Annals of Mathematics and Computer Science ISSN: 2789-7206
Vol 36 (2026) 1-15 https://doi.org/10.56947/amcs.v36.979

SURROGATE-ACCELERATED LANDWEBER ITERATION FOR STIFFNESS IDENTIFICATION

SAIDJON KAMOLOV[1*]

Abstract. Identifying stiffness degradation from sparse sensor data is a central inverse problem in structural health monitoring, and neural surrogates are increasingly used to replace the finite element solves it requires. Such replacements are rarely accompanied by guarantees that the resulting identification procedure converges or that its error is controlled. We study a Landweber iteration in which both the parameter-to-observation map and its derivative are replaced by a neural surrogate trained on finite element data. For an elliptic structural model with a finite-dimensional damage parameterization, we establish explicit Lipschitz bounds on the forward map and its derivative in terms of material bounds, load and sensor functionals. We then show that a surrogate whose empirical value and derivative errors are small on a random design attains uniform accuracy on the whole admissible set, with an explicit dependence on the fill distance of the design and on the mesh size. Under a tangential cone condition, which we prove holds locally whenever the sensor layout renders the linearized problem identifiable, the surrogate-based iteration is monotone, terminates after finitely many steps under a discrepancy principle that accounts for the surrogate error, and returns an estimate whose error is bounded by a constant multiple of the measurement noise, the empirical training error, the fill distance and the squared mesh size, divided by the smallest singular value of the linearized sensor map. The analysis yields quantitative criteria for sensor placement and for derivative-informed training. A numerical study on a bridge-deck model problem confirms the predicted trends and quantifies the conservatism of the certificate.

Keywords. Inverse problems, Landweber iteration, neural surrogate, structural health monitoring, tangential cone condition, derivative-informed training
2020 Mathematics Subject Classification. Primary 65J20, 65N21; Secondary 68T07, 35R30, 74G75

1. Introduction

Ageing bridges, tunnels and building frames are increasingly instrumented with strain gauges, accelerometers and fibre-optic sensors. Turning these measurements into a map of local stiffness loss is an inverse problem governed by a

Date: Received: Aug 2, 2026; Revised: Sep 12, 2026; Accepted: Sep 19, 2020.
* Corresponding author

partial differential equation, and it is typically solved by iterative methods that require one or more finite element solves per step. When monitoring must run continuously or over many structures, these solves dominate the cost, which has motivated the replacement of the finite element model by a trained neural surrogate.

Replacing the forward model changes the mathematical object being iterated. A surrogate that is accurate in mean square on its training distribution need not be accurate along the trajectory of an iterative solver, its derivative may be poorly approximated even when its values are not, and classical convergence results for regularizing iterations are stated for the exact operator. The practitioner is therefore left without an answer to three questions. How accurate must the surrogate be, and in which norm? When does the surrogate-based iteration stop? How large is the final identification error?

This paper answers these questions for a Landweber iteration applied to a model elliptic problem of stiffness identification. Our contributions are as follows.

- We derive explicit Lipschitz constants for the parameter-to-observation map and its derivative in terms of the lower and upper stiffness bounds, the load and the sensor functionals.
- We prove a deterministic transfer from empirical errors on a finite design to uniform errors on the admissible set, together with a high-probability bound on the fill distance of a random design. The resulting accuracy bound separates the training error, the sampling error and the finite element discretization error.
- We show that a tangential cone condition for the exact map is inherited by the surrogate up to an additive term proportional to the uniform value and derivative errors, and that the exact map satisfies the condition locally whenever the linearized sensor map has full column rank.
- We prove that the surrogate-based Landweber iteration, stopped by a discrepancy principle whose threshold includes the surrogate error, is monotone, terminates after an explicitly bounded number of steps and attains an error bound that is linear in the noise level and in each component of the surrogate error.
- We translate these bounds into criteria for sensor placement and for derivative-informed training, and we verify each of them numerically on a bridge-deck model problem.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 introduces the structural model and establishes regularity of the forward map. Section 4 analyses the accuracy of the surrogate. Section 5 contains the convergence analysis. Section 6 discusses consequences for monitoring practice, Section 7 reports the numerical study and Section 8 concludes.

2. Related work

Model updating and damage identification. Finite element model updating calibrates stiffness parameters against measured response and is a mature tool in structural engineering [8]. Data-driven damage detection has progressed

from feature-based classifiers to deep networks trained on vibration and strain records [2, 7]. These works are largely empirical, and the identification step is seldom analysed as a regularization method.

Iterative regularization. The Landweber iteration for nonlinear ill-posed problems was analysed under the tangential cone condition in [11], and the theory is surveyed in [6, 12]. These results assume exact access to the forward operator and its derivative. Perturbations of the operator are treated in the literature on learned operator correction [21] and on data-driven inverse problems more broadly [1], but explicit iteration counts and error bounds that decompose the surrogate error into training, sampling and discretization components are, to our knowledge, not available for the setting considered here. The Landweber iteration is a fixed point scheme for the normal equations, and the broader fixed point literature is surveyed in [13].

Operator learning. Universal approximation of nonlinear operators by neural networks goes back to [4] and underlies DeepONet [20] and neural operators [18]. Error estimates for operator surrogates of parametric PDEs are developed in [19]. Training on derivative information was proposed in [5] and specialized to parametric PDE maps in [22], where it was shown to improve derivative accuracy for downstream optimization. Our analysis explains why derivative accuracy, and not only value accuracy, enters the convergence guarantee of a gradient-based identification scheme.

Physics-informed learning. Physics-informed networks [23] embed the governing equation into the training loss. They address a complementary question, namely how to fit a single solution without mesh-based data, whereas we study the reuse of a trained parameter-to-observation surrogate inside an iterative inversion.

3. Problem setting

3.1. **Structural model and damage parameterization.** Let $\Omega \subset \mathbb{R}^d$ with $d \in \{1, 2, 3\}$ be a bounded convex polygonal or polyhedral domain representing the mid-surface of a deck slab or the axis of a girder. We consider the scalar elliptic model

$$-\nabla \cdot \big(\kappa_\vartheta \nabla u\big) = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{3.1}$$

which describes the out-of-plane displacement of a prestressed membrane or the anti-plane shear response of an elastic body, and which serves as a scalar prototype of linear elasticity. The load $f \in L^2(\Omega)$ is known, for instance from a controlled load test. Stiffness loss is represented through

$$\kappa_\vartheta(x) = \kappa_{\mathrm{ref}}(x)\Big(1 - \sum_{j=1}^{p} \vartheta_j\, \varphi_j(x)\Big), \qquad \vartheta \in \Theta := [0, \vartheta_{\max}]^p, \tag{3.2}$$

where the design stiffness $\kappa_{\mathrm{ref}} \in W^{1,\infty}(\Omega)$ satisfies $\kappa_{\min} \le \kappa_{\mathrm{ref}} \le \kappa_{\max}$ almost everywhere, $\vartheta_{\max} < 1$, and the zone functions $\varphi_1, \dots, \varphi_p \in W^{1,\infty}(\Omega)$ are nonnegative with $\varphi_1 + \cdots + \varphi_p \le 1$. The coordinate ϑ_j is the fraction of stiffness lost in zone j, which may correspond to a deck panel, a girder segment or a joint region.

