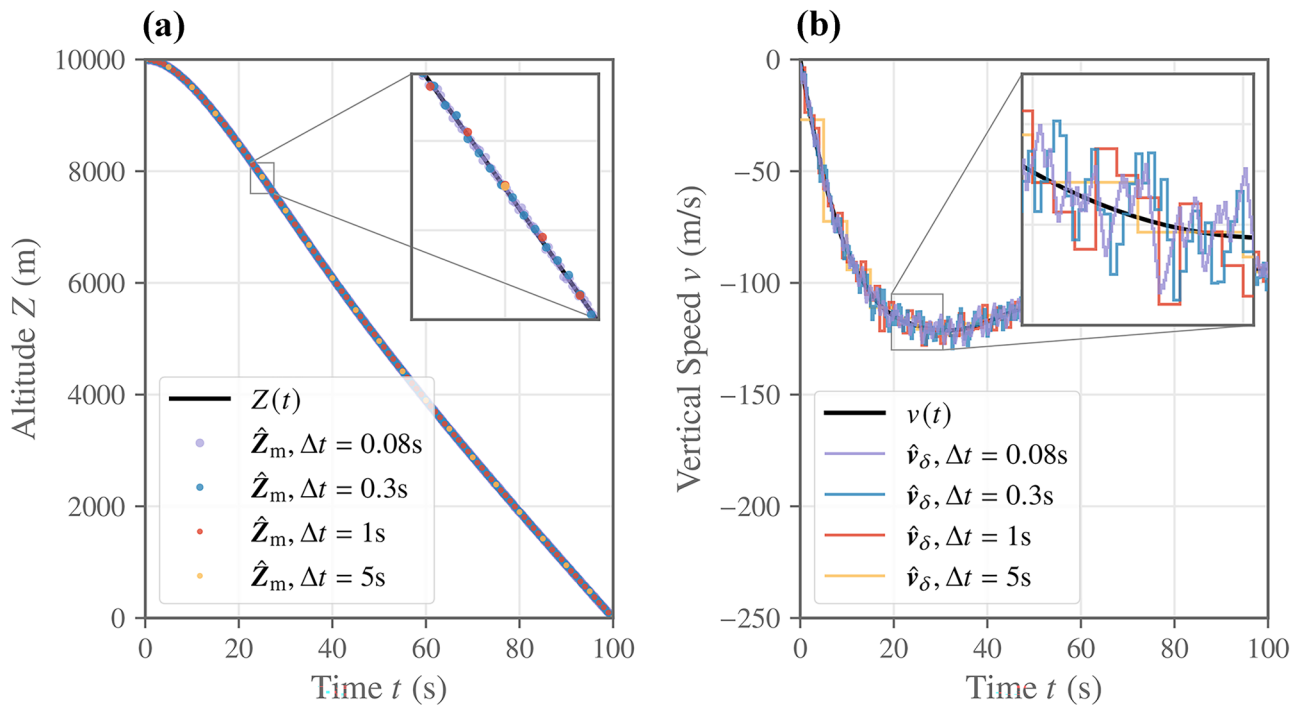
$$v_\delta = \arg \min_v \sum_{i=1}^p \left( v_i - \frac{Z_{m,i} - Z_{m,i-1}}{\Delta t} \right)^2 \tag{9}$$

subject to  $-2g \leq \frac{v_i - v_{i-1}}{\Delta t} \leq +g$   
 $\forall i \in \{2, \dots, p\}$

The solutions of equation (9) corresponding to different GPS sample rates are shown in Fig. 6. The introduced constraints *implicitly filter* the input data and the corresponding solutions: limits on the vertical acceleration exclude the highly oscillating components (generated by measurement errors) that were impairing the unconstrained solutions in Fig. 4, so the resulting variance does not diverge anymore as the discretization scheme is refined. The ill-posed problem of equation (4) was successfully transformed into the well-posed problem of equation (9) by exploiting information that was not available in the original problem.

The solutions of equation (9) reported in Fig. 6 are still quite noisy, with respect to the rather smooth (but unknown) *true* solution  $v(t)$ . However, unless additional constraints or hypotheses are introduced, there is *no way* of telling whether their roughness is an artifact of measurement noise or a true physical phenomenon.

In the mathematical literature, the process of taming an ill-posed problem through additional assumptions to obtain a meaningful solution is called *regularization*. Since the



**Fig. 6** Solution examples corresponding to the same noisy data as Fig. 4, obtained by imposing that the vertical acceleration lies between  $-2g$  and  $+g$ . The variance of the *perturbed* solution  $\hat{v}_\delta$  does not diverge anymore when  $\Delta t$  is reduced. In fact, the new constraints implicitly fil-

ter the highly oscillating components, which impair the unregularized solutions of Fig. 4. Without additional constraints, there is no way to further distinguish the true signal from undesired noise still obtained in the regularized solutions

**Table 1** A non-exhaustive list of common regularization strategies

Regularization technique	Tunable parameters	Implicit assumptions
Discretizing the solution in steps and controlling their size	Step size, variability of steps	“Solution has no meaningful content inside a step”
Expressing the solution as a sum of polynomials	Maximum order of the polynomial basis	“Solution has negligible components on higher order polynomials”
Low-pass filtering input data or the solution	Low-pass frequency	“High-frequency component of the solution can be neglected”
Tikhonov regularization with the Morozov discrepancy principle	Regularization parameter $\alpha$	“For a given penalty functional, the chosen solution is the least penalized among the statistically admissible ones”

unregularized solution  $\hat{v}_\delta$  is typically unacceptably noisy, regularization shows a smoothening and filtering effect on the solution. It shall be clearly stated that the filtering effect is a necessary consequence and not an aim of properly applied regularization. *Regularization does not modify an authentic solution to obtain a more presentable one.* An authentic solution does not even exist, as every solution to the unregularized ill-posed problem comes with an infinite error bound. Regularization defines a new problem, where additional assumptions, constraints or physical inputs overcome ill-posedness and allow one to obtain a manageable solution with bounded errors, not dominated by noise. However, if the solution does not comply with the assumed hypotheses, an unknown yet potentially huge and unwanted bias is introduced into the solution.

In any case, one must always recall that regularization solves a slightly modified problem: *if the regularizing hypotheses were true, what would the solution be?*

For example, numerical derivatives of measurements are almost never computed directly through equation (5). On the contrary, a signal is usually filtered before undergoing differentiation. In this case, it is not the true derivative that is being calculated, but rather the numerical derivative obtained by *assuming that everything above the low-pass frequency does not contain significant information.* In fact, the (arbitrary!) choice of the low-pass frequency implicitly conveys all the prior knowledge of the user about the properties of the signal that is being differentiated.

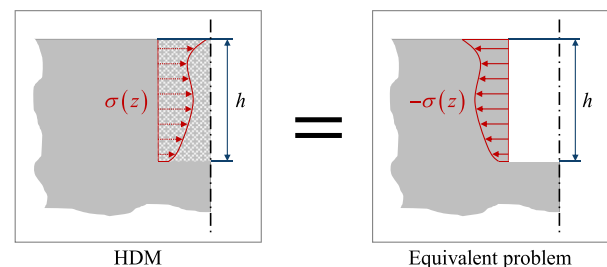
In Table 1 a non-exhaustive list of regularization strategies is shown. The focus is not on the mathematical details, but rather on the implicit assumptions that lie behind every technique. Tikhonov regularization with the Morozov discrepancy principle will be analyzed in the “[Tikhonov Regularization](#)” section, since it is a suggested strategy in the latest ASTM E837 standard for the hole drilling method [43] and it is often used with other methods [54].

## The Hole Drilling Method

### Problem Statement

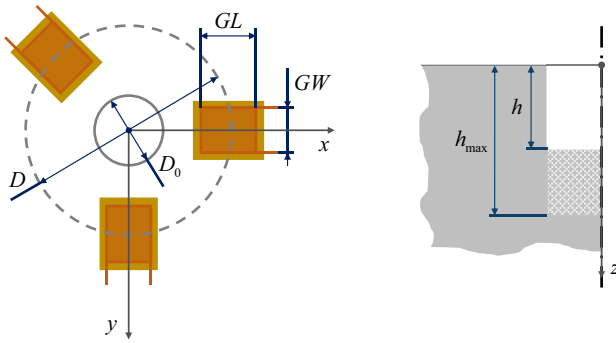
The Hole Drilling Method (HDM), one of the most widely used techniques to measure variable residual stresses in a component, is an interesting inverse problem in the field of mechanical engineering. The method consists of drilling a relatively small hole in progressive steps, and in measuring the corresponding relaxed strains on the surface surrounding the hole. By properly analyzing the measured relaxed strains, the residual stresses affecting the component before the drilling process can be deduced. For an exhaustive introduction and a thorough review of the literature, see Chapter 2 of Schajer’s book [55].

If the drilling induces elastic strains only in a linear elastic material, the problem can be considered linear and Bueckner’s superposition principle [21] holds. Schajer [42] showed that the hole drilling method has an interesting (and computationally easier) equivalent problem under the assumption of plane stress, depicted in Fig. 7 for a representative case of a uniform



**Fig. 7** Application of superposition to define the equivalent problem of the hole drilling method. The removal of a cylinder of material from a stressed body produces stress-strain fields equal to those generated by a traction distribution opposite to the stress distributions previously acting on the newly created surfaces





**Fig. 8** Hole drilling method: variables and geometrical quantities. Note that  $h$  refers to the hole physical depth, while  $z$  is a spatial coordinate

equi-biaxial residual stresses (acting on the plane normal to the hole axis) on a semi-infinite domain. The solution to the equivalent problem (namely, the predicted readings of one or more strain gauges glued on the body’s surface) can be obtained with commercially available finite element software. The only assumption is that the residual stress field at a given depth is uniform in the plane parallel to the specimen surface ( $xy$  plane in Fig. 8), at least on a region having the size of the hole diameter  $D_0$ .

