# On existence, stability, accuracy and learning of approximate decoders for ill-posed inverse problems

Nina Maria Gottschling


University of Cambridge
Department of Applied Mathematics and Theoretical Physics
Peterhouse College

April 2022

This dissertation is submitted for
the degree of Doctor of Philosophy

# Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any dissertation that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. Parts of this thesis are based on work in a series of articles to appear. In particular,

- Chapter 2 is based on my work in the manuscript, 'On non-robustness, hallucinations and unpredictability in AI for imaging', undergoing final revisions in *Nature*.
- Chapter 3 is based on my work in [83] and currently being revised titled 'The troublesome kernel - on AI generated hallucinations and the accuracy-stability trade-off in inverse problems' for submission to *SIAM Review*.
- Chapter 4 is based on my work in a preprint, 'Accuracy bounds and approximate decoders for ill-posed inverse problems', being prepared for submission to *FOCM*.

# On existence, stability, accuracy and learning of approximate decoders for ill-posed inverse problems

Nina Maria Gottschling

## Summary

Artificial intelligence (AI) methods are changing the sciences and scientific computing, in particular also in the field of inverse problems. In inverse problems, for example in imaging AI based methods seemingly achieve higher reconstruction accuracy than standard methods, such as compressed sensing. However, trustworthiness has become a serious issue as there is empirical evidence that deep learning (DL) may lead to unstable methods in inverse problems. Recently, in inverse problems in imaging, another phenomenon of DL decoders yielding false yet realistic looking artefacts, coined AI generated hallucinations, has been reported on. This thesis explores the use of DL in inverse problems and aims at providing a theoretical basis for assessing the stability and accuracy of such methods. In the second chapter, a fully learned neural network approach for image reconstruction, which was introduced in [192] and coined 'automated transform by manifold approximation', is examined. In particular, its potential benefits with respect to accuracy and disadvantages with respect to stability and robustness compared to standard methods for image reconstruction are investigated. We show that without further conditions on the sampling operator, such fully learned approaches to solving inverse problems become unstable. In the third chapter, we present a comprehensive mathematical analysis explaining different causes of AI generated hallucinations and the links to instabilities. Our results establish four crucial issues for AI methods in inverse problems. Firstly, overly accurate AI methods will wrongly transfer details from one image to another reconstructed image creating a hallucination. Secondly, there is an accuracy-hallucination trade-off. Thirdly, there is an accuracy-stability trade-off, and optimising these trade-offs through standard training processes is difficult. And lastly, hallucinations can occur due to any kind of noise model and probability distribution used on the training set. In the last chapter, we investigate how DL based methods for solving inverse problems can perform better than standard methods. Thus, we establish fundamental accuracy bounds for solving ill-posed inverse problems. This is achieved by obtaining upper and lower bounds on a universal optimality constant, that includes the best worst-case noise, the average and the statistical reconstruction error for the reconstruction of an ill-posed inverse problem. This framework encompasses non-linear and linear inverse problems with different noise models and allows to assess stability, accuracy and learning of approximate decoders for ill-posed inverse problems.

# Acknowledgments

Firstly, I would like to thank my supervisor Professor Anders C. Hansen who suggested the initial research project that led to this thesis and for his encouragement and insight in the fields of applied and pure mathematics. In particular, I am very grateful for his generosity with his time and his useful advice and vision in discovering the mathematics of deep learning. I very much enjoyed being his student while doing my PhD and look forward to many more years of friendship and collaboration. Secondly, I gratefully acknowledge the financial support of the Engineering and Physical Sciences Research Council under the grant EP/L016516/1 for the Cambridge Centre for Analysis. Thirdly, I would like to thank my collaborators. Vegard Antun for countless fruitful and insightful discussions on various topics and his numerous experiments yielding insight and providing evidence for our results. Paolo Campodonico for helpful discussions and comments. Ben Adcock for his guidance in academic insight and writing advice and comments. Ronald DeVore for interesting discussions and his great work in approximation theory which has been very inspiring for me. Moreover, I would like to thank Laura Thesing and Simon Becker for their guidance throughout my time as a PhD student and their kind welcome to the Applied Functional and Harmonic Analysis group under Anders. Lastly, I would like to thank everyone who supported me during my time as a PhD student and my friends for their support and listening to me speaking about my PhD, even if as Aleksandra they are studying something completely different.

# Contents

# Chapter 1

# Introduction

In the last years, the importance and impact of deep learning (DL), neural networks (NNs) and artificial intelligence (AI) have risen in many areas. These methods have entered our daily lives, impacting computer vision, driving assistance and natural language processing and translations. For example, DeepL is claimed to be *"[the] world's most accurate translator"* on its website, in 2022. However, findings suggest that DL and AI methods in various areas are prone to instabilities and artefacts, for example, in image reconstruction and natural language processing [6, 14, 16, 43, 45, 97, 99, 157]. These findings give rise to the need for foundations of AI that describe its methodological limitations. This realisation is now becoming increasingly apparent:

> *"2021 was the year in which the wonders of artificial intelligence stopped being a story [...] Many of this year's top articles grappled with the limits of deep learning (today's dominant strand of AI) and spotlighted researchers seeking new paths."*
>
> – From "7 Revealing Ways AIs Fail: Neural Networks can be Disastrously Brittle, Forgetful, and Surprisingly Bad at Math" (Dec. 2021) [43].

The large impact of AI and DL in various areas has sparked concern within legal frameworks for the use of AI in the European sphere:

> *"In the light of the recent advances in artificial intelligence (AI), the serious negative consequences of its use for EU citizens and organisations have led to multiple initiatives from the European Commission to set up the principles of a trustworthy and secure AI. Among the identified requirements, the concepts of robustness and explainability of AI systems have emerged as key elements for a future regulation of this technology."* – Europ. Comm. JCR Tech. Rep. (Jan. 2020).

The European Commission has stated that AI methods should provide high levels of robustness, security and accuracy, in a statement concerning the outline for legal AI (April -21). DL and AI based methods not only impact our daily lives, but also current research. This has led to

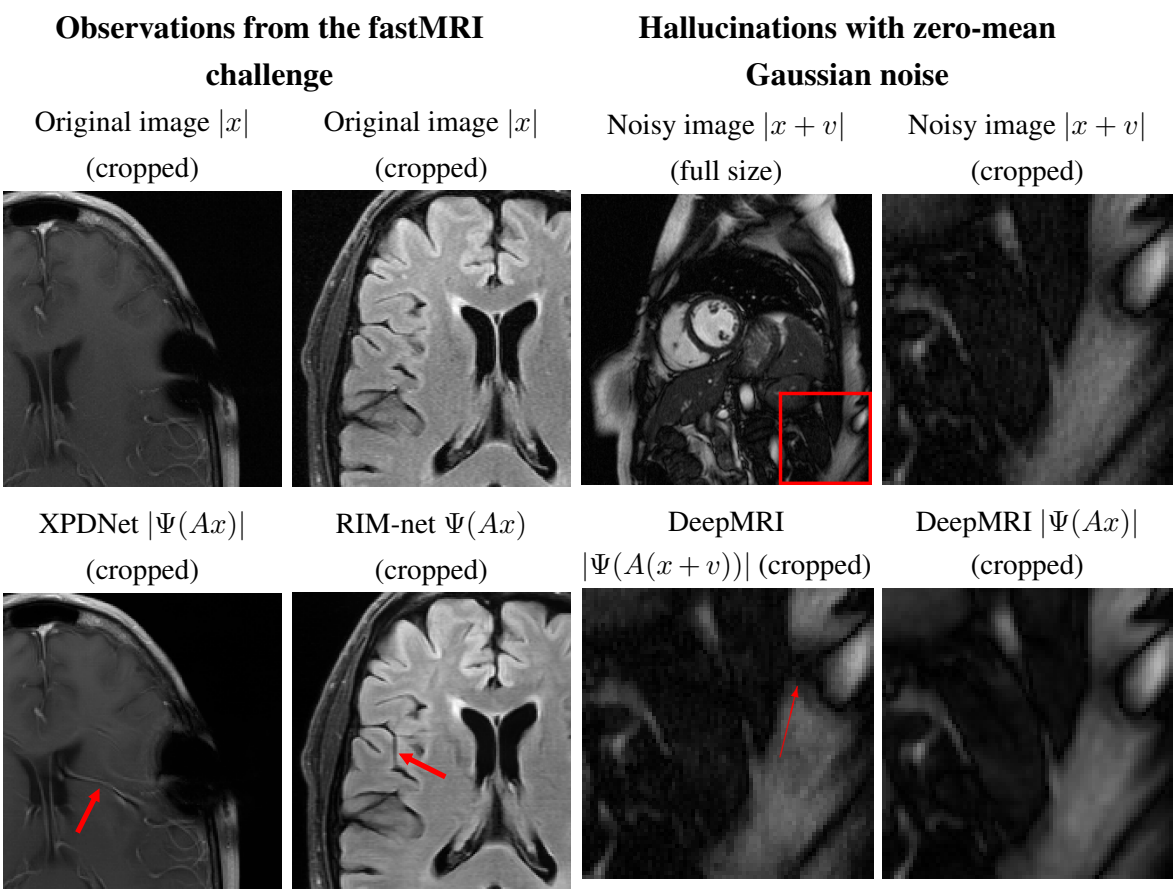


Figure 1.1: **DL methods for image reconstruction may hallucinate or exhibit instabilities.** The figure shows how two high performing neural networks [125, 155] from the fastMRI challenge [138] hallucinate by adding realistic-looking features (not present in the original image) into the reconstructed image. The same effect can be seen when adding zero-mean random Gaussian noise $v$ in the image $x$, which yields Gaussian noise in the measurements. Here the DeepMRI network [161] introduces a hallucination due to instability, and the hallucination is not present without the added noise.

empirical evidence [43, 97] that modern AI is often not robust, as in unstable and produces AI generated hallucinations, and can thus produce nonsensical output with high prediction confidence. The most tangible examples of AI generated instabilities and hallucinations can be seen in image reconstruction and natural language processing. In computational linguistics that is concerned with translating information, which may only be machine understandable, to information that is understandable for humans:

> *"[...] state-of-art neural models include misleading statements - usually called hallucinations - in their outputs."* – From "Controlling hallucinations at word level in data-to-text generation" (2022) [157].

