Nr. 97
12. Sep 2023

universität innsbruck
Institut für Mathematik

Preprint-Series: Department of Mathematics - Applied Mathematics

Data-proximal null-space networks for inverse problems

S. Göppel, J. Frikel, M. Haltmeier

AppliedMathematics

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

Data-proximal null-space networks for inverse problems

Simon Göppel

Department of Mathematics, University of Innsbruck
Technikerstrasse 13, 6020 Innsbruck, Austria
E-mail: simon.goeppel@uibk.ac.at

Jürgen Frikel

Department of Computer Science and Mathematics, OTH Regensburg
Galgenbergstraße 32, 93053 Regensburg, Germany
E-mail: juergen.frikel@oth-regensburg.de

Markus Haltmeier

Department of Mathematics, University of Innsbruck
Technikerstrasse 13, 6020 Innsbruck, Austria
E-mail: markus.haltmeier@uibk.ac.at

September 12, 2023

Abstract

Inverse problems are inherently ill-posed and therefore require regularization techniques to achieve a stable solution. While traditional variational methods have well-established theoretical foundations, recent advances in machine learning based approaches have shown remarkable practical performance. However, the theoretical foundations of learning-based methods in the context of regularization are still underexplored. In this paper, we propose a general framework that addresses the current gap between learning-based methods and regularization strategies. In particular, our approach emphasizes the crucial role of data consistency in the solution of inverse problems and introduces the concept of data-proximal null-space networks as a key component for their solution. We provide a complete convergence analysis by extending the concept of regularizing null-space networks with data proximity in the visual part. We present numerical results for limited-view computed tomography to illustrate the validity of our framework.

Keywords: Regularization, null-space network, data-proximal network, convergence analysis, data consistency

1 Introduction

Inverse problems arise in all kinds of practical applications, such as medical imaging, signal processing, astronomy, computer vision, and more. In this paper, we combine learning-

based methods with established regularization concepts for solving inverse problems. Mathematically, an inverse problem can be expressed as the problem of recovering the unknown $x \in \mathbb{X}$ from noisy data

$$y^\delta = \mathbf{A}x + \eta\,, \tag{1.1}$$

where $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ is a linear operator between separable Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$, $\eta \in \mathbb{Y}$ with $\|\eta^\delta\| \leq \delta$ is the unknown data error, and $\delta \geq 0$ is the known noise bound [10].

1.1 Regularization methods

A major feature of inverse problems is their ill-posedness, so that exact solutions of $\mathbf{A}x = y$ are either not unique or unstable with respect to data perturbations. Non-uniqueness, on the other hand, may even cause $\mathbf{A}^+\mathbf{A}x^\star$ to be different from $x^\star$. To obtain reliable reconstructions, one must use regularization techniques that adopt a stable approach to solving (1.1) and account for instability and non-uniqueness. Regularization methods consist of a family of continuous mappings $\mathbf{R}_\gamma\colon \mathbb{Y} \to \mathbb{X}$ for $\gamma \in \Gamma$ which, together with a suitable parameter choice strategy $\gamma^\star(\delta, y^\delta)$, are convergent in the sense specified in Definition 2.1. Note that, for simplicity, we work with single-valued regularization methods. For the definition of set-valued regularization methods, see [5]. In classical regularization methods, $\Gamma = (0, \infty)$ is a directed interval for which we denote the regularization parameter by α; see [10]. A regularization method is said to be linear if $\mathbf{R}_\alpha$ is linear.

Prominent examples of classical linear regularization methods are Tikhonov regularization and more general spectral filtering methods [10]. A class of non-linear regularization methods is variational regularization, where $\mathbf{R}_\alpha y^\delta$ is a minimizer of the generalized Tikhonov functional

$$\mathcal{T}_\alpha^\delta(x) = \frac{1}{2}\,\|y^\delta - \mathbf{A}x\|^2 + \alpha\mathcal{P}(x)\,. \tag{1.2}$$

Here $\|y^\delta - \mathbf{A}x\|^2/2$ is the data fidelity term that enforces data proximity between $\mathbf{A}x$ and y^δ, while the functional $\mathcal{P}$ incorporates prior information about the underlying signal class. The regularization parameter $\alpha > 0$ acts as a trade-off between proximity to the data and regularity. The regularization approach (1.2) offers great flexibility because it is easily tailored to the forward operator, the underlying signal, and the given perturbations. Common selections for $\mathcal{P}$ include the TV penalty, Sobolev norms, or sparsity priors [30]. Additionally, variational regularization technique has a solid theoretical foundation. In particular, under certain weak additional assumptions, one obtains convergence $\mathbf{R}_\alpha y^\delta \to x^\star$ and data proximity $\mathbf{A}\mathbf{R}_\alpha y^\delta \to \mathbf{A}x^\star$ as $\delta \to 0$.

1.2 Learned reconstructions

Major drawbacks of variational regularization are the challenging design of penalties $\mathcal{P}$ well tuned to the signals of interest, and the time-consuming minimization of (1.2). To overcome these issues, data-driven methods for solving inverse problems have been developed recently [1,2,4,16,20,24,37]. In these methods, a class $(\mathbf{R}_\theta)_{\theta\in\theta}$ of reconstruction operators $\mathbf{R}_\theta\colon \mathbb{Y} \to \mathbb{X}$ is designed to perform well on a class of training data $(x_1, y_1^\delta), \ldots, (x_N, y_N^\delta) \in \mathbb{X} \times \mathbb{Y}$ consisting of pairs of desired reconstructions and noisy data. The class of reconstruction operators should then be both large enough to include reasonable reconstruction methods, and sufficiently constrained to account for the limited amount of training data and computational resources.