Since $0 \le \varphi_j \le 1$, we have $\varphi_1^2 + \cdots + \varphi_p^2 \le 1$ pointwise, and every κ_ϑ with $\vartheta \in \Theta$ satisfies $\kappa_- \le \kappa_\vartheta \le \kappa_{\max}$ with $\kappa_- := \kappa_{\min}(1 - \vartheta_{\max}) > 0$.

Let $V := H_0^1(\Omega)$ with the norm $\|v\|_V := \|\nabla v\|_{L^2(\Omega)}$, and let C_P denote the Poincaré constant, so that $\|v\|_{L^2(\Omega)} \le C_P \|v\|_V$. For $\vartheta \in \Theta$ let $u_\vartheta \in V$ be the unique solution of

$$a_\vartheta(u_\vartheta, v) := \int_\Omega \kappa_\vartheta \nabla u_\vartheta \cdot \nabla v \, dx = (f, v)_{L^2} \qquad \text{for all } v \in V. \tag{3.3}$$

3.2. **Sensors and the inverse problem.** A strain gauge or displacement transducer is modelled as a local weighted average. Let $s_1, \dots, s_M \in L^2(\Omega)$ be sensor kernels and define $S : L^2(\Omega) \to \mathbb{R}^M$ componentwise by

$$(Sv)_m := (s_m, v)_{L^2}, \qquad \sigma_S := \Big(\sum_{m=1}^{M} \|s_m\|_{L^2}^2 \Big)^{1/2}, \qquad A : \Theta \to \mathbb{R}^M, \quad A(\vartheta) := S u_\vartheta, \tag{3.4}$$

so that $|Sv| \le \sigma_S \|v\|_{L^2(\Omega)}$, and A is the parameter-to-observation map. Given data $y^\delta \in \mathbb{R}^M$ with $|y^\delta - A(\vartheta^\dagger)| \le \delta$ for an unknown true state $\vartheta^\dagger \in \Theta$, we seek to recover $\vartheta^\dagger$. Throughout, $|\cdot|$ is the Euclidean norm on $\mathbb{R}^p$ and $\mathbb{R}^M$, and $\|\cdot\|$ applied to a linear map is the induced operator norm. Although Θ is finite dimensional, the problem is severely ill-conditioned when zones are small relative to the sensor spacing, and regularizing iterations remain the method of choice.

3.3. **Regularity of the forward map.** Set

$$L_A := \frac{\sigma_S \, C_P^2 \, \kappa_{\max} \, \|f\|_{L^2}}{\kappa_-^2}, \qquad L_A' := \frac{2 \, \sigma_S \, C_P^2 \, \kappa_{\max}^2 \, \|f\|_{L^2}}{\kappa_-^3}. \tag{3.5}$$

Lemma 3.1. *The map A is continuously differentiable on the interior of Θ and its derivative extends continuously to Θ. For $\xi \in \mathbb{R}^p$ one has $A'(\vartheta)\xi = Sw$, where $w \in V$ solves*

$$a_\vartheta(w, v) = \Big(\kappa_{\mathrm{ref}} \Big(\sum_{j=1}^{p} \xi_j \varphi_j \Big) \nabla u_\vartheta, \nabla v \Big)_{L^2} \qquad \textit{for all } v \in V. \tag{3.6}$$

Moreover, for all $\vartheta, \tilde\vartheta \in \Theta$,

$$|A(\vartheta) - A(\tilde\vartheta)| \le L_A |\vartheta - \tilde\vartheta|, \qquad \|A'(\vartheta)\| \le L_A, \qquad \|A'(\vartheta) - A'(\tilde\vartheta)\| \le L_A' |\vartheta - \tilde\vartheta|.$$

Proof. Testing (3.3) with $v = u_\vartheta$ and using the Poincaré inequality gives $\|u_\vartheta\|_V \le C_P \|f\|_{L^2} / \kappa_-$. For $\vartheta, \tilde\vartheta \in \Theta$, subtracting the variational equations yields

$$a_\vartheta(u_\vartheta - u_{\tilde\vartheta}, v) = -\big((\kappa_\vartheta - \kappa_{\tilde\vartheta}) \nabla u_{\tilde\vartheta}, \nabla v \big), \qquad \|\kappa_\vartheta - \kappa_{\tilde\vartheta}\|_{L^\infty} \le \kappa_{\max} |\vartheta - \tilde\vartheta|,$$

where the Cauchy–Schwarz inequality in $\mathbb{R}^p$ together with $\varphi_1^2 + \cdots + \varphi_p^2 \le 1$ controls the coefficient difference. Testing with the difference of the states gives

$$\|u_\vartheta - u_{\tilde\vartheta}\|_V \le \kappa_{\max} \, C_P \, \|f\|_{L^2} \, \kappa_-^{-2} \, |\vartheta - \tilde\vartheta|,$$

and the first inequality follows from $|Sv| \le \sigma_S C_P \|v\|_V$.

For differentiability, let ϑ and $\vartheta+\xi$ lie in Θ, let w solve (3.6) and put $r := u_{\vartheta+\xi} - u_\vartheta - w$. The identities above give

$$a_\vartheta(r,v) = \Big(\kappa_{\mathrm{ref}}\Big(\sum_j \xi_j\varphi_j\Big)\nabla(u_{\vartheta+\xi}-u_\vartheta),\nabla v\Big), \qquad \|r\|_V \le \frac{\kappa_{\max}|\xi|}{\kappa_-}\,\|u_{\vartheta+\xi}-u_\vartheta\|_V \le \frac{\kappa_{\max}^2 C_P\|f\|_{L^2}}{\kappa_-^3}\,|\xi|^2.$$

Thus $A'(\vartheta)\xi = Sw$. Testing (3.6) with $v = w$ gives $\|w\|_V \le \kappa_{\max}|\xi|\|u_\vartheta\|_V/\kappa_-$, which yields $\|A'(\vartheta)\| \le L_A$. Finally, let w and $\tilde w$ solve (3.6) at ϑ and $\tilde\vartheta$ for the same ξ. Subtracting the two equations gives

$$a_\vartheta(w-\tilde w,v) = \Big(\kappa_{\mathrm{ref}}\Big(\sum_j \xi_j\varphi_j\Big)\nabla(u_\vartheta - u_{\tilde\vartheta}),\nabla v\Big) - \big((\kappa_\vartheta-\kappa_{\tilde\vartheta})\nabla\tilde w,\nabla v\big),$$

and the preceding bounds imply $\|w-\tilde w\|_V \le 2\kappa_{\max}^2 C_P\|f\|_{L^2}\kappa_-^{-3}|\xi||\vartheta-\tilde\vartheta|$. Applying S yields the last inequality and, in particular, the continuity of A' up to the boundary of Θ. □

3.4. **Finite element approximation.** Let $V_h \subset V$ be the space of continuous piecewise linear functions on a shape-regular triangulation of Ω with mesh size h, let $u_{\vartheta,h}$ be the Galerkin approximation of u_ϑ, and let $A_h(\vartheta) := Su_{\vartheta,h}$. The derivative $A_h'(\vartheta)$ is defined by (3.6) with u_ϑ replaced by $u_{\vartheta,h}$ and V replaced by V_h. These are the quantities available for generating training data.