When residual stresses are thought to show significant variations in the  $z$  direction, the traction distribution  $\sigma(z)$  applied on the hole surface can be expressed as a combination  $\sigma_n(z)$  of  $n$  linearly independent stress functions of depth  $z$  (namely, a basis) with the implicit assumption that they can span the domain of residual stress distributions of interest. For example, by choosing a  $n$ -dimensional polynomial basis for a general stress distribution  $\sigma(z)$ , the Power Series Method [23] is defined whose unknowns are the coefficients of a polynomial function.

If the interval  $[0, h_{\max}]$  is divided into a finite number of subintervals  $[h_{i-1}, h_i]$  and their related indicator functions  $\chi_{[h_{i-1}, h_i]}(z)$  (1 in the subinterval  $[h_{i-1}, h_i]$  and 0 outside) are used as a basis, the Integral Method is obtained [22, 56]. In this case, the residual stress distribution is represented by a piecewise-constant function and the unknowns are the residual stress values  $s_i$  within each interval:

$$\sigma_n(z) = \sum_{i=1}^n s_i \chi_{[h_{i-1}, h_i]}(z) \tag{10}$$

As described by the theorems of classical linear elasticity [57], for every admissible traction distribution in the equivalent problem, a solution in terms of relaxed strains exists, is unique, and is continuous with respect to a superposition of a non-null traction distribution. Hence, the *direct* problem of HDM is well-posed.

By indefinitely increasing the discretization refinement in equation (10), the solution to the direct problem eventually becomes a definite integral of the contributions from each infinitesimal subdomain of the traction distribution.

If the residual stress field (hence the equivalent problem too) is axisymmetric, the following equation relates the equi-biaxial stress component with the strain measured by a strain gauge on the surface.

$$\varepsilon(h) = \int_0^h A(z, h) \sigma(z) dz \tag{11}$$

where  $A(z, h)$  is the corresponding *influence function*, depending on the elastic material properties and on the geometry of the hole and the strain gauges.

In the general plane stress case with unequal principal stresses, (at least) three strain gauges are needed and Schajer [22, 23, 32] provided a method with variable transformations such that three decoupled equations (similar to equation (11)) can be individually solved. Without loss of generality, the equibiaxial stress case will be considered here as a typical representative of the three decoupled equations.

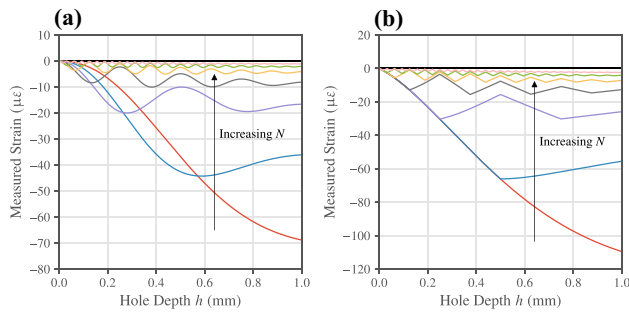
All relaxation methods require the inversion of an equation like equation (11). The Deep-hole drilling [58] and the Contour method [24] are a little bit different in the sense that they measure a displacement distribution *after* the cut is completed; nevertheless, the displacement at any point is the result of the whole residual stress distribution, so an analogous integral equation can be obtained. In addition, although the Contour method does not require the user to carry out the inversion, the FEM software actually solves a similar equation in an already discretized setting.

Equation (11) is a classical Volterra equation of the first kind [59]. Remarkably, a closed form for  $A(z, h)$  is currently known just for a few practical cases, such as Sachs’ Method [60] and the Layer Removal method [61]. In all other cases, only numerical evaluations are available, while some properties of influence functions can be deduced from physical considerations. In particular, they are reasonably smooth [62, 63].

### Ill-posedness of the HDM Inverse Problem

#### The direct problem is a low-pass filter

Like the integration of a function (studied in “[The Simplest Ill-posed Problem](#)” section), the direct problem of HDM works as a *low-pass filter*, though in the *spatial* domain. This effect is shown in Fig. 9(a). As a physical interpretation, opposite peaks in oscillating stress profiles cancel each other’s effects out as they form groups of self-balanced loads on a decreasing length scale. It can also be seen as a consequence of Saint Venant’s principle [64].



**Fig. 9** Measured strains corresponding to zero-mean periodic equibiaxial residual stress distributions with decreasing wavelengths. (a) Sine waves:  $\tilde{\sigma}(z) = 100 \sin\left(N \frac{\pi}{h_{\max}} z\right)$  (b) Square waves:  $\tilde{\sigma}(z) = 100 \operatorname{sgn}\left[\sin\left(N \frac{\pi}{h_{\max}} z\right)\right]$

In rigorous terms, the filtering properties of the direct operator can be proven with the Riemann-Lebesgue lemma (see Sect. 12.5 in [65]), which states that, if  $A(z, h)$  is integrable on its domain of definition, then:

$$\lim_{N \rightarrow \infty} \int_0^h A(z, h) \sin(Nz) dz = 0 \quad (12)$$

$$\forall h \in [0, h_{\max}]$$

It shall come with no surprise that the inversion of equation (11) shares many mathematical properties with the differentiation of a function, even if this time an inverse operator is not explicitly available in closed form.

### The inverse problem is a high-frequency amplifier

Not all linear systems admit a single unique solution, as sometimes no solutions or infinitely many can be found. Actually, in those cases one could look for *least-squares solutions* and pick the one having minimum norm. This process is known as a Moore-Penrose generalized inverse operator (often called *pseudoinverse*), and it *always* leads to a *unique* result. This solution is commonly referred to with a dagger superscript  $\sigma^\dagger(z)$ . Hence, existence and uniqueness of a result obtained with this strategy can be taken for granted.

Common numerical software such as MATLAB or Mathematica have built-in functions (respectively, `pinv()` and `PseudoInverse[]`) to evaluate the pseudoinverse of a matrix, when a finite-dimensional problem is solved. Recall that if a matrix is invertible, its pseudoinverse actually correspond to its classical inverse.

The main observation of this article can now be stated: the HDM inverse problem is ill-posed. The inversion of a low-pass filter yields a *high-frequency amplifier* in the spatial domain, and the amplification gain grows to infinity as the frequency increases. Hence, the solution to the inverse problem of equation (11) can be sent arbitrarily far

from its *true* value even with a negligible perturbation of the measured strains, as long as it has a sufficiently high frequency content.

The intuitive idea of how this impairs the practical inversion process was given in “[The Simplest Ill-posed Problem](#)” section with the differentiation of a function. Noise spans all available frequencies, and the high ones can dominate the solution through excessive amplification. This is the main reason why authors [29, 32] have observed the appearance of high oscillatory components as a result of actual or simulated errors in strain measurements.

A rigorous proof of ill-posedness can be sketched as follows. In  $L^2$  norm:

$$\left\| \sin\left(N \frac{\pi}{h_{\max}} z\right) \right\| = \sqrt{\frac{h_{\max}}{2}} \quad \forall N \in \mathbb{N} \quad (13)$$

Therefore, for every arbitrarily large  $\kappa > 0$  and any arbitrarily low  $\xi > 0$ , a residual stress distribution  $\tilde{\sigma}(z)$  and a corresponding measured strain function  $\tilde{\varepsilon}(h)$  such that  $\|\tilde{\sigma}(z)\| = \kappa$  and  $\|\tilde{\varepsilon}(h)\| < \xi$  can be obtained by choosing  $\tilde{\sigma}(z) = \kappa \sqrt{\frac{2}{h_{\max}}} \sin\left(N \frac{\pi}{h_{\max}} z\right)$  and  $N$  sufficiently high. This implies that for any neighborhood of a given measured strain function  $\varepsilon(h)$  there is a real number  $\kappa$  and sufficiently high  $N$  such that a perturbed function:

$$\varepsilon_\delta(h) = \varepsilon(h) + \int_0^h A(z, h) \tilde{\sigma}(z) dz \quad (14)$$

lies in the neighborhood  $\|\varepsilon_\delta(h) - \varepsilon(h)\| < \xi$  while the corresponding solution:

$$\sigma_\delta(z) = \sigma^\dagger(z) + \tilde{\sigma}(z) \quad (15)$$

has arbitrarily large norm and therefore lies outside of any given neighborhood of the exact solution  $\sigma^\dagger(z)$ . In equivalent terms, the inverse operator is not continuous. No matter how accurate the instrumentation is, the solution error can jump from 0 to arbitrarily huge values as soon as the measured strains deviate from their ideal values. The third Hadamard condition is violated, so the inverse problem is ill-posed.

Note that this happens while a unique *perturbed* solution always exists and can be found, at worst in a least-squares sense; however, it can potentially be infinitely far from the *true* stress profile, in the sense that no guess nor bounds on that distance can be built. It is then clear that a single value with potentially infinite error has absolutely no practical significance.

The effect is not peculiar to sinusoidal perturbations alone, as it concerns the frequency *content* of any function. For example, by substituting sinusoids with square waves similar results are obtained (see Fig. 9(b)).

To limit high-frequency content and avoid this effect, the problem DOFs are often decreased, but that comes at the cost of bias in the solution. Again, a *bias-variance*

tradeoff is encountered. Its consequences will be analyzed in the “[Practical Consequences of Ill-posedness](#)” section.

### Ill-conditioned or ill-posed?

Any mathematical operator can be characterized in terms of sensitivity of outputs with respect to changes in inputs. When a linear system is solved, that property is condensed into the so-called *condition number* [66] of the coefficient matrix. There exists no clear-cut definition of what makes a problem *ill-conditioned*. Indeed, ill-conditioning is more of a technological limit than a mathematical property. Any desired output accuracy can be obtained through *sufficiently accurate* inputs. When keeping the solution variance below a desired threshold becomes a technological challenge for the required input accuracy, the problem is said to be ill-conditioned.

On the other hand, ill-posedness is a very specific mathematical property and is defined by the Hadamard conditions, discussed in the “[Basic Concepts](#)” section. The cause of ill-posedness is not technological *at all*, as it is rather a property of the underlying mathematical structure, which here is the inversion of a particular integral operator. A discontinuous ill-posed problem has *infinite* sensitivity to input errors in its functional form. In any discretized (hence regularized) form, the variance of the solution is inversely correlated with the regularization bias. Eventually, all ill-posed problems become practically ill-conditioned when the solution has sufficiently many DOFs. Bias-variance tradeoffs are the most notable signature of ill-posed problems. Depending on the specific combination of input errors, solution DOFs and required accuracy, an ill-posed problem can actually turn out to be well-conditioned.