Popular architectures for inverse problems are two-step residual networks,

$$\mathbf{R}_\theta(y^\delta) \coloneqq (\mathrm{Id}_\mathbb{X} + \mathbf{W}_\theta) \circ \mathbf{B}_\alpha(y^\delta)\,, \tag{1.3}$$

where $\mathbf{B}_\alpha \colon \mathbb{Y} \to \mathbb{X}$ is an initial classical reconstruction method and $(\mathbf{W}_\theta)_{\theta\in\Theta}$ is an image-to-image architecture such as the U-net [29]. Given the initial reconstructions $z_n^\delta = \mathbf{B}_\alpha(y_n^\delta)$, the network $\mathbf{W}_\theta$ is trained independent of the forward operator, by minimizing the empirical risk $\mathcal{L}_N(\theta) = (1/N) \cdot \sum_{n=1}^N \|x_n - (\mathrm{Id}_\mathbb{X} + \mathbf{W}_\theta)(z_n^\delta)\|^2$. Empirically such approaches have been proven to provide excellent results [3, 15–17, 19, 21, 28]. However, from a regularization point of view, (1.3) lacks theoretical justification. Even if $\mathbf{B}_\alpha$ together with parameter choice $\alpha = \alpha(\delta, y^\delta)$ is a regularization method, convergence of neither $\mathbf{R}_{\alpha,\theta}(y^\delta)$ nor $\mathbf{A} \circ \mathbf{R}_{\alpha,\theta}(y^\delta)$ is granted as $\delta \to 0$. In particular, they even lack data consistency in the sense that there is no control over the proximity between the reconstruction $\mathbf{R}_{\alpha,\theta}(y^\delta)$ and y^δ which limits applicability in safety-critical applications such as medical imaging.

To enforce data consistency several approaches integrating the forward operator into the network architecture have been proposed including variational, iterative networks or network cascades [13, 14, 18, 31, 39]. However such architectures still do not automatically provide theoretical reconstruction guarantees. A strategy to overcome this issue has been presented in [32, 33], where the use of so-called null-space networks has been proposed. Null-space networks are a special form of (1.3) where $\mathbf{W}_\theta$ is restricted to have values only in the kernel of $\mathbf{A}$. It can be taken as $\mathbf{W} = P_{\mathcal{N}(\mathbf{A})}\mathbf{U}_\theta$ where $(\mathbf{U}_\theta)_{\theta\in\Theta}$ is any architecture. In [32] it has been shown that null-space network provable convergent regularization method for (1.1) adapted to the training data. As a drawback, they only modify the initial reconstruction $\mathbf{B}_\alpha(y^\delta)$ on the kernel $\ker(\mathbf{A})$ and keep the part in the complement unchanged. Moreover only linear $\mathbf{B}_\alpha$ have been included in the analysis in [32]. The regularizing networks [33] relaxes the null-space assumption but the design of suitable architectures is less obvious.

1.3 Main contributions

In this paper we propose and analyze an architecture which allows an update of the component in the complement of the null space, but is limited in the data domain by the noise level. The general architecture for which we design a rigorous analysis takes the form of the residual network (1.3) with

$$\mathbf{W}_{\theta,\beta} = P_{\mathcal{N}(\mathbf{A})}\mathbf{U}_\theta + \mathbf{A}^+\Phi_\beta\mathbf{A}\mathbf{V}_\theta\,, \tag{1.4}$$

where $[\mathbf{V}_\theta, \mathbf{W}_\theta]_{\theta\in\Theta}$ is is an image-to-image architecture with two output channels and Φ_β is a function with $\|\Phi_\beta z\| \le \beta$. We will show that data-proximal networks (1.4) together with (1.3) yields a data-proximal regularization method together with convergence rates. In particular we are show rate-r data proximity which we refer to as data-error estimates of the form $\|\mathbf{A}x_\alpha^\delta - y^\delta\| = \mathcal{O}(\delta^r)$. Note the architecture (1.4) in particular uses an explicit decomposition in the null-space $\mathcal{N}(\mathbf{A})$ and its complement $\mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^+)$.

This paper generalizes the regularization results of null-space networks of [32] (which are of the form (1.4) with $\mathbf{V}_\theta = 0$) to data-proximal networks. Furthermore, opposed to [32, 33] our analysis $(\mathbf{B}_\alpha)_{\alpha>0}$ does not have to be linear and for example can be of variational type (1.2). The idea to only learn the null-space components has also been used in [7, 22, 35]. In the finite dimensional setting, learning the null-space and its complement has been proposed in [9]. A regularization approach using approximate null-space networks has been

proposed in [33] without explicitly splitting into null-space component and complements. In contrast, our architecture also allows learning updates in the orthogonal complement of the kernel of $\mathbf{A}$ and has a specific form, which allows to include data consistency easily.

1.4 Outline

The remainder of this paper is organized as follows. In Section 2 we present the theoretical analysis. In particular, we introduce the background and the concept of data-proximal networks, followed by a rigorous convergence analysis. In Section 3 we apply the framework to the limited-view problem in computed tomography. We test the method with FBP and TV regularization as initial reconstruction and compare it with plain null-space learning. The paper ends with section 4 where we give a short summary and discuss some generalizations and lines of potential future research.

2 Theory

Throughout this paper let $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ be a linear bounded operator between separable Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$. We use $\mathcal{N}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})$ to the denote the null-space and range of $\mathbf{A}$, respectively. The inversion of (1.1) is unstable if $\mathcal{R}(\mathbf{A})$ is non-closed and non-unique if $\mathcal{N}(\mathbf{A}) \neq \{0\}$. Our goal is the stable solution of (1.1) in such situations, by combining regularization methods with learned networks.

2.1 Data-proximal regularization

In order to solve (1.1) we use regularization methods with general parameter sets including the classical setting as well as learned reconstruction as special cases.

