Nr. 50
23. July 2018

universität innsbruck
Institut für Mathematik

Preprint-Series: Department of Mathematics - Applied Mathematics

Deep Null Space Learning for Inverse Problems: Convergence Analysis and Rates

J. Schwab, S. Antholzer, M. Haltmeier

APPLIEDMATHEMATICS

Technikerstraße 13 - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
https://applied-math.uibk.ac.at

Deep Null Space Learning for Inverse Problems: Convergence Analysis and Rates

Johannes Schwab, Stephan Antholzer, and Markus Haltmeier*

Department of Mathematics, University of Innsbruck
Technikerstrasse 13, 6020 Innsbruck, Austria

*Correspondence: markus.haltmeier@uibk.ac.at

July 22, 2018

Abstract

Recently, deep learning based methods appeared as a new paradigm for solving inverse problems. These methods empirically show excellent performance but lack of theoretical justification; in particular, no results on the regularization properties are available. In particular, this is the case for two-step deep learning approaches, where a classical reconstruction method is applied to the data in a first step and a trained deep neural network is applied to improve results in a second step. In this paper, we close the gap between practice and theory for a new network structure in a two-step approach. For that purpose, we propose so called null space networks and introduce the concept of $\mathcal{M}$-regularization. Combined with a standard regularization method as reconstruction layer, the proposed deep null space learning approach is shown to be a $\mathcal{M}$-regularization method; convergence rates are also derived. The proposed null space network structure naturally preserves data consistency which is considered as key property of neural networks for solving inverse problems.

Keywords: inverse problems, null space networks, deep learning, $\mathcal{M}$-regularization, convergence analysis, convolutional neural networks, convergences rates

AMS subject classifications: 65J20, 65J22, 45F05

1 Introduction

We study the solution of inverse problems of the form

$$\text{Estimate } x \in X \text{ from data} \quad y^\delta = \mathbf{A}x + \xi\,. \tag{1.1}$$

Here $\mathbf{A}\colon X \to Y$ is a linear operator between Hilbert spaces X and Y, and $\xi \in Y$ models the unknown data error (noise), which is assumed to satisfy the estimate $\|\xi\| \leq \delta$ for some noise level $\delta \geq 0$. We thereby allow a possibly infinite-dimensional function space setting, but clearly the approach and results apply to a finite dimensional setting as well.

We focus on the ill-posed (or ill-conditioned) case where, without additional information, the solution of (1.1) is either highly unstable, highly undetermined, or both. Many inverse problems in biomedical imaging, geophysics, engineering sciences, or elsewhere can be written in such a form (see, for example, [7, 19]). For its stable solution one has to employ regularization methods, which are based on approximating (1.1) by neighboring well-posed problems, which enforce stability, accuracy, and uniqueness.

1.1 Regularization methods

Any method for the stable solution of (1.1) uses, either implicitly or explicitly, a-priori information about the unknowns to be recovered. Such information can be that x belongs to a certain set of admissible elements $\mathcal{M}$ or that it has small value of some regularizing functional. The most basic regularization method is probably Tikhonov regularization, where the solution is defined as a minimizer of the quadratic Tikhonov functional

$$\mathcal{T}_{\alpha;y_\delta}(x) := \frac{1}{2}\|\mathbf{A}(x) - y_\delta\|^2 + \frac{\alpha}{2}\|x\|^2 \tag{1.2}$$

Other classical regularization methods for solving linear inverse problems are filter based methods [7], which include Tikhonov regularization as special case.

In the last couple of years variational regularization methods including TV regularization or ℓ^q regularization became popular [19]. They also include classical Tikhonov regularization as special case. In the general version, the regularizer $\frac{1}{2}\|\cdot\|^2$ is replaced by general convex and lower semi-continuous functionals.

In this paper, we develop a new regularization concept that we name $\mathcal{M}$-regularization method. Roughly spoken, an $\mathcal{M}$-regularization method is a tuple $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ where (for a precise definition see Definition 2.3)

- $\mathcal{M} \subseteq X$ is the set of admissible elements;
- $\mathbf{R}_\alpha \colon Y \to X$ are continuous mappings;
- $\alpha^\star = \alpha^\star(\delta, y^\delta)$ is a suitable parameter choice;
- For any $x \in \mathcal{M}$ we have $\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) \to x$ as $\delta \to 0$.

Note that for some cases is might be reasonable to take $\mathbf{R}_\alpha$ multivalued. For the sake of simplicity here we only consider the single-valued case. Classical regularization methods are special cases of $\mathcal{M}$-regularization methods in Hilbert spaces where $\mathcal{M} = \ker(\mathbf{A})^\perp$. A typical regularization method is in this case given by Tikhonov regularization, where $\mathbf{R}_\alpha = (\mathbf{A}^*\mathbf{A} + \alpha \operatorname{Id}_X)^{-1}\mathbf{A}^*$.

1.2 Solving inverse problems by neural networks

Very recently, deep learning approaches appeared as alternative, very successful methods for solving inverse problems (see, for example, [1, 2, 3, 4, 9, 14, 8, 5, 11, 23, 22, 24, 25]). In most of these approaches, a reconstruction network $\mathbf{R}\colon Y \to X$ is trained to map measured data to the desired output image.

