

TO BE OR NOT TO BE STABLE, THAT IS THE QUESTION: UNDERSTANDING NEURAL NETWORKS FOR INVERSE PROBLEMS*

DAVIDE EVANGELISTA[†], ELENA LOLI PICCOLOMINI[†], ELENA MOROTTI[‡],
AND JAMES G. NAGY[§]

Abstract. The solution of linear inverse problems arising, for example, in signal and image processing is a challenging problem since the ill-conditioning amplifies, in the solution, the noise present in the data. Recently introduced algorithms based on deep learning overwhelm the more traditional model-based approaches in performance, but they typically suffer from instability with respect to data perturbation. In this paper, we theoretically analyze the trade-off between stability and accuracy of neural networks when used to solve linear imaging inverse problems for cases that are not underdetermined. Moreover, we propose different supervised and unsupervised solutions to increase the network stability and maintain a good accuracy, by means of regularization properties inherited from a model-based iterative scheme during the network training. Extensive numerical experiments on image deblurring confirm the theoretical results and the effectiveness of the proposed deep learning-based approaches to handle noise on the data.

Key words. neural networks stability, linear inverse problems, deep learning algorithms, image deblurring, trade-off accuracy stability

MSC codes. 65K10, 68T07, 68U10

DOI. 10.1137/23M1586872



See reproducibility of
computational results
at end of the article.

1. Introduction. Linear inverse problems of the form

$$(1.1) \quad \mathbf{y} = \mathbf{A}\mathbf{x},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a full-rank matrix discretizing a linear operator, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ with $m \geq n$, arise in various image processing tasks, such as deblurring or tomographic reconstruction [14, 15, 23]. It is well known that in these applications, (1.1) represents the discretization of an ill-posed problem. Following the well-known Hadamard definition, a problem is ill-posed if either a solution does not exist, the solution is not unique, or it does not continuously depend on the data \mathbf{y} . In the case considered in (1.1) the third condition holds; thereby the computation of the solution

*Submitted to the journal's Machine Learning Methods for Scientific Computing section July 17, 2023; accepted for publication (in revised form) August 13, 2024; published electronically January 29, 2025.

<https://doi.org/10.1137/23M1586872>

Funding: This work was partially supported by the GNCS—Gruppo Nazionale per il Calcolo Scientifico (“*Apprendimento automatico e tecniche variazionali per la tomografia*” INdAM GNCS Project, grant CUP_E55F55000270001), by the U.S. National Science Foundation, grants DMS-2038118 and DMS-2208294, and by the PRIN 2022 project “STILE: Sustainable Tomographic Imaging with Learning and rEgularization,” project 20225STXSB, funded by the European Commission under the NextGeneration EU program.

[†]Department of Computer Science and Engineering, University of Bologna, Bologna, Italy (davide.evangelista5@unibo.it, elena.loli@unibo.it).

[‡]Department of Political and Social Sciences, University of Bologna, Bologna, Italy (elena.morotti4@unibo.it).

[§]Department of Mathematics, Emory University, Atlanta, GA USA (jnagy@emory.edu).

becomes very challenging when noise affects the data. In this work, we consider data corrupted by Gaussian noise, i.e.,

$$(1.2) \quad \mathbf{y}^\delta = \mathbf{A}\mathbf{x}^{gt} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \delta^2 \mathbf{I}),$$

where δ denotes the standard deviation of the white additive Gaussian noise, \mathbf{I} is the identity matrix, and \mathbf{x}^{gt} is the ground truth, clean image.

Traditional regularization approaches tackle problem (1.2) as the minimization of an objective function containing a data-fit term and a regularization prior, with possible further constraints on the solution [4, 9]. These terms theoretically grant stability, but, in general, the computational time required by solvers is high, and it may be necessary to choose many parameters, tuning them by trial and error on the data.

In the last few years, neural networks have been introduced with great success for the solution of problem (1.2), since they are capable in many cases of achieving greater accuracy than iterative regularized methods [1, 6, 18]. However, their high accuracy is obtained at the expense of robustness against noise in the input data, hence noise-related issues still persist. Specifically, these networks frequently yield suboptimal results when applied to data contaminated with noise that differs from that encountered during the training phase. This tendency is commonly referred to as network *instability*.

Some authors have already studied the behavior of neural networks in the presence of noise on the data, focusing on the solution of underdetermined imaging inverse problems (i.e., when $m < n$ in (1.1)) [12, 30, 31, 2, 8, 17, 19, 21, 22, 25, 27, 32, 33, 13, 7]. We note that the paper [13] offers a comprehensive bibliography on this topic, with the authors remarking that “*stability implies a universal barrier on performance.*” However, to the best of our knowledge, a mathematically grounded understanding is still lacking and no works address the case of $m \geq n$.

Contributions. In this work, we look at neural networks as solvers of discrete ill-posed problems, and we contribute to the state of the art as follows.

First, we adapt the regularization theory presented by Engl, Hanke, and Neubauer in [9] for solving discrete inverse problems using neural networks. It is noteworthy that Engl et al. examined regularization in Hilbert spaces, while our focus is on discrete inverse problems. Prior to introducing neural networks as solvers, we present a more general theory encompassing a broader class of functions, termed *reconstructors*, designed for addressing problem (1.2). Within this framework, we first formalize the two fundamental concepts of η^{-1} -accuracy and ϵ -stability, and then we present significant findings for a class of functions called *stabilizers*. We establish a mathematical relationship that quantifies the trade-off between stability and accuracy, demonstrating that enhancing a solver’s stability is impossible without compromising its accuracy. In this theoretical approach, neural networks have been analyzed as formal mathematical operators, shedding light on their widely discussed “black-box” nature/misinterpretation.

Second, we propose a new ground truth-free scheme for reconstructors based on neural networks. We refer to this approach as the regularized neural network (ReNN), as the target images used in the training procedure are solutions of (1.2) computed through a regularization method. Beyond being more stable than commonly used neural networks as reconstructors, it is applicable in scenarios where collecting a set of ground-truth solutions is challenging or impossible, such as in medical imaging.

Finally, we present a novel stabilization strategy tailored for deep learning-based solvers, which incorporates a stabilizer within a preprocessing operator plugged into a neural network reconstructor. This approach demonstrates substantial efficacy in

handling elevated noise levels in data. We have termed this methodology stabilized neural network (StNN). Furthermore, and when integrated with the ReNN scheme, it evolves into the StReNN framework.

Structure of the paper. The paper is structured as follows. In section 2, we introduce theoretical concepts related to reconstructors for solving an inverse problem of the form presented in (1.2). In section 3, we present stabilizers and elucidate their effectiveness by stating their properties, then section 4 is dedicated to reconstructors based on neural networks and presents our proposals. Following that, in section 5, we describe our experimental setup, whereas section 6 showcases numerical results pertaining to deblurring and denoising. Finally, section 7 comprises conclusions and outlines potential directions for future work.

2. Reconstructors for the solution of linear inverse problems. This section establishes the theoretical background of the paper, providing essential definitions and preliminary results. To improve the readability of the work, however, we start by introducing the notation we will use throughout the paper. We always consider \mathbf{x}^{gt} to lie in a subset \mathcal{X} of \mathbb{R}^n , the set of admissible data. We denote by $\mathcal{Y} = \text{Rg}(\mathbf{A}, \mathcal{X})$ the range of \mathbf{A} over \mathcal{X} , where \mathbf{A} is a continuous linear operator. We assume \mathcal{Y} to be dense-in-itself (i.e., with no isolated points) so that, for any admissible $\mathbf{x}^{gt} \in \mathcal{X}$ and any neighborhood V of $\mathbf{y} = \mathbf{A}\mathbf{x}^{gt}$, there is at least an $\mathbf{x}' \in \mathcal{X}$, $\mathbf{x}' \neq \mathbf{x}^{gt}$, such that $\mathbf{y}' = \mathbf{A}\mathbf{x}' \in V$. When $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ or $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^m$, then $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ will be Euclidean norms. For any $\epsilon > 0$, we also define $\mathcal{Y}^\epsilon = \{\mathbf{y} + \mathbf{e}; \mathbf{y} \in \mathcal{Y}, \|\mathbf{e}\| \leq \epsilon\}$. With the following definitions, we can formalize the concept of reconstructor to solve problem (1.1) accurately.

DEFINITION 2.1. Any continuous function $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$, mapping \mathbf{y} to $\mathbf{x} = \Psi(\mathbf{y})$, is called a reconstructor.

DEFINITION 2.2. A reconstructor $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is said to be η^{-1} -accurate, with $\eta > 0$, if

$$\eta = \sup_{\mathbf{x}^{gt} \in \mathcal{X}} \|\Psi(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\|.$$

We define the set $\mathcal{R}_\eta = \{\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^n; \Psi \text{ is a reconstructor with accuracy } \eta^{-1}\}$.

We observe that without any other restriction, η could be infinite.

Example 2.3. An ∞ -accurate reconstructor of problem (1.1) is given by

$$\Psi^\dagger(\mathbf{y}) = \mathbf{A}^\dagger \mathbf{y} = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \mathbf{y},$$

where \mathbf{A}^\dagger is the pseudo-inverse matrix, \mathbf{A}^* is the transpose of \mathbf{A} , and the last equality holds since \mathbf{A} is assumed to be full-rank. In this case $\Psi^\dagger : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is ∞ -accurate, as

$$\|\Psi^\dagger(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\| = \|(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\| = \|(\mathbf{A}^* \mathbf{A})^{-1} (\mathbf{A}^* \mathbf{A}) \mathbf{x}^{gt} - \mathbf{x}^{gt}\| = 0.$$

To avoid any issue, we will always consider reconstructors with finite η in the following. However, reconstructors are rarely applied to noise-free data, hence a focus on the robustness of reconstructors with respect to noise is necessary.

DEFINITION 2.4. Let $\epsilon > 0$, and let Ψ be an η^{-1} -accurate reconstructor applied to problem (1.2). We define the ϵ -stability constant C_Ψ^ϵ of Ψ as

$$C_\Psi^\epsilon = \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|}.$$

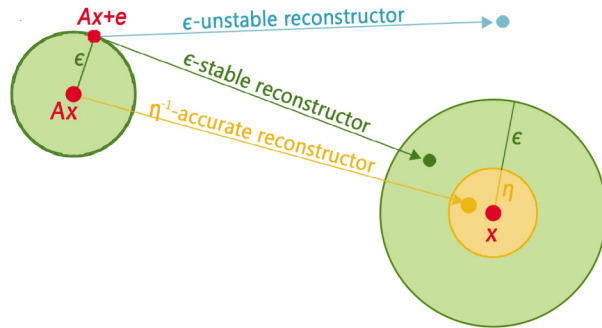


FIG. 2.1. Graphical representation of the ϵ -stability and ϵ -instability for an η^{-1} -accurate reconstructor.

We will consider in the following the realistic case of $C_{\Psi}^{\epsilon} < \infty$.

DEFINITION 2.5. The reconstructor Ψ is said to be ϵ -stable for a given $\epsilon > 0$ if $C_{\Psi}^{\epsilon} \in [0, 1)$. Otherwise, Ψ is said to be ϵ -unstable.

An ϵ -stable reconstructor Ψ does not amplify corruptions having norm less than ϵ (as graphically represented in Figure 2.1), since Definition 2.4 implies

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \leq \eta + C_{\Psi}^{\epsilon}\|\mathbf{e}\| \quad \forall \mathbf{x}^{gt} \in \mathcal{X}, \forall \mathbf{e} \in \mathbb{R}^m, \|\mathbf{e}\| \leq \epsilon.$$

DEFINITION 2.6. We define the stability radius ρ of Ψ as

$$\rho = \sup\{\epsilon > 0; C_{\Psi}^{\epsilon} \in [0, 1)\}.$$

Example 2.7. A reconstructor with an infinite stability radius is the following. Given $\epsilon > 0$, if μ is a probability distribution over \mathcal{X} (for example, μ is the normalized Lebesgue distribution over \mathcal{X}), the reconstructor defined as

$$\Psi^{\mathcal{X}, \epsilon}(\mathbf{y}^{\delta}) = \int_{\mathcal{X}} \mathbf{x}\mu(d\mathbf{x}) \quad \forall \mathbf{y}^{\delta} \in \mathcal{Y}^{\epsilon}$$

is ϵ -stable independently from the value of $\epsilon > 0$. Indeed,

$$\|\Psi^{\mathcal{X}, \epsilon}(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| = \left\| \int_{\mathcal{X}} \mathbf{x}\mu(d\mathbf{x}) - \mathbf{x}^{gt} \right\| \leq \rho(\mathcal{X}),$$

where $\rho(\mathcal{X})$ is the radius of \mathcal{X} , defined as $\rho(\mathcal{X}) = \inf\{r > 0 : \mathcal{X} \subseteq \mathcal{B}(\int_{\mathcal{X}} \mathbf{x}\mu(d\mathbf{x}); r)\}$. As a consequence the stability constant is infinite regardless of ϵ , and $\Psi^{\mathcal{X}(\mathbf{y}), \epsilon}$ has accuracy $\rho(\mathcal{X})^{-1}$.