Lemma 3.2. *There exists a constant C_{FE}, depending on Ω, $\kappa_{\min}$, $\kappa_{\max}$, $\vartheta_{\max}$, the $W^{1,\infty}$ norms of κ_{ref} and $\varphi_1,\dots,\varphi_p$, $\|f\|$ and σ_S, but not on h or ϑ, such that for all $\vartheta\in\Theta$*

$$|A(\vartheta)-A_h(\vartheta)| \le C_{\mathrm{FE}}\,h^2, \qquad \|A'(\vartheta)-A_h'(\vartheta)\| \le C_{\mathrm{FE}}\,h^2.$$

Proof. Since Ω is convex and κ_ϑ belongs to $W^{1,\infty}(\Omega)$ with norm bounded uniformly in $\vartheta\in\Theta$, the solution of (3.3) and the solutions z_m of the dual problems with right-hand sides s_m belong to $H^2(\Omega)$ with norms bounded uniformly in ϑ [9]. The first estimate is the Aubin–Nitsche duality argument [3]. For the second, the right-hand side of (3.6) has the form $(g\nabla u_\vartheta,\nabla v)$ with $g\in W^{1,\infty}(\Omega)$ and $\|g\|$ bounded by a multiple of $|\xi|$, so $w\in H^2(\Omega)$. The discrete derivative differs from the Galerkin projection of w only through the replacement of u_ϑ by $u_{\vartheta,h}$ in the data. The Galerkin part contributes $O(h^2|\xi|)$ to each sensor functional by duality. The data perturbation contributes the term $(g\nabla(u_\vartheta-u_{\vartheta,h}),\nabla z_{m,h})$; adding and subtracting z_m and integrating by parts against $z_m\in H^2(\Omega)$ bounds it by a multiple of $|\xi|$ times the sum of the L^2 error and h times the energy error of $u_{\vartheta,h}$, which is $O(h^2|\xi|)$. All constants are uniform in ϑ, which gives the claim. □

4. Uniform accuracy of the neural surrogate

A surrogate trained on finitely many finite element solves is accurate on the whole admissible set, in value and in derivative, up to its empirical error on the design, a term proportional to the fill distance of the design, and the squared mesh size.

4.1. **Surrogate class and training data.** Let $G : \Theta \to \mathbb{R}^M$ be continuously differentiable with Lipschitz constants L_G for G and L'_G for G' on Θ. Let $D_N = \{\vartheta^1, \dots, \vartheta^N\} \subset \Theta$ be a design at which the labels $A_h(\vartheta^i)$ and $A'_h(\vartheta^i)$ have been computed. The derivative label costs M adjoint solves or p tangent solves per design point, each sharing the stiffness matrix of the forward solve. Define

$$e_0 := \max_i |G(\vartheta^i) - A_h(\vartheta^i)|, \quad e_1 := \max_i \|G'(\vartheta^i) - A'_h(\vartheta^i)\|, \quad h_N := \sup_{\vartheta\in\Theta} \min_i |\vartheta - \vartheta^i|,$$
$$\varepsilon_0 := \sup_{\vartheta\in\Theta} |G(\vartheta) - A(\vartheta)|, \qquad \varepsilon_1 := \sup_{\vartheta\in\Theta} \|G'(\vartheta) - A'(\vartheta)\|. \tag{4.1}$$

The quantities e_0 and e_1 are computable after training, whereas ε_0 and ε_1 are those that enter the convergence analysis of Section 5.

4.2. **From empirical to uniform accuracy.**

Proposition 4.1. *Under the assumptions of Lemmas 3.1 and 3.2,*

$$\varepsilon_0 \le e_0 + (L_G + L_A)\, h_N + C_{\mathrm{FE}}\, h^2, \qquad \varepsilon_1 \le e_1 + (L'_G + L'_A)\, h_N + C_{\mathrm{FE}}\, h^2. \tag{4.2}$$

Proof. Fix $\vartheta \in \Theta$ and choose $\vartheta^i \in D_N$ with $|\vartheta - \vartheta^i| \le h_N$, which is possible because Θ is compact and the minimum in the definition of h_N is attained. The triangle inequality gives

$$|G(\vartheta)-A(\vartheta)| \le |G(\vartheta)-G(\vartheta^i)|+|G(\vartheta^i)-A_h(\vartheta^i)|+|A_h(\vartheta^i)-A(\vartheta^i)|+|A(\vartheta^i)-A(\vartheta)|.$$

The four terms are bounded by $L_G h_N$, e_0, $C_{\mathrm{FE}}h^2$ by Lemma 3.2 and $L_A h_N$ by Lemma 3.1, respectively. The derivative estimate follows in the same way from the Lipschitz continuity of G' and A'. Taking the supremum over ϑ proves (4.2). □

Lemma 4.2. *Let $\vartheta^1, \dots, \vartheta^N$ be independent and uniformly distributed on Θ, and let $\beta \in (0,1)$ satisfy $N \ge \log(N/\beta) \ge 1$. Then, with probability at least $1-\beta$,*

$$h_N \le 2\sqrt{p}\,\vartheta_{\max}\Big(\frac{\log(N/\beta)}{N}\Big)^{1/p}. \tag{4.3}$$

Proof. Put $x := (N/\log(N/\beta))^{1/p} \ge 1$ and $m := \lfloor x \rfloor \ge 1$, so that $m \ge x/2$. Partition Θ into m^p closed cubes of side $\vartheta_{\max}/m$ and diameter $\sqrt{p}\,\vartheta_{\max}/m$. Each cube receives a given sample with probability m^{-p}, so the probability that some cube is empty is at most

$$m^p(1-m^{-p})^N \le m^p \exp(-Nm^{-p}) \le m^p \exp\big(-\log(N/\beta)\big) = \frac{m^p\beta}{N} \le \beta,$$

where we used $m^p \le N/\log(N/\beta) \le N$. On the complementary event every $\vartheta \in \Theta$ lies in a cube containing a design point, whence $h_N \le \sqrt{p}\,\vartheta_{\max}/m \le 2\sqrt{p}\,\vartheta_{\max}/x$. □

Corollary 4.3. *Under the assumptions of Proposition 4.1 and Lemma 4.2, with probability at least $1-\beta$,*

$$\varepsilon_k \le e_k + 2\sqrt{p}\,\vartheta_{\max}\,\Lambda_k\Big(\frac{\log(N/\beta)}{N}\Big)^{1/p} + C_{\mathrm{FE}}\, h^2, \qquad k \in \{0,1\}, \tag{4.4}$$

with $\Lambda_0 := L_G + L_A$ and $\Lambda_1 := L'_G + L'_A$.