Various reconstruction networks have been introduced in the literature. In the two-step approach, the reconstruction networks take the form $\mathbf{R} = \mathbf{L} \circ \mathbf{B}$ where $\mathbf{B}\colon Y \to X$ maps the data to the reconstruction space (reconstruction layer or backprojection; no free parameters) and $\mathbf{L}\colon X \to X$ is a neural network (NN) whose free parameters are adjusted to the training data. In particular, so called residual networks $\mathbf{L} = \mathrm{Id}_X + \mathbf{N}$ where only the residual part $\mathbf{N}$ is trained [10] showed very accurate results for solving inverse problems ([2, 6, 11, 12, 15, 17, 18, 23]). Here and in the following Id_X denotes the identity on X. Another class of reconstruction networks learns free parameters in iterative schemes. In such approaches, a sequence of reconstruction networks $\mathbf{R} = \mathbf{R}^{(k)}$ is defined by some iterative process $\mathbf{R}^{(k)}(y) = \mathbf{N}_k(\Phi_k(x_{k-1}, \dots, x_0, y))$ where x_0 is some the initial guess, $\mathbf{N}_k\colon X \to X$ are networks that can be adjusted to available training data, and Φ_k are updates based on the data and the previous iterates [1, 13, 14, 20].

Further existing deep learning approaches for solving inverse problems are based on trained projection operators [4, 8], or use neural networks as trained regularization term [16].

While the above deep learning based reconstruction networks empirically yield good performance, none of them is known to be a convergent regularization method. In this paper we introduce a new network structure (null space network) that, when combined with a classical regularization of the Moore Penrose inverse is shown to provide a convergent $\mathcal{M}$-regularization method with rates.

1.3 Proposed null space networks and main results

As often argued in the recent literature, deep learning based reconstruction approaches (especially using two-stage networks) lack data consistency, in the sense that outputs of existing reconstruction networks fail to accurately predict the given data. In order to overcome this issue, in this paper, we introduce a new network, that we name null space network. The propose null space network takes the form (see Definition 3.2)

$$\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N} \quad \text{for a network function } \mathbf{N}\colon X \to X\,. \tag{1.3}$$

Note that $\mathrm{Id}_X - \mathbf{A}^+\mathbf{A} = \mathbf{P}_{\ker(\mathbf{A})}$ equals the projector onto the null space $\ker(\mathbf{A})$ of $\mathbf{A}$. Consequently, the null space network $\mathbf{L}$ satisfies the property $\mathbf{A}\mathbf{L}x = \mathbf{A}x$ for all $x \in X$. This yields that data consistency, which means that $\mathbf{A}x = y$ is invariant among application of a null space network (compare Figure 1.1).

Suppose $x_1, \dots, x_N$ are some desired output images and let $\mathbf{L}$ be a trained null space network that approximately maps $\mathbf{A}^+\mathbf{A}x_n$ to x_n. (See Subsection 3.2 for a possible training strategy.) In this paper, we show that if $(\mathbf{B}_\alpha)_{\alpha>0}$ is any classical $\ker(\mathbf{A})^\perp$-regularization, then the two-stage reconstruction network

$$\mathbf{R}_\alpha \coloneqq \mathbf{L} \circ \mathbf{B}_\alpha \quad \text{for } \alpha > 0 \tag{1.4}$$

yields a $\mathcal{M}$-regularization with $\mathcal{M} := \mathbf{L}\operatorname{ran}(\mathbf{A}^+)$. To the best of our knowledge, these are first results for regularization by neural networks. Additionally we will derive convergence rates for $(\mathbf{R}_\alpha)_{\alpha>0}$ on suitable function classes.

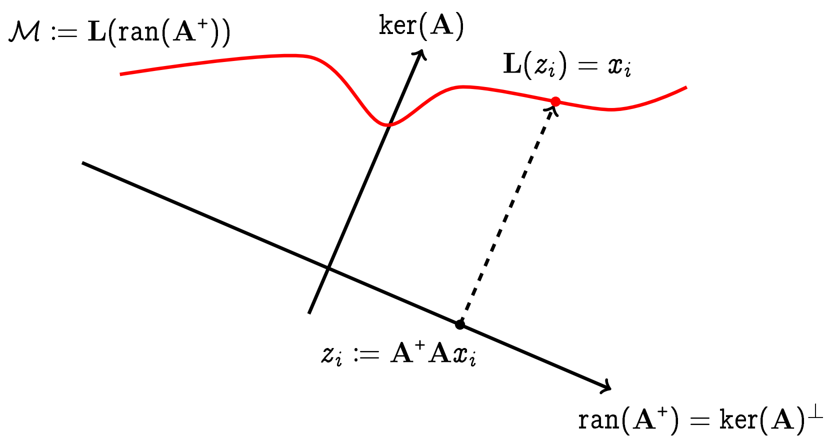

Figure 1.1: Sketch of the action of a null space network $\mathbf{L}$ that maps points $z_i \in \operatorname{ran}(\mathbf{A}^+)$ to more desirable elements in $z_i + \ker(\mathbf{A})$ along the null space of $\mathbf{A}$.

1.4 Outline

This paper is organized as follows. In Section 2 we develop a general theory of $\mathcal{M}$-regularization and introduce the notion of $\mathcal{M}$-generalized inverse (Definition 2.1) and $\mathcal{M}$-regularization methods (Definition 2.3) generalizing the classical Moore-Penrose generalized inverse and regularization concept. We show convergence (see Theorem 2.4) and derive convergence rates (Theorem 2.8) that include regularization via the null space networks as a special case. In Section3 we introduce the null-space networks, describe possible training and extend the convergence results in the special case of the null space network (Theorems 3.3 and 3.4). The paper concludes with an outlook presented in Section 4.

2 A theory of $\mathcal{M}$-regularization

In this section, we introduce the novel concepts of $\mathcal{M}$-generalized inverse and $\mathcal{M}$-regularization. We derive a general class of $\mathcal{M}$-regularization for which we show convergence and derive convergence rates.