Example 2.8. The pseudo-inverse reconstructor $\Psi^{\dagger}(\mathbf{y})$ in Example 2.3 is unstable for any $\epsilon > 0$ when \mathbf{A} is ill-conditioned. Indeed,

$$\begin{aligned} \|\Psi^{\dagger}(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| &= \|(\mathbf{A}^* \mathbf{A})^{-1}(\mathbf{A}^* \mathbf{A})\mathbf{x}^{gt} + (\mathbf{A}^* \mathbf{A})^{-1}\mathbf{A}^* \mathbf{e} - \mathbf{x}^{gt}\| \\ &= \|(\mathbf{A}^* \mathbf{A})^{-1}\mathbf{A}^* \mathbf{e}\|. \end{aligned}$$

If $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ is the singular value decomposition (SVD) of \mathbf{A} , then

$$(\mathbf{A}^* \mathbf{A})^{-1}\mathbf{A}^* \mathbf{e} = (\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^*)^{-1}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^* \mathbf{e} = \mathbf{V}\mathbf{\Sigma}^{\dagger}\mathbf{U}^* \mathbf{e} = \sum_{i=1}^n \frac{\mathbf{u}_i^T \mathbf{e}}{\sigma_i} \mathbf{v}_i,$$

which implies that $\|(\mathbf{A}^* \mathbf{A})^{-1}\mathbf{A}^* \mathbf{e}\| \gg \|\mathbf{e}\|$ when \mathbf{A} has singular values close to zero.

These examples shed light on a possible conflict between accuracy and stability for a given reconstructor Ψ . In the next paragraphs, we study this relationship.

2.1. Accuracy vs. stability trade-off. We can derive a relation between accuracy and stability, which becomes particularly interesting when \mathbf{A} is ill-conditioned.

LEMMA 2.9. *Let $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be an η^{-1} -accurate reconstructor. Then, for any $\mathbf{x}^{gt} \in \mathcal{X}$ and for any $\epsilon > 0$, $\exists \tilde{\mathbf{e}} \in \mathbb{R}^m$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$ such that*

$$(2.1) \quad \|\Psi(\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}}) - \mathbf{x}^{gt}\| \geq \|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \eta.$$

Proof. Since $\mathbf{A}\mathbf{x}^{gt} \in \mathcal{Y}$ for any $\mathbf{x}^{gt} \in \mathcal{X}$, and since \mathcal{Y} has no isolated points, then for any $\epsilon > 0$ there is an $\tilde{\mathbf{e}} \in \mathbb{R}^m$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$ such that $\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}} \in \mathcal{Y}$. Thus, $\exists \mathbf{x}' \in \mathcal{X}$ such that $\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}} = \mathbf{A}\mathbf{x}'$. Consequently,

$$\begin{aligned} \|\Psi(\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}}) - \mathbf{x}^{gt}\| &= \|\Psi(\mathbf{A}\mathbf{x}') - \mathbf{x}^{gt}\| \geq \|\mathbf{x}' - \mathbf{x}^{gt}\| - \|\Psi(\mathbf{A}\mathbf{x}') - \mathbf{x}'\| \\ &\geq \|\mathbf{x}' - \mathbf{x}^{gt}\| - \eta. \end{aligned}$$

Since $\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}} = \mathbf{A}\mathbf{x}'$ by construction, it holds that $\tilde{\mathbf{e}} = \mathbf{A}(\mathbf{x}^{gt} - \mathbf{x}')$, which implies that $\mathbf{x}^{gt} - \mathbf{x}' = \mathbf{A}^\dagger \tilde{\mathbf{e}}$. To conclude,

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}}) - \mathbf{x}^{gt}\| \geq \|\mathbf{x}' - \mathbf{x}^{gt}\| - \eta = \|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \eta. \quad \square$$

We define the following set:

$$(2.2) \quad E(\mathbf{x}^{gt}, \Psi) = \{\mathbf{e} \in \mathbb{R}^m; \text{equation (2.1) holds for } \mathbf{e}, \text{ for some } \epsilon > 0\}.$$

THEOREM 2.10 (trade-off theorem). *Under the assumptions of Lemma 2.9 it holds that, for any $\mathbf{x}^{gt} \in \mathcal{X}$ and for any $\tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$,*

$$(2.3) \quad C_\Psi^\epsilon \geq \frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - 2\eta}{\|\tilde{\mathbf{e}}\|}.$$

Proof. For any $\mathbf{x}^{gt} \in \mathcal{X}$,

$$C_\Psi^\epsilon = \sup_{\substack{\mathbf{x} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{A}\mathbf{x} + \mathbf{e}) - \mathbf{x}\| - \eta}{\|\mathbf{e}\|} \geq \sup_{\|\mathbf{e}\| \leq \epsilon} \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|}.$$

If $\tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$, $\|\tilde{\mathbf{e}}\| \leq \epsilon$, is a perturbation defined in Lemma 2.9, we have

$$\begin{aligned} C_\Psi^\epsilon &\geq \sup_{\|\mathbf{e}\| \leq \epsilon} \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|} \\ &\geq \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \tilde{\mathbf{e}}) - \mathbf{x}^{gt}\| - \eta}{\|\tilde{\mathbf{e}}\|} \\ &\geq \frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - 2\eta}{\|\tilde{\mathbf{e}}\|}, \end{aligned}$$

which concludes the proof. □

COROLLARY 2.11. *Given the assumptions of Theorem 2.10, if $\mathcal{X} = \mathbb{R}^n$ and if the smallest singular value $\sigma_n < 1$, then there is a constant $C(\mathbf{A}) > 0$, which depends only on \mathbf{A} , such that*

$$\rho \leq \frac{2}{\eta^{-1}C(\mathbf{A})}.$$

Proof. Consider a reconstructor Ψ . By Theorem 2.10, for any $\epsilon > 0$, any $\mathbf{x}^{gt} \in \mathcal{X}$, and any $\tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$,

$$(2.4) \quad C_{\Psi}^{\epsilon} \geq \frac{\|\mathbf{A}^{\dagger} \tilde{\mathbf{e}}\| - 2\eta}{\|\tilde{\mathbf{e}}\|}.$$

We first observe that if $\mathcal{X} = \mathbb{R}^n$, then $E(\mathbf{x}^{gt}, \Psi) = \mathcal{Y} := Rg(\mathbf{A}, \mathcal{X})$ for any $\mathbf{x}^{gt} \in \mathcal{X}$. Indeed, $\tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$ if and only if there exists $\mathbf{x}' \in \mathcal{X}$ such that $\tilde{\mathbf{e}} = \mathbf{A}(\mathbf{x}^{gt} - \mathbf{x}')$. Since $\mathcal{X} = \mathbb{R}^n$ is closed under addition, then $\mathbf{x}^{gt} - \mathbf{x}' \in \mathcal{X}$, which implies that $\tilde{\mathbf{e}} \in \mathcal{Y}$, and thus $E(\mathbf{x}^{gt}, \Psi) \subseteq \mathcal{Y}$. Conversely, if $\mathbf{y} \in \mathcal{Y}$, then by definition there exists $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$. By defining $\mathbf{x}' = \mathbf{x}^{gt} - \mathbf{x}$, then $\mathbf{y} = \mathbf{A}(\mathbf{x}^{gt} - \mathbf{x}')$, which implies that $\mathbf{y} \in E(\mathbf{x}^{gt}, \Psi)$ and consequently $E(\mathbf{x}^{gt}, \Psi) = \mathcal{Y}$.

Now, let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ be the SVD of \mathbf{A} and define $\tilde{\mathbf{e}} = \mathbf{A}\left(\frac{\epsilon}{\sigma_n}\mathbf{v}_n\right)$, where σ_n and \mathbf{v}_n are the smallest singular value of \mathbf{A} and its associated right-singular vector, respectively. Note that $\tilde{\mathbf{e}} \in \mathcal{Y} = E(\mathbf{x}^{gt}, \Psi)$ by definition. Moreover,

$$\begin{aligned} \tilde{\mathbf{e}} &= \mathbf{A}\left(\frac{\epsilon}{\sigma_n}\mathbf{v}_n\right) = \mathbf{U}\Sigma\mathbf{V}^*\left(\frac{\epsilon}{\sigma_n}\mathbf{v}_n\right) = \frac{\epsilon}{\sigma_n} \sum_{i=1}^n \sigma_i \mathbf{u}_i (\mathbf{v}_i^T \mathbf{v}_n) \\ &= \frac{\epsilon}{\sigma_n} \sigma_n \mathbf{u}_n = \epsilon \mathbf{u}_n, \end{aligned}$$

from which $\|\tilde{\mathbf{e}}\| = \epsilon \|\mathbf{u}_n\| = \epsilon \leq \epsilon$. Consequently, (2.3) holds for $\tilde{\mathbf{e}}$. Additionally,

$$\mathbf{A}^{\dagger} \tilde{\mathbf{e}} = \mathbf{A}^{\dagger} \mathbf{A} \left(\frac{\epsilon}{\sigma_n} \mathbf{v}_n \right) = \frac{\epsilon}{\sigma_n} \mathbf{v}_n,$$

hence $\|\mathbf{A}^{\dagger} \tilde{\mathbf{e}}\| = \frac{\epsilon}{\sigma_n}$. Given that, (2.3) reads as

$$C_{\Psi}^{\epsilon} \geq \frac{\|\mathbf{A}^{\dagger} \tilde{\mathbf{e}}\| - 2\eta}{\|\tilde{\mathbf{e}}\|} = \frac{\frac{\epsilon}{\sigma_n} - 2\eta}{\epsilon}.$$

As a consequence of the above relationship, if $\frac{\frac{\epsilon}{\sigma_n} - 2\eta}{\epsilon} > 1$, then $C_{\Psi}^{\epsilon} > 1$, i.e., $\rho \leq \epsilon$. A simple computation shows that

$$\begin{aligned} \frac{\frac{\epsilon}{\sigma_n} - 2\eta}{\epsilon} > 1 &\iff \frac{\epsilon}{\sigma_n} - 2\eta > \epsilon \\ &\iff \frac{\epsilon}{\sigma_n} - \epsilon > 2\eta \\ &\iff \epsilon \left(\frac{1}{\sigma_n} - 1 \right) > 2\eta \\ &\iff \epsilon \left(\frac{1 - \sigma_n}{\sigma_n} \right) > 2\eta \\ &\iff \epsilon > \frac{2\eta}{\frac{1 - \sigma_n}{\sigma_n}} = \frac{2}{\eta^{-1} C(\mathbf{A})}, \end{aligned}$$

where we used that $\sigma_n < 1$ and we defined $C(\mathbf{A}) := \frac{1 - \sigma_n}{\sigma_n}$. Thus,

$$\rho := \sup\{\epsilon > 0 : C_{\Psi}^{\epsilon} < 1\} \leq \frac{2}{\eta^{-1} C(\mathbf{A})},$$

concluding the proof. \square

The relation in Corollary 2.11 between the stability radius ρ and the accuracy η^{-1} suggests that there exists a trade-off between accuracy and stability, showing that a very accurate reconstructor is unstable for noise corruption larger than $\frac{2}{\eta^{-1}C(\mathbf{A})}$. We remark that for ill-conditioned problems, $C(\mathbf{A}) = \frac{1-\sigma_n}{\sigma_n}$ can be very large, making the radius potentially very small.