In image reconstruction, AI generated hallucinations are generally regarded as:

> *"[...] hallucinatory features [that] are not acceptable and especially problematic if they mimic normal structures that are either not present or actually abnormal.*

> *Neural network models can be unstable as demonstrated via adversarial perturbation studies [6].”* – Evaluation of the Facebook fastMRI challenge (2020) [138].

Fig. 1.1, shows examples of AI hallucinations and instabilities in DL methods for image reconstruction. Despite these pitfalls AI and DL based methods have appeal. There is a vast potential of DL and AI in scientific computing which is emphasized by many classical approximation theorems. In [54], it is shown that any continuous function can be approximated arbitrarily well by a NN. However, scientific computing is based on stability and accuracy [98], and often there is a trade-off between the two [46]. In particular, there may be barriers preventing the existence of simultaneously very accurate and very stable algorithms. These issues lead to the list of Smale's mathematical problems for the 21st century. Namely,

> Smale's 18th problem: *“Limits of intelligence. What are the limits of intelligence, both artificial and human?”* – From the list of mathematical problems for the 21st century (1998) [168].

In Smale's 18th problem, the question is posed what role learning and problem-solving play together with the role of mathematics. In this thesis we only consider a fraction of the first part of the problem. Namely, the limits of AI used to solve inverse problems. More specifically, the aim of this thesis is to assess the limits of accuracy and stability of AI and DL applied to ill-posed inverse problems and investigate the potential benefits and difficulties these methods present.

## 1.1 Overview of ill-posed inverse problems

In the following section, we give define the main problem investigated in this thesis with introduction to ill-posed inverse problems. Generally an inverse problem is formalized as solving an equation of the form

$$y = A(x) + e, \tag{1.1}$$

where $y \in Y$ is the measurement data, which are assumed to be given by noisy measurements obtained by a map $A : X \to Y$, from the object that is to be recovered $x \in X$. The map $A$ is typically referred to as the forward operator, and in the following we will mainly consider linear forward operators, hence, linear inverse problems. Moreover, $e \in Y$ is the measurement noise that can be modelled as a random variable or deterministic noise. For example, $Y$ and $X$ can be infinite dimensional spaces or also finite dimensional metric spaces.

There are various areas in which inverse problems arise, such as MRI, parallel MRI [53], some instances of fluorescence microscopy [117], structured illumination in temporal compressive microscopy [187], computer tomography (CT) [166] and positron emission tomography (PET)

scans [160]. Further details are outlined in Section 1.5. Of particular interest to ongoing research are ill-posed inverse problems, which are by definition difficult to solve.

We follow the notion of ill-posedness postulated by Hadamard [90, 91]. An inverse problem, as in 1.1, is ill-posed if one of the following criteria is not satisfied. Otherwise it is referred to as well-posed.

(1) **Existence:** foy $y \in Y$ and $e \in Y$, there exists a solution $x \in X$ such that $y = Ax + e$.

(2) **Uniqueness:** the solution in $(1)$ is unique.

(3) **Stability:** the solution in $(1)$ depends continuously on $y$.

According to Hadamard ill-posed problems should be modelled differently in order to make them well-posed. Concerning well-posed inverse problems, depending on the forward operator $A$, there also exist methods of direct inversion, in the case of infinite dimensional spaces $X$ and $Y$. Yet, when discretizing and subsampling the data obtained by a discretized version of the inverse problem, it often becomes ill-posed. Such ill-posed problems do arise in relevant mathematical problems, as shown by Calderon and Zygmund [26, 27]. Moreover, in 1955 and 1985 it was shown that there exists a numerical solution to specific ill-posed problems [107, 108].

An approach to describing an ill-posed inverse problem with noise, as in (1.1) where $(X, d_X)$ and $(Y, d_Y)$ are metric spaces, is with respect to the set $F_y^e := \{x \in X : d_Y(A(x), y) \leq d_Y(e, 0)\}$ given $y, e \in Y$. This is an unbounded set, for $A$ linear with a non-trivial null space and a $X$ linear unbounded vector space or for $A$ a continuous operator with non-closed range [8]. With respect to point $(2)$, this means that there exists a solution, yet it may not be unique and there might be an unbounded set of different possible solutions to (1.1).

The aim of solving the inverse problem is to obtain a accurate, stable and robust decoder

$$\varphi : Y \to X.$$

For simplicity we assume that the noise level is bounded by $\epsilon \geq 0$, meaning that $d_y(e, 0) \leq \epsilon$. If (1.1) is ill-posed, in the sense that the set $\{x \in X : d_Y(A(x), y) \leq \epsilon\}$ given $y \in Y$ is unbounded [8], then, it may not be possible to achieve exact reconstruction. Hence, one may try to to minimise the reconstruction error $d_X(\varphi(y), x)$ for a subset of elements $(y, x)$ of $Y \times X$ in order to solve (1.1). This distance also determines the *accuracy* of the decoder.

**Accuracy of a decoder**

A decoder $\varphi : Y \to X$ is considered to be *accurate*, if $d_X(\varphi(y), x)$ is zero or relatively small.

With *stable*, we refer to the decoder $\varphi$ yielding a continuous approximation to a vector in $X$ from perturbed measurements $y \in Y$. Note that the definition of stability varies depending on

the context. For example, in [74] stability is defined as the decoder yielding a reconstruction that has an error controlled by the distance to sparse vectors. This notion of stability is also referred to as instance optimality, [44]. In the following, we use the notion of stability below.

**Stability of a decoder**

A decoder $\varphi : Y \to X$ is considered to be *stable*, if its Lipschitz constant is relatively small and bounded.

The *robustness* is determined by the continuity of the decoder $\varphi$ and thus closely related to its stability. This is determined by the distance $d_X(\varphi(y), x)$, where $\varphi(y) \in X$ for $y \in Y$ and $x \in X$. More, specifically in [74] robustness is defined by the distance $d_X(\varphi(y), x)$ being bound from above by the noise level, which in [74] is referred to as the measurement error $d_y(Ax, y)$.

For choosing a specific solution to (1.1) from or near the set $F_y^e$, there exist different approaches which will be summarised in the following subsections. These include variational methods 1.3.2, Tikhonov regularization 1.3.2, Bayesian formulation 1.3.2, iterative reconstruction 1.3.2 and sparsity based reconstruction 1.3.3. Of course different methods choose a solution with different properties to (1.1) from the set $F_y^e$. These properties of the different solutions will be outlined with respect to the main topic of this thesis. Namely, can learned reconstruction methods for ill-posed inverse problems perform in some sense better than the above mentioned classical methods, and, if so, to what extent? For example, better performance with respect to accuracy, in many recent publications learning based methods have shown to obtain superior accuracy compared to standard methods. Examples, where DL is currently used in research, include numerical PDEs [60, 185], discovering PDE dynamics [158], uncertainty quantification and high-dimensional approximation [162]. However, as mentioned a concern regarding learning based methods is the lack of a theoretical framework with precise error bounds such as they exist for standard methods [8]. Moreover, learning based methods have shown to produce severe artefacts, and AI hallucinations, that appear realistic in the absence of contradicting information. Such hallucinations and instabilities have become a serious issue, for example, in the fastMRI challenge [68, 138, 188] and in other applications [14, 16, 99]. The necessity for further research for DL used in MRI is outlined in [41].

## 1.2 Structure of the thesis

The first chapter of this thesis gives a short summary of standard methods for solving ill-posed inverse problems with a focus on image reconstruction. The second chapter is concerned with a specific learning based approach for image reconstruction, which was introduced in [192] and coined 'automated transform by manifold approximation'. This approach is a fully learned neural network and we highlight its potential benefits and potential disadvantages compared to standard methods for image reconstruction. The third chapter of this thesis is concerned with

finding sufficient mathematical conditions for AI hallucinations to occur and how to protect against them. The fourth chapter, then, presents a more extensive analysis of both linear and non-linear inverse problems, equipped with any kind of noise model, including multiplicative noise. It relies on an approximation theoretic framework and aims at finding the solution to (1.1) in the set $F_y^e$ or an approximation to it that obtains the smallest reconstruction error. For example, the decoders that obtain the pointwise best worst-case noise reconstruction error and the best average reconstruction error are derived. Furthermore, we provide lower and upper bounds to these errors in order to provide a possible theoretical guideline as to what and how much learned methods can improve standard methods. Lastly, each chapter is written to be self-contained and, hence, the necessary notation is introduced at the beginning of each chapter.