Throughout this section, let $\mathbf{A}\colon X \to Y$ be a linear bounded operator and $\mathbf{\Phi}\colon X \to \ker(\mathbf{A}) \subseteq X$ be Lipschitz continuous and define

$$\mathcal{M} := (\mathrm{Id}_X + \mathbf{\Phi})\operatorname{ran}(\mathbf{A}^+). \tag{2.1}$$

The prime example is $\mathbf{\Phi} = \mathbf{P}_{\ker(\mathbf{A})} \circ \mathbf{N}$ being a null space network with a neural network function $\mathbf{N}\colon X \to X$. This case will be studied in the following section.

The results presented in this section apply to general Lipschitz continuous functions $\mathbf{\Phi}$ whose image is contained in $\ker(\mathbf{A})$.

2.1 $\mathcal{M}$-regularization methods

In the following we denote by $\mathbf{A}^+\colon \operatorname{dom}(\mathbf{A}^+) \subseteq Y \to X$ the Moore-Penrose generalized inverse of $\mathbf{A}$, defined by $\operatorname{dom}(\mathbf{A}^+) := \operatorname{ran}(\mathbf{A}) \oplus \operatorname{ran}(\mathbf{A})^\perp$ and

$$\mathbf{A}^+(y) := \arg\min\left\{\|x\|^2 \mid x \in X \wedge \mathbf{A}^*\mathbf{A}x = \mathbf{A}^*y\right\} \tag{2.2}$$

It is well known [7] that $\{x \in X \mid \mathbf{A}^*\mathbf{A}x = \mathbf{A}^*y\} \neq \emptyset$ if and only if $y \in \operatorname{ran}(\mathbf{A}) \oplus \operatorname{ran}(\mathbf{A})^\perp$. In particular, $\mathbf{A}^+y$ is well defined, and can be found as the unique minimal norm solution of the normal equation $\mathbf{A}^*\mathbf{A}x = \mathbf{A}^*y$

Classical regularization methods aim for approximating $\mathbf{A}^+y$. In contrast, the null space network will recover different solutions of the normal equation. For that purpose we introduce the following concept.

Definition 2.1 ($\mathcal{M}$-generalized inverse)**.** We call $\mathbf{A}^{\mathcal{M}}\colon \operatorname{dom}(\mathbf{A}^+) \subseteq Y \to X$ the $\mathcal{M}$-generalized inverse of $\mathbf{A}$ if

$$\forall y \in \operatorname{dom}(\mathbf{A}^+)\colon \quad \mathbf{A}^{\mathcal{M}}y = (\mathrm{Id}_X + \mathbf{\Phi})(\mathbf{A}^+y)\,. \tag{2.3}$$

Recall that for any $y \in \operatorname{dom}(\mathbf{A}^+)$, the solution set of the normal equation $\mathbf{A}^*\mathbf{A}x = \mathbf{A}^*y$ is given by $\mathbf{A}^+y + \ker(\mathbf{A})$. Hence $\mathbf{A}^{\mathcal{M}}y$ gives a particular solution of the normal equation, that can be adapted to a training set. The $\mathcal{M}$-generalized inverse coincides with the Moore-Penrose generalized inverse if and only if $\mathbf{\Phi}(x) = 0$ for all $x \in X$ in which case $\mathcal{M} = \ker(\mathbf{A})^\perp = \operatorname{ran}(\mathbf{A}^+)$.

Lemma 2.2. *The $\mathcal{M}$-generalized inverse is continuous if and only if* $\operatorname{ran}(\mathbf{A})$ *is closed.*

Proof. If $\operatorname{ran}(\mathbf{A})$ is closed, then classical results show that $\mathbf{A}^+$ is bounded (see for example [7]). Consequently, $(\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{A}^+$ is bounded too. Conversely, if $\mathbf{A}^{\mathcal{M}}$ is continuous, then the identity $\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)}\mathbf{A}^{\mathcal{M}} = \mathbf{A}^+$ implies that the Moore-Penrose generalized inverse $\mathbf{A}^+$ is bounded and therefore that $\operatorname{ran}(\mathbf{A})$ is closed. □

Lemma 2.2 shows that as in the case of the classical Moore-Penrose generalized inverse, the $\mathcal{M}$-generalized inverse is discontinuous in the case that $\operatorname{ran}(\mathbf{A})$ is not closed. In order to stably solve the equation $\mathbf{A}x = y$ we therefore require bounded approximations of the $\mathcal{M}$-generalized inverse. For that purpose, we introduce the following concept of regularization methods adapted to $\mathbf{A}^+$.

Definition 2.3 ($\mathcal{M}$-regularization method)**.** Let $(\mathbf{R}_\alpha)_{\alpha>0}$ be a family of continuous (not necessarily linear) mappings $\mathbf{R}_\alpha\colon Y \to X$ and let $\alpha^\star\colon (0,\infty) \times Y \to (0,\infty)$. We call the pair $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ a $\mathcal{M}$-regularization method for the equation $\mathbf{A}x = y$ if the following hold:

- $\forall y \in Y\colon\ \lim_{\delta\to 0} \sup\left\{\alpha^\star(\delta, y^\delta) \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta\right\} = 0$.

- $\forall y \in Y\colon \lim_{\delta\to 0} \sup \left\{ \|\mathbf{A}^{\mathcal{M}} y - \mathbf{R}_{\alpha^\star(\delta,y^\delta)} y^\delta\| \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta \right\} = 0$.