The fact that some relaxation methods have low measurement sensitivity (because of the distance between the stress location and the strain probe) provides an *additional* challenge and worsens the compromise between bias and variance, but,

again, that is not the *true cause* of ill-posedness. For instance, the Contour method is still an ill-posed problem despite the fact that deformations are measured precisely where the desired stress fields lie.

To sum up, ill-conditioning and ill-posedness are two very distinct properties and should not be confused. *Relaxation methods are ill-posed problems*. Due to their typical experimental configuration and due to the current technological state of the art, in practice they are often ill-conditioned too, even when solutions DOFs are kept reasonably low.

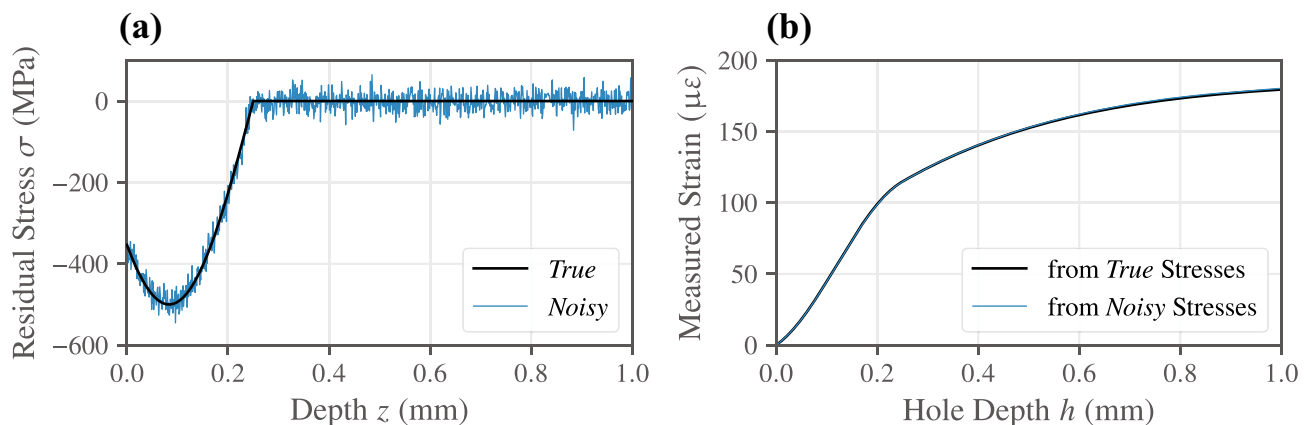
Ill-posed problems also show a high sensitivity to an imperfect knowledge of the *true* direct operator. Since influence functions of relaxation methods are seldom known in an exact closed form, this fact provides an additional source of error, which for brevity will not be analyzed here. The interested reader can find more details in a specialized text [67]. Neglecting this error in numerical analyses causes an underestimation of both the solution bias and variability. The mathematical literature coined a specific (and quite ironic) term for this kind of mistake, by calling it an *inverse crime* [68].

### A shot peening example

The discussion on ill-posedness will be carried out using a practical example of a residual stress distribution hypothetically produced by a shot peening treatment on a thick steel component ( $E = 206 \text{ GPa}, \nu = 0.3$ ). Considering reasonable numerical values, expressing  $z$  in mm and  $\sigma(z)$  in MPa, the following cosine-like equi-biaxial residual stresses distribution is assumed:

$$\sigma(z) = \begin{cases} -500 \cos \left[ \frac{\pi}{4}(12z - 1) \right] & z \leq 0.12 \\ 0 & z > 0.12 \end{cases} \quad (16)$$

in accordance with the classical textbook of Schulze [69]. The distribution is plotted in Fig. 10. For simplicity, having



**Fig. 10** (a) Hypothetical equi-biaxial residual stress distribution produced by a shot peening treatment, used as an example of the inversion process. The residual stress distribution is shown with and without an additional Gaussian noise having standard deviation  $\zeta = 20 \text{ MPa}$ . (b)

The simulated relaxed strains measured by the strain rosette, arising from the equi-biaxial residual stress fields of equation (16), are virtually identical

assumed a sufficiently thick specimen, tensile residual stresses for  $z > 0.12$  mm are neglected. The polynomial fitting reported in [62] was adopted for the influence function. A standard type-A strain rosette with diameter  $D = 5.13$  mm was considered, according to ASTM E837 standard [43], and a hole with diameter  $D_0 = 2.05$  mm was assumed.

Combining equations (11) and (16), the direct problem can be solved. The corresponding strains  $\varepsilon(h)$  measured by the rosette are shown in Fig. 10(b). It can be verified that the direct problem is highly robust to noise-like variations of input data, due to its low-pass filtering properties. As shown in Fig. 10, the superposition of a severe noise on the assumed residual stresses produced an almost negligible effect on the strain curves.

**Practical consequences of ill-posedness**

As described for the example in “The Simplest Ill-posed Problem” section, the HDM problem is never solved in its functional form. Measurements are sampled at a finite number  $p$  of hole drilling steps  $h_i$ , which produce an array  $e_m$ , modeled as the superposition of *ideal discrete* values  $e$  with measurement errors. Consequently, equation (11) cannot be solved, due to a partial and imperfect knowledge of  $\varepsilon(h)$ . The collocation method [70] is usually employed to transform equation (11) into a linear system. To do this, the space of residual stress functions is approximated by the span of a  $n$ -dimensional basis  $\beta = [\beta_1(z), \beta_2(z) \dots \beta_n(z)]$ . This operation is a *projection* of the *true* solution  $\sigma(z)$  onto its approximation  $\sigma_n^\dagger(z) = \sum_{i=1}^n s_i \beta_i(z)$  in a finite-dimensional space, which can then be *represented* by the coefficients  $s = [s_1, s_2 \dots s_n]$  with respect to the chosen basis. After that, the elements  $A_{ij}$  of a  $p \times n$  matrix  $A$  are built by evaluating the effect of an element of the basis  $\beta_j$  on the sample  $e_i$  through the direct integral operator:

$$A_{ij} = \int_0^{h_i} A(z, h) \beta_j(z) dz \tag{17}$$

Eventually, a linear system with  $n$  variables and  $p$  known constants is obtained:

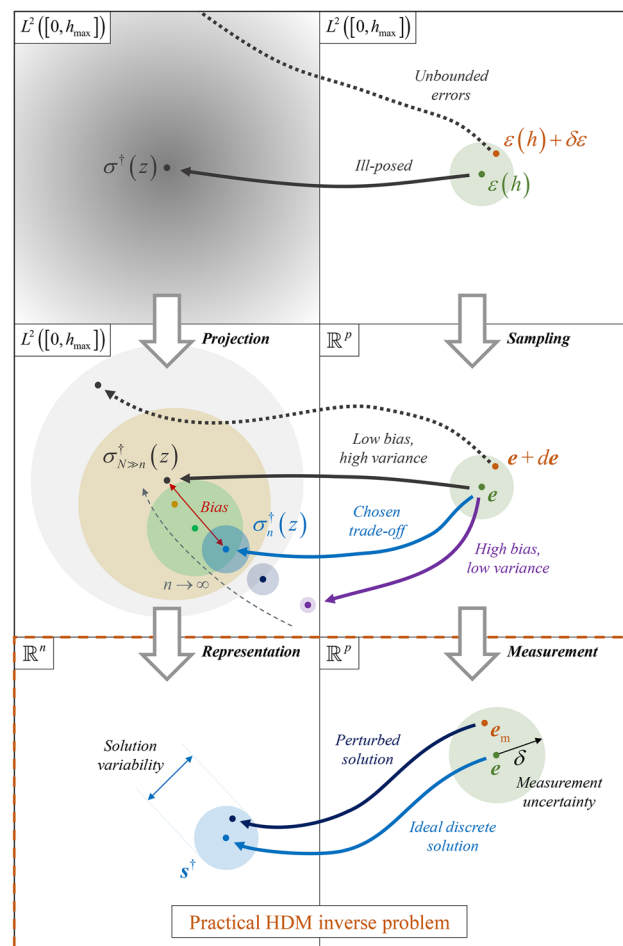
$$As = e \tag{18}$$

By projecting the original functional stress space onto a span of a finite-dimensional basis, a bias is introduced in the solution (unless the *true* solution can be proven to *actually* belong to that span). On the other hand, by restricting the solution DOFs to a finite number  $n$ , its high-frequency content is *implicitly* limited. This operation has a stabilizing effect on the solution variance, similarly as shown in the “Regularization” section.

Discretization transformed an ill-posed discontinuous problem into a well-posed continuous problem, though cursed by a bias-variance tradeoff: no matter how refined, *discretization is always a regularizing operation*. It is usually assumed that a

more refined numerical scheme leads to a better approximation of the underlying functional problem, and in a broad sense that is still true. Recall that, since the functional inverse problem is not continuous, it has infinite sensitivity to measurement noise. This is consistent with the fact that the ill-conditioning of the linear system in equation (18) diverges as the solution DOFs are increased. By refining the numerical scheme, the discretized problem better mimics the actually undesired discontinuity of the functional problem. An equivalent behavior can be observed when the maximum degree of polynomials in the Power Series method is increased [41]: like a reduction of the integration step size, it is an extension of the solution DOFs.