## 1.3 Classical image reconstruction as an inverse problem

In the subsequent sections, we consider linear forward operators $A : X \to Y$. Concerning classical image reconstruction, there are two main approaches. Firstly, direct methods that are derived in the continuous domain, where $X$ and $Y$ are function spaces. An example is the filtered back projection (FBP) algorithm for X-ray computed tomography (CT), which is stated in Section 1.3.1. Secondly, there are variational methods, Section 1.3.2, which obtain a decoder for (1.1) by minimizing an objective functional. These methods can also be referred to as functional analytic inversion [8]. Examples of variational methods include regularized methods such as Tikhonov regularization, Section 1.3.2, and iterative methods, Section 1.3.2. Variational methods are often more robust than direct methods, according to [132].

### 1.3.1 Direct inversion

Different imaging models, such as magnetic resonance imaging, Brightfield microscopy or X-ray computed tomography can be formalized for $X$ and $Y$ being infinite dimensional vector spaces and by defining $A$ as a composition of linear operators, as nicely summarised in [132]. Then, without the presence of noise, which means letting $e = 0$ in (1.1), these can be directly and analytically inverted. An example of such a direct inversion is the FBP algorithm for X-ray CT, initially formalized in 1971 [154], for further details see for instance Example 7 [132]. Many of these direct inversion theorems are based on the Fourier transform being invertible:

In the case that $X = Y = L^2(\mathbb{R}^d)$ where $\mathbb{R}^d$ is equipped with the Lebesgue measure $\mu$ and that $A : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is the Fourier transform, it is formally given by

$$A(f)(\omega) = \int f(x) \exp^{-i\langle \omega, x \rangle} \, d\mu(x),$$

where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product and $i$ the imaginary unit. Note that initially the Fourier transform can be defined on $L^1(\mathbb{R}^d)$ and then uniquely extended to $L^2(\mathbb{R}^d)$ by

Plancherel's Theorem, Theorem 5.3 [124]. The inverse Fourier transform, Theorem 5.5 [124], is then given by, $A^{-1} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$A^{-1}(g)(\omega) = \frac{1}{(2\pi)^d} \int g(\omega) \exp^{i\langle \omega, x \rangle} \, d\mu(x).$$

Evidently, the inverse Fourier transform involves an integral over a function. In practice this would correspond to taking infinitely many samples. Yet, in most cases this is often not even approximately possible due to various constraints, a good example is medical imaging. Usually, there are only a finite amount of measurement points available. Thus, the continuous case is discretized. The integral is replaced by sums either already in the forward model or in the direct analytic inversion if it exists. For the discretization of the analytic inversion, the integral is replaced by sums over the obtained samples. If there are not enough samples available, one approach is to use interpolation to approximate a quantity between known samples [121].

### 1.3.2 Discretized variational methods

Instead of obtaining a solution to (1.1) by direct inversion of an infinite dimensional model, the linear forward operator $A$ can be discretized such that it can be represented by a matrix. As mentioned, many practical scenarios will only have a limited amount of measurements available while the object we wish to reconstruct has more components. This can be formalized by assuming that $A \in \mathbb{C}^{m \times N}$, with $m < N$. Yet, even if $A$ is full-rank and we consider the noiseless case, $m < N$ means that the inverse problem (1.1) is ill-posed. An approach to solve this ill-posed inverse problem is to reformulate it as an optimization problem of the form, for $y \in Y$

$$\operatorname{argmin}_{x \in X} d_Y(Ax, y).$$

Such approaches are also referred to as variational methods. Unfortunately, the above minimization problem does not have a unique minimizer in many cases. A classical example, as also presented in [132], of this setting is for $X = \mathbb{R}^N$, $Y = \mathbb{R}^m$ being equipped with the $\ell_2$-norm, $\|\cdot\|_{\ell^2}$,

$$\operatorname{argmin}_{x \in \mathbb{R}^N} \|Ax - y\|_{\ell^2}^2. \tag{1.2}$$

For fixed $y \in \mathbb{R}^m$, in this example there exist infinitely many solutions, which are an affine subspace of $\mathbb{R}^N$, $F_y = \mathcal{N}(A) + \{z\}$. Where $\mathcal{N}(A) = \{x \in \mathbb{R}^N : Ax = 0\}$ is the null space of $A$ and $z \in \mathbb{R}^N$ such that $Az = y$. Thus, the optimization problem is still ill-posed. Adding a regularization term to this initial variational formulation may yield a well-posed problem. Examples of such regularization terms will be presented in the following subsections.

#### Tikhonov regularization

A prominent example in order to obtain a well-posed problem, as mentioned in [132], is Tikonov regularization. It was first introduced by Tikhonov in 1943 [175] and [176], Phillips [147], and

Tikhonov and Arsenin [177] for solving ill-posed inverse problems. A common variational approach with a regularization term $\mathcal{R}(x)$ for $x \in \mathbb{R}^N$ and $\lambda > 0$ is the following

$$\operatorname{argmin}_{x\in\mathbb{R}^N} \|Ax - y\|_{\ell^2}^2 + \lambda\mathcal{R}(x), \tag{1.3}$$

for fixed $y = A\tilde{x} + e$, with $\tilde{x} \in \mathbb{R}^N$ and $e \in \mathbb{R}^m$ with $\|e\|_{\ell^2} \leq \epsilon$, where $\epsilon \geq 0$ is the noise level. Where for Tikhonov regularization $\mathcal{R}(x) = \|Lx\|_{\ell^2}^2$ and $L \in \mathbb{R}^{N\times N}$ is either the identity matrix or a discrete approximation to some derivative operator [95]. The overall aim of variational methods is to choose $\lambda > 0$ and the regulaization $\mathcal{R}$ such that when letting the noise level $\epsilon \to 0$ the solution obtained by (1.3) converges in some sense to a solution of the noiseless problem. An example of a theorem with such a result, is Theorem 2.4 [64]. Under certain assumptions it can be proved, that a so called minimum norm least-squares solution, where $\mathcal{R}(x) = \|x\|_{\ell^2}^2$, for the noisy case, converges to the same notion of solution in the noiseless case when the noise level $\epsilon \to 0$.

### Iterative reconstruction

In iterative reconstruction methods, often algorithms based on gradient descent for the term $\|Ax-y\|_{\ell^2}^2$ are used [8]. Based on (1.2), the minimizer is approximated and the iterative method is stopped after a certain number of iterations which prevents the reconstruction error from diverging. Hence, the stopping serves as some kind of regularization. There are many different iteration schemes, well known example can be found in [113, 141]. For further reference see [25, 28, 75, 93, 94, 164].

An example of an iterative approach to solving (1.3) in the case of Tikhonov regularization is the following. For fixed $y \in \mathbb{R}^m$ let $f(x) = \|Ax - y\|_{\ell^2}^2 + \lambda\|Lx\|_{\ell^2}^2$ for $x \in \mathbb{R}^N$ and $L \in \mathbb{R}^{N\times N}$ and using the gradient $\nabla f(x) = -2A^*y + 2(A^*A + \lambda L^*L)x$. Here $L^*$ denotes the transpose of the matrix $L$ and in the complex case the adjoint. Then for $k \in \mathbb{N}$ a solution can be iteratively approached by,

$$x^{k+1} = x^k - \gamma^{(k)}\nabla f(x^k),$$

where $\gamma^{(k)} \in \mathbb{R}$ is chosen to obtain a convergent algorithm, see [132] and for a more in depth treatment [22]. Moreover, note that additional assumptions are necessary for iterative approaches such that the obtained solution is in the set $F_y^e$.

## Bayesian formulation

Opposed to the deterministic model in (1.1) considered so far, on can consider a statistical measurement model [8, 132]. For simplicity consider the setting that $X = \mathbb{R}^N$ and $Y = \mathbb{R}^m$ being equipped with the $\ell_2$-norm. Here the sampling operator $A \in \mathbb{R}^{m\times N}$ is still a deterministic matrix, yet $x$ and $e$, hence the measurements $y$, are now considered to be random variables. Assuming that the distribution of the noise $e$ is given and omitting further details, one can

for example aim at obtaining a solution to the inverse problem by determining the probability density function of the measurements $y$ of a fixed input $x$ given by $p(y|x)$. A more rigorous definition is given in Definition 3.1. [8]. Formally, then one can determine the conditional probability density of inputs $x$ given a measurement $y$ by Bayes rule,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

With this statistical model one can seek a solition which obtaines the minimum mean square error (MMSE), also coined MMSE solution. This can be formally written as,

$$\operatorname{argmin}_{\tilde{x}} \mathbb{E}_{p(x,y)}(\|\tilde{x}(y) - x\|_{\ell^2}^2) = \operatorname{argmin}_{\tilde{x}} \int \|\tilde{x}(y) - x\|_{\ell^2}^2 p(x,y) dx dy. \tag{1.4}$$

here $\tilde{x}(y)$ is the attained reconstruction and $x$ is the ground truth and $p(x, y)$ is the joint probability density function [86]. A possible interpretation, offered by [132], of the MMSE solution is that it obtains the lowest average reconstruction error, under the assumption that all the model assumptions are correct. Formally, the MMSE solution of (1.4) is given by the condtional expectation $\tilde{x}(y) = \mathbb{E}_{p(x|y)}(x|y) = \int x p(x|y) dx$. The conditional probability densities $p(x|y)$ can be obtained by Bayes rule, under sufficiently strong assumptions for existence. However, a prior distribution $p(x)$ is necessary and unless all random variables are multivariate Gaussians constructing algorithms obtaining the MMSE solution is hard [132].