In the case that $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ is a $\mathcal{M}$-regularization method for $\mathbf{A}x = y$, then we call the family $(\mathbf{R}_\alpha)_{\alpha>0}$ a regularization of $\mathbf{A}^{\mathcal{M}}$ and $\alpha^\star$ an admissible parameter choice.

In our generalized notation, a classical regularization method for the equation $\mathbf{A}x = y$ corresponds to a $\operatorname{ran}(\mathbf{A}^+)$-regularization method for $\mathbf{A}x = y$

2.2 Convergence analysis

The following theorem shows that the combination of a null space network and a regularization method of $\mathbf{A}^+$ yields a regularization of $\mathbf{A}^{\mathcal{M}}$.

Theorem 2.4. *Suppose $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ is any classical regularization method for $\mathbf{A}x = y$. Then, the pair $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ with $\mathbf{R}_\alpha := (\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{B}_\alpha$ is a $\mathcal{M}$-regularization method for $\mathbf{A}x = y$. In particular, the family $(\mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$.*

Proof. Because $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ is a $\operatorname{ran}(\mathbf{A}^+)$-regularization method, it holds that $\lim_{\delta\to 0} \sup\{\alpha^\star(\delta, y^\delta) \mid y^\delta \in Y \text{ und } \|y^\delta - y\| \leq \delta\} = 0$. Let L be a Lipschitz constant of $\mathrm{Id}_X + \mathbf{\Phi}$. For any y^δ we have

$$\left\|\mathbf{A}^{\mathcal{M}} y - \mathbf{R}_{\alpha^\star(\delta,y^\delta)} y^\delta\right\| = \left\|(\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{A}^+ y - (\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{B}_{\alpha^\star(\delta,y^\delta)} y^\delta\right\| \\ \leq L \left\|\mathbf{A}^+ y - \mathbf{B}_{\alpha^\star(\delta,y^\delta)} y^\delta\right\| .$$

Consequently

$$\sup \left\{ \|\mathbf{A}^{\mathcal{M}} y - (\mathrm{Id}_X + \mathbf{\Phi}) \mathbf{B}_{\alpha^\star(\delta,y^\delta)} y^\delta\| \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta \right\} \\ \leq L \sup \left\{ \|\mathbf{A}^+ y - \mathbf{B}_{\alpha^\star(\delta,y^\delta)} y^\delta\| \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta \right\} \to 0 .$$

In particular, $((\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{B}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$. □

A wide class of $\mathcal{M}$-regularization methods can be defined by a regularizing filter.

Definition 2.5. A family $(g_\alpha)_{\alpha>0}$ of functions $g_\alpha\colon [0, \|\mathbf{A}^*\mathbf{A}\|] \to \mathbb{R}$ is called a regularizing filter if it satisfies

- For all $\alpha > 0$, g_α is piecewise continuous;
- $\exists C > 0\colon \sup \{ |\lambda g_\alpha(\lambda)| \mid \alpha > 0 \wedge \lambda \in [0, \lambda_{\max}] \} \leq C$.
- $\forall \lambda \in (0, \|\mathbf{A}^*\mathbf{A}\|]\colon \lim_{\alpha\to 0} g_\alpha(\lambda) = 1/\lambda$.

Corollary 2.6. *Let $(g_\alpha)_{\alpha>0}$ be a regularizing filter and define $\mathbf{B}_\alpha := g_\alpha(\mathbf{A}^*\mathbf{A})\,\mathbf{A}^*$. Then $((\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{B}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$.*

Proof. The family $(\mathbf{B}_\alpha)_\alpha$ is a regularization of $\mathbf{A}^+$; see [7]. Therefore, according to Theorem 2.4, $((\mathrm{Id}_X + \mathbf{\Phi}) \circ \mathbf{B}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$. □

Basic examples of filter based regularization methods are Tikhonov regularization, where $g_\alpha(\lambda) = 1/(\alpha+\lambda)$, and truncated singular value decomposition where

$$g_\alpha(\lambda) := \begin{cases} 0 & \text{if } \lambda < \alpha \\ 1/\lambda & \text{if } \lambda \geq \alpha\,. \end{cases}$$

Classical regularization methods are based on approximating the Moore-Penrose inverse. In our notation, this corresponds to a $\operatorname{ran}(\mathbf{A}^+)$-regularization methods. The following result shows that $\mathcal{M}$-regularization methods are essentially continuous approximations of $\mathbf{A}^{\mathcal{M}}$.

Proposition 2.7. *Let $(\mathbf{R}_\alpha)_{\alpha>0}$ be a family of continuous mappings $\mathbf{R}_\alpha \colon Y \to X$.*

(a) If $\mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} \to \mathbf{A}^{\mathcal{M}}$ pointwise as $\alpha \to 0$, then the family $(\mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$.

(b) Suppose that $(\mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$ and that there exists a parameter choice $\alpha^\star$ that is continuous in the first argument. Then $\mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} \to \mathbf{A}^{\mathcal{M}}$ pointwise as $\alpha \to 0$.

Proof. (a) If $\mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} \to \mathbf{A}^{\mathcal{M}}$ pointwise, then $\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} \to \mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{A}^{\mathcal{M}} = \mathbf{A}^+$ pointwise. Hence, classical regularization theory implies that $\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha$ is a regularization of $\mathbf{A}^+$. We have $\mathbf{R}_\alpha = (\operatorname{Id}_X + \mathbf{\Phi}) \circ \mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha$ and, according to Theorem 2.4, the family $(\mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$.