Definition 2.1 (Regularization method)**.** *Let Γ be an index set, $\mathbb{M} \subseteq \mathbb{X}$, $(\mathbf{R}_\gamma)_{\gamma\in\Gamma}$ a family of continuous mappings $\mathbf{R}_\gamma\colon \mathbb{Y} \to \mathbb{X}$, and $\gamma^\star\colon (0,\infty) \times \mathbb{Y} \to \Gamma$. The pair $((\mathbf{R}_\theta)_{\theta\in\Theta}, \gamma^\star)$ is said to be a (convergent) regularization method for $\mathbf{A}x = y$ over $\mathbb{M}$, if*

$$\forall x \in \mathbb{M}\colon \quad \lim_{\delta\to 0}\Big(\sup\{\|x - \mathbf{R}_{\gamma^\star(\delta,y^\delta)}(y^\delta)\| \mid y^\delta \in \mathbb{Y} \wedge \|y^\delta - \mathbf{A}x\| \leq \delta\}\Big) = 0\,. \tag{2.1}$$

Classical regularization methods use $\Gamma = (0,\infty)$ in which case we denote its elements by α and the parameter choice by $\alpha^\star$. In this situation one usually additionally assumes $\gamma^\star(\delta, y^\delta) \to 0$ uniformly in y^δ as $\delta \to 0$. Many classical methods are further based on the pseudoinverse $\mathbf{A}^+$ where the set of limiting solutions is given by $\mathbb{M} = \mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^+)$; see for example [10]. However also different limiting solutions are frequently used, in particular in variational regularization.

Remark 2.2 (Variational regularization)**.** *The prime example with different limiting solutions is variational regularization* (1.2) *which together with $\alpha, \delta^2/\alpha \to 0$ gives a regularization method over the set of $\mathcal{P}$-minimizing solutions defined by* $\arg\min_x\{\mathcal{P}(x) \mid \mathbf{A}x = y\}$ *with $y \in \mathcal{R}(\mathbf{A})$. This implicitly requires unique minimizers of* (1.2)*. For some regularization methods, $\mathbf{B}_\alpha$ should be taken set-valued and* (2.1) *adjusted accordingly (see [5]). For the sake of simplicity in the presented theory we restrict to the single-valued case.*

We are in particular interested in regularization methods $((\mathbf{R}_\gamma)_{\gamma\in\Gamma}, \gamma^\star)$ that are rate-r data proximal in the following sense.

Definition 2.3 (Data-proximal regularization method)**.** *Let $r \in (0,1]$. A regularization method $((\mathbf{R}_\gamma)_{\gamma\in\Gamma}, \gamma^\star)$ for $\mathbf{A}x = y$ over $\mathbb{M}$ is called rate-r data proximal, if for some $\tau > 0$,*

$$\forall x \in \mathbb{M}\, \forall y^\delta \in \mathbb{Y}\colon \quad \|y^\delta - \mathbf{A}x\| \le \delta \Rightarrow \|y^\delta - \mathbf{A}\mathbf{R}_{\gamma^\star(\delta,y^\delta)}(y^\delta)\| \le \tau\delta^r\,. \tag{2.2}$$

Data proximity of a regularization method seems a reasonable condition as the true solution is known to satisfy the data proximity condition $\|y^\delta - \mathbf{A}x\| \le \delta$. Thus any potential reconstruction without data proximity lacks the only information provided by the noisy data y^δ. Even though this is such an important property we are not aware of an explicit definition in the literature. This may partially be due to the fact that it is automatically satisfied by common regularization methods. For example rate-1 data proximity is satisfied by filter based methods under the source condition $x^+ \in (\mathbf{A}^*\mathbf{A})^\mu$ for any $\mu > 0$ as well as for variational regularization under the source condition $\partial\mathcal{P}(x^+) \in \mathcal{R}(\mathbf{A}^*)$. The following example shows that for filter based methods rate-r data proximity for all $r < 1$ even holds without a source condition.

Example 2.4 (Data proximity without source condition)**.** *Consider a filter based regularization method $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ where $\mathbf{B}_\alpha = g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*$ for filter functions $g_\alpha\colon \mathbb{R} \to \mathbb{R}$ and $\alpha^\star, \delta^2/\alpha^\star \to 0$; see [10] for precise definitions. Then*

$$\begin{aligned}
\|\mathbf{A}\mathbf{B}_\alpha y^\delta - y^\delta\| &= \|(\mathbf{A}g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* - \mathrm{Id}_\mathbb{Y})(y^\delta)\| \\
&\le \|(\mathbf{A}g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* - \mathrm{Id}_\mathbb{Y})(\mathbf{A}x - y^\delta)\| + \|(\mathbf{A}g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* - \mathrm{Id}_\mathbb{Y})\mathbf{A}(x)\| \\
&\le \|\mathbf{A}g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* - \mathrm{Id}_\mathbb{Y}\|\,\delta + \|\mathbf{A}(g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*\mathbf{A} - \mathrm{Id}_\mathbb{X})\|\,\|x\| \\
&\le \|\mathbf{A}g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^* - \mathrm{Id}_\mathbb{Y}\|\,\delta + \alpha^{1/2}\,\|x\|\,,
\end{aligned}$$

where the latter inequality used that the filter has at least qualification $1/2$. Noting that $\|\mathbf{A}(g_\alpha(\mathbf{A}^*\mathbf{A})\mathbf{A}^*\mathbf{A} - \mathrm{Id}_\mathbb{Y})\|$ is bounded for any filter, this shows that for all $r < 1$ and $R > 0$ with $\alpha^\star \asymp \delta^{2r}$ and $\|x\| \le R$ we get $\|\mathbf{A}\mathbf{B}_{\alpha^\star}y^\delta - y^\delta\| \le \tau\delta^r$.

Variational regularization approximates solutions of $\mathbf{A}x = y$ with minimal value of $\mathcal{P}$. In particular for $\mathcal{P} = \|\cdot\|^2/2$ this minimal norm solution is given by the Moore-Penrose inverse $\mathbf{A}^+(y) \in \ker(\mathbf{A})^\perp$. The same holds true for other spectral filtering methods. The concepts of null-space networks [32] addresses potentially suboptimal solution selection by approximating elements in a general set parameterized by $\mathcal{R}(\mathbf{A}^+)$.