4.3. **Remarks.**

Remark 4.4 (Computable Lipschitz constants). For a network with one hidden layer, $G(\vartheta) = W_2\sigma(W_1\vartheta + b)$, and an activation with $|\sigma'| \le 1$ and $|\sigma''| \le c_2$, one has $G'(\vartheta) = W_2\,\mathrm{diag}(\sigma'(W_1\vartheta + b))W_1$. It follows that $L_G \le \|W_2\|\|W_1\|$ and

$$L_G' \le c_2\,\|W_2\|\,\|W_1\|\,\max_k |w_k|,$$

where w_k denotes the k-th row of W_1. Analogous products of layerwise norms bound L_G and L_G' for deeper networks, so that the right-hand side of (4.2) is a fully computable certificate once C_{FE} is estimated, for example by comparing two mesh levels.

Remark 4.5 (Dimension dependence). The rate $N^{-1/p}$ in (4.4) reflects the absence of structural assumptions and is sharp for fill distances. The quantity controlled in Lemma 4.2 is the largest spacing of a uniform random sample; spacings of random samples are studied by means of order statistics in other settings, for instance for synthetically generated samples in [14, 15]. Quasi-uniform deterministic designs remove the logarithm but not the exponent. For many zones one must exploit structure, such as the decay of the influence of a zone on distant sensors, and we return to this point in Section 8.

Remark 4.6 (Role of the derivative error). A surrogate trained only on values controls ε_0 but not ε_1, since two maps can be uniformly close while their derivatives differ by a quantity of order one. Section 5 shows that ε_1 enters the stopping threshold and the final error on the same footing as ε_0. This supplies a mathematical justification for the derivative-informed training of [5, 22], in which e_1 is part of the loss. Derivatives of a trained network with respect to its input are used for a different purpose in attribution methods, where they are integrated along a path from a baseline to the input [16]; the accuracy of those derivatives is not certified there, whereas it is precisely the quantity bounded by ε_1 here.

5. Surrogate-based Landweber iteration

The surrogate-based iteration behaves like the exact Landweber iteration applied to data whose noise level is inflated by the uniform value and derivative errors of the surrogate, provided the stopping threshold accounts for this inflation.

5.1. **Iteration and standing assumptions.**

Let $B_\rho(\vartheta^\dagger)$ denote the closed Euclidean ball of radius $\rho > 0$ about $\vartheta^\dagger$, and assume that it lies in the interior of Θ. Given $\vartheta_0 \in B_\rho(\vartheta^\dagger)$ and a step size $\omega > 0$, the iteration reads

$$\vartheta_{k+1} = \vartheta_k - \omega\,G'(\vartheta_k)^\top\big(G(\vartheta_k) - y^\delta\big), \qquad k = 0, 1, 2, \dots \tag{5.1}$$

Each step requires one evaluation of the network and one reverse-mode differentiation, in place of one forward and one adjoint finite element solve. We assume that

$$\omega \sup_{\vartheta \in B_\rho(\vartheta^\dagger)} \|G'(\vartheta)\|^2 \le 1, \tag{5.2}$$

and that the exact map satisfies the tangential cone condition on the ball, namely that for some $\eta \in [0, 1/2)$ and all $\vartheta, \tilde{\vartheta} \in B_\rho(\vartheta^\dagger)$

$$\big|A(\vartheta) - A(\tilde{\vartheta}) - A'(\vartheta)(\vartheta - \tilde{\vartheta})\big| \le \eta \,\big|A(\vartheta) - A(\tilde{\vartheta})\big|. \tag{5.3}$$

5.2. **Inheritance and verification of the cone condition.**

Lemma 5.1. *Under (5.3), for all $\vartheta, \tilde{\vartheta} \in B_\rho(\vartheta^\dagger)$,*

$$\big|G(\vartheta) - G(\tilde{\vartheta}) - G'(\vartheta)(\vartheta - \tilde{\vartheta})\big| \le \eta \,\big|G(\vartheta) - G(\tilde{\vartheta})\big| + 2(1+\eta)\,\varepsilon_0 + \varepsilon_1|\vartheta - \tilde{\vartheta}|.$$

Proof. Write the left-hand side as the sum of the exact Taylor remainder, the difference of the value errors $G - A$ at ϑ and at $\tilde{\vartheta}$, and the term $(G'(\vartheta) - A'(\vartheta))(\vartheta - \tilde{\vartheta})$. By (5.3) and (4.1) its norm is at most $\eta|A(\vartheta) - A(\tilde{\vartheta})| + 2\varepsilon_0 + \varepsilon_1|\vartheta - \tilde{\vartheta}|$. Inserting $|A(\vartheta) - A(\tilde{\vartheta})| \le |G(\vartheta) - G(\tilde{\vartheta})| + 2\varepsilon_0$ completes the proof. □

Condition (5.3) is a hypothesis in the infinite-dimensional theory. In the present setting it follows from identifiability of the linearized sensor map, which is a verifiable property of the sensor layout.

Proposition 5.2. *Assume $M \ge p$, let $\sigma > 0$ be the smallest singular value of $A'(\vartheta^\dagger)$, and let $\rho < \sigma/(3L'_A)$. Then (5.3) holds on $B_\rho(\vartheta^\dagger)$ with*

$$\eta = \frac{L'_A\,\rho}{\sigma - L'_A\,\rho} < \frac{1}{2}, \qquad \text{and} \qquad |A(\vartheta) - A(\tilde{\vartheta})| \ge (\sigma - L'_A\rho)\,|\vartheta - \tilde{\vartheta}| \quad \text{for all } \vartheta, \tilde{\vartheta} \in B_\rho(\vartheta^\dagger). \tag{5.4}$$