(b) We have

$$\begin{aligned} \sup \Big\{ \|\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)}(\mathbf{A}^{\mathcal{M}} y - \mathbf{R}_{\alpha^\star(\delta, y^\delta)} y^\delta)\| \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta \Big\} \\ \leq \sup \Big\{ \|\mathbf{A}^{\mathcal{M}} y - \mathbf{R}_{\alpha^\star(\delta, y^\delta)} y^\delta\| \mid y^\delta \in Y \wedge \|y^\delta - y\| \leq \delta \Big\} \to 0\,, \end{aligned}$$

which shows that $(\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^+ = \mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{A}^{\mathcal{M}}$. Together with standard regularization theory this shows that $\mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} \to \mathbf{A}^+$ pointwise as $\alpha \to 0$. Consequently, $\mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)} = (\operatorname{Id}_X + \mathbf{\Phi}) \circ \mathbf{P}_{\operatorname{ran}(\mathbf{A}^+)} \circ \mathbf{R}_\alpha|_{\operatorname{dom}(\mathbf{A}^+)}$ converges pointwise to $\mathbf{A}^{\mathcal{M}} = (\operatorname{Id}_X + \mathbf{\Phi}) \circ \mathbf{A}^+$. □

2.3 Convergence rates

Next we derive quantitative error estimates. For that purpose, we assume in the following that $\mathbf{B}_\alpha = g_\alpha(\mathbf{A}^*\mathbf{A})\,\mathbf{A}^*$ is defined by the regularizing filter $(g_\alpha)_{\alpha>0}$. We use the notation $\alpha^\star \asymp (\delta/\rho)^a$ as $\delta \to 0$ where $\alpha^\star \colon Y \times (0,\infty) \to (0,\infty)$ and $a, \rho > 0$ to indicate there are positive constants d_1, d_2 such that $d_1(\delta/\rho)^a \leq \alpha^\star(\delta) \leq d_2(\delta/\rho)^a$.

Theorem 2.8. *Suppose $\mu, \rho > 0$ and let $(g_\alpha)_{\alpha>0}$ be a regularizing filter such that there exist constants $\alpha_0, c_1, c_2 > 0$ with*

- $\forall \alpha > 0\ \forall \lambda \in [0, \|\mathbf{A}^*\mathbf{A}\|] \colon \lambda^\mu \,|1 - \lambda g_\alpha(\lambda)| \leq c_1 \alpha^\mu$;
- $\forall \alpha \in (0, \alpha_0) \colon \|g_\alpha\|_\infty \leq c_2/\alpha$.

Consider the $\mathcal{M}$-regularization method $\mathbf{R}_\alpha := (\mathrm{Id}_X + \mathbf{\Phi}) \circ g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*$ and set

$$\mathcal{M}_{\mu,\rho,\mathbf{\Phi}} := (\mathrm{Id}_X + \mathbf{\Phi})\,(\mathbf{A}^*\mathbf{A})^\mu \left(\overline{B_\rho(0)}\right) . \tag{2.4}$$

Moreover, let $\alpha^\star \colon (0,\infty) \times Y \to (0,\infty)$ be a parameter choice (possible depending on the source set $\mathcal{M}_{\mu,\rho,\mathbf{\Phi}}$) that satisfies $\alpha^\star \asymp (\delta/\rho)^{\frac{2}{2\mu+1}}$ as $\delta \to 0$. Then there exists a constant $c > 0$ such that

$$\sup\left\{\|\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) - x\| \mid x \in \mathcal{M}_{\mu,\rho,\mathbf{\Phi}} \wedge y^\delta \in Y \wedge \|\mathbf{A}x - y^\delta\| \leq \delta\right\} \\ \leq c\delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} . \tag{2.5}$$

In particular, for any $x \in \mathrm{ran}((\mathrm{Id}_X + \mathbf{\Phi}) \circ (\mathbf{A}^*\mathbf{A})^\mu)$ we have the convergence rate result $\|\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) - x\| = \mathcal{O}(\delta^{\frac{2\mu}{2\mu+1}})$.

Proof. We have $\mathbf{P}_{\mathrm{ran}(\mathbf{A}^+)}\mathcal{M}_{\mu,\rho,\mathbf{\Phi}} = (\mathbf{A}^*\mathbf{A})^\mu \left(\overline{B_\rho(0)}\right)$ and $\mathbf{P}_{\mathrm{ran}(\mathbf{A}^+)}\mathbf{R}_\alpha = g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*$. Suppose $x \in \mathcal{M}_{\mu,\rho,\mathbf{\Phi}}$ and $y^\delta \in Y$ with $\|\mathbf{A}x - y^\delta\| \leq \delta$. Under the given assumptions, $g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*$ is an order optimal regularization method on $(\mathbf{A}^*\mathbf{A})^\mu \left(\overline{B_\rho(0)}\right)$, which implies (see [7])

$$\|g_{\alpha^\star(\delta,y^\delta)}(\mathbf{A}^*\mathbf{A})\mathbf{A}^*(y^\delta) - \mathbf{P}_{\mathrm{ran}(\mathbf{A}^+)}x\| \leq C\,\delta^{\frac{2\mu}{2\mu+1}}$$

for some constant $C > 0$ independent of x, y^δ. Consequently, we have

$$\begin{aligned}\|\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) - x\| &= \|(\mathrm{Id}_X + \mathbf{\Phi}) \circ g_{\alpha^\star(\delta,y^\delta)}(\mathbf{A}^*\mathbf{A})\mathbf{A}^*(y^\delta) - (\mathrm{Id}_X + \mathbf{\Phi})\mathbf{P}_{\mathrm{ran}(\mathbf{A}^+)}x\| \\ &\leq L\|g_{\alpha^\star(\delta,y^\delta)}(\mathbf{A}^*\mathbf{A})\mathbf{A}^*(y^\delta) - \mathbf{P}_{\mathrm{ran}(\mathbf{A}^+)}x\| \\ &\leq LC\delta^{\frac{2\mu}{2\mu+1}}\rho^{\frac{1}{2\mu+1}} ,\end{aligned}$$

where L is the Lipschitz constant of $\mathrm{Id}_X + \mathbf{\Phi}$. Taking the supremum over all $x \in \mathcal{M}_{\mu,\rho,\mathbf{\Phi}}$ and $y^\delta \in Y$ with $\|\mathbf{A}x - y^\delta\| \leq \delta$ yields (2.5). □