Example 2.5 (Null-space networks)**.** *The regularizing null-space networks analyzed in [32] take the form*

$$\forall \alpha > 0\colon \qquad \mathbf{R}_\alpha \coloneqq (\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U}) \circ \mathbf{B}_\alpha\,, \tag{2.3}$$

where $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ is a regularization method with admissible set $\mathcal{R}(\mathbf{A}^+)$, and $\mathbf{U}$ a Lipschitz function. The function $\mathbf{U}$ can be selected from any network architecture $(\mathbf{U}_\theta)_{\theta\in\Theta}$ based on training data $x_1, \dots, x_N$. Except of being Lipschitz, no other assumptions are required from a theoretical point of view. In [32] it is shown that $((\mathbf{R}_\alpha)_{\alpha>0}, \alpha^\star)$ is a regularization method with $\mathbb{M} \coloneqq (\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(\mathcal{R}(\mathbf{A}^+))$. Further, $(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})})\mathbf{U}$ preserves data proximity of $\mathbf{B}_\alpha$ in the sense that $\|\mathbf{A}\mathbf{B}_\alpha(y^\delta) - y^\delta\| \le \tau\delta^r \Rightarrow \|\mathbf{A}\mathbf{R}_\alpha(y^\delta) - y^\delta\| \le \tau\delta^r$.

Data consistency of null-space networks comes at the cost that the component of $\mathbf{B}_\alpha(y^\delta)$ in the range $\mathcal{R}(\mathbf{A}^+)$ remains unchanged by $(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})$. Allowing a network to also act in $\mathcal{R}(\mathbf{A}^+)$ is however beneficial, if the forward operator $\mathbf{A}$ contains many small singular values. How to obtain data consisten regularizations for networks that also act in $\mathcal{R}(\mathbf{A}^+)$ are studied in the present paper.

2.2 Data-proximal networks

Throughout this section assume that we have given a data-proximal, potentially non-linear, regularization method $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$.

Definition 2.6 (Data-proximal null-space networks)**.** *Let $[\mathbf{U}_\theta, \mathbf{V}_\theta]_{\theta\in\Theta}$ be a family of Lipschitz mappings $\mathbf{U}_\theta, \mathbf{V}_\theta \colon \mathbb{X} \to \mathbb{X}$ and $(\Phi_\beta)_{\beta>0}$ a family of mappings $\Phi_\beta \colon \mathcal{R}(\mathbf{A}) \to \mathcal{R}(\mathbf{A})$ such that $\forall \beta > 0\ \forall z \in \mathcal{R}(\mathbf{A}) \colon \|\Phi_\beta y\|_2 \leq \beta$. We call the family of mappings*

$$\mathbf{D}_{\theta,\beta} \coloneqq \mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})} \circ \mathbf{U}_\theta + \mathbf{A}^+ \circ \Phi_\beta \circ \mathbf{A} \circ \mathbf{V}_\theta \tag{2.4}$$

data-proximal null-space network defined by $\mathbf{U}_\theta, \mathbf{V}_\theta, \Phi_\beta, \mathbf{A}$.

Note that for the special case $\mathbf{V}_\theta = 0$ we obtain a null-space network $\mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})} \circ \mathbf{U}_\theta$. The latter obeys strict data consistency in the sense that $\mathbf{A} \circ (\mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})} \circ \mathbf{U}_\theta) = \mathbf{A}$. Data-proximal networks relax the strict data consistency to the data proximity condition

$$\|\mathbf{A} \circ (\mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})} \circ \mathbf{U}_\theta + \mathbf{A}^+ \circ \Phi_\beta \circ \mathbf{A} \circ \mathbf{U}_\theta)(x) - \mathbf{A}x\|_2 = \|(\Phi_\beta \circ \mathbf{A} \circ \mathbf{V}_\theta)(x)\| \leq \beta\,.$$

In particular, if $\|\mathbf{A}x - y^\delta\| \leq \delta$, then $\|\mathbf{A}\mathbf{D}_{\theta,\beta}x - y^\delta\| \leq \delta + \beta$ independent of the selected θ. Any reconstruction method without such an estimate seems unreasonable as $\|\mathbf{A}x - y^\delta\| \leq \delta$ is the information provided by the noise data and therefore should be respected.

Remark 2.7 (Special cases)**.** *The proposed architecture includes many image reconstruction methods as special case:*

- Classical regularization: *With $\mathbf{U}_\theta = \mathbf{V}_\theta = 0$ we have classical regularization $\mathbf{R}_\alpha = \mathbf{B}_\alpha$. For example, in convex variational regularization elements $\mathbf{B}_\alpha$ converge to $\mathcal{P}$-minimizing solutions [30]. No trained network can be included (expect of course learning the regularizer and the regularization parameter).*
- Standard residual networks: *With $\mathbf{U}_\theta = \mathbf{V}_\theta$ and $\Phi_\beta = \mathrm{Id}_{\mathbb{Y}}$ (thus formally using $\beta = \infty$) and $\mathbf{B}_\alpha = \mathbf{A}^+$ we obtain the residual network $\mathbf{R}_\theta = (\mathrm{Id}_{\mathbb{X}} + \mathbf{U}_\theta) \circ \mathbf{A}^+$ of [16, 19]. However it lacks data consistency for which we need $\beta < \infty$.*
- Regularized null-space networks: *With $\mathbf{V}_\theta = 0$, $\mathbf{U}_\theta = \mathbf{U}$ and $\mathbf{B}_\alpha$ a regularization for $\mathbf{A}^+$ we obtain the regularized null-space networks $\mathbf{R}_\alpha = (\mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U}) \circ \mathbf{B}_\alpha$ of [32]. This network architecture does not allow to learn anything orthogonal to the null-space.*
- Range-nullspace decomposition: *With $\Phi_\beta = \mathrm{id}$ and $\mathbf{B}_\alpha = \mathbf{A}^+$ we range-nullspace decomposition $\mathbf{R}_\theta = (\mathrm{Id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U}_\theta + P_{\mathcal{R}(\mathbf{A}^+)}\mathbf{V}_\theta) \circ \mathbf{A}^+$ considered in [9]. This architecture does not include data consistency and moreover $\mathbf{A}^+$ might be unstable.*
- Regularizing networks: *With $\mathbf{U}_\theta = \mathbf{U}_\theta$, proper choice $\theta = \theta(\alpha)$ taking $\mathbf{B}_\alpha$ as filter based regularization and under certain convergence properties, we get the regularizing networks $\mathbf{R}_\alpha = (\mathrm{Id}_{\mathbb{X}} + \mathbf{U}_{\theta(\alpha)}) \circ \mathbf{B}_\alpha$ of [33]. This architecture does not explicitly include data consistency.*