Proof. Put $\vartheta_t := \tilde{\vartheta} + t(\vartheta - \tilde{\vartheta})$, which lies in the convex set $B_\rho(\vartheta^\dagger)$ for $t \in [0, 1]$. By Lemma 3.1,

$$A(\vartheta) - A(\tilde{\vartheta}) - A'(\vartheta)(\vartheta - \tilde{\vartheta}) = \int_0^1 \big(A'(\vartheta_t) - A'(\vartheta)\big)(\vartheta - \tilde{\vartheta})\,dt,$$

and since $|\vartheta_t - \vartheta| = (1-t)|\vartheta - \tilde{\vartheta}|$, the norm of the right-hand side is at most $(L'_A/2)|\vartheta - \tilde{\vartheta}|^2 \le L'_A\rho|\vartheta - \tilde{\vartheta}|$. Similarly,

$$A(\vartheta) - A(\tilde{\vartheta}) = A'(\vartheta^\dagger)(\vartheta - \tilde{\vartheta}) + \int_0^1 \big(A'(\vartheta_t) - A'(\vartheta^\dagger)\big)(\vartheta - \tilde{\vartheta})\,dt,$$

and since $|\vartheta_t - \vartheta^\dagger| \le \rho$, the second inequality in (5.4) follows. Combining the two estimates yields (5.3) with the stated η, and $\eta < 1/2$ is equivalent to $\rho < \sigma/(3L'_A)$. □

5.3. **Convergence, stopping and error estimate.** Set

$$\bar{\varepsilon} := 2(1+\eta)\,\varepsilon_0 + \rho\,\varepsilon_1, \qquad \gamma := (1+\eta)(\delta + \varepsilon_0) + \bar{\varepsilon}, \qquad c_\tau := 1 - 2\eta - \frac{2}{\tau}, \tag{5.5}$$

fix $\tau > 2/(1-2\eta)$, so that $c_\tau > 0$, and define the stopping index by the discrepancy principle

$$k_* := \min\big\{k \ge 0 : |G(\vartheta_k) - y^\delta| \le \tau\gamma\big\}. \tag{5.6}$$

Theorem 5.3. *Assume (5.2) and (5.3), and let $\gamma > 0$. Then the following hold.*

(1) *The iterates of (5.1) are well defined for $k \le k_*$, remain in $B_\rho(\vartheta^\dagger)$, and satisfy $|\vartheta_{k+1} - \vartheta^\dagger|^2 \le |\vartheta_k - \vartheta^\dagger|^2 - \omega c_\tau |G(\vartheta_k) - y^\delta|^2$ for $0 \le k < k_*$.*

(2) *The stopping index is finite and satisfies* $k_* < \rho^2/(\omega c_\tau \tau^2 \gamma^2)$ *whenever* $k_* \geq 1$.
(3) *If in addition the assumptions of Proposition 5.2 hold, then*

$$|\vartheta_{k_*} - \vartheta^\dagger| \leq \frac{\tau\gamma + \delta + \varepsilon_0}{\sigma - L'_A \rho}. \tag{5.7}$$

Proof. Write $e_k := \vartheta_k - \vartheta^\dagger$, $r_k := G(\vartheta_k) - y^\delta$ and $J_k := G'(\vartheta_k)$, and suppose that $\vartheta_k \in B_\rho(\vartheta^\dagger)$ for some $k < k_*$. Expanding the square and using (5.2),

$$|e_{k+1}|^2 = |e_k|^2 - 2\omega\langle J_k e_k, r_k\rangle + \omega^2 |J_k^\top r_k|^2 \leq |e_k|^2 - 2\omega|r_k|^2 + 2\omega\langle r_k - J_k e_k, r_k\rangle + \omega|r_k|^2.$$

We decompose $r_k - J_k e_k$ into the surrogate Taylor remainder $G(\vartheta_k) - G(\vartheta^\dagger) - J_k e_k$ and the term $G(\vartheta^\dagger) - y^\delta$. The latter has norm at most $\varepsilon_0 + \delta$. By Lemma 5.1 with $\tilde\vartheta = \vartheta^\dagger$ and $|e_k| \leq \rho$, the former is bounded by $\eta|G(\vartheta_k) - G(\vartheta^\dagger)| + \bar\varepsilon$, and $|G(\vartheta_k) - G(\vartheta^\dagger)| \leq |r_k| + \varepsilon_0 + \delta$. Hence $|r_k - J_k e_k| \leq \eta|r_k| + \gamma$ and

$$|e_{k+1}|^2 - |e_k|^2 \leq \omega\, |r_k|\big((2\eta - 1)|r_k| + 2\gamma\big).$$

For $k < k_*$ we have $|r_k| > \tau\gamma$, so that $2\gamma < 2|r_k|/\tau$ and the right-hand side is at most $-\omega c_\tau |r_k|^2 \leq 0$, which proves the monotonicity assertion and, inductively, that all iterates remain in $B_\rho(\vartheta^\dagger)$. Summing the inequality over $k < k_*$ and using $|r_k| > \tau\gamma$ gives $\omega c_\tau k_* \tau^2\gamma^2 < |e_0|^2 \leq \rho^2$, which is the second assertion. For the third, the second inequality in (5.4) with $\tilde\vartheta = \vartheta^\dagger$ gives

$$(\sigma - L'_A\rho)|\vartheta_{k_*} - \vartheta^\dagger| \leq |A(\vartheta_{k_*}) - A(\vartheta^\dagger)| \leq |G(\vartheta_{k_*}) - y^\delta| + \varepsilon_0 + \delta \leq \tau\gamma + \varepsilon_0 + \delta. \qquad \square$$

Inserting (4.2) into (5.7) and using $\gamma \leq (1+\eta)(\delta + \varepsilon_0) + 2(1+\eta)\varepsilon_0 + \rho\varepsilon_1$ gives the error bound in terms of the four sources,

$$|\vartheta_{k_*} - \vartheta^\dagger| \lesssim \frac{\delta + e_0 + \rho e_1 + (\Lambda_0 + \rho\Lambda_1)h_N + (1+\rho)C_{\mathrm{FE}}h^2}{\sigma - L'_A\rho}, \tag{5.8}$$

in which the implied constant depends only on τ and η.

6. Implications for sensor design and training

The estimate (5.8) converts three engineering decisions, namely where to place sensors, how many finite element solves to buy and how to weight the training loss, into explicit terms of a single error bound.

6.1. **Sensor placement.**

The constant σ in (5.4) and (5.8) is the smallest singular value of the linearized sensor map. Since $\vartheta^\dagger$ is unknown, a robust criterion follows from Lemma 3.1 and Weyl's inequality for singular values, which give

$$\sigma_{\min}\big(A'(\vartheta^\dagger)\big) \geq \sigma_{\min}\big(A'(\vartheta_{\mathrm{ref}})\big) - L'_A\,|\vartheta^\dagger - \vartheta_{\mathrm{ref}}| \tag{6.1}$$

for any reference state ϑ_{ref}, such as the undamaged state $\vartheta_{\mathrm{ref}} = 0$. Maximizing the smallest singular value of $A'(\vartheta_{\mathrm{ref}})$ over admissible sensor kernels is the E-optimality criterion of experimental design. Choosing M sensors from a finite candidate set is a subset selection problem, and criteria for selecting subsets of variables under a spectral objective are reviewed, in the statistical learning context, in [17]. The analysis thus identifies E-optimality, rather than trace or determinant criteria, as the design objective that directly controls both the size of

the admissible region $\rho < \sigma/(3L'_A)$ and the error amplification factor $1/(\sigma - L'_A\rho)$. A layout with $M < p$ sensors, or one that leaves a zone unobserved, gives $\sigma = 0$ and voids the guarantee.