Note that the filters $(g_\alpha)_{\alpha>0}$ of the truncated SVD and the Landweber iteration satisfy the assumptions of Theorem 2.8. In the case of Tikhonov regularization the assumptions are satisfied for $\mu \leq 1$. In particular under the assumption (resembling the classical source condition)

$$x \in (\mathrm{Id}_X + \mathbf{\Phi})(\mathrm{ran}(\mathbf{A}^+))$$

we obtain the convergence rate $\|\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) - x\| = \mathcal{O}(\delta^{1/2})$.

3 Deep null space learning

Throughout this section let $\mathbf{A} \colon X \to Y$ be a linear bounded operator. In this case, we define $\mathcal{M}$-regularizations by null-space networks. We describe a possible training strategy and derive regularization properties and rates. For the following recall that the projector onto the kernel of $\mathbf{A}$ is given by $\mathbf{P}_{\ker(\mathbf{A})} = \mathrm{Id}_X - \mathbf{A}^+\mathbf{A}$.

3.1 Null space networks

For simplicity. we work with layered feed forward networks, although more complicated networks can be applied as long as their Lipschitz constant is not too large. While their notation is standard in a finite-dimensional setting, no formal definitions seems available for general Hilbert spaces. We introduce the following Hilbert space notion.

Definition 3.1 (Layered feed forward network)**.** Let X and Z be Hilbert spaces. We call a function $\mathbf{N}\colon X \to Z$ defined by

$$\mathbf{N} := \sigma_L \circ \mathbf{W}_L \circ \sigma_{L-1} \circ \mathbf{W}_{L-1} \circ \cdots \circ \sigma_1 \circ \mathbf{W}_1 \,, \tag{3.1}$$

a layered feed forward neural network function of depth $L \in \mathbb{N}$ with activations $\sigma_1, \dots, \sigma_L$ if

(N1) X_ℓ are Hilbert spaces with $X_0 = X$ and $X_L = Z$;

(N2) $\mathbf{W}_\ell\colon X_{\ell-1} \to X_\ell$ are affine, continuous;

(N3) $\sigma_\ell\colon X_\ell \to X_\ell$ are continuous.

Usually the nonlinearities σ_ℓ are fixed and the affine mappings $\mathbf{W}_\ell$ are trained. In the case that X_ℓ is a function space, then a standard operation for σ_ℓ is the ReLU (the rectified linear unit), $\mathrm{ReLU}(x) := \max\{x, 0\}$, that is applied componentwise, or ReLU in combination with max pooling which takes the maximum value $\max\{|x(i)| : i \in I_k\}$ within clusters of transform coefficients. The network in Definition 3.1 may in particular be a convolutional neural network (CNN); see [16] for a definition even in Banach spaces. In a similar manner one could define more general feed forward networks in Hilbert spaces, for example following the notion of [21] in the finite dimensional case.

We are now able to formally define the concept of a null space network.

Definition 3.2. A function $\mathbf{L}\colon X \to X$ is a null space network if it has the form $\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ where $\mathbf{N}\colon X \to X$ is a neural network function as in (3.1).

3.2 Network training

We train the null space network $\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ to (approximately) map elements to the desired class of training phantoms. For that purpose, we fix the following:

- $\mathcal{M}_N = \{x_1, \dots, x_N\}$ is a class of training phantoms;
- For all ℓ fix the nonlinearity $\sigma_\ell\colon X_\ell \to X_\ell$;
- $\mathcal{W}_\ell$ are finite-dimensional spaces of affine continuous mappings;
- $\mathcal{N}$ is the set of all NN functions of the form (3.1) with $\mathbf{W}_\ell \in \mathcal{W}_\ell$.

We then consider null space network $\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ where $\mathbf{N} \in \mathcal{N}$. To train the null space networks we propose to minimize the regularized error functional $E \colon \mathcal{N} \to \mathbb{R}$ defined by

$$E(\mathbf{N}) := \frac{1}{2} \sum_{\ell=1}^{L} \|x_n - (\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N})(\mathbf{A}^+\mathbf{A}x_n)\|^2 + \mu \prod_{\ell=1}^{L} \|\mathbf{L}_\ell\| \tag{3.2}$$

where $\mathbf{N}$ is of the form (3.1) and $\mathbf{L}_\ell$ is the linear part of $\mathbf{W}_\ell$ and μ is a regularization parameter.