Thus our architecture might be seen as generalization of ones of [9, 33]. Extending [9] we allow $\mathbf{A}^+$ to be replaced by a regularization $\mathbf{B}_\alpha$, include the data proximity function in the range network $P_{\mathcal{R}(\mathbf{A}^+)}\mathbf{V}_\theta$ and allow the ill-posed case. We extend [33] by treating the range and null-space components separately, include the data proximity function in the range network and treat θ as independent parameter in the architecture.

Extending the concept of regularizing null-space networks, our aim is to show that $\mathbf{R}_{\alpha,\beta,\theta} = \mathbf{D}_{\beta,\theta} \circ \mathbf{B}_\alpha$ yields a convergent data consistent regularization method in the sense of Definitions 2.1 and 2.4 with parameter selections $\alpha^\star, \beta^\star, \theta^\star$. Our strategy assuring this is simple. Starting with an rate-r data-proximal regularization method $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ we select $\theta^\star$ and $\beta^\star$ such that convergence is preserved, however to an element different to $x^\star$ selected by a limiting null-space networks. The network $\mathbf{A}^+ \circ \Phi_\beta \circ \mathbf{A} \circ \mathbf{V}_\theta$ is especially relevant in the noisy case in order to obtain improved denoising properties on specific sets and Φ_β is used to preserve data-proximity. Note that the parameter β in data-proximal null-space network directly allows to control the data proximity between any x and $\mathbf{D}_{\theta,\beta}(x)$. Opposed to θ It is not intended to be subject to the training process.

2.3 Convergence analysis

Throughout this section let $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ be a regularization method over $\mathbb{M}$ and $\mathbf{D}_{\theta,\beta}$ be a data-proximal null-space network defined by $\mathbf{U}_\theta, \mathbf{V}_\theta, \Phi_\beta, \mathbf{A}$. The goal is to show that $\mathbf{D}_{\theta,\beta} \circ \mathbf{B}_\alpha$ gives a convergent (data-proximal) regularization method with rates.

Theorem 2.8 (Convergence)**.** *Suppose there a Lipschitz function $\mathbf{U}\colon \mathbb{X} \to \mathbb{X}$ and $\beta^\star = \beta^\star(\delta, y^\delta)$, $\theta^\star = \theta^\star(\delta, y^\delta)$ such that $(\mathbf{D}_{\theta,\beta})_{\theta,\beta}$ are uniformly Lipschitz on bounded sets and*

$$\forall z \in \mathbb{M}\colon \mathbf{D}_{\theta^\star,\beta^\star}(z) \to (\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z) \quad \text{as } \delta \to 0\,. \tag{2.5}$$

Then with $(\mathbf{R}_\gamma)_\gamma \coloneqq (\mathbf{D}_{\theta,\beta} \circ \mathbf{B}_\alpha)_{\alpha,\beta,\theta}$ and $\gamma^\star \coloneqq (\alpha^\star, \beta^\star, \theta^\star)$, the following hold:

1. *$((\mathbf{R}_\gamma)_\gamma, \gamma^\star)$ is a convergent regularization method on $(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(\mathbb{M})$.*
2. *If $\beta^\star = \mathcal{O}(\delta^r)$ then $((\mathbf{R}_\gamma)_\gamma, \gamma^\star)$ is r-rate data proximal, provided $((\mathbf{B}_\alpha)_\alpha, \alpha^\star)$ is.*

Proof. Let $x^* = (\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star)$ with $z^\star \in \mathbb{M}$. Then

$$\begin{aligned}
\|x^\star - \mathbf{R}_\gamma(y^\delta)\| &= \|(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star) - \mathbf{D}_{\beta,\theta}(\mathbf{B}_\alpha y^\delta)\| \\
&\leq \|\mathbf{D}_{\beta,\theta}(\mathbf{B}_\alpha(y^\delta) - z^\star)\| + \|(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star) - \mathbf{D}_{\beta,\theta}(z^\star)\| \\
&\leq L\|\mathbf{B}_\alpha(y^\delta) - z^\star)\| + \|(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star) - \mathbf{D}_{\beta,\theta}(z^\star)\|
\end{aligned}$$

With the convergence of $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ and (2.5) this shows the convergence of $((\mathbf{R}_\gamma)_\gamma, \gamma^\star)$. Now let $((\mathbf{B}_\alpha)_\alpha, \alpha^\star)$ be r-rate data proximal and $\beta^\star = \mathcal{O}(\delta^r)$. By the definition of $\mathbf{R}_\gamma$ we have $\|y^\delta - \mathbf{A}\mathbf{R}_\gamma(y^\delta)\| \leq \|y^\delta - \mathbf{A}(\mathbf{B}_\alpha(y^\delta)\| + \beta$ which gives the r-rate data proximity of $((\mathbf{R}_\gamma)_\gamma, \gamma^\star)$ be □

Under additional assumptions we also obtain convergence rates.