6.2. **Severity of damage.** By (3.5), L'_A grows like $\kappa_-^{-3} = \kappa_{\min}^{-3}(1 - \vartheta_{\max})^{-3}$. Admitting more severe damage in the model therefore shrinks the region in which the cone condition is certified, and it tightens the requirement on the initial guess. In practice this suggests a continuation strategy in which the identification is first run with a moderate $\vartheta_{\max}$ and the admissible set is enlarged only in zones where the estimate approaches its bound.

6.3. **Allocation of the offline budget.** To reach a target accuracy ϵ in (5.8), the three surrogate terms should be of comparable size. Balancing the sampling and discretization terms gives $h^2 \asymp N^{-1/p}$, up to logarithmic factors. With a finite element solve costing of order h^{-d} operations for an optimal solver, the offline cost of the training set scales like $Nh^{-d} \asymp \epsilon^{-p-d/2}$, whereas each online step of the finite element Landweber iteration costs of order $h^{-d} \asymp \epsilon^{-d/2}$. The surrogate is therefore advantageous when the number of identification runs, summed over monitoring epochs and over structures of a common design, exceeds a threshold of order ϵ^{-p} divided by the number of iterations per run. This quantifies the common observation that surrogates pay off for fleets of similar structures, such as standardized highway overpasses, and for continuous monitoring.

6.4. **Training loss.** Estimate (5.8) weights the value error and the derivative error as $e_0 + \rho e_1$. This suggests the Sobolev loss

$$\mathcal{L}(G) = \frac{1}{N}\sum_{i=1}^{N}\Big(|G(\vartheta^i) - A_h(\vartheta^i)|^2 + \rho^2\,\|G'(\vartheta^i) - A'_h(\vartheta^i)\|_F^2\Big), \tag{6.2}$$

in which the derivative weight equals the squared radius of the region of interest, and it suggests reporting the maximum errors e_0 and e_1 over a held-out design rather than mean squared errors, since only the former enter the certificate. Spectral normalization of the layers keeps L_G and L'_G small and hence reduces the sampling term $\Lambda_0 + \rho\Lambda_1$.

7. Numerical experiments

The experiments test each term of (5.8) separately on a bridge-deck model problem. All computations use P1 finite elements, the Adam optimizer and double precision for the solver; the reported errors are maxima, as required by the certificate, not mean squared errors.

7.1. **Test problem.** The domain is a nondimensionalized deck panel $\Omega = (0,3)\times(0,1)$ with clamped edges, $\kappa_{\mathrm{ref}} \equiv 1$ and a load f equal to the indicator of the wheel-lane strip $0.35 \le y \le 0.65$. The deck is divided into $p = 6$ zones arranged as a 3×2 grid of panels, and the zone functions are a smooth partition of unity subordinate to this grid. Damage states are drawn from $\Theta = [0, 0.5]^6$. Sensors are $M = 12$ Gaussian averages of width 0.05 representing strain-gauge footprints, placed on a 4×3 grid. Training data are generated with $h \in \{1/20, 1/40\}$ and

the reference solution uses $h = 1/80$; these mesh sizes are chosen so that the strip edges $y = 0.35$ and $y = 0.65$ fall on mesh lines, which is necessary for the second-order rate of Lemma 3.2 to be observed.

A mesh-refinement study at a fixed ϑ confirms Lemma 3.2: the errors $|A_h - A|$ are $6.18 \cdot 10^{-4}$, $1.53 \cdot 10^{-4}$, $3.64 \cdot 10^{-5}$ for $h = 1/20, 1/40, 1/80$, a factor of 4.05 and 4.20 per refinement, and the derivative errors decrease by the same factors. This gives $C_{\mathrm{FE}} \approx 0.25$, the value used in the certificates below. The Jacobian assembled from (3.6) agrees with central differences of A_h to $2.2 \cdot 10^{-11}$.

Table 1 compares the constants of (3.5) with the values attained on Θ. The analytic constants exceed the sampled ones by roughly two orders of magnitude, which is the dominant source of conservatism in what follows.

TABLE 1. Constants of Lemma 3.1 for the test problem. Sampled values are maxima of difference quotients over 400 random states at $h = 1/40$.

	from (3.5)	sampled
L_A	6.76	0.107
L'_A	27.1	0.183
$\sup_\Theta \|A'\|$	6.76	0.154

7.2. **Surrogates and iteration.** The surrogate is a fully connected network with two hidden layers of width 128 and tanh activation, with affine input and output normalization folded into the architecture so that values and Jacobians are returned in physical units. We compare a value-only loss with the Sobolev loss (6.2) at $\rho = 0.15$, using designs of size $N \in \{2^8, 2^{10}, 2^{12}\}$ drawn uniformly from Θ. Derivative labels are obtained from $p = 6$ tangent solves that reuse the factorization of the forward solve. Synthetic data are generated from the reference solver with Gaussian noise rescaled so that its Euclidean norm equals 1% of the norm of the exact data. The iteration (5.1) uses ω equal to the reciprocal of the estimated maximum of $\|G'\|^2$ near the true state and $\tau = 1.1 \cdot 2/(1 - 2\eta)$, with η from Proposition 5.2.

7.3. **E1: tightness of the certificate.** Table 2 reports, for each design and mesh, the empirical errors e_0, e_1 on the design, the uniform errors $\varepsilon_0, \varepsilon_1$ estimated on 300 independent states, and the certificate (4.2) evaluated with the analytic constants, the spectral-norm bounds on L_G, L'_G of Remark 4.4 and the observed fill distance.