Instead of aiming to solve (1.1) by the MMSE solution, another approach is to find the solution that in some sense most likely has yielded the measurements. This is referred to as the maximum a posteriori (MAP) solution and is obtained by

$$\operatorname{argmax}_x p(x|y) = \operatorname{argmax}_x p(y|x)p(x) \tag{1.5}$$

These two approaches are related. In the case that we only consider Gaussian random variables, the MAP and MMSE solutions are the same. However, in general the solution obtained by (1.5) and (1.4) are not the same, see [86, 143] for an in depth discussion.

### 1.3.3 Sparsity based reconstruction

Another approach using the variational framework to regularize an ill-posed inverse problem is sparsity based reconstruction. The main idea is that a high-dimensional image can be represented by a small number of non-zero coefficients in an appropriate basis. This approach was introduced in [29, 30, 57]. The corresponding optimization problem can be stated as (1.3), with $\mathcal{R}(x) = \|x\|_{\ell^1}$ being the $\ell_1$-norm, for $x \in \mathbb{R}^N$. Under sufficient conditions on the sampling matrix $A \in \mathbb{R}^{m \times N}$ the solution to the optimization problem is stable and accurate in a specific sense [57]. Moreover, for general $A$ and under some additional assumptions, there exists an explicit characterization of solutions, as stated in the following theorem.

**Theorem 1.3.1** (Theorem 6 (Convex Problem With $\ell_1$ Minimization) [180]). *Let $A \in \mathbb{R}^{m\times N}$, $m < N$, $\mathcal{C} \subseteq \mathbb{R}^m$ be closed and convex, such that $A^{-1}(\mathcal{C}) \subseteq \mathbb{R}^N$ is non-empty, which is the feasibility hypothesis. Then,*

$$V = \operatorname{argmin}_{x\in\mathbb{R}^N} \|x\|_{\ell^1}, \quad \textit{such that} \quad Ax \in \mathcal{C}$$

*is a non-empty, convex, compact subset of $\mathbb{R}^N$ with extreme points $x_{\text{sparse}}$ of the form*

$$x_{\text{sparse}} = \sum_{k=1}^{K} a_k e_{n_k}$$

*with $K \leq m$, $\{e_n\}_{n=1}^N$ the canonical basis of $\mathbb{R}^N$, $n_k \in \{1, ..., N\}$ for $k = 1, ..., K$ and $a \in \mathbb{R}^K$ are suitable coefficients.*

Note that the ill-posedness of the inverse problem is apparent in the above theorem, as there is a set of possible solutions, in particular the solution must not be unique. However, for the standard Tikhonov regularization, with $L = 1$ and $\lambda = 1$, the solution is unique as stated in the following theorem. This is related to the fact that the $\ell_2$ norm is strictly convex, whereas the $\ell_1$-norm is not.

**Theorem 1.3.2** (Theorem 5 (Convex Problem With $\ell_2$ Minimization) [180]). *Let $A \in \mathbb{R}^{m\times N}$, $m < N$, $\mathcal{C} \subseteq \mathbb{R}^m$ be closed and convex, such that $A^{-1}(\mathcal{C}) \subseteq \mathbb{R}^N$ is non-empty, which is the feasibility hypothesis. Then,*

$$V = \operatorname{argmin}_{x\in\mathbb{R}^N} \|x\|_{\ell^2}, \quad \textit{such that} \quad Ax \in \mathcal{C}$$

*has a unique extreme point of the form*

$$x_{\text{LS}} = \sum_{k=1}^{m} a_k a^k = A^\top a$$

*with $a \in \mathbb{R}^m$ a suitable coefficient and $(a^k)_{k=1}^m \in \mathbb{R}^N$ are the row vectors of $A$.*

More details on sparsity based reconstruction methods are given in the following chapters. For example, Section 2.3.1 presents an overview of compressive imaging and undersampled acquisition. Section 2.3.2 presents standard methods used for image reconstruction, with a focus on methods used in this thesis.

## 1.4 Possible advantages through learning

In the following section we give a brief overview of some DL techniques with a specific focus on methods used for solving inverse problems, see [132, 156] for a more detailed overview. The main difference from classical schemes presented above to learning schemes, is that for learning training data $\mathcal{T} = \{y_k, x_k\}_{k=1}^T \subseteq \mathbb{R}^m \times \mathbb{R}^N$ with $y_k = Ax_k + e_k$ for $k \in \{1, \cdots T\}$ and

$T \in \mathbb{N}$ is used in order to obtain a decoder for (1.1). There is a range of learned approaches, where given an initial structure of a neural network, with is a specific kind of non-linear function, this function is fitted to the training data. These are referred to as fully learned approaches. Moreover, these can be divided into different classes, according to [156]. Firstly, there is supervised learning, where the training data pairs $\mathcal{T} = \{y_k, x_k\}_{k=1}^T$ are given. Supervised learning includes image domain learning [42, 92, 106, 111], hybrid domain learning [89, 173] and, for example, 'automated transform by manifold approximation' short AUTOMAP [192]. As a specific example of a fully learned supervised approach for solving (1.1), AUTOMAP will be more closely investigated with respect to stability and robustness in Chapter 2. Secondly, there is semi-supervised learning [49, 144] and unsupervised learning [34]. However, various disadvantages of such fully learned methods have been observed, for instance, the lack of sufficient training data for learned methods:

> *"However, in many scientific applications, the solution method needs to be robust and there is insufficient training data to support an entirely data-driven approach. This seriously limits the use of entirely data-driven approaches for solving problems in the natural and engineering sciences, in particular for inverse problems."* – From "Solving inverse problems using data-driven models" (2019) [8].

Thus, opposed to fully learned approaches, there also exist semi-learned methods, where the training data is for example used for learning the regularizer, as in [47, 115, 127, 139]. Other numerous different semi-learned approaches are mentioned in [8]. Yet, there are also various possible advantages of learned methods. Firstly, the design of the network can be conducted rather decoupled from the problem at hand and possibly re-used for other applications. One such example, are convolutional neural networks (CNNs), as indicated in [131]. Moreover, the main time demand is used when training learning based methods. Once the training is finished this means that compared to state of the art methods running time and computational costs are far less [131]. Moreover, learned methods have demonstrated higher accuracy in ill-posed settings and in some cases it is claimed that performance surpasses standard methods in many different settings [52, 171, 183, 192].

### 1.4.1 Deep learning for linear inverse problems

In the following we outline the application of deep learning methods to inverse problems. In general, the objective of DL is to construct a neural network that approximates a map $f : \mathbb{C}^m \to \mathbb{C}^N$, with $m, N \in \mathbb{N}$, from *samples*, i.e. pairs $(y, f(y))$, where $y = Ax + e$. In the following, consider the vanilla case of *feedforward* neural networks, although many of the results presented in this thesis also apply to more exotic setups. Now let $L, m', N' \in \mathbb{N}$ and consider the reals. An $L$-layer feedforward neural network is a function $\Psi : \mathbb{R}^{m'} \to \mathbb{R}^{N'}$ of the form

$$\Psi(y) = V_L(\rho(V_{L-1}(\rho(\dots\rho(V_1(y)))))), \quad y \in \mathbb{R}^{m'},$$

where each $V_j : \mathbb{R}^{n_{j-1}} \to \mathbb{R}^{n_j}$ is an affine map

$$V_j(y^{j-1}) = W_j y^{j-1} + b_j, \qquad W_j \in \mathbb{R}^{n_j \times n_{j-1}}, \quad b_j \in \mathbb{R}^{n_j},$$

and

$$y^j = \rho(V_j(y^{j-1})) \in \mathbb{R}^{n_j},$$

and $\rho : \mathbb{R} \to \mathbb{R}$ is a non-linear function, which is applied pointwise to $y$ by $\rho(y) = (\rho(y_i))$ for $y = (y_i)$, and $n_0 = m'$, $n_L = N'$. The $W_j$'s are referred to as *weights* and the $b_j$'s as *biases*. The number $L$ is the *depth* of the network, and $n_l$ is the *width* of its $l$th layer. The function $\rho$ is the *activation function*. Typical choices for $\rho$ are the *Rectified Linear Unit (ReLU)*, for $z \in \mathbb{R}$ defined by $\rho(z) = \max\{0, z\}$, or the *sigmoid*, defined by $\rho(z) = \frac{1}{1+e^{-z}}$, where $e$ is the exponential function. The *architecture* of a neural network refers to choice of the depth $L$, widths $n_1, \ldots, n_{L-1}$ and activation function $\rho$. Let $n = (n_0, n_1, \ldots, n_L) \in \mathbb{N}^{L+1}$. Then, the class of neural networks with a given architecture is denoted as $\mathcal{NN}$ or $\mathcal{NN}_n$. An *adaptive* neural network is a map $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ that depending on the input $y \in \mathbb{C}^m$ yields a neural network $\Psi' : \mathbb{C}^m \to \mathbb{C}^N$.

**Remark 1.4.1.** In inverse problems, it is common to deal with complex input $y \in \mathbb{C}^m$ and output $x \in \mathbb{C}^N$. A standard procedure is to associate $y \in \mathbb{C}^m$ with a vector $y' \in \mathbb{R}^{2m}$ consisting of the real and imaginary parts of $y$ and, then, apply a real-valued network $\Psi : \mathbb{R}^{2m} \to \mathbb{R}^{2N}$. Similarly, $x' = \Psi(y')$ is associated with a complex image $x \in \mathbb{C}^N$. In the above this corresponds to letting $m' = 2m$ and $N' = 2N$. We assume the complex case is treated in this way throughout this paper. We simply write

$$\Psi : \mathbb{C}^m \to \mathbb{C}^N$$

for a *network or decoder taking complex inputs and outputs*, with the assumption that it has this form.