This behavior is summarized in Fig. 11 and is perfectly acknowledged in the field, as it has been studied as an



**Fig. 11** Graphic explanation of the regularization properties of discretization. Single elements are reported as solid circles, while their uncertainties are represented as colored shadows. When the solution  $\sigma^\dagger(z)$  is projected onto the span of a finite-dimensional basis, a well-posed problem is obtained and the number of DOFs is chosen to obtain a manageable variance while limiting the representation bias. Projected stresses  $\sigma_n^\dagger(z)$  are then represented as a vector  $s^\dagger$  and obtained as a solution to a linear system. Its numerical conditioning allows to estimate the solution variability, but the representation bias is inaccessible to the user



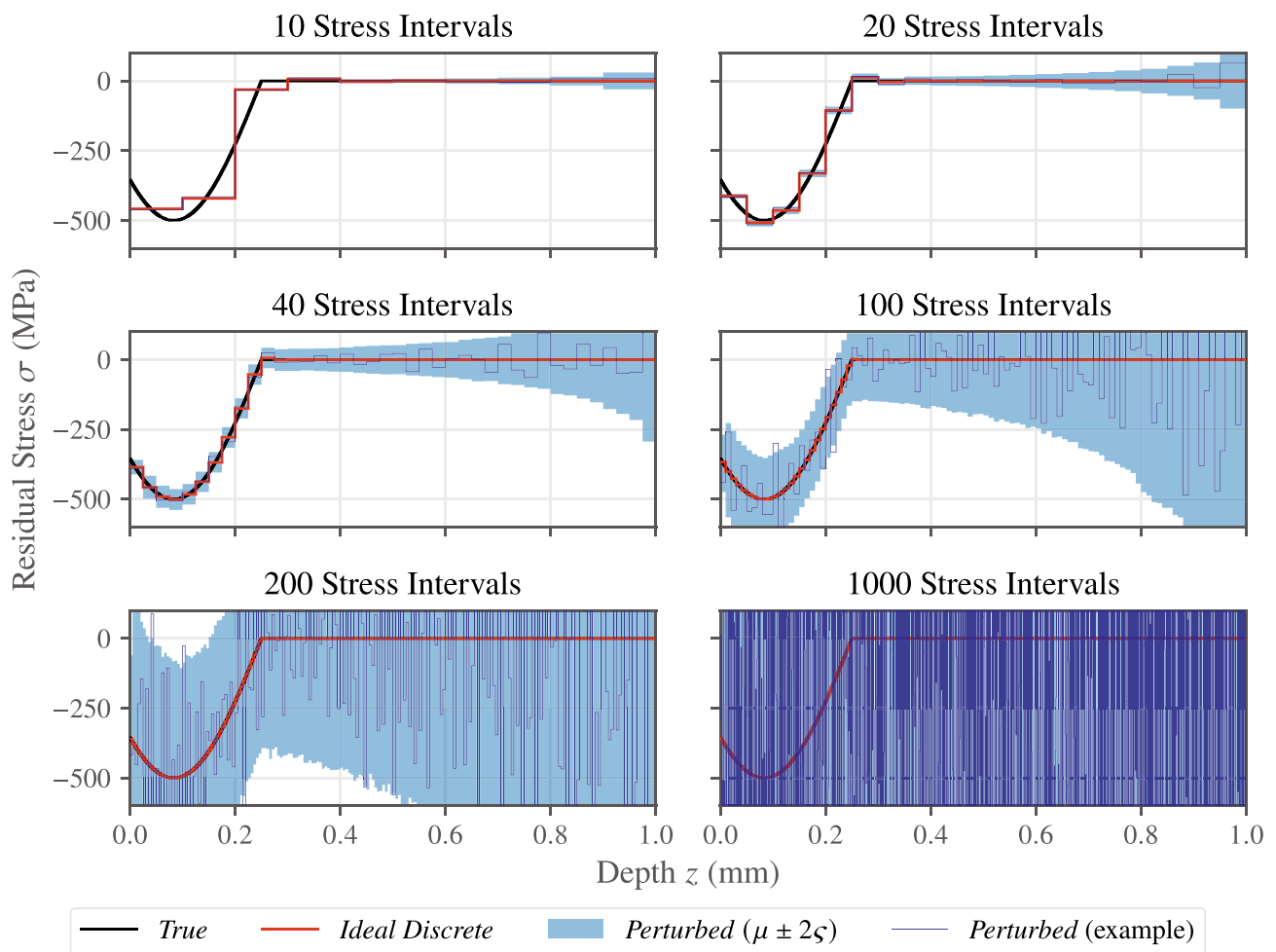
optimization [29, 71–73]: a smaller integration step yields a better spatial resolution but a higher ill-conditioning, a greater integration step leads to a well-conditioned problem but with a poorer spatial resolution. Schajer and Altus [27] defined this as the “Heisenberg Uncertainty Principle” of residual stress measurements. Non-uniform integration steps have also been explored to fully exploit depth-dependent error sensitivities.

Recall the shot peening example from the “A Shot Peening Example” section. Since this example is purely theoretical, the *true* residual stress distribution and its corresponding continuous relaxed strains are available. To simulate a practical measurement, relaxed strains are sampled. The inverse problem is solved starting from a 1000-points sampling of the theoretical  $\varepsilon(h)$  with uniform depth steps. This uncommonly high value with respect to typical practical measurements is used to get very close to an analytical solution. Although technologically questionable, this very high

number of intervals is actually beneficial to the solution, as will be shown in the “Discussion” section. An independent Gaussian random error with a standard deviation of  $1 \mu\epsilon$  is then added to measurements. The influence functions are assumed to be exact, so an *inverse crime* is actually committed. That is fine for illustration purposes.

To apply the collocation method, the stress basis of the Integral Method (piecewise constant functions) is used. The solution domain is divided in equally spaced sub-intervals. Various refinement levels are attempted; the spatial resolution is proportional to the number of calculation points. Note that only the number of calculation points is changed, while the number of strain samples is kept constant. To explore the solution variability, 1000 different realization of Gaussian noise are created and just as many *perturbed* solutions are obtained.

Results are shown in Fig. 12. The *ideal discrete* solution converges to the *true* stresses as the spatial refinement of the



**Fig. 12** Solution to the HDM inverse problem, trying to obtain the shot peening residual stress distribution reported in equation (16) from simulated strain measurements. Plots are reported for increasing solution DOF. The theoretical relaxed strains are sampled at 1000 equidistant points, and the *true* and the *ideal discrete* solution are

reported. Then a Gaussian noise with standard deviation  $\zeta = 1 \mu\epsilon$  is added and 1000 different realizations of the *perturbed* solution are created. They are reported as  $\pm 2\zeta$  intervals, together with an example. The solution variance diverges as the DOFs increase





solution space basis is increased. On the contrary, when noisy strains are fed into the problem, convergence is not achieved anymore. As the degrees of freedom of the solution increase, the accuracy of the *perturbed* solution gets severely impaired.

If one wanted to minimize the *perturbed* solution error with respect to the *true* solution, it may be tempting to choose the degree of refinement to minimize the observed solution variability, but it would actually be a very poor choice. Due to the bias-variance tradeoff, the least variable solution is *always* the one with least DOFs, which in this case is a constant stress distribution. However, this would lead to an unacceptable bias in many practical cases. Variance is observable and/or simulable from input data, bias is not, as Fig. 11 points out.

An experienced operator incorporates his prior knowledge of the specific practical situation to choose DOFs wisely, so as to limit variance within an acceptable bias. The achieved solution variability is therefore a combined effect of two distinct factors: the measurement error *and* the chosen discretization scheme. If and only if the latter did not introduce any significant bias, then the solution variability provides useful information on the solution *global* error.

For the above-mentioned reasons:

- Choosing between different regularization strategies applied to a common residual stress experiment by comparing the variability of their solutions is completely meaningless. The actual error has also a bias component, which is always compromised at the expense of variance. A constant stress solution is always the least variable, and any robust solution could be the result of a greater (and *hidden*) bias.
- In the authors' opinion, the output of a residual stress measurement obtained through relaxation methods shall not be limited to the stress results together with their expected variability. The chosen regularization scheme shall be clearly stated too, as it discloses any potential bias. Different experimental setups can be compared only if the regularization strategy is kept the same; only in this case would a mitigation of the solution variability signal a practical improvement, as it would not be achieved through the introduction of additional biases.

## Tikhonov Regularization

### Formulation in functional spaces

It is often desirable to be able to choose “how much” the solution has to be regularized, and to do it in a computationally fast way. As seen in the “[Practical Consequences of Ill-posedness](#)” section, the discretization of a problem is itself a form of regularization; however, changing the discretization scheme involves a lot of numerical integrations to evaluate the coefficients matrix  $A$  in equation (18), and that is usually not a computationally easy task.

The mathematical theory of ill-posed problems provides a solution to this challenge, usually (but not solely) by approximating the ill-posed inverse operator with a well-posed one and by controlling the degree of approximation through an appropriate choice of a continuously variable regularization parameter. For an overview of classical methods, see [46–49]. Tikhonov regularization [67] is one of the most-widely used techniques; it has been previously applied to the hole drilling method inverse problem [32, 38, 74] and it is currently included in the ASTM E837 standard [43] as an approved procedure for the calculation of non-uniform residual stresses with the HDM.

An unregularized least-squares solution to equation (11), corresponding to the given measured strains  $\varepsilon(h)$ , is obtained as:

$$\sigma^\dagger(z) \triangleq \arg \min_{\sigma} \|T(\sigma(z)) - \varepsilon(h)\|^2 \quad (19)$$

Therefore, it looks for a minimum of the so-called squared *discrepancy* (also known as *residuals*)  $\delta^2 \triangleq \|T(\sigma(z)) - \varepsilon(h)\|^2$  of the measured strains from the strains related to the computed solution, and picks the corresponding argument  $\sigma^\dagger(z)$ . Note that the squared discrepancy  $\delta^2$  is a function of the solution  $\sigma(z)$ .