Three observations follow. First, the certificate is correct but very loose: it exceeds the estimated uniform error by factors between 94 and 3411, and the gap widens as the empirical error decreases, because the sampling term then dominates. Second, the sampling term is responsible for almost all of the gap. With $p = 6$ the fill distance falls only from 0.30 to 0.15 as N grows from 256 to 4096, consistent with the $N^{-1/6}$ rate of Lemma 4.2, and it is multiplied by $\Lambda_0 \approx 7.9$ or $\Lambda_1 \approx 32$. Third, the analytic constants of Table 1 inflate Λ_k by about

TABLE 2. E1 and E2. Empirical and uniform errors, and the ratio of the certificate (4.2) to the estimated uniform error. S = Sobolev loss, V = value-only loss.

h	loss	N	e_0	e_1	ε_0	ε_1	$\mathrm{cert}_0/\varepsilon_0$	$\mathrm{cert}_1/\varepsilon_1$
1/20	S	256	$1.39 \cdot 10^{-3}$	$4.26 \cdot 10^{-2}$	$2.35 \cdot 10^{-3}$	$3.57 \cdot 10^{-2}$	1002	275
1/20	V	256	$4.16 \cdot 10^{-3}$	$9.17 \cdot 10^{-2}$	$6.15 \cdot 10^{-3}$	$8.60 \cdot 10^{-2}$	366	109
1/20	S	1024	$3.12 \cdot 10^{-4}$	$9.60 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$2.45 \cdot 10^{-2}$	1001	255
1/20	V	1024	$1.09 \cdot 10^{-3}$	$3.86 \cdot 10^{-2}$	$3.61 \cdot 10^{-3}$	$5.90 \cdot 10^{-2}$	412	104
1/20	S	4096	$3.21 \cdot 10^{-4}$	$1.33 \cdot 10^{-2}$	$9.02 \cdot 10^{-4}$	$8.37 \cdot 10^{-3}$	1364	598
1/20	V	4096	$1.58 \cdot 10^{-3}$	$4.96 \cdot 10^{-2}$	$1.59 \cdot 10^{-3}$	$3.49 \cdot 10^{-2}$	751	141
1/40	S	256	$9.02 \cdot 10^{-4}$	$2.75 \cdot 10^{-2}$	$1.76 \cdot 10^{-3}$	$4.43 \cdot 10^{-2}$	1242	205
1/40	V	256	$3.05 \cdot 10^{-3}$	$7.79 \cdot 10^{-2}$	$5.54 \cdot 10^{-3}$	$9.16 \cdot 10^{-2}$	377	94
1/40	S	1024	$4.03 \cdot 10^{-4}$	$1.32 \cdot 10^{-2}$	$5.22 \cdot 10^{-4}$	$1.54 \cdot 10^{-2}$	2830	393
1/40	V	1024	$1.21 \cdot 10^{-3}$	$4.35 \cdot 10^{-2}$	$1.83 \cdot 10^{-3}$	$5.11 \cdot 10^{-2}$	784	116
1/40	S	4096	$4.09 \cdot 10^{-4}$	$1.75 \cdot 10^{-2}$	$3.56 \cdot 10^{-4}$	$1.11 \cdot 10^{-2}$	3411	445
1/40	V	4096	$1.67 \cdot 10^{-3}$	$5.84 \cdot 10^{-2}$	$1.15 \cdot 10^{-3}$	$4.06 \cdot 10^{-2}$	1026	120

two orders of magnitude; replacing L_A and L'_A by their sampled values reduces the certificates by the same factor and still leaves a gap of one to two orders of magnitude from the fill distance alone. The certificate is therefore useful as a qualitative guide to which quantity to control, and not as a numerical error bar at these design sizes.

7.4. **E2: derivative-informed training.** The paired rows of Table 2 isolate the effect of the loss. At equal N and h, Sobolev training reduces the empirical derivative error e_1 by a factor between 2.1 and 4.0 and the empirical value error e_0 by a factor between 3.0 and 5.2; on the independent test set the reduction in ε_1 ranges from 1.9 to 4.2. The effect is systematic across every design size and both meshes. Sobolev training is also between 4 and 7 times more expensive per epoch in our implementation, since each step differentiates the network once per output component. Since ε_1 enters γ through $\rho\varepsilon_1$ and enters the final bound (5.8) on the same footing as ε_0, this supports the recommendation of Section 6, and it confirms Remark 4.6: the value-only surrogate at $N = 4096$ still has $\varepsilon_1 = 3.5 \cdot 10^{-2}$, four times the $8.4 \cdot 10^{-3}$ of the Sobolev surrogate trained on the same design.

7.5. **E3: the stopping threshold.** For six true states drawn from $[0.05, 0.45]^6$ with 1% noise, the median constants are $\sigma = 6.7 \cdot 10^{-3}$, $\rho = 1.7 \cdot 10^{-2}$, $\eta = 0.43$, $\tau = 15.4$, $\delta = 1.7 \cdot 10^{-3}$ and $\gamma = 4.3 \cdot 10^{-3}$, giving thresholds $\tau\gamma = 6.6 \cdot 10^{-2}$ and $\tau\delta = 2.6 \cdot 10^{-2}$. Starting from the undamaged state $\vartheta_0 = 0$, which is the initial guess available in practice, the certified rule (5.6) stops at $k_* = 0$ in every run: the threshold $\tau\gamma$ exceeds the discrepancy $|A(0) - y^\delta|$ produced by the damage itself, so the rule accepts the undamaged state and returns a median parameter error of 0.655. The same iteration stopped at the uninflated threshold $\tau\delta$ runs for one to three steps and returns a median error of 0.340.

This is a negative result for the guarantee as a practical stopping rule at this noise level, and it has a clear cause. The inflation γ is proportional to $(1+\eta)(\delta+$

$\varepsilon_0) + 2(1+\eta)\varepsilon_0 + \rho\varepsilon_1$, and τ diverges as $\eta \to 1/2$; here $\eta = 0.43$ already forces $\tau = 15.4$. The certified error bound (5.7) evaluates to a median of 14.8, which is vacuous, since $|\vartheta| \le \sqrt{6}\,\vartheta_{\max} = 1.22$ for every admissible state. Two conclusions follow for practice: the inflated threshold should be used only when $\varepsilon_0, \varepsilon_1$ and δ are small relative to the signal produced by the damage of interest, and the theory is informative about the *ordering* of the error sources rather than about the size of the final error.

7.6. **E4: sensor layout.** Three layouts of $M = 12$ sensors were compared at $h = 1/40$, $N = 1024$ with the Sobolev loss: the uniform 4×3 grid of Section 7.1, a greedy E-optimal layout maximizing $\sigma_{\min}(A'(0))$ over a candidate grid, and a random selection from the same candidates. The greedy layout roughly doubles the design criterion, $\sigma_{\min}(A'(0)) = 7.66 \cdot 10^{-3}$ against $4.09 \cdot 10^{-3}$ for the uniform grid and $4.37 \cdot 10^{-3}$ for the random one, and the same ordering holds at the true states. Over five true states with the $\tau\delta$ rule, the median identification errors are 0.321 (E-optimal), 0.327 (random) and 0.332 (uniform), with overlapping interquartile ranges. The design criterion therefore moves in the direction predicted by (6.1), and the identification error follows the same ordering, but with five states the difference in final error is within the spread and this experiment does not establish it. A larger study, and layouts optimized over a finer candidate set, would be needed to resolve the effect.