Network training aims at making $E(\mathbf{N})$ small, for example, by gradient descent. Clearly $\prod_{\ell=1}^{L} \|\mathbf{L}_\ell\|$ is an upper bound on the Lipschitz constant of $\mathbf{N}$. Therefore, the Lipschitz constant of the finally trained network will stay reasonably small. Note that it is not required that (3.2) is exactly minimized. Any trained network where $\frac{1}{2} \sum_{\ell=1}^{L} \|x_n - (\mathrm{Id}_X + \mathbf{A}^+\mathbf{A}\mathbf{N})(\mathbf{A}^+\mathbf{A}x_n)\|^2$ is small yields a null space network $\mathrm{Id}_X + \mathbf{A}^+\mathbf{A}\mathbf{N}$ that does, at least on the training set, a better job in estimating x_n from $\mathbf{A}^+\mathbf{A}x_n$ than the identity.

Alternatively, we may train a regularized null space network $\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{B}_\alpha\mathbf{A})\mathbf{N}$ to map the regularized data $\mathbf{B}_\alpha\mathbf{A}x_n$ (instead of $\mathbf{A}^+\mathbf{A}x_n$) to the outputs x_n. This yields the modified error functional

$$E_\alpha(\mathbf{N}) := \frac{1}{2} \sum_{\ell=1}^{L} \|x_n - (\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{B}_\alpha\mathbf{A})\mathbf{N})(\mathbf{B}_\alpha x_n)\|^2 + \mu \prod_{\ell=1}^{L} \|\mathbf{L}_\ell\| \,. \tag{3.3}$$

Trying to minimize E_α may be beneficial in the case that many singular values are small but do not vanish exactly. The regularized version $\mathbf{B}_\alpha$ might be defined by truncated SVD or Tikhonov regularization.

3.3 Convergence and convergence rates

Let $\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ be a null-space network, possibly trained as described in Section 3.2 by approximately minimizing (3.2). Any such network belongs to the class of functions $\mathrm{Id}_X + \mathbf{\Phi}$ by taking $\mathbf{\Phi} = (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$. Consequently, the convergence theory of Section 2 applies. In particular, Theorem 2.4 shows that a regularization $(\mathbf{B}_\alpha)_{\alpha>0}$ of the Moore-Penrose generalized inverse defines a $\mathcal{M}$-regularization method via $\mathbf{R}_\alpha := (\mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N})\mathbf{B}_\alpha$. Additionally, Theorem 2.8 yields convergence rates for the regularization $(\mathbf{R}_\alpha)_{\alpha>0}$ of $\mathbf{A}^{\mathcal{M}}$.

In some cases, the projection $\mathbf{P}_{\ker(\mathbf{A})} = \mathrm{Id}_X - \mathbf{A}^+\mathbf{A}$ might be costly to be computed exactly. For that purpose, in this section we derive more general regularization methods that include approximate evaluations of $\mathbf{A}^+\mathbf{A}$.

Theorem 3.3. *Let $\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ be a null space network and set $\mathcal{M} := \mathrm{ran}(\mathbf{L})$. Suppose $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ is a regularization method for $\mathbf{A}x = y$. Moreover, let $(\mathbf{Q}_\alpha)_{\alpha>0}$ be a family of bounded operators on X with $\|\mathbf{Q}_\alpha - \mathbf{P}_{\ker(\mathbf{A})}\| \to 0$ as $\alpha \to 0$. Then, the pair $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ with*

$$\mathbf{R}_\alpha := (\mathrm{Id}_X + \mathbf{Q}_\alpha\mathbf{N}) \circ \mathbf{B}_\alpha \tag{3.4}$$

is a $\mathcal{M}$-regularization method for $\mathbf{A}x = y$. In particular, the family $(\mathbf{R}_\alpha)_{\alpha>0}$ is a regularization of $\mathbf{A}^{\mathcal{M}}$.

Proof. We have

$$\begin{aligned}&\left\|(\mathrm{Id}_X + \mathbf{Q}_\alpha \mathbf{N}) \circ \mathbf{B}_\alpha(y^\delta) - \mathbf{A}^{\mathcal{M}} y\right\| \\ &\quad\leq \left\|(\mathrm{Id}_X + \mathbf{P}_{\ker(\mathbf{A})}\mathbf{N}) \circ \mathbf{B}_\alpha(y^\delta) - \mathbf{A}^{\mathcal{M}} y\right\| + \|\mathbf{Q}_\alpha - \mathbf{P}_{\ker(\mathbf{A})}\| \, \|\mathbf{N}\mathbf{B}_\alpha(y^\delta)\| . \end{aligned} \tag{3.5}$$

The claim follows from Theorem 2.4. □

Theorem 3.4. *Let* $\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$ *be a null space network and set* $\mathcal{M} := \mathrm{ran}(\mathbf{L})$. *Let* $\mu > 0$, *suppose* $(g_\alpha)_{\alpha>0}$ *satisfies the assumptions of Theorem 2.8, and let* $(\mathbf{Q}_\alpha)_{\alpha>0}$ *be a family of bounded operators on* X *with* $\|\mathbf{Q}_\alpha - \mathbf{P}_{\ker(\mathbf{A})}\| = \mathcal{O}(\delta^{\frac{2\mu}{2\mu+1}})$. *Consider the regularization* $(\mathbf{R}_\alpha)_{\alpha>0}$ *with*

$$\mathbf{R}_\alpha := (\mathrm{Id}_X + \mathbf{Q}_\alpha \mathbf{N}) \circ g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* . \tag{3.6}$$

Then, the parameter choice $\alpha^\star \asymp (\delta/\rho)^{\frac{2}{2\mu+1}}$ *yields the convergence rate results* $\|\mathbf{R}_{\alpha^\star(\delta,y^\delta)}(y^\delta) - x\| = \mathcal{O}(\delta^{\frac{2\mu}{2\mu+1}})$ *for any* $x \in \mathrm{ran}(\mathbf{L}(\mathbf{A}^*\mathbf{A})^\mu)$.