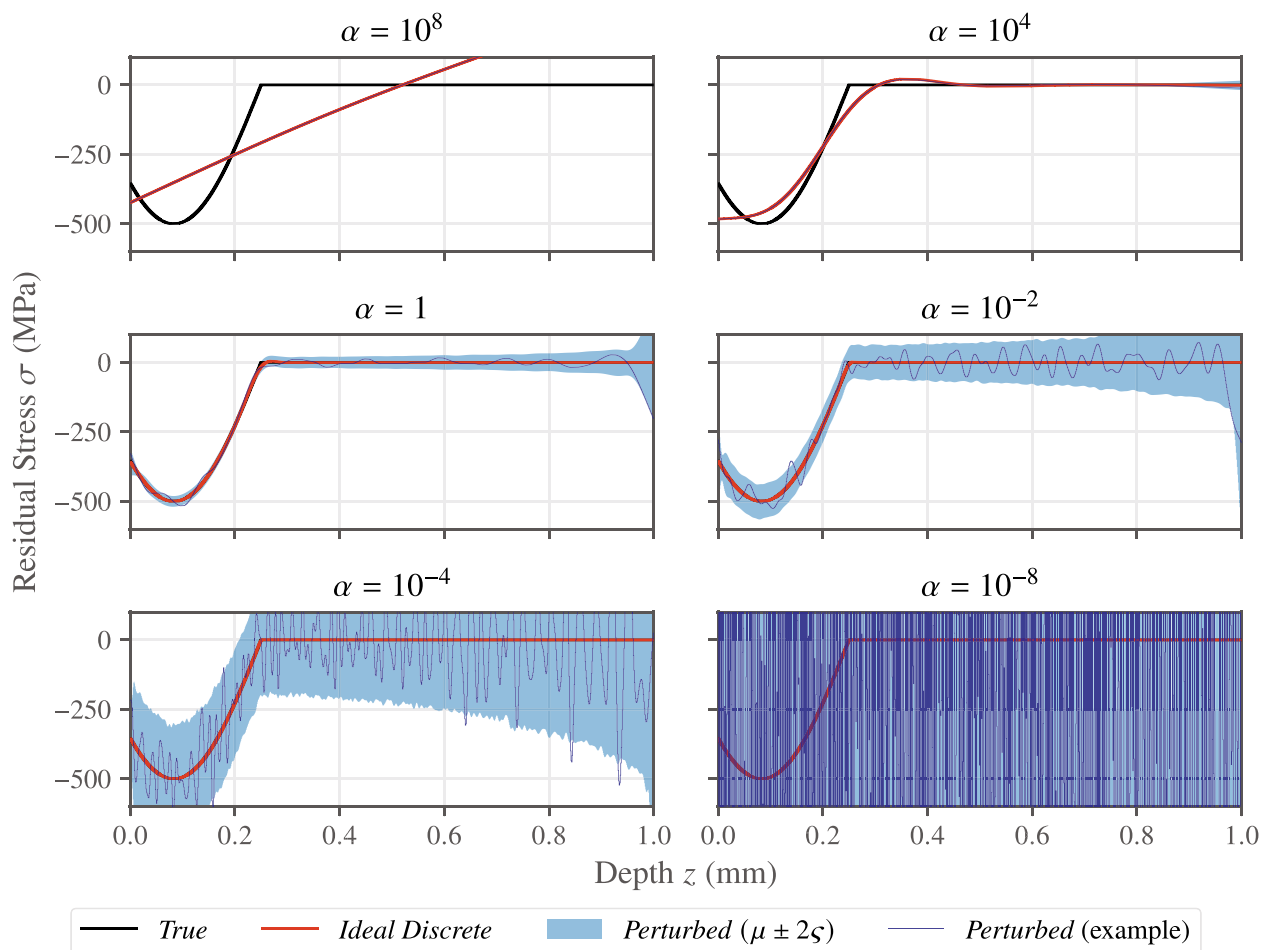
On the other hand, in its general form Tikhonov regularization seeks the minimum of another functional, obtained by adding a penalty term, proportional to a regularization parameter  $\alpha$ :

$$\sigma_\alpha^\dagger(z) \triangleq \arg \min_{\sigma} \left( \|T(\sigma(z)) - \varepsilon(h)\|^2 + \alpha \|\Gamma(\sigma(z))\|^2 \right) \quad (20)$$

where  $\Gamma(\cdot)$  is a user-defined operator. When solving Problem 20 it is not the original Problem 19 that is being solved, but rather an approximation of it, whose refinement is controlled by the regularization parameter  $\alpha$ . The operator  $\Gamma(\cdot)$  of the penalty term is usually a high-pass filter, to penalize the roughness of the solution. A common choice in relaxation methods is to penalize the  $L^2$  norm of the solution or of its higher derivatives, as will be analyzed in the “[Practical Application](#)” section. In this case, it can be proven that equation (20) defines a well-posed continuous problem  $\forall \alpha > 0$  (see [46]).

When  $\alpha \rightarrow 0$ , the regularized problem tends to the original discontinuous one. As  $\alpha \rightarrow +\infty$  the penalty functional alone is actually minimized, so in general equation (20) represents a completely different (and usually well-conditioned) problem. The parameter  $\alpha$  defines another typical bias-variance trade-off. It can be observed in Fig. 13, using the same shot peening example of Fig. 12. However, if compared to the number of solution DOFs,  $\alpha$  has a less obvious physical meaning as a regularization parameter. Indeed, without a *physically driven* choice for  $\alpha$ , equation (20) has no practical sense.





**Fig. 13** Solution to the HDM inverse problem through second-order Tikhonov regularization, trying to obtain the shot peening residual stress distribution reported in equation (16) from hypothetical strain measurements in the same manner as Fig. 12. Different values of

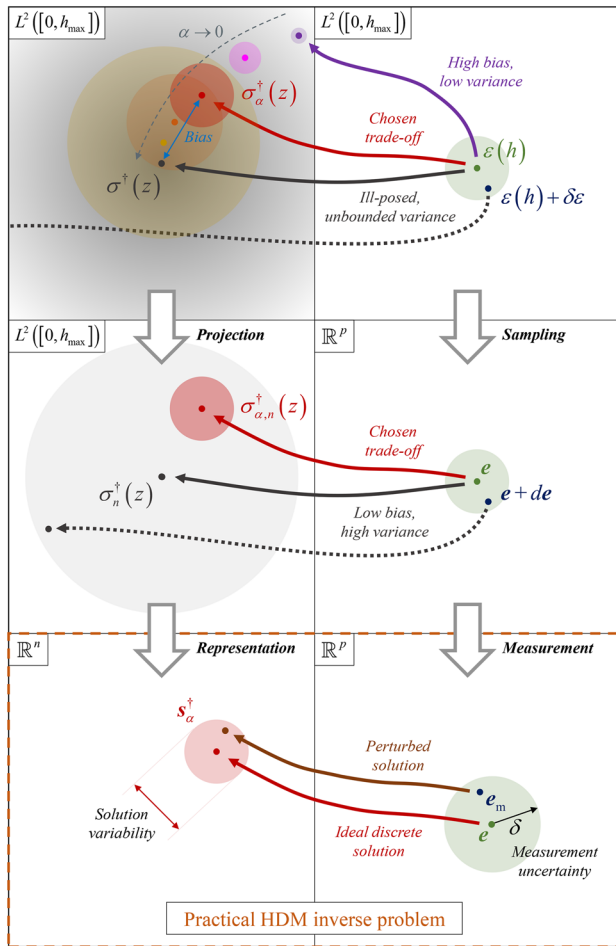
the regularization parameter  $\alpha$  are explored, while the discretization scheme is kept fixed at 1000 stress intervals. Another bias-variance tradeoff is obtained, controlled by the value of  $\alpha$

Since the problem is already regularized in its functional form (before discretization!), there is no need to compromise the number of DOFs anymore, yet a discretization scheme is anyway needed for a numerical solution  $\sigma_{\alpha,n}^\dagger$ , which is the projection of  $\sigma_\alpha^\dagger$  onto the span of the chosen finite-dimensional basis  $\beta$ . In practice, a very refined discretization is usually used, in order to leave the regularization level to the parameter  $\alpha$  only. This process is summarized in Fig. 14. Tikhonov regularization can, however, be combined with other regularization strategies, including but not limited to a coarse discretization scheme.

A reasonable choice for  $\alpha$  (and in general, of any regularization parameter) stems from the accuracy of the specific instrumentation that is being used. This is reflected in the strategy known as the Morozov Discrepancy Principle [75]. As seen before,  $\alpha = 0$  corresponds to the least-squares solution. Since data are affected by errors, the tightest match will most likely

be an overfitting of noise. If an oracle would be able to tell the *true* stress distribution, its corresponding relaxed strains would actually be different from the measured ones, due to experimental errors. The expected squared value of that difference is taken as a criterion to tune the regularization level.

In practice, the Morozov discrepancy principle states that  $\alpha$  may be increased until the squared discrepancy matches its expected value  $\bar{\delta}^2$  (in a deterministic or, more often, in a statistical sense). Recall that  $\delta^2$  is a function of  $\sigma(z)$ , while  $\bar{\delta}^2$  is a real positive value that is assumed to be known. The Morozov discrepancy principle assumes knowledge on how far the measured strains are from their *ideal* values, which may be difficult to determine and thus adds another error source. In the case of HDM, a simple and statistically justified method was proposed in [31, 32], which then was merged into the ASTM E837 standard; this procedure can easily be extended from hole drilling to general strain measurements.



**Fig. 14** Graphic explanation of Tikhonov regularization. Regularization occurs in the underlying functional spaces at the expense of a bias in  $\sigma_\alpha^\dagger$ . Then, a very refined discretization (which would lead to an excessive variance without Tikhonov regularization) can be adopted. Projected stresses  $\sigma_{\alpha,n}^\dagger$  are then represented as a vector  $s_\alpha^\dagger$  and obtained as solution to a linear system. The solution variability can be estimated through the condition number of the coefficient matrix, but the regularization bias is inaccessible to the user

It shall be emphasized that the Morozov discrepancy principle is only one among many strategies proposed in the mathematical literature for the choice of the regularization parameter, though it is one of the most used. In addition, its application is not limited to Tikhonov regularization, since it may be used with any other strategy (such as discretization itself). As with *all* other strategies, there is no guarantee that the chosen parameter is the *best* in terms of solution error. Nevertheless, some statistically optimal properties can be proven, which the interested reader can find in a specialized textbook such as [47].

When the measurement error is not known at all, the mathematical literature provided ill-posed problems with regularization parameter choice criteria that are specifically designed to work without that piece of information, such as: *Generalized Cross-Validation* (GCV) [76], the *L-Curve* criterion

[77], *Iterative Predictive Risk Optimization* (I-PRO) [78] or the *Quasi-Optimality* principle [79].