7.7. **E5: cost.** One surrogate step, comprising a network evaluation and one reverse-mode differentiation, takes 2.56 ms, against 32.0 ms for a finite element step at $h = 1/40$ comprising one forward solve and six tangent solves, a speed-up of 12.5 per step. Generating $N = 1024$ labels and training the Sobolev surrogate costs about 137 s offline. With the median $k_* = 2$ steps observed in E3, the saving per identification run is only 59 ms, so the surrogate pays for its offline cost after roughly $2.3 \cdot 10^3$ runs. The break-even point is this high precisely because the iteration terminates after very few steps on this problem; it falls in proportion to the number of iterations per run, so a problem needing 10^2 steps per run would break even after some tens of runs. This is consistent with the scaling argument of Section 6.3, and it sharpens it: the relevant quantity is the total number of *iterations* served by one surrogate, not the number of runs.

7.8. **Summary.** The numerical study confirms the qualitative content of the analysis: the finite element error obeys Lemma 3.2, derivative-informed training reduces ε_1 by factors of two to four as Remark 4.6 predicts, and E-optimal placement increases σ as (6.1) predicts. It also delimits the quantitative reach of the result: with the analytic constants of (3.5) and a six-dimensional design the certificate (4.2) is two to three orders of magnitude loose, the certified stopping rule is too conservative to be used at 1% noise, and the error bound (5.7) is vacuous on this problem. Sharper constants, structural assumptions that defeat the $N^{-1/p}$ fill-distance rate, or a posteriori estimates of ε_0 and ε_1 on a held-out design would each narrow the gap, and we regard the last as the most promising route for practice.

8. Limitations and conclusion

8.1. **Limitations. Structural model.** The analysis is carried out for a scalar elliptic model. For linear elasticity the argument of Lemma 3.1 carries over with the Poincaré inequality replaced by Korn's inequality, and Lemma 3.2 carries over on convex domains with smooth coefficients. Kirchhoff plates and Euler–Bernoulli beams lead to fourth-order problems, for which the finite element rates, and hence the exponent of h in (5.8), depend on the chosen conforming or mixed discretization.

Measurement modality. Static load tests are assumed. Vibration-based monitoring replaces the solution map by eigenvalue and mode-shape maps, whose derivatives are Lipschitz only away from eigenvalue crossings, and the constants of Lemma 3.1 must then be expressed in terms of a spectral gap.

Locality and dimension. The guarantee is local, requiring an initial guess within ρ of the true state, and the sampling term deteriorates like $N^{-1/p}$ in the number of zones. Removing the latter requires structural assumptions, for instance that each sensor depends only weakly on distant zones, under which sparse or compositional surrogates may achieve rates independent of p. Newton-type schemes such as the regularizing Levenberg–Marquardt method [10] would enlarge the region of convergence and can be analysed along the same lines.

Model error. Deviations of the real structure from the model, such as uncertain boundary conditions or temperature effects, are not represented in (3.1). They can be absorbed into δ when a bound is available, but a systematic treatment calls for a Bayesian formulation [24] or for learned model corrections [21].

8.2. **Conclusion.** We have shown that replacing the forward model of a structural identification problem by a neural surrogate preserves the convergence guarantees of the Landweber iteration, provided that the surrogate is accurate in value and in derivative and that the stopping rule accounts for its error. The resulting bound separates measurement noise, training error, sampling density and discretization, and it is governed by the smallest singular value of the linearized sensor map. These results give a mathematical basis for three practical choices in data-driven structural health monitoring: E-optimal sensor placement, derivative-informed training with a weight equal to the squared radius of the search region, and a stopping threshold inflated by the certified surrogate error. The numerical study of Section 7 confirms the qualitative predictions and quantifies the conservatism of the certificate.

References

[1] S. Arridge, P. Maass, O. Öktem and C.-B. Schönlieb, Solving inverse problems using data-driven models, *Acta Numerica* 28 (2019), 1–174.

[2] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj and D. J. Inman, A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications, *Mechanical Systems and Signal Processing* 147 (2021), 107077.

[3] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd edition, Springer, 2008.

[4] T. Chen and H. Chen, Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems, *IEEE Transactions on Neural Networks* 6(4) (1995), 911–917.
[5] W. M. Czarnecki, S. Osindero, M. Jaderberg, G. Świrszcz and R. Pascanu, Sobolev training for neural networks, *Advances in Neural Information Processing Systems* 30 (2017).
[6] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, 1996.
[7] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*, Wiley, 2013.
[8] M. I. Friswell and J. E. Mottershead, *Finite Element Model Updating in Structural Dynamics*, Kluwer, 1995.
[9] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, 1985.
[10] M. Hanke, A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems, *Inverse Problems* 13(1) (1997), 79–95.
[11] M. Hanke, A. Neubauer and O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numerische Mathematik* 72 (1995), 21–37.
[12] B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, de Gruyter, 2008.
[13] F. Kamalov and H. H. Leung, Fixed point theory: a review, *arXiv preprint* arXiv:2309.03226 (2023).
[14] F. Kamalov, Asymptotic behavior of SMOTE-generated samples using order statistics, *Gulf Journal of Mathematics* 17(2) (2024), 327–336.
[15] F. Kamalov, H. Sulieman and W. Pedrycz, Theoretical convergence of SMOTE-generated samples, *arXiv preprint* arXiv:2601.01927 (2026).
[16] F. Kamalov, F. Thabtah, R. Sivaraj and N. Abdelhamid, Path-sampled integrated gradients, *Gulf Journal of Mathematics* 22(1) (2026), 1–10.
[17] F. Kamalov, H. Sulieman, A. Alzaatreh, M. Emarly, H. Chamlal and M. Safaraliev, Mathematical methods in feature selection: a review, *Mathematics* 13(6) (2025), 996.
[18] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart and A. Anandkumar, Neural operator: learning maps between function spaces with applications to PDEs, *Journal of Machine Learning Research* 24(89) (2023), 1–97.
[19] S. Lanthaler, S. Mishra and G. E. Karniadakis, Error estimates for DeepONets: a deep learning framework in infinite dimensions, *Transactions of Mathematics and Its Applications* 6(1) (2022), tnac001.
[20] L. Lu, P. Jin, G. Pang, Z. Zhang and G. E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nature Machine Intelligence* 3 (2021), 218–229.
[21] S. Lunz, A. Hauptmann, T. Tarvainen, C.-B. Schönlieb and S. Arridge, On learned operator correction in inverse problems, *SIAM Journal on Imaging Sciences* 14(1) (2021), 92–127.
[22] T. O'Leary-Roseberry, P. Chen, U. Villa and O. Ghattas, Derivative-informed neural operator: an efficient framework for high-dimensional parametric derivative learning, *Journal of Computational Physics* 496 (2024), 112555.
[23] M. Raissi, P. Perdikaris and G. E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics* 378 (2019), 686–707.
[24] A. M. Stuart, Inverse problems: a Bayesian perspective, *Acta Numerica* 19 (2010), 451–559.

[1]Faculty of Engineering, Tajik Technical University, Dushanbe, Tajikistan.
Email address: said.kamolov@yahoo.com; said.kamolov@ttu.tj