Similarly, Theorem 2.10 shows that a reconstructor Ψ can be ϵ -stable only if its accuracy is bounded.

COROLLARY 2.12. *Given the assumptions of Theorem 2.10, there exists $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X}) \in \mathbb{R} \cup \{+\infty\}$, such that any reconstructor Ψ with accuracy $\eta^{-1} \geq \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})^{-1}$ is ϵ -unstable, i.e., $C_\Psi^\epsilon \geq 1$. Moreover, if $\mathcal{X} = \mathbb{R}^n$ and $\eta^{-1} \geq \frac{2}{C(\mathbf{A})\epsilon}$, where $C(\mathbf{A}) = \frac{1-\sigma_n}{\sigma_n}$, then Ψ is ϵ -unstable.*

Proof. From Theorem 2.10, Ψ is ϵ -unstable for a given $\epsilon > 0$ if $\frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - 2\eta}{\|\tilde{\mathbf{e}}\|} \geq 1$. Such a condition holds if and only if

$$\eta \leq \frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \|\tilde{\mathbf{e}}\|}{2}.$$

Thus, for any $\Psi \in \mathcal{R}_\eta$, if $\exists \tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$ such that $\eta \leq \frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \|\tilde{\mathbf{e}}\|}{2}$, then Ψ is ϵ -unstable. In particular, if we define

$$(2.5) \quad \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X}) = \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \Psi \in \cup_{\eta>0} \mathcal{R}_\eta \\ \tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi) \\ \|\tilde{\mathbf{e}}\| \leq \epsilon}} \frac{\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \|\tilde{\mathbf{e}}\|}{2},$$

we get the result. Note that, in general, $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$ could be infinite. In the assumption of $\mathcal{X} = \mathbb{R}^n$, we proved in Corollary 2.11 that for any $\epsilon > 0$ and any $\mathbf{x}^{gt} \in \mathcal{X}$, we can always choose $\tilde{\mathbf{e}} \in E(\mathbf{x}^{gt}, \Psi)$ with $\|\tilde{\mathbf{e}}\| \leq \epsilon$ such that $\|\mathbf{A}^\dagger \tilde{\mathbf{e}}\| - \|\tilde{\mathbf{e}}\| = \frac{1-\sigma_n}{\sigma_n} \epsilon = C(\mathbf{A})\epsilon$. Thus, Ψ is ϵ -unstable if

$$\eta \leq \frac{C(\mathbf{A})\epsilon}{2},$$

which proves the corollary. □

2.2. A sufficient condition for stability. Whenever a reconstructor is (locally) Lipschitz continuous, we can also derive conditions assessing its stability. First of all, we recall the definition of locally Lipschitz continuous reconstructors.

DEFINITION 2.13. *Given $\mathcal{Y} \subseteq \mathbb{R}^m$ and $\epsilon > 0$, we define the ϵ -Lipschitz (also called local Lipschitz) constant of Ψ over \mathcal{Y} as*

$$L^\epsilon(\Psi, \mathcal{Y}) = \sup_{\substack{\mathbf{y} \in \mathcal{Y}, \mathbf{z} \in \mathbb{R}^m \\ \|\mathbf{z} - \mathbf{y}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{z}) - \Psi(\mathbf{y})\|}{\|\mathbf{z} - \mathbf{y}\|}.$$

If $L^\epsilon(\Psi, \mathcal{Y}) < \infty$ for some $\epsilon > 0$, then Ψ is said to be locally Lipschitz continuous.

Focusing on our problem (1.2), we remark that we are interested in the cases where $\mathcal{Y} = Rg(\mathbf{A}, \mathcal{X})$. In this case, $\mathbf{y} \in \mathcal{Y}$ implies that $\exists \mathbf{x}^{gt} \in \mathcal{X}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}^{gt}$ and each $\mathbf{z} \in \mathbb{R}^m$ with $\|\mathbf{z} - \mathbf{y}\| \leq \epsilon$ can be characterized by $\mathbf{z} = \mathbf{A}\mathbf{x}^{gt} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\| \leq \epsilon$. Thus, the definition of $L^\epsilon(\Psi, \mathcal{Y})$ can be rewritten as

$$L^\epsilon(\Psi, \mathcal{Y}) = \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \Psi(\mathbf{A}\mathbf{x}^{gt})\|}{\|\mathbf{e}\|}.$$

The importance of the local Lipschitz constant $L^\epsilon(\Psi, \mathcal{Y})$ lies in its strong relationship to the stability constant C_Ψ^ϵ of the reconstructor. Indeed, if $\mathbf{y} = \mathbf{A}\mathbf{x}^{gt} \in \mathcal{Y}$ gets corrupted by additional noise \mathbf{e} with $\|\mathbf{e}\| \leq \epsilon$, then $L^\epsilon(\Psi, \mathcal{Y})$ represents the maximum possible variation of the reconstruction obtained by Ψ around the corrupted \mathbf{y} , as stated by the following proposition.

PROPOSITION 2.14. *If $\Psi \in \mathcal{R}_\eta$ has local Lipschitz constant $L^\epsilon(\Psi, \mathcal{Y})$, then, for any $\|\mathbf{e}\| \leq \epsilon$, the following holds:*

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \leq \eta + L^\epsilon(\Psi, \mathcal{Y})\|\mathbf{e}\|.$$

Proof. By the triangle inequality, it follows that

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \leq \|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \Psi(\mathbf{A}\mathbf{x}^{gt})\| + \|\Psi(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\|.$$

Since $\|\mathbf{e}\| \leq \epsilon$, the definition of local Lipschitz constant implies that

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \Psi(\mathbf{A}\mathbf{x}^{gt})\| \leq L^\epsilon(\Psi, \mathcal{Y})\|\mathbf{A}\mathbf{x}^{gt} + \mathbf{e} - \mathbf{A}\mathbf{x}^{gt}\| = L^\epsilon(\Psi, \mathcal{Y})\|\mathbf{e}\|,$$

whereas the accuracy of Ψ gives

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\| \leq \eta.$$

Thus, we can conclude that

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \leq L^\epsilon(\Psi, \mathcal{Y})\|\mathbf{e}\| + \eta. \quad \square$$

COROLLARY 2.15. *Under the assumptions of Proposition 2.14, it holds that*

$$C_\Psi^\epsilon \leq L^\epsilon(\Psi, \mathcal{Y}).$$

Proof. From the inequality in Proposition 2.14, we have

$$\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \leq \eta + L^\epsilon(\Psi, \mathcal{Y})\|\mathbf{e}\| \iff L^\epsilon(\Psi, \mathcal{Y}) \geq \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|}$$

for any $\mathbf{x}^{gt} \in \mathcal{X}$ and any $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\| \leq \epsilon$. Consequently, $L^\epsilon(\Psi, \mathcal{Y})$ is a majorant of the set

$$\left\{ \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|}; \mathbf{x}^{gt} \in \mathcal{X}, \|\mathbf{e}\| \leq \epsilon \right\}.$$

Since C_Ψ^ϵ is defined as the supremum of this set, by the minimality of the supremum we have $C_\Psi^\epsilon \leq L^\epsilon(\Psi, \mathcal{Y})$. \square

We remark that Corollary 2.15 proves that Ψ is ϵ -stable if $L^\epsilon(\Psi, \mathcal{Y}) < 1$, yielding a useful sufficient condition to the assessment of stability.

Example 2.16. Under suitable parameter choices, the Tikhonov reconstructor is ϵ -stable for any $\epsilon > 0$. The Tikhonov reconstructor is built on the Tikhonov method [26, 28] and defined as

$$(2.6) \quad \Psi^{\lambda, \mathbf{L}}(\mathbf{y}^\delta) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}^\delta\|^2 + \frac{\lambda}{2} \|\mathbf{L}\mathbf{x}\|^2,$$

where $\lambda > 0$ is the regularization parameter and $\mathbf{L} \in \mathbb{R}^{d \times n}$ is a matrix such that $\ker(\mathbf{A}) \cap \ker(\mathbf{L}) = \{\mathbf{0}\}$. \mathbf{L} is usually chosen as the identity or the forward-difference operator. We can prove the following proposition regarding Tikhonov stability.

PROPOSITION 2.17. *Let $\epsilon > 0$ and $\mathbf{L} \in \mathbb{R}^{d \times n}$. Then $\exists \lambda > 0$ such that*

$$L^\epsilon(\Psi^{\lambda, \mathbf{L}}, \mathcal{Y}) < 1.$$

Proof. For any $\lambda > 0$ and any $\mathbf{y}^\delta \in \mathcal{Y}^\epsilon$, it can be shown, by considering the normal equations of (2.6), that

$$\Psi^{\lambda, \mathbf{L}}(\mathbf{y}^\delta) = \left(\mathbf{A}^* \mathbf{A} + \lambda \mathbf{L}^* \mathbf{L}\right)^{-1} \mathbf{A}^* \mathbf{y}^\delta = \frac{1}{\lambda} \left(\frac{1}{\lambda} \mathbf{A}^* \mathbf{A} + \mathbf{L}^* \mathbf{L}\right)^{-1} \mathbf{A}^* \mathbf{y}^\delta.$$

Consequently, for any $\mathbf{y}^\delta \in \mathcal{Y}^\epsilon$, it holds that $\Psi^{\lambda, \mathbf{L}}(\mathbf{y}^\delta) \rightarrow 0$ for $\lambda \rightarrow \infty$. Then

$$L^\epsilon(\Psi^{\lambda, \mathbf{L}}, \mathcal{Y}) = \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi^{\lambda, \mathbf{L}}(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \Psi^{\lambda, \mathbf{L}}(\mathbf{A}\mathbf{x}^{gt})\|}{\|\mathbf{e}\|} \rightarrow 0 \quad \text{for } \lambda \rightarrow \infty,$$

which implies that, for all $\alpha > 0$, there exists $\bar{\lambda} > 0$ such that for any $\lambda > \bar{\lambda}$, $L^\epsilon(\Psi^{\lambda, \mathbf{L}}, \mathcal{Y}) < \alpha$. Choosing $\alpha = 1$ we obtain the required result. \square

Corollary 2.15 and Proposition 2.17 demonstrate that it is always possible to build a stable Tikhonov reconstructor. Such a property will play a crucial role in subsection 4.3 where we will explain our proposed ReNN approach.

3. Stabilizers in the solution of linear inverse problems. In this section, we delve into additional properties pertinent to stable reconstructors by introducing the novel concept of *stabilizer* which will be exploited in subsection 4.4 to define our StNN and StReNN approaches.

3.1. Stabilizers and properties.

DEFINITION 3.1. *A continuous function $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^t$ is an ϵ -stabilizer of a reconstructor $\Psi: \mathbb{R}^m \rightarrow \mathbb{R}^n$ if*

1. $\forall \mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\| \leq \epsilon$, $\exists C_\phi^\epsilon \in [0, 1)$ and $\exists \mathbf{e}' \in \mathbb{R}^t$ with $\|\mathbf{e}'\| = C_\phi^\epsilon \|\mathbf{e}\|$ such that

$$\phi(\mathbf{A}\mathbf{x} + \mathbf{e}) = \phi(\mathbf{A}\mathbf{x}) + \mathbf{e}';$$

2. $\exists \gamma: \mathbb{R}^t \rightarrow \mathbb{R}^n$ such that $\Psi = \gamma \circ \phi$.

The smallest constant C_ϕ^ϵ for which the definition holds is defined as the stability constant of the stabilizer ϕ .

We also define the set

$$\mathcal{S}_\eta^\epsilon = \{\Psi \in \mathcal{R}_\eta; \exists \gamma: \mathbb{R}^t \rightarrow \mathbb{R}^n, \exists \phi \text{ } \epsilon\text{-stabilizer, s.t. } \Psi = \gamma \circ \phi\}.$$

Whenever $t = m$ and $\gamma: \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a reconstructor, the reconstructor Ψ is said to be ϵ -stabilized with respect to γ .