**Practical application**

Assuming that matrix  $A$  has full rank (which is usually the case in relaxation methods), the least squares solution to equation (18) is given by:

$$s^\dagger = \arg \min_s \|As - e_m\|^2 \tag{21}$$

In the context of relaxation methods, Tikhonov regularization is usually realized by penalizing the norm of the solution second derivative [32]; this strategy is usually referred to as *second-order Tikhonov regularization*. Denoting with  $C$  the discrete operator that approximates the second derivative of a function (namely, the second-order finite difference operator corresponding to the chosen stress basis), Problem 20 in its discretized form can be written as:

$$s_\alpha^\dagger = \arg \min_s \left( \|As - e_m\|^2 + \alpha \|Cs\|^2 \right) = (A^T A + \alpha C^T C)^{-1} A^T e_m \tag{22}$$

A useful reference on how to build finite difference approximations can be found in [80]. When hole steps are uniform with interval  $\Delta h$ , matrix  $C$  has a simple form:

$$C = \frac{1}{\Delta h^2} \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix} \tag{23}$$

Eventually, the application of the Morozov discrepancy principle can be formulated as a constrained optimization problem:

$$(\alpha^*, s_\alpha^\dagger) = \arg \min_{\alpha, s} \|As - e_m\|^2 + \alpha \|Cs\|^2 \tag{24}$$

subject to  $\|As - e_m\|^2 \leq \bar{\delta}^2$

The target value  $\alpha = \alpha^*$  and the corresponding regularized solution  $s_\alpha^\dagger$  are found by numerical means. As it is stated, Problem 24 does not convey a clear idea of how the regularized solution is chosen. Nevertheless, it has an interesting equivalent problem:

$$s_\alpha^\dagger = \arg \min_s \|Cs\|^2 \tag{25}$$

subject to  $\|As - e_m\|^2 \leq \bar{\delta}^2$

The equivalence can be proven by enforcing Karush-Kuhn-Tucker (KKT) conditions [81] in Problem 25 and excluding some trivial cases of no practical significance. Problem 25 has a clear physical meaning:

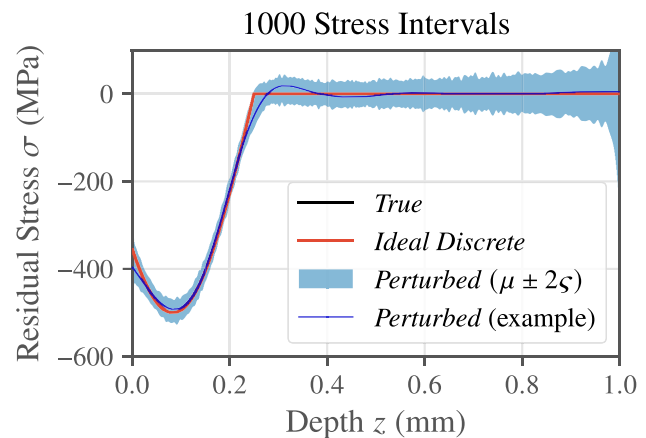


- $\|As - e_m\|^2 \leq \bar{\delta}^2$  defines an *admission criterion* between solutions. Instead of an infinite refinement, which would lead to an excessive variance, all solutions that are statistically compatible with the obtained measurements and their errors are considered as equivalent.
- $\arg \min \|Cs\|^2$  defines a *preference criterion* between solutions. Among all the equivalent solutions, the minimizer of a penalty functional that disadvantages those with high second derivatives (hence including highly oscillatory solutions) is chosen.

When coupled with the Morozov discrepancy principle, Tikhonov regularization changes the original ill-posed question of finding an exact solution with another query: for a given penalty functional, which is the *least penalized* solution among a set of admissible ones, built from knowledge of measurement uncertainties? In this case, the penalty functional *favors* low second derivatives, hence smooth solutions with gently varying gradients. It does not strictly assume anything *at all*, since it simply provides an answer to a different question. Due to this, it is a very valuable strategy when nothing about the solution can be strictly assumed *a priori*. It is also quite reasonable: since slight experimental noise always affects results through highly oscillatory terms, it is more likely that a smooth admissible solution is closer to the *true* solution, than being coincidentally a perturbation of a very rough *true* solution. Indeed, a very interesting statistical interpretation of Tikhonov regularization is available [82], as a *maximum a posteriori* estimation corresponding to a prior distribution of the possible solutions. Whether the *least penalized admissible solution* is actually close to the *true* solution, it depends on the properties of the latter, and that is where physical considerations must come into play. Solution variability will not help in quantifying the global solution error, as was the case for discretization. This fact is summarized in Fig. 14.

Recalling the shot peening example, second-order Tikhonov regularization with the Morozov discrepancy principle has been applied to its inverse problem. Results are reported in Fig. 15. Since the problem defined by equation (20) is well-posed in its functional form, the solution is convergent as the integration step is reduced and its variance does not diverge anymore. The compromise between bias and variance is already controlled by  $\alpha$ , as shown in Fig. 13: in fact, the attached *perturbed* solution example is slightly smoother than the *true* solution, (i.e. the sharp corner point is not well reproduced).

As measurements are increasingly affected by noise, which means  $\bar{\delta}$  is increased, the admissible set of solutions in Problem 25 grows, hence second-order Tikhonov regularization and the Morozov discrepancy principle actually pick a smoother solution (as it has a lower penalty value). This is of great significance for the practical user: poor measurement accuracy will generally yield *smoother* regularized solutions (so with



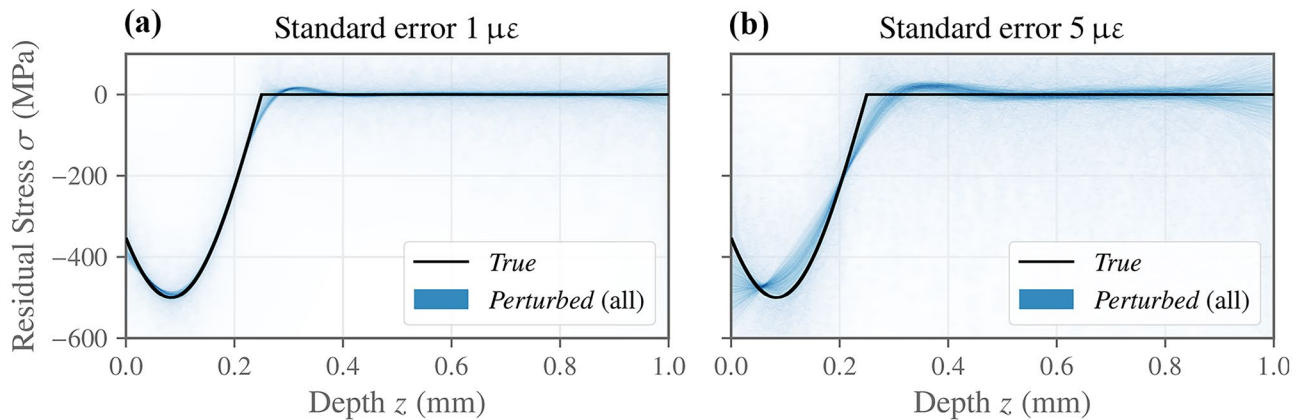
**Fig. 15** Solution to the HDM inverse problem through second-order Tikhonov regularization and the Morozov discrepancy principle, trying to obtain the shot peening residual stress distribution reported in equation (16) from hypothetical strain measurements in the same manner as Fig. 12. Compared to Fig. 12, the variance no longer diverges when a very refined spatial resolution is adopted

poorer spatial resolution) than those arising from precise instrumentation, and that may be quite misleading. For example, the 1000 shot peening cases were also simulated with measurement noise having a standard error of  $5 \mu\epsilon$ , which corresponds to a quite low signal-to-noise ratio. The solutions are compared with the previous ones in Fig. 16. It is then clear why an accurate estimate of the noise properties is at least as important as the noise level itself: a poor estimation leads to an inaccurate admission criterion, which can then result in dangerous under- or over-regularization.

## Discussion

Any process that limits the solution DOFs and/or filters high frequencies has a regularizing effect. A variably refined discretization and Tikhonov regularization are only two among many strategies to deal with ill-posed problems. As long as problem linearity can be assumed (as in most relaxation methods), applying a low-pass linear filter to input strains or to output stresses yields the same results, as *they are actually the same operation*.

When prior knowledge suggests that the solution coefficients with respect to a given basis will be significantly *sparse*, it is certainly convenient to adopt that basis and enforce sparsity by truncating the corresponding solution expansion at a limited number of elements. Besides the Integral Method, several different basis have been explored: polynomials [41, 83, 84], splines [72, 85], Fourier series [24] and wavelets [28, 86]. It cannot be stressed enough that all these methods should not be compared by their achieved sensitivity to noise, and that the selection should rather be made based on physical arguments and real-world cases. If



**Fig. 16** Comparison of solutions obtained through second-order Tikhonov regularization and the Morozov discrepancy principle, arising from measurements having different noise levels. A discretization scheme with 1000 stress intervals is used. 1000 random realizations

no assumptions can be made, it can actually be shown (see Sect. 3.3 in [46]) that none of the above-mentioned methods is optimal: for a given number of DOFs, the minimum variability is achieved with a *truncated singular value decomposition* (TSVD) [87] of the integral operator.