Theorem 2.9 (Convergence Rates)**.** *In the situation of Theorem 2.8, let $((\mathbf{B}_\alpha)_{\alpha>0}, \alpha^\star)$ be rate-r data consistent regularization method over $\mathbb{M}_s \subseteq \mathbb{M}$ that is convergent of rate s. Then under the approximation assumption $\|(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star) - \mathbf{D}_{\beta,\theta}(z^\star)\| = \mathcal{O}(\delta^r)$ on $\mathbb{M}_s$ we have, for $\beta^\star = \mathcal{O}(\delta^r)$ and with $x^\star \in ((\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(\mathbb{M}_s)$ with $\|y^\delta - \mathbf{A}x^\star\| \leq \delta$*

$$\|x^\star - \mathbf{B}_\gamma(y^\delta)\| = \mathcal{O}(\delta^s) \quad \text{as } \delta \to 0\,, \tag{2.6}$$

$$\|y^\delta - \mathbf{A}\mathbf{R}_\gamma(y^\delta)\| = \mathcal{O}(\delta^r) \quad \text{as } \delta \to 0\,. \tag{2.7}$$

That is, the regularization method $((\mathbf{R}_\gamma)_\gamma, \gamma^\star)$ is rate-r data proximal and rate-s convergent on $(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(\mathbb{M}_s)$.

Proof. Condition (2.7) follows from Theorem 2.8. Moreover according to the proof of the theorem we have $\|x^\star - \mathbf{R}_\gamma(y^\delta)\| \leq L\|\mathbf{B}_\alpha(y^\delta) - z^\star)\| + \|(\mathrm{Id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U})(z^\star) - \mathbf{D}_{\beta,\theta}(z^\star)\|$. This gives the claim by the parameter choice and the made approximation assumption. □

3 Application

In this section we present a numerical example for our proposed data-proximal regularization approach. We consider limited angle computed tomography (CT) modeled by the Radon transform as forward problem.

3.1 The Radon transform

The Radon transform of a compactly supported smooth function $u\colon \mathbb{R}^2 \to \mathbb{R}$ is defined by $\mathbf{K}x(\theta, s) \coloneqq \int_{L(\theta,s)} u(x)\,\mathrm{d}L(x)$ for $(\theta, s) \in [-\pi/2, \pi/2) \times \mathbb{R}$. Here $L(\theta, s) \coloneqq \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1\cos(\theta) + x_2\sin(\theta) = s\}$ denotes the line in $\mathbb{R}^2$ with singed distance $s \in \mathbb{R}$ from the origin and direction $(\cos(\theta), \sin(\theta))^T$ with $\theta \in [-\pi/2, \pi/2)$. In limited angle CT, the data is only known within a limited subset $\Omega \subseteq [-\pi/2, \pi/2)$ of the full angular range. The limited angle Radon transform is then defined as

$$\mathbf{K}_\Omega\colon \mathcal{D}(\mathbf{K}_\Omega) \subseteq L^2(\mathbb{R}^2) \to L^2(S^1 \times \mathbb{R})\colon u \mapsto \chi_{\Omega\times\mathbb{R}}\mathbf{K}u.$$

The well known filtered back-projection (FBP) inversion formula for the full data Radon transform reads $u = \mathbf{K}^*\mathbf{I}(\mathbf{K}u)$, where $\mathbf{I}$ is the so-called Riesz-potential and defined in the Fourier domain by $\mathbf{F}_2(\mathbf{I}u) \coloneqq \|\cdot\|(\mathbf{F}_2u)/(4\pi)$ where $\mathbf{F}_2$ is the Fourier transform in the second component; see [23]. The application of the FBP formula to limited angular data is known to cause prominent streak artifacts which can obscure important information [26,27]. While these artifacts have been characterized by methods from microlocal analysis [6, 11, 12], finding suitable reconstruction strategies is still an ongoing challenge. Thus, we will employ our proposed data-proximal null-space network to obtain a reliable and data-proximal reconstruction.

In our simulations we use synthetic Shepp-Logan type phantoms supported within the ball of radius one where u is represented by discrete image $x \in \mathbb{R}^{N\times N}$ with $N = 128$. To obtain a discretized versions for the forward operator, we evaluate the limited angle Radon transform at $N_s = 128$ equidistant distances in $[-1, 1]$ and $N_\Omega = 120$ equidistant angles in $[-\pi/3, \pi/3)$. More details on the implementation of the discretized version of the Radon transform which we used in our experiments can be found in the repository `https://github.com/drgHannah/Radon-Transformation`. The discretized limited angle Radon transform and FBP formula are denoted by $\mathbf{A}$ and $\mathbf{A}^\sharp$, respectively.

3.2 Network design and training

Throughout all numerical calculations, the networks architectures for $\mathbf{U}_\theta$ and $\mathbf{V}_\theta$ are taken as the basic U-net [29], which is still considered a state-of-the-art model due to its ability to reliably learn image features. Based on the U-net we then consider the architectures

$$\begin{aligned}\mathbf{M}_\theta^{(1)} &= \mathrm{id}_\mathbb{X} + \mathbf{U}_\theta\,,\\ \mathbf{M}_\theta^{(2)} &= \mathrm{id}_\mathbb{X} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U}_\theta\,,\end{aligned}$$

$$\mathbf{M}_\theta^{(3)} = \mathrm{id}_{\mathbb{X}} + P_{\mathcal{N}(\mathbf{A})}\mathbf{U}_\theta + \mathbf{A}^\sharp \Phi_\beta \mathbf{A}\mathbf{V}_\theta \,,$$

where $\mathbf{M}_\theta^{(1)}$ is the plain residual U-net, $\mathbf{M}_\theta^{(2)}$ the null-space architecture and $\mathbf{M}_\theta^{(3)}$ the prosed data-proximal null-space network. The data-proximal network uses shared weights and divides the output in two streams via projections onto the kernel and the orthogonal complement, respectively.