Note that, in the definition of ϵ -stabilizer, we only require a stability condition for ϕ in the first item.

Interestingly, given a reconstructor $\Psi = \gamma \circ \phi$, we can estimate its ϵ -stability constant C_Ψ^ϵ by means of the constant C_ϕ^ϵ and the local Lipschitz constant of γ , as proved in the following proposition.

PROPOSITION 3.2. *Let $\Psi: \mathbb{R}^m \rightarrow \mathbb{R}^n$, $\Psi = \gamma \circ \phi$, with ϕ being an ϵ -stabilizer. If C_ϕ^ϵ is the constant mentioned in Definition 3.1, $L^\epsilon(\gamma, \mathcal{T})$ is the local Lipschitz constant of γ with $\mathcal{T} = \phi(\mathcal{Y})$, it holds that*

$$C_\Psi^\epsilon \leq L^\epsilon(\gamma, \mathcal{T}) C_\phi^\epsilon.$$

Proof. Let $\mathbf{x}^{gt} \in \mathcal{X}$ and $\|\mathbf{e}\| \leq \epsilon$. Then

$$\|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| = \|\gamma(\phi(\mathbf{Ax}^{gt} + \mathbf{e})) - \mathbf{x}^{gt}\|.$$

Since ϕ is a stabilizer, $\phi(\mathbf{Ax}^{gt} + \mathbf{e}) = \phi(\mathbf{Ax}^{gt}) + \mathbf{e}'$ with $\|\mathbf{e}'\| \leq C_\phi^\epsilon \|\mathbf{e}\|$. Thus,

$$\begin{aligned} \|(\gamma(\phi(\mathbf{Ax}^{gt} + \mathbf{e})) - \mathbf{x}^{gt})\| &= \|\gamma(\phi(\mathbf{Ax}^{gt}) + \mathbf{e}') - \mathbf{x}^{gt}\| \\ &\leq \|\gamma(\phi(\mathbf{Ax}^{gt}) + \mathbf{e}') - \gamma(\phi(\mathbf{Ax}^{gt}))\| + \|\gamma(\phi(\mathbf{Ax}^{gt})) - \mathbf{x}^{gt}\| \\ &\leq \eta + L^\epsilon(\gamma, \mathcal{T})\|\mathbf{e}'\| = \eta + L^\epsilon(\gamma, \mathcal{T})C_\phi^\epsilon\|\mathbf{e}\|, \end{aligned}$$

which implies that $L^\epsilon(\gamma, \mathcal{T})C_\phi^\epsilon$ is a majorant of the set

$$\left\{ \frac{\|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|}; \mathbf{x}^{gt} \in \mathcal{X}, \|\mathbf{e}\| \leq \epsilon \right\}.$$

Since C_Ψ^ϵ is defined as the supremum of the same set, by the minimality of the supremum we have $C_\Psi^\epsilon \leq L^\epsilon(\gamma, \mathcal{T})C_\phi^\epsilon$. \square

Proposition 3.2 implies the following important result.

THEOREM 3.3. *For any $\epsilon > 0$, $\eta_1, \eta_2 > 0$, let $\Psi_1 = \gamma_1 \circ \phi_1 \in \mathcal{S}_{\eta_1}^\epsilon$, and $\Psi_2 \in \mathcal{R}_{\eta_2}$. If*

$$(3.1) \quad C_{\phi_1}^\epsilon \in \left[0, \frac{C_{\Psi_2}^\epsilon}{L^\epsilon(\gamma_1, \mathcal{T})} \right],$$

then

$$C_{\Psi_1}^\epsilon \leq C_{\Psi_2}^\epsilon.$$

Proof. Since (3.1) holds by hypothesis and $C_{\Psi_1}^\epsilon \leq L^\epsilon(\gamma_1, \mathcal{T})C_{\phi_1}^\epsilon$ for $\Psi_1 \in \mathcal{S}_{\eta_1}^\epsilon$ by Proposition 3.2, we get

$$C_{\Psi_1}^\epsilon \leq L^\epsilon(\gamma_1, \mathcal{T})C_{\phi_1}^\epsilon \leq L^\epsilon(\gamma_1, \mathcal{T}) \frac{C_{\Psi_2}^\epsilon}{L^\epsilon(\gamma_1, \mathcal{T})} = C_{\Psi_2}^\epsilon,$$

which concludes the proof. \square

The theorem yields interesting consequences for the special case where Ψ_1 and Ψ_2 share the same accuracy. For instance, when $\Psi_1 = \gamma_1 \circ \phi_1 \in \mathcal{S}_\eta^\epsilon$ and $\Psi_2 \in \mathcal{R}_\eta$, if (3.1) holds, the theorem suggests that Ψ_1 is preferable to Ψ_2 , as Ψ_1 is more stable than Ψ_2 . In addition, we can state the following result, whose proof is trivial.

COROLLARY 3.4. *Let $\Psi_1 = \gamma_1 \circ \phi_1 \in \mathcal{S}_\eta^\epsilon$ and $\Psi_2 = \gamma_2 \circ \phi_2 \in \mathcal{S}_\eta^\epsilon$. If (3.1) holds, then $C_{\phi_1}^\epsilon \leq C_{\phi_2}^\epsilon$.*

In the next proposition, we show a result linking the accuracy of a reconstructor $\Psi \in \mathcal{S}_\eta^\epsilon$ to a characterization of its ϵ -stabilizer ϕ .

PROPOSITION 3.5. *Let $\Psi = \gamma \circ \phi \in \mathcal{S}_\eta^\epsilon$. Let*

$$(3.2) \quad \sigma(\phi) := \sup\{\|\mathbf{x}_1 - \mathbf{x}_2\|; \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \phi(\mathbf{Ax}_1) = \phi(\mathbf{Ax}_2)\}.$$

Then

$$\eta^{-1} \leq \frac{2}{\sigma(\phi)}.$$

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ such that $\phi(\mathbf{Ax}_1) = \phi(\mathbf{Ax}_2)$. Then

$$\begin{aligned} \|\mathbf{x}_1 - \mathbf{x}_2\| &\leq \|\phi(\mathbf{Ax}_1) - \mathbf{x}_1\| + \|\phi(\mathbf{Ax}_1) - \mathbf{x}_2\| \\ &= \|\phi(\mathbf{Ax}_1) - \mathbf{x}_1\| + \|\phi(\mathbf{Ax}_2) - \mathbf{x}_2\| \leq 2\eta, \end{aligned}$$

which implies that

$$\eta \geq \frac{\|\mathbf{x}_1 - \mathbf{x}_2\|}{2}.$$

Since the estimation above holds for any $\mathbf{x}_1, \mathbf{x}_2$ with $\phi(\mathbf{Ax}_1) = \phi(\mathbf{Ax}_2)$, it holds for $\sigma(\phi)$, thus concluding the proof. \square

As a consequence of Proposition 3.5, if ϕ is the constant operator (having $C_\phi^\epsilon = 0$ as observed in Example 2.7), it results in $\sigma(\phi) = \infty$, which implies that for any γ , the accuracy of $\Psi = \gamma \circ \phi$ will be zero whenever \mathcal{X} is unbounded.

The question now is, how can I construct an ϵ -stabilizer? In the following proposition, we show that a sequence of functions $\{\phi_k\}_{k \in \mathbb{N}}$ approximating $\Psi \in \mathcal{R}_\eta$, i.e.,

$$\lim_{k \rightarrow \infty} \sup_{\mathbf{y}^\delta \in \mathcal{Y}^\epsilon} \|\phi_k(\mathbf{y}^\delta) - \Psi(\mathbf{y}^\delta)\| = 0,$$

can be exploited to construct a good stabilizer.

PROPOSITION 3.6. *Given a reconstructor $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ with local Lipschitz constant $L^\epsilon(\Psi, \mathcal{Y}) < 1$ and a sequence of functions $\{\phi_k\}_{k \in \mathbb{N}}$ approximating Ψ , there exists $K \in \mathbb{N}$ such that for any $k \geq K$, $C_{\phi_k}^\epsilon < 1$.*

Proof. Consider $\mathbf{x}^{gt} \in X$ and $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\| \leq \epsilon$. To prove the result, we need to show that

$$\phi_k(\mathbf{Ax}^{gt} + \mathbf{e}) = \phi_k(\mathbf{Ax}^{gt}) + \mathbf{e}' \text{ for } k \geq K,$$

with $\|\mathbf{e}'\| = C_\phi^\epsilon \|\mathbf{e}\|$ and $C_\phi^\epsilon \in [0, 1)$.

Let $\mathbf{e}' := \phi_k(\mathbf{Ax}^{gt} + \mathbf{e}) - \phi_k(\mathbf{Ax}^{gt})$. Then

$$\begin{aligned} \|\mathbf{e}'\| &= \|\phi_k(\mathbf{Ax}^{gt} + \mathbf{e}) - \phi_k(\mathbf{Ax}^{gt})\| \\ &\leq L^\epsilon(\phi_k, \mathcal{Y}) \|\mathbf{Ax}^{gt} + \mathbf{e} - \mathbf{Ax}^{gt}\| = L^\epsilon(\phi_k, \mathcal{Y}) \|\mathbf{e}\|, \end{aligned}$$

which implies that $C_{\phi_k}^\epsilon \leq L^\epsilon(\phi_k, \mathcal{Y})$. Since $\{\phi_k\}_{k \in \mathbb{N}}$ is a sequence of approximators of Ψ , for any $k \in \mathbb{N}$ there is a constant c_k such that $\|\phi_k(\mathbf{y}^\delta) - \Psi(\mathbf{y}^\delta)\| \leq c_k$ and $c_k \rightarrow 0$ as $k \rightarrow \infty$. Consequently, it holds that

$$\begin{aligned} L^\epsilon(\phi_k, \mathcal{Y}) &= \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\phi_k(\mathbf{Ax}^{gt} + \mathbf{e}) - \phi_k(\mathbf{Ax}^{gt})\|}{\|\mathbf{e}\|} \\ &\leq \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\phi_k(\mathbf{Ax}^{gt} + \mathbf{e}) - \Psi(\mathbf{Ax}^{gt} + \mathbf{e})\| + \|\phi_k(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\| + \|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \Psi(\mathbf{Ax}^{gt})\|}{\|\mathbf{e}\|} \\ &\leq \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \Psi(\mathbf{Ax}^{gt})\| + 2c_k}{\|\mathbf{e}\|}, \end{aligned}$$

which implies that $L^\epsilon(\phi_k, \mathcal{Y}) \rightarrow L^\epsilon(\Psi, \mathcal{Y})$ as $k \rightarrow \infty$. Since $L^\epsilon(\Psi, \mathcal{Y}) < 1$, $\exists K \in \mathbb{N}$ such that for any $k \geq K$, $L^\epsilon(\phi_k, \mathcal{Y}) < 1$. For those values of k , $C_{\phi_k}^\epsilon \leq L^\epsilon(\phi_k, \mathcal{Y}) < 1$. \square

3.2. Tikhonov stabilizers. In particular, we can apply the results of Proposition 3.6 by considering Ψ as the Tikhonov reconstructor $\Psi^{\lambda, L}$ introduced in Example 2.16. A simple way to generate the sequence $\{\phi_k\}_{k \in \mathbb{N}}$ is to consider a convergent iterative algorithm for the solution of (2.6) such as

$$\begin{cases} \mathbf{x}^0 \in \mathbb{R}^n, \\ \mathbf{x}^{k+1} = \mathcal{T}_k(\mathbf{x}^k, \mathbf{y}^\delta), \end{cases}$$

where $\mathcal{T}_k(\mathbf{x}^k, \mathbf{y}^\delta)$ models the application of the k th iterate of the algorithm, starting from \mathbf{x}^k and with datum \mathbf{y}^δ . To set an example, the conjugate gradient for least squares (CGLS) algorithm is an iterative method solving the normal equations associated with (2.6). Now, for any $k \in \mathbb{N}$ we can define the *Tikhonov stabilizers* ϕ_k to be the composition of the first k iterations of the algorithm, i.e.,

$$(3.3) \quad \phi_k(\cdot) = \bigcirc_{i=1}^k \mathcal{T}_i(\cdot, \mathbf{y}^\delta).$$

For the convergence property of the algorithm, $\{\phi_k\}_{k \in \mathbb{N}}$ is a sequence of functions approximating $\Psi^{\lambda, L}$ and satisfying $C_{\phi_k}^\epsilon < 1$ for suitable $k \geq K$. Such a property will be fundamental for the stabilization technique we propose in subsection 4.4.