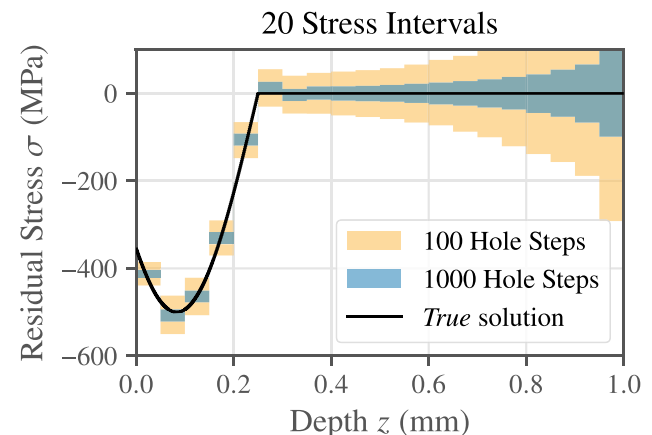
An ill-posed problem is regularized by a limitation of the solution DOFs, and the latter in turn are limited by the measurement samples, so reducing the dimension of input data seems to improve the ill-conditioning of the problem [29, 71]. However, for a given regularization bias (governed by the chosen solution basis), having more measurements is *always* beneficial, as it reduces the solution variance. This happens because an overdetermined system exploits redundancy and implicitly averages out random errors by fitting a solution having less DOFs than measurement points. This effect is shown in Fig. 17 and is of fundamental significance for the design of relaxation experiments. Similar properties hold for Tikhonov regularization with the Morozov discrepancy principle [47]. Therefore, compatibly with technological and timing constraints, *the highest number of measurements should be taken*, as observed in [30]. Then, a suitable regularization level should be tuned according to prior knowledge and physical principles.

Standard uncertainty estimates only capture the variance and fail to capture the bias. Since residual stress measurements can be used for life critical structural integrity assessments, such a non-conservative uncertainty estimate is cause for concern. This issue has been recognized under the name of *model error* by Prime and Hill [30], with an attempt to estimate the additional uncertainty for the case of a series expansion method applied to incremental slitting. While adequate for their example problems, the method cannot be conservative for the general case. As a matter of fact, bias-variance tradeoffs and underestimated uncertainties affect

of a *perturbed* solution are reported in the plots. A greater measurement noise is fairly rejected, though at the cost of an increased solution bias towards smoothness. (a)  $\zeta = 1 \mu\epsilon$  (b)  $\zeta = 5 \mu\epsilon$

measurements throughout experimental mechanics, because many of them include an inverse problem such as fitting a diffraction pattern to a Gaussian peak for neutron diffraction measurements of residual stress.

Olson et al. [88, 89] proposed a method to evaluate the regularization uncertainty in relaxation methods, similar to the model error estimation of Prime and Hill [30]. The uncertainty due to a given regularization level is estimated by varying the regularization parameter in an appropriate interval and evaluating the corresponding change in the solution. The specific properties of that interval are tuned on both numerical and practical experiments. Then, the



**Fig. 17** Comparison of *perturbed* solutions with 20 DOFs, arising from two different number of strain samples (100 and 1000 hole steps), which share the same Gaussian noise with standard deviation  $\zeta = 0.1 \mu\epsilon$ . 1000 different realizations of the *perturbed* solution are created and reported as  $\pm 2\zeta$  intervals. Tikhonov regularization is not applied. For given stress DOFs (hence solution bias), having more measurement points is beneficial to the solution variance



regularization uncertainty is combined with the uncertainty due to input noise in order to pick the regularization parameter at the minimum of the total combined error. Since an optimal regularization level corresponds to a minimum of the distance between the *perturbed* and the *true* solution, it is definitely reasonable to expect that it occurs at a place where the solution is not very sensitive to changes in the regularization level and to input noise. A similar parameter choice strategy is known in the mathematical literature as *Quasi-Optimality* principle [79], and its properties are well known. Nevertheless, the uncertainty value that is obtained through the method should not be interpreted as a general estimator of the solution error, because the full extent of the bias component remains unknown. Anyway, in the study of problems that are similar to the ones used to calibrate the procedure, it is reasonable to expect that the estimated uncertainty is a valid heuristic.

In a recent article, Tognan et al. [90] applied Gaussian Process Regression (GPR) to the Contour method to perform input data filtering and obtain residual stress uncertainties through a procedure that does not depend on the user's discretion. This promising approach shares many properties with Tikhonov regularization and the Morozov discrepancy principle. It assumes that close samples of the *true* solution are highly correlated in space, according to a specific covariance kernel that is governed by a finite set of variable *hyperparameters* that can tune this relation. Like Tikhonov regularization, it generally penalizes roughness, while controlling the regularization level through tunable parameters. Like the Morozov discrepancy principle, the hyperparameters are chosen on the basis of input data, by assuming a given measurement noise. In this case, they are chosen by maximizing likelihood over input data. Once again, the resulting stress uncertainty is actually the solution variance corresponding to the chosen regularization bias, which depends on the kernel choice and on the estimation of measurement variance. If the *true* solution does not follow the assumed kernel (e.g. due to discontinuities), the introduced bias may be arbitrarily high.

As a final consideration on this ill-posed problem, it should be noted that its undesirable mathematical properties are a side effect of the continuum assumption, which is instead fundamental for the direct problem and for the whole theory of elasticity. An unbounded sensitivity to perturbations is achieved only if stress distributions are allowed to include arbitrarily high frequency contents, but the latter get physically meaningless as their wavelengths become smaller than the typical grain size. This may suggest a sort of “natural” regularization, obtained by keeping its characteristic length scale above some threshold where the chosen continuum model fails to describe the problem. However, in most applications this strategy is currently unfeasible. For example, a discretization having a resolution in the order of  $10\mu\text{m}$  (a typical grain size of structural steel) would generate a numerical problem

whose ill-conditioning is incompatible with the accuracy of presently available measurement setups, and the obtained solutions would be unacceptably noisy. Still, one may reconsider whether stresses at low spatial scales are really that important: many failure processes involve an integral effect of stresses over some sort of *critical distance* (as defined by Taylor [91]), which may be chosen as a target length scale.

## Conclusions

A mathematical framework for the analysis of the inverse problem of relaxation methods has been provided, together with a proof of its ill-posedness. The analysis applies to any residual stress technique that requires an inversion of a similar integral operator.

Any discretized solution of the inverse problem comes with an inherent bias-variance tradeoff. Therefore, a regularization strategy is unavoidable to obtain a meaningful solution. An effective regularization strategy adds additional hypotheses on the solution to reduce variance but balances those with the introduced bias and the ill-conditioning of the resulting problem.

Numerical discretization is itself a form of regularization, and the resulting bias has quite an intuitive form: the solution cannot have components outside of the chosen basis. Unfortunately, increasing the resolution of the discretization (i.e., its degrees of freedom) to reduce the bias causes an unacceptable increase in the variance. When Tikhonov regularization is added, the inverse problem no longer suffers from increasing ill-conditioning as the solution DOFs increase, though a slightly biased problem is solved. The Morozov discrepancy principle provides a strategy for choosing the regularization parameter  $\alpha$ , which balances ill-conditioning with the introduced bias on a reasonable statistical basis, assuming that the input errors are known.

The bias-variance tradeoff is primarily controlled by the solution DOFs and not by the number of measurements points. Indeed, the highest number of measurements should always be taken, as it reduces the variance for a given regularization scheme.

The variability of the regularized solution *cannot be used* to build a valid confidence interval, since an unknown bias term is also included in the true overall error. The only exception to this curse occurs when the regularizing assumptions can be proven true. In particular, physical arguments that provide a characterization of the *true* solution would be of utmost importance, as they would regularize the problem without introducing an otherwise unknown bias.

This kind of *certain* information is seldom available, and that is quite common in the field of inverse problems. In these cases, the only way to evaluate the performance of different regularization strategies is through *ground truth* data,

by trying to identify an accurately known stress distribution while having real-world error sources. For example, in the field of tomographic imaging the obtained reconstructions are usually compared with a physical sectioning of the object of interest (interestingly, that object is often a walnut [92]).

This paper is primarily intended to clarify some misleading concepts and to foster a correct interpretation of residual stress measurements through relaxation methods. It should not be interpreted in any way as a declaration of distrust towards these techniques. On the contrary, in the authors' experience, skilled operators are *already able* to use their engineering judgment and prior knowledge about the problem to obtain accurate residual stress distributions, although this regularization process is often not reported. In fact, since inverse method of one sort or another are ubiquitous in data analysis, many other experimental mechanics methods likely contain similar issues that are not yet acknowledged.

It is hoped that this work will pave the way for a fruitful interaction between residual stress measurement methods and the mathematical theory of ill-posed problems, whose arguably biggest contribution is to provide a structured way to apply prior knowledge about the problem. Therefore, it could profitably serve as an extension of the tool set available to the residual stress engineer. As a matter of fact, many useful results of that theory have not yet been explored in the field of residual stresses.

## Declarations

**Conflicts of Interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Timoshenko S (1983) History of strength of materials: with a brief account of the history of theory of elasticity and theory of structures. Courier Corporation
- 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>
- 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>
- 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>
- Hagara M, Trebuña 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>
- Salehi SD, Rastak MA, Shokrieh MM, Barrallier L, Kubler R (2020) Full-field measurement of residual stresses in composite materials using the incremental slitting and digital image correlation techniques. *Exp Mech* 60(9):1239–1250. ISBN: 1741-2765. Publisher: Springer
- 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. ISBN: 1741-2765. Publisher: Springer
- Gubbels WL, Schajer GS (2016) Development of 3-D digital image correlation using a single color-camera and diffractive speckle projection. *Exp Mech* 56(8):1327–1337. <https://doi.org/10.1007/s11340-016-0173-y>
- Withers PJ, Bhadeshia H (2001) Residual stress. Part 1-measurement techniques. *Mater Sci Technol* 17(4):355–365. ISBN: 0267-0836. Publisher: Taylor & Francis
- Withers PJ, Bhadeshia H (2001) Residual stress. Part 2-Nature and origins. *Mater Sci Technol* 17(4):366–375. ISBN: 0267-0836. Publisher: Taylor & Francis
- Ueda Y, Fukuda K, Tanigawa M (1979) New measuring method of three dimensional residual stresses based on theory of inherent strain (welding mechanics, strength & design). *Trans JWRI* 8(2):249–256
- Ueda Y, Fukuda K (1989) New measuring method of three-dimensional residual stresses in long welded joints using inherent strains as parameters-Lz method. *J Eng Mater Technol* 111(1):1–8. <https://doi.org/10.1115/1.3226427>
- Korsunsky AM (2009) Eigenstrain analysis of residual strains and stresses. *J Strain Anal Eng Des* 44(1):29–43. <https://doi.org/10.1243/03093247JSA423>
- 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>
- Jun T-S, Venter AM, Korsunsky AM (2011) Inverse Eigenstrain analysis of the effect of non-uniform sample shape on the residual stress due to shot peening. *Exp Mech* 51(2):165–174. <https://doi.org/10.1007/s11340-010-9346-2>
- Noyan IC, Cohen JB (2013) Residual stress: measurement by diffraction and interpretation. Springer
- Hutchings MT, Withers PJ, Holden TM, Lorentzen T (2005) Introduction to the characterization of residual stress by neutron diffraction. CRC Press
- Gur CH (2018) Review of residual stress measurement by magnetic Barkhausen noise technique. *Mater Perform Charact* 7(4):504–525. ISBN: 2379-1365. Publisher: ASTM International
- Crecraft DI (1967) The measurement of applied and residual stresses in metals using ultrasonic waves. *J Sound Vib* 5(1):173–192. ISBN: 0022-460X. Publisher: Elsevier
- Schajer GS (2010) Relaxation methods for measuring residual stresses: techniques and opportunities. *Exp Mech* 50(8):1117–1127. <https://doi.org/10.1007/s11340-010-9386-7>
- Bueckner HF (1970) Novel principle for the computation of stress intensity factors. *Zeitschrift fuer Angewandte Mathematik & Mechanik* 50(9)
- 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>