The data-proximity function is taken as radial function

$$\Phi_\beta(x) \coloneqq \begin{cases} x, & \|x\|_2 \leq \beta, \\ \beta \cdot x/\|x\|_2, & \text{else}\,. \end{cases} \tag{3.1}$$

In the numerical simulations we use $\beta \coloneqq \delta \sum_{i=1}^N \|\eta_i\|_2/N$. This way, we obtain an estimate of the magnitude of the perturbations present in the data domain.

As initial reconstruction method we use the FBP operator as well as total-variation (TV) regularization, which is known to be a good prior for the missing data setup [25, 34, 36, 38]. The family $(\mathbf{B}_\alpha)_{\alpha>0}$ is then given by $\mathbf{B}_\alpha y^\delta \coloneqq \arg\min_x \|\mathbf{A}x - y^\delta\|_2^2/2 + \alpha\|\nabla x\|_1$ and numerically solved with the Chambolle-Pock algorithm [8].

For training the networks we generate data pairs $(x_i, y_i^\delta)_{i=1}^{600}$ with $y_i^\delta = \mathbf{A}x_i + \delta\eta_i$ where $\delta = 0.05$ and $\eta_i \sim \|\mathbf{A}x_i\|_\infty \cdot \mathcal{N}(0,1)$. All networks $\mathbf{M}_\theta^{(i)}$ are trained by minimizing

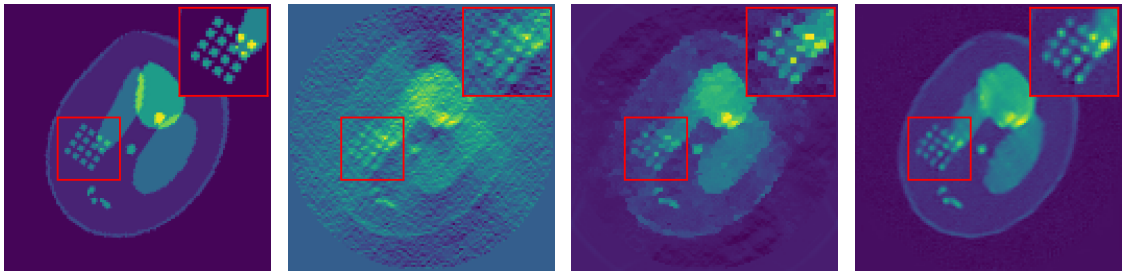
$$\mathcal{L}_N(\theta) \coloneqq \frac{1}{N}\sum_{i=1}^N \|\mathbf{M}_\theta \mathbf{B}_\theta(y_i^\delta) - x_i\|_2^2 \tag{3.2}$$

by the Adam optimizer with learning rate of 0.001. We trained each network for a total 50 epochs, and chose the learned network parameters with minimal validation error during training as our final network weights. We split our dataset into 500 training and 100 test samples. For further implementation details regarding our experiments we refer to our github repository `https://github.com/sgoep/data_proximal_networks`.

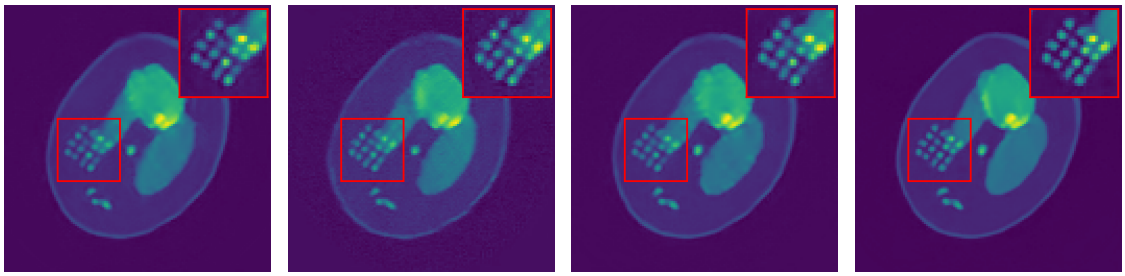
3.3 Results

For the presented results we write $x^\star$ for the ground truth image, x_{FBP} for the FBP reconstruction, x_{TV} for the TV-regularized solution and add the superscripts RES, NSN, DP for subsequent residual network, null-space network and data-proximal network, respectively. All reconstructions are compared to the ground truth via the mean squared error (MSE), the peak-signal-to-noise-ratio (PSNR) and the structural similarity index measure (SSIM).

Reconstruction results are shown in Figure 3.1. We see that all data-driven modalities overall yield rather good results. Looking closely at the fine grid like features in the magnified section, we can observe that reconstructions shown in Figure 3.1d-3.1g tend to differ in their extent of their expression. However, these details appear to be recovered more accurately by our proposed data proximity approach as shown in Figure 3.1h. Here, all dots are of similar intensity and shape. Furthermore, the intersecting part of the upper and left bigger ellipse like features inside the phantom are recovered more precisely. We attribute these improvements to our data-proximal architecture, which with which the output of the residual network is constraint to the correct energy level and able to converge faster to a suitable solution. A quantitative error comparison is shown in Table 1. We see that our proposed data proximity reconstruction performs best in the chosen metrics. This is in line with our visual inspection above.



(a) $x^\star$ (b) x_{FBP} (c) x_{TV} (d) $x_{\mathrm{FBP}}^{\mathrm{RES}}$



(e) $x_{\mathrm{TV}}^{\mathrm{RES}}$ (f) $x_{\mathrm{FBP}}^{\mathrm{NSN}}$ (g) $x_{\mathrm{TV}}^{\mathrm{NSN}}$ (h) $x_{\mathrm{TV}}^{\mathrm{DP}}$

Figure 3.1: Exemplary reconstructions from limited angle data for the methods described above. All data-driven approaches give visually good results. Looking closely at the figure 3.1h, we see that our proposed method is able to recover the fine details in the magnified region with the highest accuracy.