4. Neural networks for the solution of linear inverse problems. In this section, the theoretical results previously outlined are applied to scenarios where reconstructors are operationalized through neural networks. Concurrently, we delineate our methodologies aimed at advancing current state-of-the-art approaches. Figure 4.1 offers a detailed schematic that encapsulates all the approaches considered within this study. The ‘‘Tik’’ label refers to the Tikhonov reconstructor $\Psi^{\lambda, L}$, defined in Example 2.16.

4.1. Parameter-dependent families of reconstructors. We now consider a family of reconstructors $\{\Psi_\Theta\}_{\Theta \in \mathbb{R}^s}$, depending on a vector of parameters Θ , approximating a reconstructor Ψ to solve problem (1.1). We prove in the following theorem that the stability of Ψ_Θ is strongly related to the stability of Ψ .

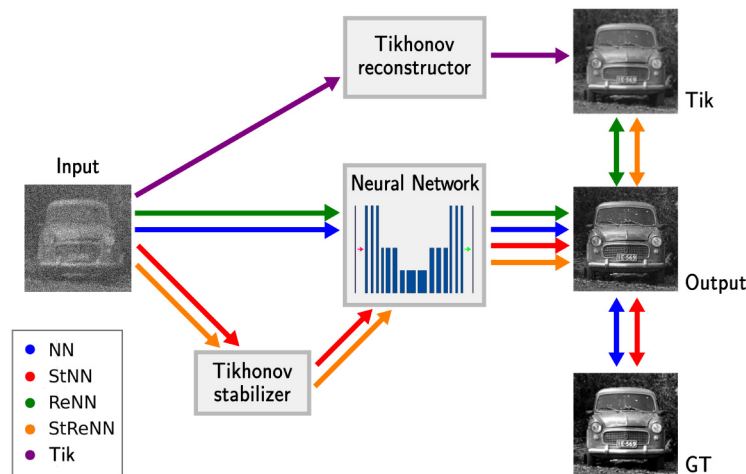


FIG. 4.1. A schematic representation of the proposed methods.

THEOREM 4.1 (approximation theorem for reconstructors). *Let Ψ be an η^{-1} -accurate reconstructor, and let $\{\Psi_\Theta\}_{\Theta \in \mathbb{R}^s}$ be a set of reconstructors with accuracy η_Θ^{-1} for any Θ . We define, for any $\Theta \in \mathbb{R}^s$,*

$$\Delta(\Theta) := \sup_{\mathbf{x}^{gt} \in \mathcal{X}} \|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\|$$

and

$$\Delta_\epsilon(\Theta) := \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| < \epsilon}} \|\Psi_\Theta(\mathbf{Ax}^{gt} + \mathbf{e}) - \Psi(\mathbf{Ax}^{gt} + \mathbf{e})\|.$$

If $\Delta(\Theta) \rightarrow 0$ when $\Theta \rightarrow \Theta^*$, then

$$(4.1) \quad \lim_{\Delta(\Theta) \rightarrow 0} \eta_\Theta = \eta.$$

Moreover, if $\Delta_\epsilon(\Theta) \rightarrow 0$ when $\Theta \rightarrow \Theta_\epsilon^*$, then

$$(4.2) \quad \lim_{\Delta_\epsilon(\Theta) \rightarrow 0} C_{\Psi_\Theta}^\epsilon = C_\Psi^\epsilon.$$

Proof. Consider $\mathbf{x}^{gt} \in \mathcal{X}$. Since

$$\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \leq \|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| + \|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\|$$

and

$$\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \geq \|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| - \|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\|,$$

it holds that

$$|\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| - \|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\|| \leq \|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\| \leq \Delta(\Theta),$$

which implies that $\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \rightarrow \|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\|$ as $\Delta(\Theta) \rightarrow 0$ and consequently $\eta_\Theta \rightarrow \eta$ as $\Delta(\Theta) \rightarrow 0$.

Now, consider $\epsilon > 0$ and $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\| \leq \epsilon$. A similar computation shows that

$$|\|\Psi_\Theta(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\|| \leq \Delta_\epsilon(\Theta),$$

which implies that $\|\Psi_\Theta(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| \rightarrow \|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\|$ for $\Delta_\epsilon(\Theta) \rightarrow 0$. Consequently, for $\Delta_\epsilon(\Theta) \rightarrow 0$,

$$C_{\Psi_\Theta}^\epsilon = \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi_\Theta(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta_\Theta}{\|\mathbf{e}\|} \rightarrow \sup_{\substack{\mathbf{x}^{gt} \in \mathcal{X} \\ \|\mathbf{e}\| \leq \epsilon}} \frac{\|\Psi(\mathbf{Ax}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \eta}{\|\mathbf{e}\|} = C_\Psi^\epsilon,$$

which concludes the proof. \square

COROLLARY 4.2. *For any $\Theta \in \mathbb{R}^s$, it holds that*

$$\eta_\Theta \leq \eta + \Delta(\Theta).$$

Proof. Consider $\mathbf{x}^{gt} \in \mathcal{X}$. Then

$$\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \leq \|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\| + \|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\|.$$

Since $\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \Psi(\mathbf{Ax}^{gt})\| \leq \Delta(\Theta)$ by hypothesis and $\|\Psi(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \leq \eta$ since Ψ is η^{-1} -accurate, then

$$\|\Psi_\Theta(\mathbf{Ax}^{gt}) - \mathbf{x}^{gt}\| \leq \eta + \Delta(\Theta),$$

which shows that $\eta_\Theta \leq \eta + \Delta(\Theta)$. \square

Note that $\Delta(\Theta)$ and $\Delta_\epsilon(\Theta)$ are, in general, not independent, as proved in the following proposition.

PROPOSITION 4.3. For any $\epsilon > 0$, let $\Delta(\Theta)$ and $\Delta_\epsilon(\Theta)$ be the quantities defined in Theorem 4.1. Then

$$\Delta(\Theta) \leq \Delta_\epsilon(\Theta).$$

Proof. Observe that, by definition of \mathcal{Y} and \mathcal{Y}^ϵ , $\Delta(\Theta)$ and $\Delta_\epsilon(\Theta)$ can be rewritten as

$$\begin{aligned} \Delta(\Theta) &= \sup_{\mathbf{y} \in \mathcal{Y}} \|\Psi_\Theta(\mathbf{y}) - \Psi(\mathbf{y})\|, \\ \Delta_\epsilon(\Theta) &= \sup_{\mathbf{y} \in \mathcal{Y}^\epsilon} \|\Psi_\Theta(\mathbf{y}) - \Psi(\mathbf{y})\|, \end{aligned}$$

where $\mathcal{Y}^\epsilon = \{\mathbf{y} + \mathbf{e}; \mathbf{y} \in \mathcal{Y}, \|\mathbf{e}\| \leq \epsilon\} \supseteq \mathcal{Y}$. The result follows from the property that the supremum of a set must be larger than the supremum of its subsets. \square

An insight on the stability properties of Ψ_Θ can be obtained by the following proposition.

PROPOSITION 4.4. Let Ψ_Θ be a reconstructor parameterized by $\Theta \in \mathbb{R}^s$, approximating a reconstructor Ψ with error $\Delta(\Theta) > 0$. Let η_Θ^{-1} and η^{-1} be the accuracy of Ψ_Θ and Ψ , respectively. If

$$(4.3) \quad \Delta(\Theta) \leq \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X}) - \eta$$

for a fixed $\epsilon > 0$, where $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$ is the constant defined in Corollary 2.12, then $C_{\Psi_\Theta}^\epsilon \geq 1$.

Proof. Let $\epsilon > 0$ be fixed. By Corollary 4.2, the accuracy of Ψ_Θ can be estimated as $\eta_\Theta \leq \eta + \Delta(\Theta)$. Consequently, by Corollary 2.12, if $\eta + \Delta(\Theta) \leq \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$, then $\eta_\Theta \leq \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$, which implies that $C_{\Psi_\Theta}^\epsilon \geq 1$. \square

We note that if Ψ is ϵ -stable, then $\eta > \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$, implying that the right-hand side of inequality (4.3) is negative. Given that $\Delta(\Theta)$ is inherently positive, this indicates that the conditions of Proposition 4.4 are never met when Ψ is ϵ -stable. Conversely, if $\eta \ll \bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$, implying that Ψ is far from being ϵ -stable, then its approximation Ψ_Θ will be ϵ -unstable even if $\Delta(\Theta)$ is large. Therefore, Proposition 4.4 aligns with the intuition that the approximation Ψ_Θ behaves similarly to Ψ whenever $\Delta(\Theta)$ is sufficiently small. In the following paragraphs, we will analyze particular families of reconstructors $\{\Psi_\Theta\}_{\Theta \in \mathbb{R}^s}$ based on deep learning.

4.2. Neural networks as reconstructors: The NN approach. We begin by considering the set of neural networks defined by a fixed architecture as the family $\{\Psi_\Theta\}_{\Theta \in \mathbb{R}^s}$.

DEFINITION 4.5. Given a neural network architecture $\mathcal{A} = (\nu, S)$, where $\nu = (\nu_0, \nu_1, \dots, \nu_L) \in \mathbb{N}^{L+1}$, $\nu_0 = m, \nu_L = n$, defines the width of each layer and $S = (S_{1,1}, \dots, S_{L,L}), S_{j,k} \in \mathbb{R}^{\nu_j \times \nu_k}$ is the set of matrices representing the skip connections, we define the parametric family of neural network reconstructors with architecture \mathcal{A} , parameterized by $\Theta \in \mathbb{R}^s$, as

$$\mathcal{F}_\Theta^{\mathcal{A}} = \{\Psi_\Theta : \mathbb{R}^m \rightarrow \mathbb{R}^n; \Theta \in \mathbb{R}^s\},$$

where $\Psi_\Theta(\mathbf{y}^\delta) = \mathbf{z}^L$ is given by

$$(4.4) \quad \begin{cases} \mathbf{z}^0 = \mathbf{y} \\ \mathbf{z}^{l+1} = \rho(W^l \mathbf{z}^l + b^l + \sum_{k=1}^l S_{l,k} \mathbf{z}^k) \quad \forall l = 0, \dots, L-1 \end{cases}$$

and $W^l \in \mathbb{R}^{\nu_{l+1} \times \nu_l}$ is the weights matrix, $b^l \in \mathbb{R}^{\nu_{l+1}}$ is the bias vector.

Given $\mathcal{D} \subseteq \mathcal{X}$, consider the dataset $\mathbb{D} = \{(\mathbf{y}_i^\delta, \mathbf{x}_i^{gt}); \mathbf{x}_i^{gt} \in \mathcal{D}\}_{i=1}^{N_{\mathbb{D}}}$ of images according to (1.2). Training a neural network to solve the inverse problem (1.2) results in finding the parameters Θ^* such that the associated reconstructor $\Psi_{\Theta^*} \in \mathcal{F}_{\Theta}^A$ satisfies:

$$(4.5) \quad \Psi_{\Theta^*} \in \arg \min_{\Psi_{\Theta} \in \mathcal{F}_{\Theta}^A} \frac{1}{N_{\mathbb{D}}} \sum_{i=1}^{N_{\mathbb{D}}} \ell(\Psi_{\Theta}(\mathbf{y}_i^\delta), \mathbf{x}_i^{gt}),$$

where $\delta \geq 0$ and $\ell: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ is the loss function.