23. 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>
24. 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>
25. Schajer GS, Prime MB (2007) Residual stress solution extrapolation for the slitting method using equilibrium constraints. ISBN: 0094-4289
26. 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>
27. Schajer GS, Altus E (1996) Stress calculation error analysis for incremental hole-drilling residual stress measurements. *J Eng Mater Technol* 118(1):120–126. <https://doi.org/10.1115/1.2805924>
28. Wern H, Cavelius R, Schäfer D (1997) A new method to determine triaxial non-uniform residual stresses from measurements using the hole drilling method. *Strain* 33(2):39–46. ISBN: 0039-2103. Publisher: Wiley Online Library
29. Zuccarello B (1999) Optimal calculation steps for the evaluation of residual stress by the incremental hole-drilling method. *Exp Mech* 39(2):117–124. <https://doi.org/10.1007/BF02331114>
30. 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>
31. 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>
32. 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>
33. Tikhonov AN (1943) On the stability of inverse problems. In: *Doklady Akademii Nauk SSSR*, vol 39. pp 195–198
34. Tikhonov AN (1963) On the solution of ill-posed problems and the method of regularization. In: *Doklady Akademii Nauk*, vol 151. Russian Academy of Sciences. Issue: 3, pp 501–504
35. Tikhonov AN, Arsenin VI, Arsenin VY (1977) Solutions of ill-posed problems. Vh Winston
36. Tenorio L (2001) Statistical regularization of inverse problems. *SIAM Rev* 43(2):347–366. <https://doi.org/10.1137/S0036144500358232>
37. Kondermann D (2013) Ground truth design principles: an overview. In: *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications*. ACM, St. Petersburg Russia, pp 1–4. <https://doi.org/10.1145/2501105.2501114>
38. Tjhung T, Li K (2003) Measurement of in-plane residual stresses varying with depth by the interferometric strain/slope rosette and incremental hole-drilling. *J Eng Mater Technol* 125(2):153–162. <https://doi.org/10.1115/1.1555654>
39. Schajer GS, An Y (2009) Inverse calculation of uniaxial stress-strain curves from bending test data. *J Eng Mater Technol* 131(4):041001. <https://doi.org/10.1115/1.3120409>
40. Faghidian SA (2014) A smoothed inverse eigenstrain method for reconstruction of the regularized residual fields. *Int J Solids Struct* 51(25–26):4427–4434. <https://doi.org/10.1016/j.ijsolstr.2014.09.012>
41. 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>
42. 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>
43. 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>
44. Apostel L (1961) Towards the formal study of models in the non-formal sciences. In: *The Concept and the Role of the Model in Mathematics and Natural and Social Science*. Springer, pp 1–37
45. Hadamard J (1902) Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pp 49–52
46. Engl HW, Hanke M, Neubauer A (1996) Regularization of inverse problems, vol 375. Springer
47. Vogel CR (2002) Computational methods for inverse problems. SIAM
48. Hansen PC (2010) Discrete inverse problems: insight and algorithms. SIAM
49. Groetsch CW (1993) Inverse problems in the mathematical sciences, vol 52. Springer
50. Golub GH, van Loan CF (1980) An analysis of the total least squares problem. *SIAM J Numer Anal* 17(6):883–893. <https://doi.org/10.1137/0717073>
51. Golub GH, Hansen PC, O’Leary DP (1999) Tikhonov regularization and total least squares. *SIAM J Matrix Anal Appl* 21(1):185–194. <https://doi.org/10.1137/S0895479897326432>
52. VanHuffel S, Lemmerling P (2013) Total least squares and errors-in-variables modeling: analysis, algorithms and applications. Springer
53. Bertero M, Poggio TA, Torre V (1988) Ill-posed problems in early vision. *Proc IEEE* 76(8):869–889. <https://doi.org/10.1109/5.5962>
54. 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>
55. Schajer GS (2013) Practical residual stress measurement methods. John Wiley & Sons
56. 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>
57. Love AEH (2013) A treatise on the mathematical theory of elasticity. Cambridge University Press
58. George D, Kingston E, Smith DJ (2002) Measurement of through-thickness stresses using small holes. *J Strain Anal Eng Des* 37(2):125–139. <https://doi.org/10.1243/0309324021514899>
59. 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. [https://doi.org/10.1007/978-3-7091-6296-5\\_4](https://doi.org/10.1007/978-3-7091-6296-5_4)
60. Sachs G, Espey G (1941) The measurement of residual stresses in metal. *Iron Age* 148(63–71):148
61. 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>
62. 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>
63. Beghini M, Bertini L, Mori LF (2010) Evaluating non-uniform residual stress by the hole-drilling method with concentric and eccentric holes. Part I. Definition and validation of the influence functions: eccentric influence functions - Part I. *Strain* 46(4):324–336. <https://doi.org/10.1111/j.1475-1305.2009.00683.x>
64. von Mises R (1945) On Saint Venant’s principle. *Bull Am Math Soc* 51(8):555–562. <https://doi.org/10.1090/S0002-9904-1945-08394-3>
65. Goldberg RR (1970) Methods of real analysis. Oxford and IBH Publishing
66. Golub GH, Loan CFV (2013) Matrix computations. JHU Press. Google-Books-ID: X5YfsuCWpxMC
67. Tikhonov AN, Goncharky 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>



68. Colton D, Kress R (2019) Inverse acoustic and electromagnetic scattering theory. Applied Mathematical Sciences, vol 93. Springer, Cham. <https://doi.org/10.1007/978-3-030-30351-8>
69. Schulze V (2005) Modern mechanical surface treatment: states, stability, effects, 1st edn. Wiley. <https://doi.org/10.1002/3527607811>
70. Phillips JL (1972) The use of collocation as a projection method for solving linear operator equations. SIAM J Numer Anal 9(1):14–28. <https://doi.org/10.1137/0709003>
71. Zuccarello B (1996) Optimization of depth increment distribution in the ring-core method. J Strain Anal Eng Des 31(4):251–258. ISBN: 0309-3247 Publisher: SAGE Publications Sage UK: London, England
72. Beghini M, Bertini L, Mori LF, Rosellini W (2009) Genetic algorithm optimization of the hole-drilling method for non-uniform residual stress fields. J Strain Anal Eng Des 44(1):105–115. <https://doi.org/10.1243/03093247JSA457>
73. Stefanescu D, Truman CE, Smith DJ, Whitehead PS (2006) Improvements in residual stress measurement by the incremental centre hole drilling technique. Exp Mech 46(4):417–427. <https://doi.org/10.1007/s11340-006-7686-8>
74. Liu X, Schajer GS (1997) More reliable calculations for layer-removal residual stress measurements. In: Proceedings of SEM Spring Conference on Experimental Mechanics. pp 255–256
75. Morozov VA (1984) Methods for solving incorrectly posed problems. Springer, New York, NY. <https://doi.org/10.1007/978-1-4612-5280-1>
76. Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21(2):215–223. ISBN: 0040-1706. Publisher: Taylor & Francis
77. Hansen PC (1999) The L-curve and its use in the numerical treatment of inverse problems. Publisher: Citeseer
78. 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>
79. 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. ISBN: 0041-5553. Publisher: Elsevier
80. LeVeque RJ (2007) Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. Soc Ind Appl Math. <https://doi.org/10.1137/1.9780898717839>
81. Kuhn HW, Tucker AW (1951) Nonlinear programming. Proceedings of 2nd Berkeley Symposium
82. Kaipio J, Somersalo E (2005) Statistical and computational inverse problems. Applied Mathematical Sciences, vol v. 160. Springer, New York
83. Cao YP, Hu N, Lu J, Fukunaga H, Yao ZH (2002) An inverse approach for constructing the residual stress field induced by welding. J Strain Anal Eng Des 37(4):345–359. <https://doi.org/10.1243/030932402760074562>
84. Lee MJ, Hill MR (2007) Intralaboratory repeatability of residual stress determined by the slitting method. Exp Mech 47(6):745–752. <https://doi.org/10.1007/s11340-007-9085-1>
85. Prime MB, Sebring RJ, Edwards JM, Hughes DJ, Webster PJ (2004) Laser surface-contouring and spline data-smoothing for residual stress measurement. Exp Mech 44(2):176–184. <https://doi.org/10.1007/BF02428177>
86. Wern H (1997) A new approach to triaxial residual stress evaluation by the hole drilling method. Strain 33(4):121–126. <https://doi.org/10.1111/j.1475-1305.1997.tb01059.x>
87. Hansen PC (1990) Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank. SIAM J Sci Stat Comput 11(3):503–518. ISBN: 0196-5204. Publisher: SIAM
88. 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>
89. 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>
90. 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>
91. Taylor D (2007) The theory of critical distances: a new perspective in fracture mechanics. Elsevier, Amsterdam; Boston. OCLC: ocm78988689
92. Kumar R (2018) Tomography of a walnut. <https://doi.org/10.13140/RG.2.2.25130.52161>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