Method	MSE	PSNR	SSIM
x_{FBP}	0.0137	24.6556	0.2867
x_{TV}	0.0020	33.0772	0.6089
$x_{\mathrm{FBP}}^{\mathrm{RES}}$	0.0013	34.9414	0.8455
$x_{\mathrm{TV}}^{\mathrm{RES}}$	0.0010	35.7354	0.9032
$x_{\mathrm{FBP}}^{\mathrm{NSN}}$	0.0012	35.2030	0.8437
$x_{\mathrm{TV}}^{\mathrm{NSN}}$	0.0009	36.6717	0.9184
$x_{\mathrm{TV}}^{\mathrm{DP}}$	**0.0008**	**37.1900**	**0.9265**

Table 1: Reconstruction errors for CT reconstruction with limited angular range. The best values in each column are highlighted in bold.

4 Conclusion

In this paper, we have introduced a provably convergent data-driven regularization strategy in terms of data-proximal networks. We have demonstrated improved reconstruction properties in our numerical experiments. These experiments were performed on synthetic phantoms and for the parallel beam geometry of the Radon transform. In particular, data were generated and the noise model is explicitly known. Future work could focus on real world applications. It is possible to combine our approach with appropriate noise estimation techniques and different data proximity functions. More precise adaptation can be achieved by designing more problem-specific data proximity functions of a certain regularity. Analysis under random noise also appears to be an interesting line of research.

Acknowledgement

The contribution by S.G. is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 847476. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

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

[2] F. Altekrüger, A. Denker, P. Hagemann, J. Hertrich, P. Maass, and G. Steidl. PatchNR: learning from very few images by patch normalizing flow regularization. *Inverse Problems*, 39(6):064006, 2023.

[3] S. Antholzer, M. Haltmeier, and J. Schwab. Deep learning for photoacoustic tomography from sparse data. *Inverse Problems in Science and Engineering*, 27(7):987–1005, 2019. PMID: 31057659.

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

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

[6] L. Borg, J. S. Jørgensen, J. Frikel, and E. T. Quinto. Analyzing reconstruction artifacts from arbitrary incomplete x-ray ct data. *SIAM J. Imaging Sciences*, 11(4):2786–2814, 2018.

[7] T. A. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan. Learning the invisible: a hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6):064002, 2019.

[8] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[9] D. Chen and M. E. Davies. Deep decomposition learning for inverse imaging problems. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16*, pages 510–526. Springer, 2020.

[10] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[11] J. Frikel and E. T. Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Problems*, 29(12):125007, 2013.

[12] J. Frikel and E. T. Quinto. Limited data problems for the generalized radon transform in $\mathbb{R}^n$. *SIAM Journal on Mathematical Analysis*, 48(4):2301–2318, 2016.

[13] M. Genzel, J. Macdonald, and M. März. Solving inverse problems with deep neural networks – robustness included? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1119–1134, 2023.

[14] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

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

[17] E. Kang, J. Min, and J. Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Medical Physics*, 44(10):e360–e375, 2017.

[18] A. Kofler, M. Haltmeier, C. Kolbitsch, M. Kachelrieß, and M. Dewey. A u-nets cascade for sparse view computed tomography. In F. Knoll, A. Maier, and D. Rueckert, editors, *Machine Learning for Medical Image Reconstruction*, pages 91–99, Cham, 2018. Springer International Publishing.

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

[20] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier. NETT: solving inverse problems with deep neural networks. *Inverse Problems*, 36(6):065005, 2020.

[21] A. Majumdar. Real-time dynamic mri reconstruction using stacked denoising autoencoder, 2015.

[22] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly. Deep generative adversarial neural networks for compressive sensing mri. *IEEE transactions on medical imaging*, 38(1):167–179, 2018.

[23] F. Natterer. *The mathematics of computerized tomography*. SIAM, 2001.

[24] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

[25] M. Persson, D. Bone, and H. Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine & Biology*, 46(3):853, 2001.

[26] E. T. Quinto. Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. *SIAM J. Math. Anal.*, 24(5):1215–1225, 1993.

[27] E. T. Quinto. Artifacts and visible singularities in limited data x-ray tomography. *Sens. Imaging*, 18(1):1–14, 2017.

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

[29] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[30] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*. Springer, New York, 2009.

[31] 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 Transactions on Medical Imaging*, 37(2):491–503, 2018.

[32] J. Schwab, S. Antholzer, and M. Haltmeier. Deep null space learning for inverse problems: convergence analysis and rates. *Inverse Problems*, 35(2):025008, 2019.

[33] J. Schwab, S. Antholzer, and M. Haltmeier. Big in Japan: Regularizing networks for solving inverse problems. *Journal of Mathematical Imaging and Vision*, 62(3):445–455, 2020.

[34] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology*, 53(17):4777–4807, 2008.

[35] C. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár. Amortised map inference for image super-resolution. In *International Conference on Learning Representations*, 2017.

[36] J. Velikina, S. Leng, and G.-H. Chen. Limited view angle tomographic image reconstruction via total variation minimization. *Proc SPIE*, 6510:709–720, 2007.

[37] G. Wang, J. C. Ye, and B. De Man. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2(12):737–748, 2020.

[38] T. Wang, K. Nakamoto, H. Zhang, and H. Liu. Reweighted anisotropic total variation minimization for limited-angle ct reconstruction. *IEEE Transactions on Nuclear Science*, PP:2742–2760, 2017.

[39] G. Yiasemis, J.-J. Sonke, C. Sánchez, and J. Teuwen. Recurrent variational network: A deep learning inverse problem solver applied to the task of accelerated mri reconstruction, 2022.