In this work, we consider as reconstructors Ψ_{Θ} the neural networks trained with the mean squared error (MSE) loss. When we train the networks on noise-free data ($\delta = 0$) (4.5) corresponds to

$$(4.6) \quad \min_{\Psi_{\Theta} \in \mathcal{F}_{\Theta}^A} \sum_{i=1}^{N_{\mathbb{D}}} \|\Psi_{\Theta}(\mathbf{y}_i) - \mathbf{x}_i^{gt}\|_2^2 = \min_{\Psi_{\Theta} \in \mathcal{F}_{\Theta}^A} \sum_{i=1}^{N_{\mathbb{D}}} \|\Psi_{\Theta}(\mathbf{A}\mathbf{x}_i^{gt}) - \Psi^\dagger(\mathbf{A}\mathbf{x}_i^{gt})\|_2^2,$$

which results in the minimization of $\Delta(\Theta)$ as introduced in Theorem 4.1 with $\Psi = \Psi^\dagger$. We will name this family NN in the following.

We observe that when \mathbf{A} is ill-conditioned, the value $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$ defined in (2.5) is large. This becomes particularly apparent when $\mathcal{X} = \mathbb{R}^n$, as under these circumstances $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X})$ is bounded below by a quantity depending on $C(\mathbf{A}) = \frac{1-\sigma_n}{\sigma_n}$. Additionally, the value of $\Delta(\Theta^*)$ derived from NN training likely meets the established inequality in Proposition 4.4, which leads to instability. This confirms that effective neural network training can produce a very accurate but unstable reconstructor Ψ_{Θ} .

A widely adopted strategy to bolster robustness in neural networks is known as noise injection. This technique involves adding noise to the input of the network during its training phase. In this context, the set of reconstructors Ψ_{Θ} , referred to as iNN, is defined by a neural network trained through the following equation:

$$(4.7) \quad \min_{\Psi_{\Theta} \in \mathcal{F}_{\Theta}^A} \sum_{i=1}^{N_{\mathbb{D}}} \|\Psi_{\Theta}(\mathbf{y}_i^\delta) - \mathbf{x}_i^{gt}\|_2^2,$$

where $\delta > 0$. Research detailed in [5] has demonstrated that this approach effectively introduces a Tikhonov regularization term into the loss function. Although this technique, as described in [3], enhances the stability of the resultant network, the impact of noise injection on the accuracy of the model remains somewhat ambiguous. Furthermore, the optimal amount of noise to be added to each input to optimize the balance between stability and accuracy is still a subject of investigation.

4.3. Regularized NN-based reconstructors: The ReNN approach. To develop a reconstructor with improved stability compared to standard neural networks, we harness the properties of Tikhonov regularization. It is important to note that a Tikhonov regularized reconstructor $\Psi^{\lambda, \mathcal{L}}$ achieves stability for an appropriately chosen regularization parameter, as delineated in Proposition 2.17. Hence, the proposed reconstructor will be referred to as the regularized neural network (ReNN), denoted by $\Psi_{\Theta}^{\lambda, \mathcal{L}}$ and it is defined by training a neural network with a new loss ℓ as

$$(4.8) \quad \Psi_{\Theta}^{\lambda, \mathcal{L}} \in \arg \min_{\Psi_{\Theta} \in \mathcal{F}_{\Theta}^A} \sum_{i=1}^{N_{\mathbb{D}}} \|\Psi_{\Theta}(\mathbf{y}_i^\delta) - \Psi^{\lambda, \mathcal{L}}(\mathbf{y}_i^\delta)\|_2^2,$$

with $\delta > 0$. We underline that ReNN does not require any ground-truth solution \mathbf{x}^{gt} since the target is computed from the corrupted datum \mathbf{y}^δ via the Tikhonov-regularized reconstructor. Furthermore, in the training of ReNN, noise is present not

solely to the input of the neural network model, as is the case with iNN, but also to the input of the Tikhonov-regularized reconstructor, which is responsible for generating the target. In the following, we consider for simplicity the case $\mathcal{X} = \mathbb{R}^n$, but similar results hold for a general $\mathcal{X} \subset \mathbb{R}^n$.

Why does ReNN exhibit greater stability than NN? Starting from inequality (4.3) it is easy to notice that (4.8) corresponds to the minimization of $\Delta_\epsilon(\Theta)$ in Theorem 4.1. Moreover, by Proposition 4.3, if $\Delta_\epsilon(\Theta)$ is small, as it is common when Ψ_Θ is a neural network, then $\Delta(\Theta) \in [0, \Delta_\epsilon(\Theta)]$ is also small. Regarding the right-hand side $\bar{\eta}(\mathbf{A}, \epsilon, \mathcal{X}) - \eta$ of (4.3), it is noted that in this instance η depends on λ , and $\eta(\lambda) \rightarrow \infty$ for $\lambda \rightarrow \infty$. Consequently, for sufficiently large values of λ , it is probable that ReNN does not fulfill inequality (4.3), which would lead to instability. Moreover, minimizing $\Delta_\epsilon(\Theta)$ is crucial for enforcing the method's stability, as proven by Theorem 4.1, where we have shown that in our hypothesis the stability constant $C_{\Psi_\Theta}^\epsilon < 1$ for sufficiently small $\Delta_\epsilon(\Theta)$. Hence, effective training of ReNN should produce an accurate and stable reconstructor.

4.4. Stabilization on NN and ReNN: St-approaches. In the remainder of this section, we discuss an application of the stabilizers, introduced in section 3, to improve the stability of neural network-based reconstructors. We propose new reconstructors $\Psi \in \mathcal{S}_\eta^\epsilon$, $\Psi = \gamma \circ \phi$, where γ is a neural network-based reconstructor. In particular, we consider ϕ as the Tikhonov ϵ -stabilizer ϕ_k defined in subsection 3.2 and obtained by k iterations of the CGLS algorithm, with a suitable k . When γ is chosen as NN, iNN, and ReNN we obtain the ϵ -stabilized reconstructors StNN, StiNN, and StReNN, respectively.

Note that, in this case, we can apply Theorem 3.3 with $\Psi_1 = \gamma \circ \phi_k$ and $\Psi_2 = \gamma$, and whenever we choose ϕ_k such that

$$(4.9) \quad C_{\phi_k}^\epsilon \leq \frac{C_\gamma^\epsilon}{L^\epsilon(\gamma, \mathcal{Y})},$$

the ϵ -stabilized reconstructor Ψ_1 gets more stable than its unstabilized version Ψ_2 . We remark that it is always possible to find a Tikhonov stabilizer ϕ_k fitting (4.9), by suitably tuning λ and k . Clearly, this comes at the expense of accuracy as discussed in Proposition 3.5, but we will show that the accuracy does not suffer excessively, as evidenced by empirical results in section 6.

5. Experimental setup. To assess the theoretical issues proposed, we conducted a series of experiments. It is important to highlight that all tests were carried out utilizing the same end-to-end U-net architecture. For details on the architecture and its training, you can refer to [20, 10]. In the following experiments, the stabilizer applied to all the considered reconstructors is obtained with $k = 3$ iterations of the CGLS algorithm on (2.6). The codes can be found in our GitHub repository at <https://github.com/loibo/ToBeOrNotToBeStable>.

As a test case, we consider image deblurring [16], a common inverse problem in imaging. In this case, \mathbf{A} is a block circulant matrix with circulant blocks obtained from a convolutional kernel with periodic boundary conditions [16]. In our experiments, we use the 11×11 Gaussian blur filter \mathcal{K} :

$$(5.1) \quad \mathcal{K}_{i,j} = e^{-\frac{1}{2} \frac{i^2+j^2}{\sigma_G^2}}, \quad i, j \in \{-5, \dots, 5\}$$

with variance $\sigma_G^2 = 1.3$.

5.1. Dataset. Our results have been tested on the famous GoPro image dataset (<https://seungjunnah.github.io/Datasets/gopro>), introduced in [24], which is constituted by high-resolution RGB images. All the images have been cropped into patches of size 256×256 (without overlapping), converted into grayscale, normalized in $[0, 1]$, and labeled as $\mathbf{x}_i^{gt}, i = 1, \dots, N_{\mathbb{D}}$, with $N_{\mathbb{D}} = 3614$. We generated the blurred and noisy data $\mathbf{y}_i^\delta = \mathbf{A}\mathbf{x}_i^{gt} + \mathbf{e}$, where $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \delta^2 \mathbf{I})$. We need the following datasets to train the three considered neural networks-based reconstructors.

- For the NN training (see (4.6)) we consider the set $\mathbb{D} = \{(\mathbf{y}_i, \mathbf{x}_i^{gt})\}_{i=1}^{N_{\mathbb{D}}}$ containing the blurred noiseless datum (i.e., $\delta = 0$) and the exact \mathbf{x}_i^{gt} target picture.
- For the iNN training (see (4.7)) we consider the set $\mathbb{D}_\delta = \{(\mathbf{y}_i^\delta, \mathbf{x}_i^{gt})\}_{i=1}^{N_{\mathbb{D}}}$ containing the blurred and noisy datum \mathbf{y}_i^δ and the exact \mathbf{x}_i^{gt} target picture.
- For the ReNN training (see (4.8)) we consider the set $\mathbb{D}_\delta^{\lambda, L} = \{(\mathbf{y}_i^\delta, \Psi^{\lambda, L}(\mathbf{y}_i^\delta))\}_{i=1}^{N_{\mathbb{D}}}$ containing the blurred and noisy datum \mathbf{y}_i^δ and the target image computed by the Tikhonov reconstructor (using $\mathbf{L} = \mathbf{I}$ in (2.6)). In particular, we choose λ heuristically and computed $\Psi^{\lambda, L}(\mathbf{y}_i^\delta)$ by means of the CGLS algorithm [14] to solve the normal equations of (2.6).

We finally split the $N_{\mathbb{D}}$ data samples into train and test subsets, with $N_{train} = 2503$ and $N_{test} = 1111$.

5.2. Results evaluation. In order to estimate in our experiments the accuracy and the stability constants of a given reconstructor Ψ we compute the *empirical* accuracy $\hat{\eta}^{-1}$ and the *empirical* stability constant \hat{C}_Ψ^ϵ , over the test set \mathcal{TS} . They are, respectively, defined as

$$(5.2) \quad \hat{\eta} = \sup_{\mathbf{x}^{gt} \in \mathcal{TS}} \|\Psi(\mathbf{A}\mathbf{x}^{gt}) - \mathbf{x}^{gt}\|$$

and

$$(5.3) \quad \hat{C}_\Psi^\epsilon = \sup_{\mathbf{x}^{gt} \in \mathcal{TS}} \frac{\|\Psi(\mathbf{A}\mathbf{x}^{gt} + \mathbf{e}) - \mathbf{x}^{gt}\| - \hat{\eta}}{\|\mathbf{e}\|},$$

where $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \delta^2 \mathbf{I})$ differs for each datum $\mathbf{x}^{gt} \in \mathcal{TS}$. Finally, we compute the *empirical reconstruction error* on the test set as

$$\mathcal{E}(\Psi, \delta) = \sup_{\mathbf{x}^{gt} \in \mathcal{TS}} \|\Psi(\mathbf{y}^\delta) - \mathbf{x}^{gt}\|.$$

To evaluate a single image reconstruction, we also calculate the widely used Structural Similarity Index (SSIM) [29], taking values in $[0, 1]$.

To augment the stochastic nature of our experiments, we replicate the tests $T = 20$ times, each with different realization of noise. In the following, we report the maximum value of the computed parameters $\hat{\eta}$ and \hat{C}_Ψ^ϵ over the T experiments.

6. Numerical results. In this section, we present the outcomes achieved in terms of empirical accuracy, stability, and reconstruction error for the solvers proposed in this study. The objective of this section is twofold: first, to validate the key theoretical findings established in the previous part of the paper, with a particular emphasis on the deep learning-based reconstructors introduced in section 4; and second, to examine the impact of the stabilizer in scenarios where the noise levels exceed those the parameters were initially selected for.

6.1. Results with NN-based reconstructors. We begin by considering the NN and iNN approaches, and their stabilized counterparts, StNN and StiNN, assuming the availability of ground-truth images $\mathbf{x}_i^{gt}, i = 1, \dots, N_{\mathbb{D}}$.

The first experiment concerns NN and StNN. The first row of Figure 6.1 shows the reconstructions obtained with both the methods on one image \mathbf{y}_i from the test set (without noise added). To assess the stability of our frameworks concerning unseen noise on the data, we also tested the NN reconstructor on noisy images $\mathbf{y}_i^{\delta} = \mathbf{y}_i + \mathbf{e}_i$ with $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \delta^2 \mathbf{I})$ and $\delta = 0.01$. The second row of Figure 6.1 displays the reconstructions obtained on the same test image.