In *inverse problems*, the goal is to construct a mapping that takes the noisy measurements $y = Ax + e$ of some unknown image $x$ as input and returns $x$ (or some approximation to it) as output. In DL for inverse problems (see, e.g. [8, 106, 131]) this is achieved using a *training set*

$$\mathcal{T} = \{(y^j, x^j) : y^j = Ax^j + e^j,\ j = 1, \ldots, K\},$$

consisting of pairs of the form $(Ax + e, x)$, where $x$ is a training image and $y = Ax + e$ are its noisy measurements. Compared to the initial statement, this means that $x = f(y)$ where $f$ is the mapping that is sought to be approximated by the NN. In practice, this means that the training set consists of measurements $y$ and vectors $x$, which are for example images obtained by a standard methods. With a fixed architecture, i.e. a class $\mathcal{NN}$, *training* a neural network is the process of computing an approximation $\Psi \in \mathcal{NN}$ from the data $\mathcal{T}$. The training set is defined as $\mathcal{T} = \{(Ax^j + e^j, x^j) \in \mathbb{C}^m \times \mathbb{C}^N : j = 1, \cdots K\} \subset\subset \mathbb{C}^m \times \mathbb{C}^N$. This is typically achieved by computing a minimiser of a certain optimization problem

$$\Psi \in \operatorname*{argmin}_{\tilde{\Psi} \in \mathcal{NN}} \frac{1}{K} \sum_{j=1}^{K} \mathrm{Cost}(\tilde{\Psi}, Ax^j + e^j, x^j), \tag{1.6}$$

where $\mathrm{Cost}$ is an appropriate *cost function*. This procedure is referred to as *standard training*. A popular choice is the $\ell^2$-loss, given by

$$\mathrm{Cost}(\tilde{\Psi}, Ax^j + e^j, x^j) = \frac{1}{2}\|x^j - \tilde{\Psi}\left(Ax^j + e^j\right)\|_{\ell^2}^2.$$

However, there is a myriad of other possibilities. The optimization problem (1.6) is generally non-convex, and devising effective numerical methods for computing minima is a substantial topic in its own right. This topic is not the concern of this thesis. Instead, simply observe that the result of a such a procedure is, in general, a neural network $\Psi \in \mathcal{NN}$ with a small *training error*. That is,

$$\|\Psi(y) - x\|_{\ell^2}^2 \leq \delta, \qquad \forall (y, x) \in \mathcal{T}, \tag{1.7}$$

for some $\delta > 0$. In practice, $\delta$ will depend on the cost function and the algorithm used for approximately solving (1.6). This makes main question this that thesis aims to answer is more precise. Namely, if theoretical limits of $\delta \geq 0$ exist and if so, what these limits are? A related and common issue in DL is *overfitting*. This refers to a network $\Psi$ with small training error, but poor *generalization*: that is, the trained network performs poorly on new images $x$ that are not in the training set. Typically, generalization performance is measured using a second set of data, the *test set*, not used in the training process. The corresponding error is the *test error*. Regularization is a standard method to attempt to cure the tendency of trained NNs to overfit. Instead of (1.6), one computes

$$\Psi \in \underset{\tilde{\Psi} \in \mathcal{NN}}{\operatorname{argmin}} \frac{1}{K}\sum_{i=1}^{K} \frac{1}{2}\|x_i - \tilde{\Psi}\left(Ax_i + e_i\right)\|_{\ell^2}^2 + \lambda J(\tilde{\Psi}), \tag{1.8}$$

where $\lambda \geq 0$ and $J : \mathcal{NN} \to \mathbb{R}$ is a regularization function. Often $J$ may be chosen to penalize large weight matrices. Note that this procedure in general still yields a small training error. The standard training for neural networks used to solve inverse problems, however yields risks for producing networks that become unstable, hallucinate and display additional or removed elements in the reconstruction. There are a variety of effects occurring which we aim at explaining in the following chapters.

## 1.5 Applications

There is a wide range of applications and imaging modalities that can be directly related to (1.1). In the following a short summary of methods, such as MRI, parallel MRI, some instances of fluorescence microscopy, computer tomography (CT) and positron emission tomography (PET) scans, is given and related to the inverse problem as in (3.1). Thus, these applications yield a decoder for (1.1) for a specific sampling operator $A$.

In general, compressive imaging pertains to the accurate and stable reconstruction of images from undersampled measurements. Typically, one models this problem as the discrete linear

problem (1.1). For undersampled acquisitions, an overall goal is to not sample more data than necessary. One possibility to explicitly model the degree of undersampling one uses a mask (projection) operator $P_\Omega$ defined as follows. For a given sampling mask (set) $\Omega \subsetneq \{1, \dots, N\}$, of cardinality $|\Omega| = m < N$, we let $P_\Omega \in \mathbb{C}^{m \times N}$, denote the projection which extracts the elements of a vector, indexed by $\Omega$. That is, for $x \in \mathbb{C}^N$, we have $(P_\Omega x)_i = x_{\Omega(i)}$ for $i = 1, \dots, m$, where $\Omega(i)$, denotes the $i$'th element in $\Omega$, using the natural ordering. Another possibility for undersampling arising is in fluorescence microscopy where due to phototoxicity or other effects [117], such as interactions of the samples with photons, the number of photons is limited. This indirectly yields undersampling and, hence, the kernel of the sampling matrix satisfies $\mathcal{N}(A) \neq \{0\}$.

Concerning MRI, there are two main modalities - single coil and multi-coil or also parallel MRI. Firstly, in single-coil MRI, e.g. used for AUTOMAP [192], the standard model for the sensing matrix $A$ is

$$A = P_\Omega F, \tag{1.9}$$

where $F \in \mathbb{C}^{N \times N}$ is a 2-dimensional discrete Fourier transform (DFT) matrix. We highlight that in this model one always has $m < N$ measurements, since it is assumed that $\Omega$ is a strict subset of $\{1, \dots, N\}$. Thus, (1.9) yields a undetermined inverse problem and when adding noise can be directly written as in (1.1). Secondly, in multi-coil (parallel) MRI, multiple receiver coils simultaneously acquire $k$-space data. To acquire as much information as possible, each of the receiver coils is adjusted to be sensitive to a limited spatial region [53]. For an MRI scanner with $c$ coils, one models this by introducing a diagonal matrix $S_i \in \mathbb{C}^{N \times N}$, $i = 1, \dots, c$ for each coil. This matrix is usually called the *sensitivity matrix* or *sensitivity profile* of the coil. Given these sensitivity profiles, the full model for the acquisition matrix $A \in \mathbb{C}^{m \times N}$ is

$$A = \begin{bmatrix} P_\Omega F S_1 \\ \vdots \\ P_\Omega F S_c \end{bmatrix} \tag{1.10}$$

where $|\Omega| = m'$, $P_\Omega \in \mathbb{C}^{m' \times N}$ and $m = m'c$. Depending on the choice of $m'$ and $c$, the total number of measurements $m$ can exceed the number of pixels $N$ in the image, i.e., $m > N$. Furthermore, notice that since $|\Omega| = m' < N$ and $A$ in (1.10) is rank deficient, this is a form of undersampled acquisition. Thus, (1.10) yields an undetermined inverse problem and when adding noise it can be directly written as in (1.1).

Concerning fluorescence microscopy, there are different methods of data acquisition. For once, there is structured illumination in temporal compressive microscopy, as for example described in [187]. Here, the aim is to obtain a compressive video microscope based on structured illumination with an incoherent light source. One main functionality of structured illumination is to induce a frequency shift in the Fourier domain through a periodic illumination pattern and, hence, transforming the high spatial frequency components into the detectable range. Let $(x, y, t)$ denote position and time, $f(x, y, t)$ the fluorescence signal from the object and $h(x, y)$

the point spread function. The structured illumination imposed on the sample is $S(x, y, t)$. Then, the measurement at the detector coordinates $(x', y')$ and time point $t'_i$ is,

$$g(x', y', t'_i) = \int_{t'_i}^{t'_i+\delta t} \left[ \int \int h(x - x', y - y') S(x, y, t) f(x, y, t) dx dy \right] dt.$$

In order to obtain an inverse problem of the form (3.1), we need to discretize the above sampling scheme. The sampling can be discretized as

$$g = H \left[ S_1 S_2 \dots S_{N_T} \begin{bmatrix} f_1 \\ \vdots \\ f_{N_T} \end{bmatrix} \right], \tag{1.11}$$

where $S_i$ is the structured illumination matrix of $i$-th illumination pattern, and $H$ is the PSF matrix of the objective served as a blur kernel. Now, with $A = HS$ we can write (1.11) as $g = Af$. Thus, this directly yields (1.1). Yet, whether this yields an undersampled aqcuisition and thus an undetermined inverse problem, depends on the blur kernel $H$ and the illumination pattern $S$. Concerning CT, the transformation of the data acquisition to (1.1) is outlined in [166]. Another related application is fluorescence molecular tomography, for example, outlined in [104]. The PET inverse problem is, for instance, outlined in [160].