Proof. Follows from the estimate (3.5) with $\mathbf{B}_\alpha = g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*$ and Theorem 2.8. □

One might use $\mathbf{Q}_\alpha = \mathbf{B}_{\phi(\alpha)}\mathbf{A}$ as a possible approximation to $\mathbf{P}_{\ker(\mathbf{A})^\perp} = \mathbf{A}^+\mathbf{A}$ for some function $\phi\colon [0,\infty) \to [0,\infty)$. In such a situation, one can use existing software packages (for example, for the filtered backprojection algorithm and the discrete Radon transform in case of computed tomography) for evaluating $\mathbf{B}_{\phi(\alpha)}$ and $\mathbf{A}$.

4 Conclusion

In this paper, we introduced the concept of null space networks that have the form $\mathbf{L} = \mathrm{Id}_X + (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$, where $\mathbf{\Phi}$ is any neural network function (for example a deep convolutional neural network) and $\mathrm{Id}_X - \mathbf{A}^+\mathbf{A} = \mathbf{P}_{\ker(\mathbf{A})}$ is the projector onto the kernel of the forward operator $\mathbf{A}\colon X \to Y$ of the inverse problem to be solved. The null space network shares similarity with a residual network that takes the general form $\mathrm{Id}_X + \mathbf{N}$. However, the introduced projector $\mathrm{Id}_X - \mathbf{A}^+\mathbf{A}$ guarantees data consistency which is an important issue when solving inverse problems.

The null space networks are special members of the class of functions $\mathrm{Id}_X + \mathbf{\Phi}$ that satisfy $\mathrm{ran}(\mathbf{\Phi}) \subseteq \ker(\mathbf{A})$. For this class, we introduced the concept of $\mathcal{M}$-generalized inverse $\mathbf{A}^{\mathcal{M}}$ and $\mathcal{M}$-regularization as point-wise approximations of $\mathbf{A}^{\mathcal{M}}$ on $\mathrm{dom}(\mathbf{A}^+)$. We showed that any classical regularization $(\mathbf{B}_\alpha)_{\alpha>0}$ of the Moore-Penrose generalized inverse defines a $\mathcal{M}$-regularization method via $(\mathrm{Id}_X + \mathbf{\Phi})\mathbf{B}_\alpha$. In the case of null space networks where $\mathbf{\Phi} = (\mathrm{Id}_X - \mathbf{A}^+\mathbf{A})\mathbf{N}$, we additionally derived convergence results using only approximation of the projection operator $\mathbf{P}_{\ker(\mathbf{A})}$. Additionally, we derived convergence rates using either exact or approximate projections.

To the best of our knowledge, the obtained convergence and convergence rates are the first regularization results for solving inverse problems with neural networks. Future work has to be done to numerically test the null space networks for typical inverse problems such as limited data problems in CT or deconvolution and compare the performance with standard residual networks, iterative networks or variational networks.

Acknowledgement

The work of M.H and S.A. has been supported by the Austrian Science Fund (FWF), project P 30747-N32.

References

[1] J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl.*, 33(12):124007, 2017.

[2] S. Antholzer, M. Haltmeier, and J. Schwab. Deep learning for photoacoustic tomography from sparse data. *arXiv:1704.04587*, 2017.

[3] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numer.*, 27:1–111, 2018.

[4] J.H.R. Chang, C. Li, B. Poczos, V. Kumar, and A.C. Sankaranarayanan. One network to solve them all – solving linear inverse problems using deep projection models. *arXiv:1703.09912*, 2017.

[5] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang. Low-dose CT via convolutional neural network. *Biomed. Opt. Express*, 8(2):679–694, 2017.

[6] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang. Low-dose ct via convolutional neural network. *Biomed. Opt. Express*, 8(2):679–694, 2017.

[7] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[8] H. Gupta, Kyong H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser. Cnn-based projected gradient descent for consistent ct image reconstruction. *IEEE Trans. Med. Imag.*, 37(6):1440–1453, 2018.

[9] Y. Han, J. J. Yoo, and J. C. Ye. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis, 2016. http://arxiv.org/abs/1611.06391.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.*, 26(9):4509–4522, 2017.

[12] E. Kang, J. Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Med. Phys.*, 44(10), 2017.

[13] B. Kelly, T. P. Matthews, and M. A. Anastasio. Deep learning-guided image reconstruction from incomplete data. *arXiv:1709.00584*, 2017.

[14] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. Variational networks: connecting variational methods and deep learning. In *German Conference on Pattern Recognition*, pages 281–293. Springer, 2017.

[15] Dongwook Lee, Jaejun Yoo, and Jong Chul Ye. Deep residual learning for compressed sensing mri. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 15–18. IEEE, 2017.

[16] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier. Nett: Solving inverse problems with deep neural networks. *arXiv:1803.00092*, 2018.

[17] A. Majumdar. Real-time dynamic mri reconstruction using stacked denoising autoencoder. *arXiv:1503.06383*, 2015.

[18] Y. Rivenson, H. Göröcs, Z.and Günaydin, Y. Zhang, H. Wang, and A. Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017.

[19] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[20] J. Schlemper, J. Caballero, J. V Hajnal, A. N. Price, and D. Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE Trans. Med. Imag.*, 37(2):491–503, 2018.

[21] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[22] G. Wang. A perspective on deep imaging. *IEEE Access*, 4:8914–8924, 2016.

[23] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517, 2016.

[24] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. Deep learning computed tomography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 432–440. Springer, 2016.

[25] H Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. arXiv:1607.08707, 2016.