From the images presented in Figure 6.1 and their SSIM values, it is observable that the NN reconstructor excels in restoring the blurred image, yet it demonstrates its unreliability as soon as even a minimal amount of noise is added to the data. In contrast, StNN emerges as an effective compromise between accuracy, as evidenced by the high-quality image in the first row with noise-free data, and stability, highlighted by the superior quality of the StNN image compared to the NN one in the second row under noisy conditions. Table 6.1 reports the values of the empirical accuracy $\hat{\eta}^{-1}$ and empirical stability constant $\hat{C}_{\Psi}^{\epsilon}$ for the considered methods on the whole test set. It confirms that there is a trade-off between accuracy and stability, as proved in Theorem 2.10, and that the stabilization strategy improves the value of $\hat{C}_{\Psi}^{\epsilon}$.

To further investigate the different behavior of the two reconstructors for increasing values of δ , in Figure 6.2 we plot the reconstruction error for $\delta \in [0, 0.03]$. The value of $\delta = 0$ used in the training is indicated with a star marker. We note that

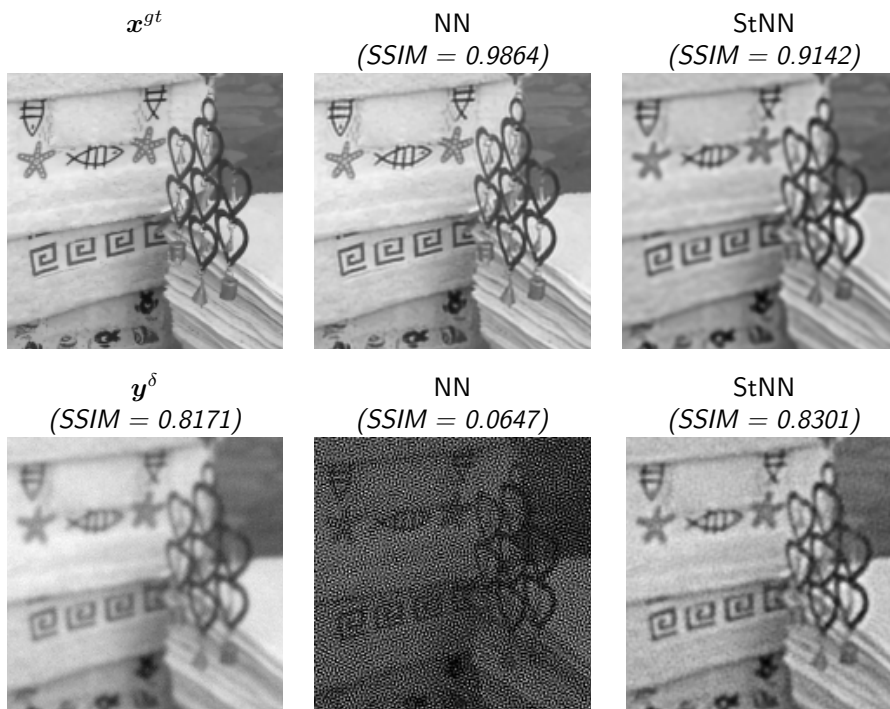


FIG. 6.1. Results obtained by the NN and StNN reconstructors on a single test image \mathbf{y}^{δ} with $\delta = 0$ (first row) and $\delta = 0.01$ (second row). The ground-truth clean image is also reported for reference.

TABLE 6.1
Values of empirical accuracy and ϵ -stability constant obtained by NN and StNN reconstructors, trained on noise-free data.

	NN	StNN
$\hat{\eta}^{-1}$	0.1203	0.0616
$\hat{C}_{\Psi}^{\epsilon}(\delta = 0.01)$	36.7298	0.1579

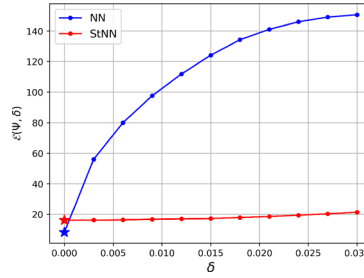


FIG. 6.2. *Plots of the empirical error yielded by NN and StNN reconstructors for increasing values of δ in the test images.*

TABLE 6.2
Values of empirical accuracy and stability constant obtained for iNN and StiNN reconstructors, trained on noisy data with $\delta = 0.025$ and tested with different values of δ .

	iNN	StiNN
$\hat{\eta}^{-1}$	0.0707	0.0606
$C_{\Psi}^{\epsilon}(\delta = 0.025)$	0.0899	0.0703
$C_{\Psi}^{\epsilon}(\delta = 0.060)$	0.4309	0.2122
$C_{\Psi}^{\epsilon}(\delta = 0.125)$	0.8385	0.6215

the StNN curve is characterized by a notably flat trajectory, in contrast to the NN curve, which exhibits a rapid increase. This observation aligns with and reinforces the insights gathered from previous analyses.

In the second experimental setting, we considered the iNN reconstructor, trained by (4.7), with $\delta = 0.025$. Table 6.2 reports the empirical accuracy and stability computed for both iNN and its stabilized version, StiNN, when the methods are tested on data \mathbf{y}_i^{δ} with $\delta = 0.025, 0.060, 0.125$, respectively. The table shows that injecting noise into the observed data during training produces slightly less accurate but far more stable reconstructors (as visible by comparing the results with unseen noise in Table 6.2 to those in Table 6.1).

6.2. Results with ReNN-based reconstructors. In this subsection, we focus on the application of the proposed ReNN reconstructor and its stabilized variant StReNN on noisy data characterized by $\delta = 0.025$. It is important to recall that ReNN is trained following the methodology outlined in (4.8) and utilizes a dataset that does not include the exact \mathbf{x}^{gt} images. The target images are the output of Tikhonov reconstructor $\Psi^{\lambda, L}$ applied to the data $\mathbf{y}_i^{\delta}, i = 1, \dots, N_{\mathbb{D}}$. The Tikhonov regularization parameter λ has been heuristically chosen as $\lambda = 0.31$ to obtain a small reconstruction error on the training set. The same value of λ has been used for computing the input and the target image of the ReNN scheme, as well as to compute the Tikhonov reconstruction. However, in experiments not documented in this paper, we observed that this is not a strict requirement; different values of λ can be selected at each step. We opted not to include these results here because using a consistent λ demonstrated slightly superior performance in our experimental setup. The methods have been tested on noisy data with $\delta = 0.025, 0.060, 0.125$, respectively. The outcomes obtained in terms of accuracy and stability constants are reported in Table 6.3. In the final column of this table, we also include the metrics pertaining to the Tikhonov reconstructor. It is observed that the accuracy of the three methods is quite comparable. Notably, the stability of the regularized NN-based reconstructors

TABLE 6.3

Values of empirical accuracy and stability constant obtained for ReNN, StReNN, and Tikhonov reconstructors, trained on noisy data with $\delta = 0.025$ and tested with different values of δ .

	ReNN	StReNN	Tik
$\hat{\eta}^{-1}$	0.0461	0.0420	0.0474
$C_{\Psi}^{\epsilon}(\delta = 0.025)$	0.0270	0.0150	0.0614
$C_{\Psi}^{\epsilon}(\delta = 0.060)$	0.0739	0.0588	0.1490
$C_{\Psi}^{\epsilon}(\delta = 0.125)$	0.2261	0.1702	0.2822

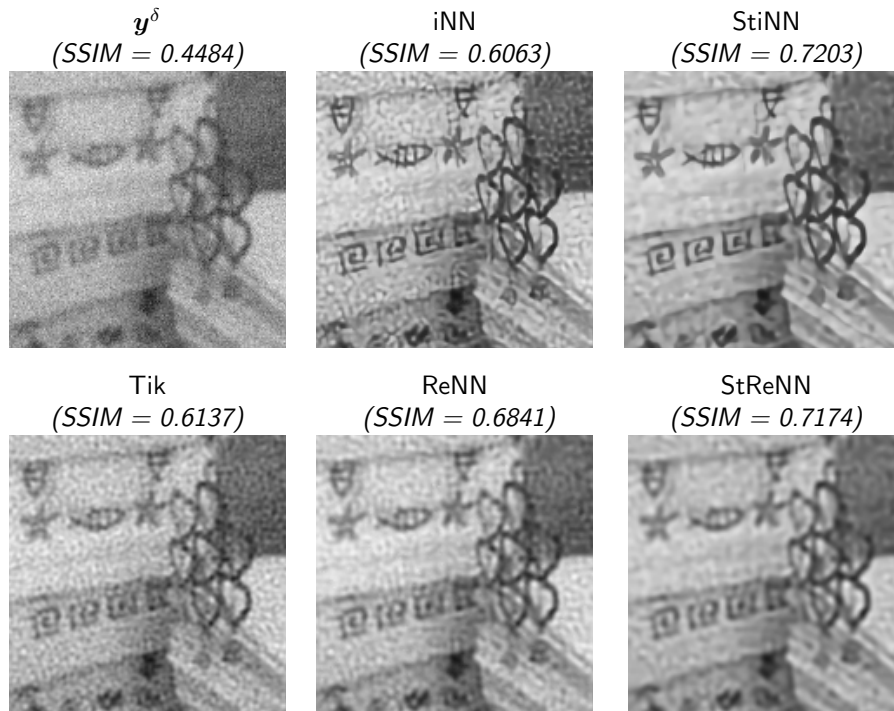


FIG. 6.3. Blurred noisy input image \mathbf{y}^{δ} ($\delta = 0.06$) on the top left and examples of reconstruction obtained by the *iNN*, *StiNN*, *ReNN*, *StReNN*, and *Tikhonov* methods on a test image.

surpasses that of the Tikhonov method. Furthermore, the application of stabilization to ReNN exhibits increasingly beneficial effects as the noise level in the data escalates, as evidenced in the table's last row.

6.3. Comparison among reconstructors. In this final subsection we provide an overview of the results and compare the NN-based reconstructors with the ReNN-based ones. Figure 6.3 shows the output images of the reconstructors trained on noisy data \mathbf{y}_i^{δ} , $\delta = 0.025$ and tested on noisy data with $\delta = 0.060$. As previously observed, the stabilization technique is effective, as demonstrated by the image quality and the SSIM value. Interestingly, comparing *iNN* and *ReNN* we observe that the *ReNN* output images inherit smoothness from the regularized images used as target in (4.8), and exhibits a higher SSIM. At last, *ReNN* also outperforms Tikhonov reconstructor in terms of SSIM.

In Figure 6.4a we plot the reconstruction error of the methods for increasing value of $\delta \in [0, 0.1]$. The value of $\delta = 0.025$ used in the training is indicated with a star

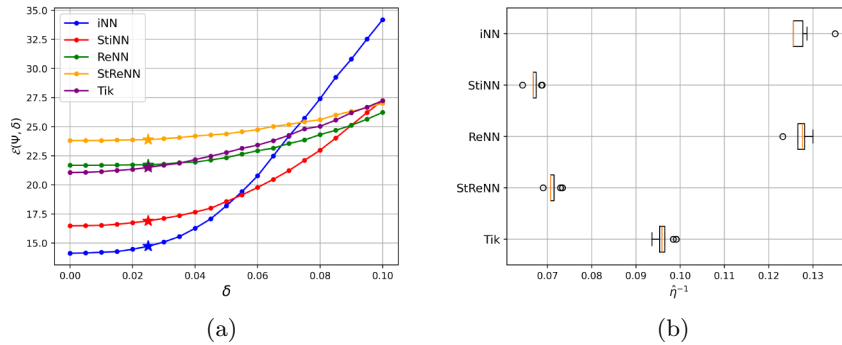


FIG. 6.4. (a) Plots of the empirical error yielded by *iNN*, *StiNN*, *ReNN*, *StReNN*, and *Tikhonov* reconstructors for increasing values of δ in the test images. (b) Boxplots over the $T = 20$ executions.

marker. It is discernible that the blue *iNN* curve demonstrates a markedly steeper gradient, commencing from the minimal error value and escalating to the maximal. The red plot, representing *StiNN*, intersects the blue *iNN* curve at approximately $\delta = 0.055$, indicating a more stable behavior at higher noise levels (color plot available in electronic version of this paper). The remaining three curves, corresponding to the regularized approaches *ReNN*, *StReNN*, and *Tikhonov*, exhibit similar slopes and behaviors. They manifest elevated errors for smaller values of δ , yet surpass the performance of *iNN* when $\delta > 0.07$, yielding results comparable to those of *StiNN*. We underline that when we use *Tikhonov* reconstruction images as targets, i.e., in *ReNN* and *StReNN* approaches, we aim at speeding the iterative process by means of the neural network. Finally, Figure 6.4b presents the boxplots of the experimental accuracy achieved across $T = 20$ executions with varied random realizations. The limited variance in these plots indicates that the values of $\hat{\eta}^{-1}$ are remarkably consistent for each individual reconstructor, thereby affirming the robustness of our accuracy definition.