# Chapter 2

# Deep learning through domain-transform manifold learning for image reconstruction is not robust

The following chapter is concerned with a fully learned neural network approach for image reconstruction, which was introduced in [192] and coined 'automated transform by manifold approximation'. In particular, its potential benefits and disadvantages compared to standard methods for image reconstruction are highlighted. This chapter is based on joint work with Vegard Antun, University of Oslo, who provided the code and figures, and Francesco Renna, University of Porto, and was supervised by Anders C. Hansen, University of Cambridge, and Ben Adock, Simon Fraser University.

## 2.1 Introduction and related work

In the last several years, artificial intelligence (AI), driven by deep learning (DL), has shown great potential for enhancing the performance of image reconstruction across a range of medical, scientific and industrial imaging modalities. However, recent studies have raised grave concerns regarding the trustworthiness of DL for image reconstruction [6, 16, 99, 138]. Existing methods based on DL are often *non-robust*. They may be unstable to perturbations or generalize in unpredictable or inconsistent ways, both of which can lead to undesirable artefacts. This chapter discusses and explains mathematically why these phenomena may occur in general for fully-learned AI-based methods for solving inverse problems. To highlight these issues, we focus on a recent work by Zhu et al. [192], which was also featured in [171]. In [192] a new AI-based framework termed AUTOMAP was introduced as an alternative to current standard techniques. This framework, termed 'automated transform by manifold approximation' (AUTOMAP), was introduced as an alternative to current standard techniques. The AUTOMAP framework promises "*superior immunity to noise and a reduction in reconstruction artefacts compared with*

*conventional handcrafted reconstruction methods*" [192] for a range of different imaging scenarios. Notably, this includes undersampled acquisition in medical imaging, one of the major areas of modern image reconstruction research, due to its potential to significantly reduce scan times in modalities such as Magnetic Resonance Imaging (MRI), and therefore ultimately reduce healthcare costs.

Based on [182], the AUTOMAP framework remodels image reconstruction as a learning problem, where a reconstruction map is learned from an appropriate amount of training data which models the relation between the image and the corresponding sampling data [192]. Precisely, a deep neural network reconstruction map is learned in a manner that is agnostic to the physical and mathematical principles underlying the specific acquisition device at hand. It therefore constitutes a significant and interesting departure from classical approaches. In these approaches, properties of the sampling process and the underlying data are used in order to obtain a reconstruction, for a general overview see for example [8, 132]. Moreover, the physical acquisition device of the imaging modality needs to be modelled in most classical approaches. Examples of such modalities, as listed in [192], include magnetic resonance imaging, X-ray computed tomography, positron emission tomography, ultrasound imaging and radio astronomy [84, 88, 189]. The AUTOMAP methodology obviates the need to mathematically model the physical acquisition device in each different imaging modality. It does so by exploiting the low-dimensional structure of medical images and the ability of convolutional layers to capture such a structure, in combination with several fully connected layers to 'learn the physics'. Indeed,

> *"AUTOMAP provides a new paradigm for image reconstruction that learns a reconstruction function for arbitrary acquisition strategies conditioned upon low-dimensional features of real-world data to improve artefact reduction and reconstruction accuracy for noisy and undersampled acquisitions."* – From "Image reconstruction by domain-transform manifold learning" (2018) [192].

Therefore, the AUTOMAP methodology is an interesting contribution to the research community's understanding of the potential and limitations of learning for image reconstruction in many practical scenarios where the acquisition device cannot be modelled accurately, or when it is undesirable to do so. For example, in the case that the acquisition device is known only through training data, a first theoretical result on convergence and stability of regularized solutions is presented in [9]. However, we show, both empirically and theoretically, that methods such as AUTOMAP are bound to produce reconstruction methods that are not robust. In particular, such methods are unstable to perturbations and may suffer from inconsistent performance and unpredictable generalization.

Concerning the application of AUTOMAP to medical imaging, there are fundamental issues with the method's reliability and robustness. These issues are rooted in the fact that medical imaging relies on reconstruction algorithms that are robust to random noise and other effects that corrupt the data, and produce reliable images without unpredictable artefacts. In the last

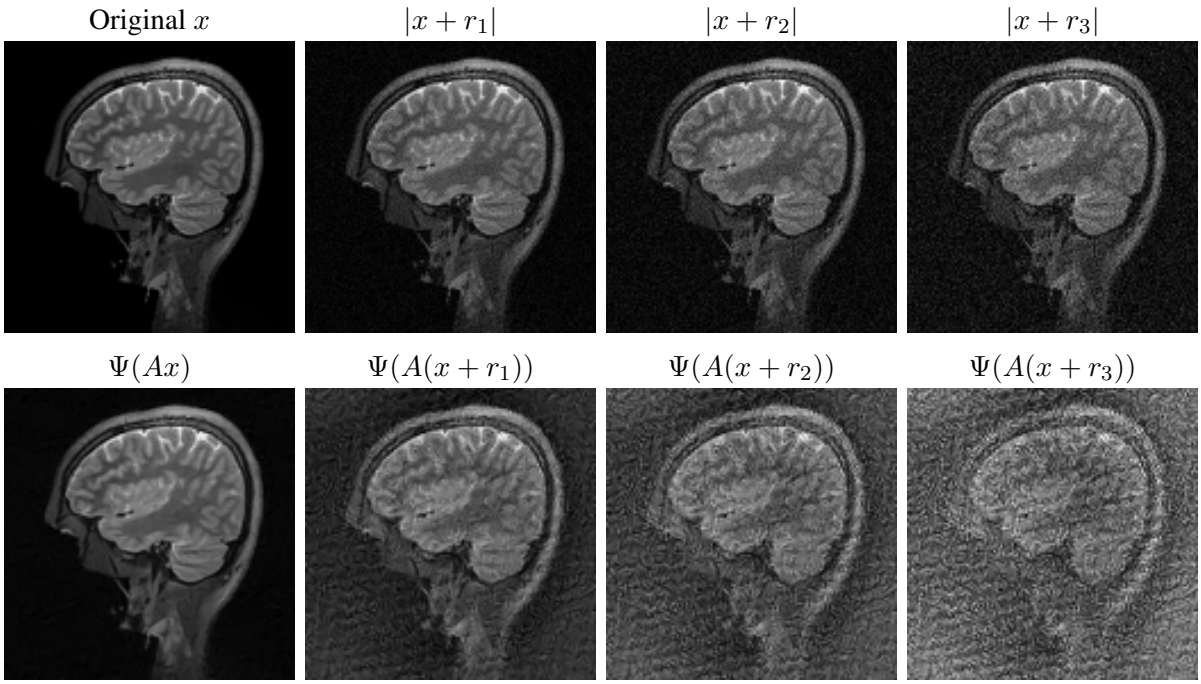


Figure 2.1: (**AUTOMAP is unstable**). This figure shows the performance of the AUTOMAP reconstruction map $\Psi\colon \mathbb{C}^m \to \mathbb{C}^N$ when applied to the undersampled discrete MRI reconstruction problem, in which the measurement matrix $A \in \mathbb{C}^{m\times N}$ is a subsampled discrete Fourier transform with 60% subsampling. Upper row: Tiny perturbations $r_1, r_2, r_3 \in \mathbb{C}^N$ are added to the image $x \in \mathbb{C}^N$ (written as a vector) to simulate worst-case effects. Lower row: Reconstruction from the perturbed measurements $A(x + r_j)$. Note that the condition number $\mathrm{cond}(AA^*) = 1$. In particular, these instabilities are caused by the methodology, not by the nature of the image reconstruction problem itself. Other non-AI based methods are perfectly stable for this problem to simulated worst-case perturbations ( Fig. 2.6). See Section 2.3.4 for details on the measurement model and the computation of the $r_j$.

several years, DL has shown great potential to improve the fidelity of medical image reconstruction. Yet there is growing concern, as indicated in the 2019 and 2020 fastMRI challenges, that AI-based reconstruction methods suffer from a lack of robustness [114, 138]. In particular, AI generated 'hallucinations' related to instability [114] have been reported to produce unacceptable false negatives and false positives. Viewed in the broader context of AI, where DL is known to produce unstable methods for problems such as image classification [67, 116, 135, 174], audio and speech recognition [32, 33, 190] and natural language processing [123], this fact is unsurprising. Yet it is a serious cause for concern in medical imaging and medicine in general [73], where robustness and reliability are critical. However, it should be noted that there has also been a significant corpus of research dedicated towards improving the stability and robustness of DL generated methods for various applications [80, 102, 153, 159]. Moreover, most approaches in imaging proposed to improve stability and robustness of DL generated image reconstruction rely on the underlying data acquisition procedure.

### 2.1.1 Overview

In the following section we present a summary of our main results. This is followed by an examination of robustness of the AUTOMAP method, Section 2.2.1, and of the performance-stability trade-off, Section 2.2.2 in which our main Theorem 2.2.1 is presented. Moreover, in Section 2.2.4, we provide an explanation for why the theoretical premise for AUTOMAP, presented in [192], is not satisfied in undersampled acquisitions. This is followed by a comparison of fully learned neural networks and AUTOMAP, Section 2.2.5, and an explanation of why the training process of AUTOMAP typically yields a small training error, Section 2.2.7. Concluding this section, Theorem 2.2.1, its consequences for AUTOMAP, and its consequences in a broader context are presented. We also explain the assumptions which make standard methods stable for these problems. This chapter is then finished with a general conclusion followed by the methods section, providing greater detail on the applied methods. The Section 2.3.1 includes an overview on compressive imaging and undersampled acquisition. The Section 2.3.2 describes standard methods used for image reconstruction. Moreover, details on computational aspects, Section 2.3.3, computing worst-case perturbations, Section 2.3.4, and the standard reconstruction method used for comparison, Section 2.3.6, are presented.