7. Conclusions. In this paper, we conducted a comprehensive theoretical analysis of a broad spectrum of reconstructors for addressing a discrete ill-posed inverse problem with noisy data. Our findings, particularly encapsulated in Theorem 2.10, establish that enhancing stability in these reconstructors invariably leads to a decrease in accuracy. Our focus was primarily on reconstructors that leverage neural networks.

In consideration of the trade-off theorem, our objective was to enhance the stability of reconstructors based on deep learning, while preserving their accuracy as much as possible. We based our analysis on the reconstructors represented by the popular end-to-end NN approach for image restoration, and we also considered the extensively utilized noise injection stabilization technique, here referred to as *iNN*. As is commonly understood, these approaches are trained using datasets that include images with known ground truth.

We have proposed new deep learning-based approaches: (i) an additional reconstructor, *ReNN*, which is trained on noisy images and increases the stability of NN by inheriting regularization from a model-based scheme in its training; (ii) a stabilization technique which stabilizes the solving process by reducing the impact of the noise with few iterations of a model-based algorithm, and it is applied to all the proposed reconstructors, resulting in *StNN*, *StiNN*, and *StReNN*.

We performed extensive numerical experiments on image deblurring and denoising, with results serving to substantiate the theoretical framework presented in our study. First, we observe, from Tables 6.1, 6.2, and 6.3, that the introduction of the

proposed stabilizers reduces the stability constants of 99.6% in StNN, of about 50% in StiNN and StReNN, with a minimal accuracy loss of about 10–20% in StiNN and StReNN and of about 50% in StNN. Second, in cases where only noisy data are available and ground-truth images are not accessible, the ReNN approach performs exceptionally well and represents a more stable alternative compared to the Tikhonov reconstructor, as demonstrated by Table 6.3 and Figure 6.4. ReNN outperforms even NN when noise impacts the data.

We believe that this new approach for solving noisy linear inverse problems with stable deep learning-based tools is relevant in this field. It can be further theoretically extended to more general problems and formally applied in real imaging applications, as, for example, in [11].

Reproducibility of computational results. This paper has been awarded the “SIAM Reproducibility Badge: Code and data available” as a recognition that the authors have followed reproducibility principles valued by SISC and the scientific computing community. Code and data that allow readers to reproduce the results in this paper are available at <https://github.com/loibo/ToBeOrNotToBeStable> and in the supplementary material (ToBeOrNotToBeStable.zip [local/web 3.76MB]).

REFERENCES

- [1] J. AMJAD, J. SOKOLIĆ, AND M. R. RODRIGUES, *On deep learning for inverse problems*, in 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1895–1899.
- [2] V. ANTUN, F. RENNA, C. POON, B. ADCOCK, AND A. C. HANSEN, *On instabilities of deep learning in image reconstruction and the potential costs of AI*, Proc. Natl. Acad. Sci. USA, 117 (2020), pp. 30088–30095, <https://doi.org/10.1073/pnas.1907377117>.
- [3] S. ARRIDGE, P. MAASS, O. ÖKTEM, AND C.-B. SCHÖNLIEB, *Solving inverse problems using data-driven models*, Acta Numer., 28 (2019), pp. 1–174, <https://doi.org/10.1017/S0962492919000059>.
- [4] M. BERTERO, P. BOCCACCI, AND C. DE MOL, *Introduction to Inverse Problems in Imaging*, CRC Press, 2021.
- [5] C. M. BISHOP, *Training with noise is equivalent to Tikhonov regularization*, Neural Comput., 7 (1995), pp. 108–116, <https://doi.org/10.1162/neco.1995.7.1.108>.
- [6] P. CASCARANO, E. L. PICCOLOMINI, E. MOROTTI, AND A. SEBASTIANI, *Plug-and-Play gradient-based denoisers applied to CT image enhancement*, Appl. Math. Comput., 422 (2022), 126967, <https://doi.org/10.1016/j.amc.2022.126967>.
- [7] M. J. COLBROOK, V. ANTUN, AND A. C. HANSEN, *The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale’s 18th problem*, Proc. Natl. Acad. Sci. USA, 119 (2022), e2107151119, <https://doi.org/10.1073/pnas.2107151119>.
- [8] M. Z. DARESTANI, A. S. CHAUDHARI, AND R. HECKEL, *Measuring robustness in deep learning based compressive sensing*, in International Conference on Machine Learning, PMLR, 2021, pp. 2433–2444.
- [9] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Math. Appl. 375, Springer Science & Business Media, 1996.
- [10] D. EVANGELISTA, E. MOROTTI, AND E. L. PICCOLOMINI, *RISING: A new framework for model-based few-view CT image reconstruction with deep learning*, Comput. Med. Imag. Grap., 103 (2023), 102156, <https://doi.org/10.1016/j.compmedimag.2022.102156>.
- [11] D. EVANGELISTA, E. MOROTTI, E. L. PICCOLOMINI, AND J. NAGY, *Ambiguity in solving imaging inverse problems with deep-learning-based operators*, J. Imaging, 9 (2023), 133, <https://doi.org/10.3390/jimaging9070133>.
- [12] M. GENZEL, J. MACDONALD, AND M. MARZ, *Solving inverse problems with deep neural networks—robustness included?*, IEEE Trans. Pattern Anal. Mach. Intell., 45 (2023), pp. 1119–1134, <https://doi.org/10.1109/TPAMI.2022.3148324>.
- [13] N. M. GOTTSCHLING, V. ANTUN, B. ADCOCK, AND A. C. HANSEN, *The Troublesome Kernel: Why Deep Learning for Inverse Problems Is Typically Unstable*, preprint, version 1, arXiv:2001.01258v1, 2020.
- [14] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, 1998, <https://doi.org/10.1137/1.9780898719697>.

- [15] P. C. HANSEN, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, 2010, <https://doi.org/10.1137/1.9780898718836>.
- [16] P. C. HANSEN, J. G. NAGY, AND D. P. O'LEARY, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, 2006, <https://doi.org/10.1137/1.9780898718874>.
- [17] Y. HUANG, T. WÜRFL, K. BREININGER, L. LIU, G. LAURITSCH, AND A. MAIER, *Some investigations on robustness of deep learning in limited angle tomography*, in Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, Part I, Springer, 2018, pp. 145–153.
- [18] C. M. HYUN, S. H. BAEK, M. LEE, S. M. LEE, AND J. K. SEO, *Deep learning-based solvability of underdetermined inverse problems in medical imaging*, *Med. Image Anal.*, 69 (2021), 101967, <https://doi.org/10.1016/j.media.2021.101967>.
- [19] P. M. JOHNSON, G. JEONG, K. HAMMERNIK, J. SCHLEMPER, C. QIN, J. DUAN, D. RUECKERT, J. LEE, N. PEZZOTTI, E. DE WEERDT, ET AL., *Evaluation of the robustness of learned MR image reconstruction to systematic deviations between training and test data for the models from the fastMRI challenge*, in Machine Learning for Medical Image Reconstruction: 4th International Workshop (MLMIR, 2021), Springer, 2021, pp. 25–34.
- [20] E. MOROTTI, D. EVANGELISTA, AND E. PICCOLOMINI, *A green prospective for learned post-processing in sparse-view tomographic reconstruction*, *J. Imaging*, 7 (2021), 139, <https://doi.org/10.3390/jimaging7080139>.
- [21] J. N. MORSHUIS, S. GATIDIS, M. HEIN, AND C. F. BAUMGARTNER, *Adversarial robustness of MR image reconstruction under realistic perturbations*, in Machine Learning for Medical Image Reconstruction: 5th International Workshop, MLMIR. 2022, held in Conjunction with Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2022, pp. 24–33.
- [22] M. J. MUCKLEY, B. RIEMENSCHNEIDER, A. RADMANESH, S. KIM, G. JEONG, J. KO, Y. JUN, H. SHIN, D. HWANG, M. MOSTAPHA, ET AL., *Results of the 2020 fastMRI challenge for machine learning MR image reconstruction*, *IEEE Trans. Med. Imaging*, 40 (2021), pp. 2306–2317, <https://doi.org/10.1109/TMI.2021.3075856>.
- [23] J. L. MUELLER AND S. SILTANEN, *Linear and Nonlinear Inverse Problems with Practical Applications*, SIAM, 2012, <https://doi.org/10.1137/1.9781611972344>.
- [24] S. NAH, T. H. KIM, AND K. M. LEE, *Deep multi-scale convolutional neural network for dynamic scene deblurring*, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 257–265, <https://doi.org/10.1109/CVPR.2017.35>.
- [25] A. PAL AND Y. RATHI, *A review and experimental evaluation of deep learning methods for MRI reconstruction*, *J. Mach. Learn. Biomed. Imaging*, 1 (2022), pp. 1–50, <https://doi.org/10.59275/j.melba.2022-3g12>.
- [26] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational Methods in Imaging*, *Appl. Math. Sci.* 167, Springer, 2009.
- [27] E. SHIMRON, J. I. TAMIR, K. WANG, AND M. LUSTIG, *Implicit data crimes: Machine learning bias arising from misuse of public data*, *Proc. Natl. Acad. Sci. USA*, 119 (2022), e2117203119, <https://doi.org/10.1073/pnas.2117203119>.
- [28] A. N. TIKHONOV, A. GONCHARSKY, V. STEPANOV, AND A. G. YAGOLA, *Numerical Methods for the Solution of Ill-Posed Problems*, *Math. Appl.* 328, Springer Science & Business Media, 1995.
- [29] Z. WANG, E. P. SIMONCELLI, AND A. C. BOVIK, *Multiscale structural similarity for image quality assessment*, in the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Vol. 2, IEEE, 2003, pp. 1398–1402.
- [30] W. WU, D. HU, W. CONG, H. SHAN, S. WANG, C. NIU, P. YAN, H. YU, V. VARDHANABHUTI, AND G. WANG, *Stabilizing deep tomographic reconstruction: Part A, Hybrid framework and experimental results*, *Patterns*, 3 (2022), 100474.
- [31] W. WU, D. HU, W. CONG, H. SHAN, S. WANG, C. NIU, P. YAN, H. YU, V. VARDHANABHUTI, AND G. WANG, *Stabilizing deep tomographic reconstruction: Part B, Convergence analysis and adversarial attacks*, *Patterns*, 3 (2022), 100475, [10.1016/j.patter.2022.100475](https://doi.org/10.1016/j.patter.2022.100475).
- [32] T. YU, T. HILBERT, G. F. PIREDDA, A. JOSEPH, G. BONANNO, S. ZENKHRI, P. OMOUMI, M. B. CUADRA, E. J. CANALES-RODRÍGUEZ, T. KOBER, ET AL., *Validation and generalizability of self-supervised image reconstruction methods for undersampled MRI*, *J. Mach. Learn. Biomed. Imaging*, 22 (2022), pp. 1–31.
- [33] C. ZHANG, J. JIA, B. YAMAN, S. MOELLER, S. LIU, M. HONG, AND M. AKÇAKAYA, *Instabilities in conventional multi-coil MRI reconstruction with small adversarial perturbations*, in 2021 55th Asilomar Conference on Signals, Systems, and Computers, IEEE, 2021, pp. 895–899.