## 2.2 Main results

Our key findings are summarised as follows:

(M1) **Instabilities due to generalization behaviour of fully learned methods (Summary Theorem 2.2.1).** The generalization behaviour of fully learned neural networks for image reconstruction for undersampled data acquisition can lead to instabilities. Our results demonstrate that there is a fundamental *performance-stability* trade-off, which is illustrated in Fig. 2.5, for image reconstruction methods, in which over-performance or inconsistent performance on certain images necessarily leads to instability. Theorem 2.2.1 shows that any stable method must have mechanisms built-in to prohibit the conditions described therein that necessarily lead to instability. In particular, 2.2.1 also shows that the learning methodology that lies at the heart of AUTOMAP, namely, a methodology that seeks to learn a reconstruction map directly from training data without any restrictions based on the acquisition process, will in general yield non-robust methods as there is no mechanism to control this trade-off. Moreover, empirically, AUTOMAP can also be shown to satisfy the conditions of Theorem 2.2.1. In other words, this theoretical result is a cause of its instability, see Table 2.1. AUTOMAP does not have mechanisms built-in to prohibit the conditions described in Theorem 2.2.1 that necessarily lead to instability. In fact, its methodology encourages such conditions to occur. In contrast, standard "*handcrafted*" approaches do have such mechanisms, and under the right conditions lead to robust methods, see Fig. 2.6 and Section 2.2.3.

(M2) **AUTOMAP is not robust (Summary Section 2.2.1 and Figs. 2.1,2.2,2.3).** AUTOMAP is not robust in undersampled acquisitions, both with respect to worst-case perturbations Fig. 2.1 and to random Gaussian noise Fig. 2.2 centered around image-independent worst-case perturbations. Moreover, it performs inconsistently and, hence, generalizes unpredictably. Specifically, measurements corresponding to similar images, for example images within, or close to, the training and test set, can result in substantial variability in the reconstruction quality, see Fig. 2.2 and Fig. 2.3. In Section 2.2.1 we provide a discussion concerning our experimental results indicating that AUTOMAP is not robust.

(M3) **Absence of the theoretical premise for the AUTOMAP methodology (Summary Section 2.2.4).** The theoretical premise underlying AUTOMAP that guarantees the method's robustness, namely, the existence of a smooth homeomorphism between image space and sample space, is typically absent in the case of undersampled acquisitions, see Section 2.2.4. This is, for example, the case in both parallel and single coil MRI.

### 2.2.1 Examination of robustness

In order to experimentally examine the robustness of the fully learned method, we investigate the quality of reconstruction of an image from its corresponding measurements compared to the reconstruction from measurements perturbed by noise. These perturbations $e_{\text{pert}}$ in the input data can come from several sources, and can be modelled as

$$e_{\text{pert}} = e^1_{\text{pert}} + e^2_{\text{pert}}. \tag{2.1}$$

Here $e^1_{\text{pert}}$ is the non-generic part of the perturbation – potentially caused by motion changes, anatomic differences, malfunctioning apparatus etc. – and $e^2_{\text{pert}}$ is the generic part, typically modelled as a mean-zero random variable (e.g. Gaussian white noise), whose distribution depends on the acquisition process.

While a model for $e^2_{\text{pert}}$ is often known, it is difficult to model $e^1_{\text{pert}}$ to account for all possible perturbations. Since there is empirical evidence that DL produces unstable methods, it is important that comprehensive stability examinations consider both mean-zero random noise and worst-case scenarios, to account for potentially hard-to-model non-generic perturbations. In order to account for worst-case scenarios, in Figs. 2.2 and 2.9, $e^1_{\text{pert}}$ is chosen to be an adversarial perturbation with respect to a different image.

In Fig. 2.1 we demonstrate the non-robustness of AUTOMAP with respect to worst-case perturbations. In Fig. 2.2 the experiment is repeated with Gaussian random noise, where $e_{\text{pert}} = e^1_{\text{pert}} + e^2_{\text{pert}}$ as in (2.1) is such that $e^1_{\text{pert}}$ is the fixed perturbation from Fig. 2.1 and $e^2_{\text{pert}}$ is a mean-zero Gaussian random variable. Thus, $e_{\text{pert}}$ is a non-zero mean Gaussian random variable. We conclude that AUTOMAP is unstable to both image-dependent, worst-case perturbations and image-independent Gaussian random noise. As image-independent Gaussian random noise is

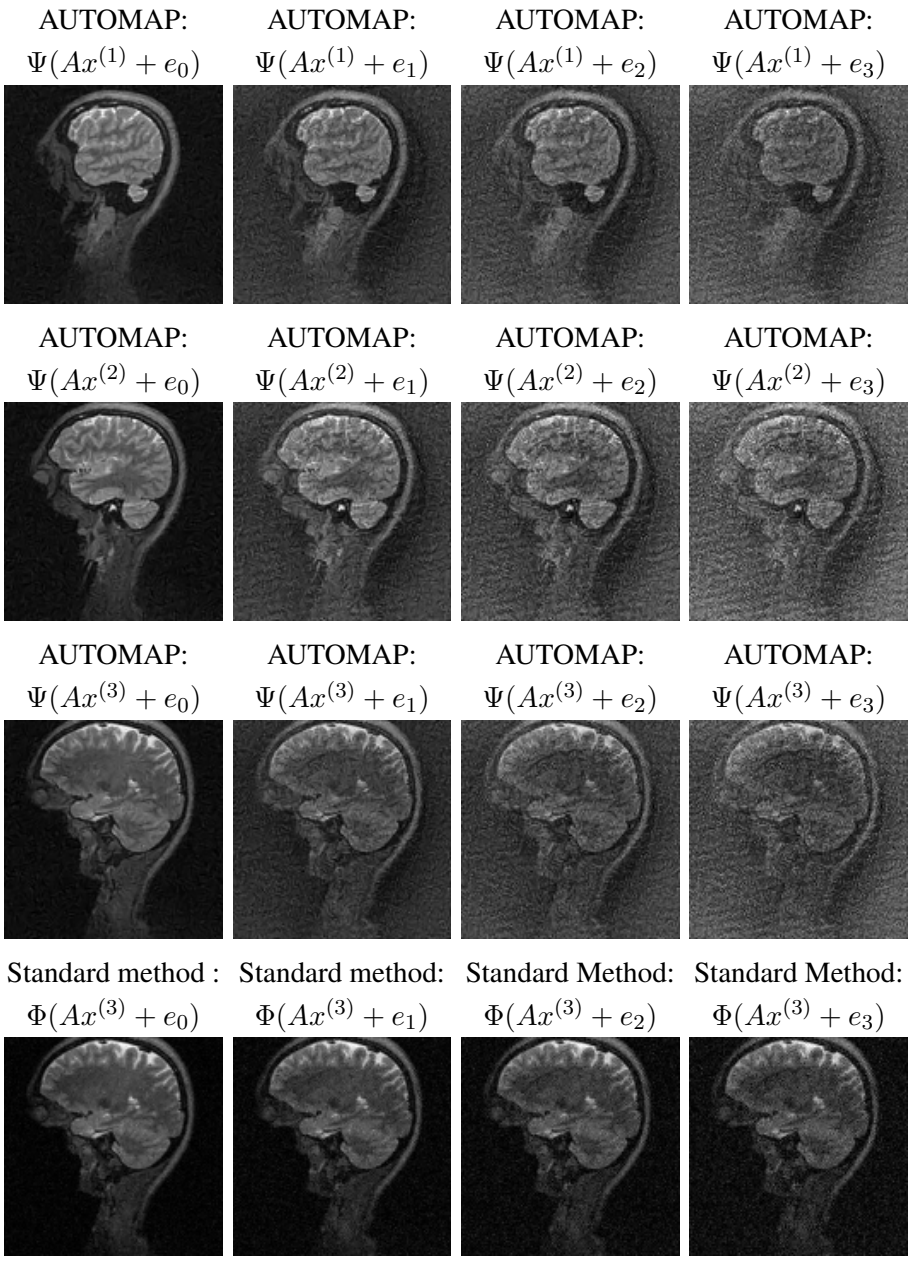


Figure 2.2: (**AUTOMAP is unstable to image-independent random Gaussian noise**). This experiment considers noise of the form $e_j = e^1_{\text{pert},j} + e^2_{\text{pert},j}$, where $j = 0, 1, 2, 3$, $e^1_{\text{pert},j}$ is constant and $e^2_{\text{pert},j}$ are different types of mean-zero Gaussian noise. The noise is generated as follows. Let $r_0 = 0$ and $r_j \in \mathbb{C}^N$, $j = 1, 2, 3$, denote the worst-case perturbations computed for AUTOMAP in Fig. 2.1. This perturbation is independent of the images $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ considered in this figure. The real and imaginary components of the sampled perturbations $\tilde{e}_j = Ar_j$ are denoted by $\tilde{e}^R_j = \Re(\tilde{e}_j)$ and $\tilde{e}^I_j = \Im(\tilde{e}_j)$, respectively. We draw new random perturbations $e^R_j \sim \mathcal{N}(\tilde{e}^R_j, 0.01^2)$ and $e^I_j \sim \mathcal{N}(\tilde{e}^I_j, 0.01^2)$ from normal distributions with means $e^R_j$ and $e^I_j$, respectively, and standard deviation 0.01. The non-mean zero random Gaussian noise is computed by $e_j = e^R_j + e^I_j \mathrm{i}$, were i is the imaginary unit. The first to third row show AUTOMAP's reconstruction of $x^{(i)}$, $i = 1, 2, 3$ from measurements $y = Ax^{(i)} + e_j$, $j = 0, 1, 2, 3$. The fourth row shows reconstructions of $x^{(3)}$ from measurements $y = Ax^{(3)} + e_j$, $j = 0, 1, 2, 3$ using a standard method based on a LASSO decoder with a wavelet transform, see Section 2.3.1. In this experiment $\|\tilde{e}_0\|_{\ell^2} < \cdots < \|\tilde{e}_3\|_{\ell^2}$, and $\{\|\tilde{e}_j - e_j\|_{\ell^2} / \|\tilde{e}_j\|_{\ell^2}\}_{j=1}^3 = \{0.708, 0.473, 0.338\}$. Here $\Psi \colon \mathbb{C}^m \to \mathbb{C}^N$ and $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ are the reconstruction maps of AUTOMAP and the standard method, respectively, and $A \in \mathbb{C}^{m \times N}$ is a subsampled discrete Fourier transform (60% subsampling).

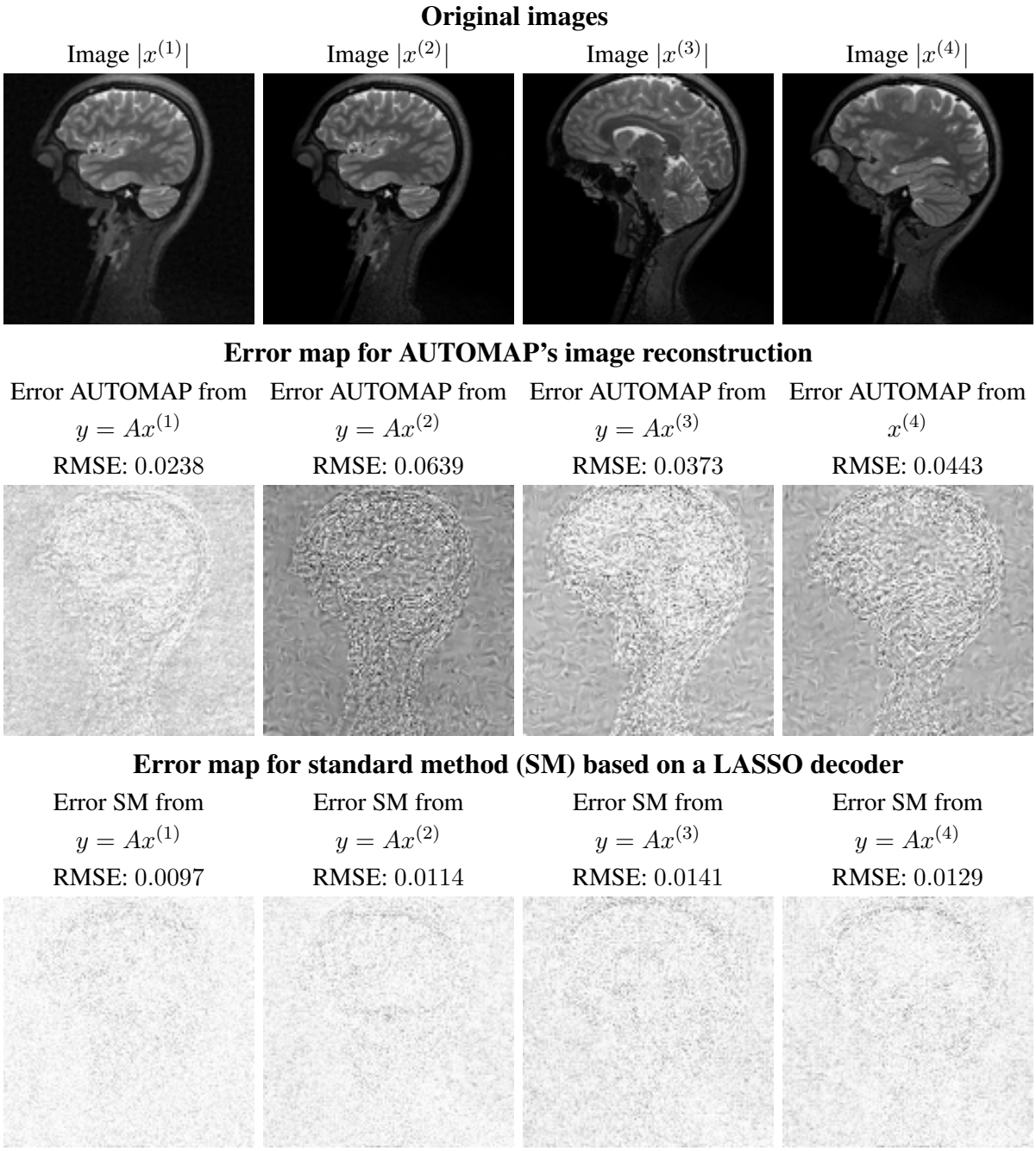


Figure 2.3: (**AUTOMAP generalizes unpredictability**). We examine AUTOMAP's performance on four different images $x^{(i)}$, $i = 1, \ldots, 4$. $x^{(2)}, x^{(3)}$ and $x^{(4)}$ are from the MGH–USC dataset [69] used for training and/or testing of AUTOMAP. Image $x^{(1)}$ is identical to $x^{(2)}$, except for a small, visually imperceivable perturbation, which has been added to it in order to increase AUTOMAP's performance. The first row shows the original images, the second row shows the error maps for the AUTOMAP reconstructions, and the third row shows the error maps for the reconstructions obtained by the standard method used in Fig. 2.2. In this experiment all the images $x^{(i)} \in \mathbb{C}^N$ are normalized to lie in the interval $[0, 1]$ and the data is given by $y = Ax^{(i)}$. To visualize the error map, $\iota = \max\{|x_j^{(i)} - \tilde{x}_j^{(i)}| : i = 1, 2, 3, 4 \text{ and } j \in \{1, \ldots, N\}\}$ denotes the maximum pixel-wise absolute error between the true and reconstructed images $\tilde{x}^{(i)}$ from both methods, and we show $\mathbb{I} - \iota^{-1}|x^{(i)} - \tilde{x}^{(i)}|$, where $\mathbb{I} \in \mathbb{C}^N$ is a vector of ones. AUTOMAP is trained to reconstruct images where the mean is subtracted. To make a fair comparison all images reconstructed by AUTOMAP are scaled to the interval $[0, 1]$.

in general not a rare event, the corresponding occurence of substantial artefacts is also not rare event. See Fig. 2.2.

In Fig. 2.3 we demonstrate AUTOMAP's inconsistent performance and unpredictable generalization. The four images considered are very similar, yet the reconstruction errors exhibit substantial variation. Notably, a small, imperceptible change in the data can cause the RMSE to change by over 250%. Note that in all the above experiments, AUTOMAP is compared to a standard "*handcrafted*" method. This comparison demonstrates that robustness (stability to perturbations and consistent performance) can be achieved in all the cases where AUTOMAP exhibits non-robustness. This comparison also demonstrates that AUTOMAP does not lead to "*a reduction in reconstruction artefacts compared with conventional reconstruction methods*" [192], neither for images within the training and test set (Fig. 2.3) nor outside of these sets (Fig. 2.4).

### 2.2.2 The performance-stability trade-off

We now theoretically examine the robustness of AUTOMAP. Before presenting the main theorem, we make a few comments on the notation. From now on, $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{C}^m$ and $\mathbb{C}^N$, $\mathcal{B}(v,\epsilon)$ denotes the open ball centred at $v$ with radius $\epsilon > 0$ (with respect to the norm $\|\cdot\|$), and $I \in \mathbb{C}^{m\times m}$ denotes the identity matrix. In the following result, $A \in \mathbb{C}^{m\times N}$ is the mathematical model of a given imaging modality.

**Theorem 2.2.1** (The performance-stability trade-off)**.** *Let $A \in \mathbb{C}^{m\times N}$, and let $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be a continuous reconstruction map. Suppose there are $x, x', x'' \in \mathbb{C}^N$, $\eta_1 > 0$ and $\eta_2 > 0$, with $\|x - x''\| >> 2\eta_1$, such that*

$$\|\Psi(Ax) - x\| < \eta_1 \tag{2.2}$$

*and*

$$\|\Psi(Ax') - x''\| < \eta_1 \quad \text{and} \quad \|Ax - Ax'\| \leq \eta_2. \tag{2.3}$$

*(1) Then there is a $\tilde{e} \in \mathbb{C}^m$ with $\|\tilde{e}\| \leq \eta_2$ and a $\varepsilon > 0$ such that for all $e \in \mathcal{B}(\tilde{e},\varepsilon)$ we have that*

$$\|\Psi(Ax + e) - \Psi(Ax)\| \geq \|x - x''\| - 2\eta_1. \tag{2.4}$$

*(2) If $e = \{e_1, \ldots, e_m\}$ is an absolutely continuous random vector, with a strictly positive probability density function, then there is a $c > 0$ such that* (2.4) *holds with probability greater than $c$.*

*(3) Moreover, for any $\epsilon > 0$, there is a Gaussian distribution on $e$ such that* (2.4) *holds with probability greater than $1 - \epsilon$.*

Theorem 2.2.1 has three main conclusions:

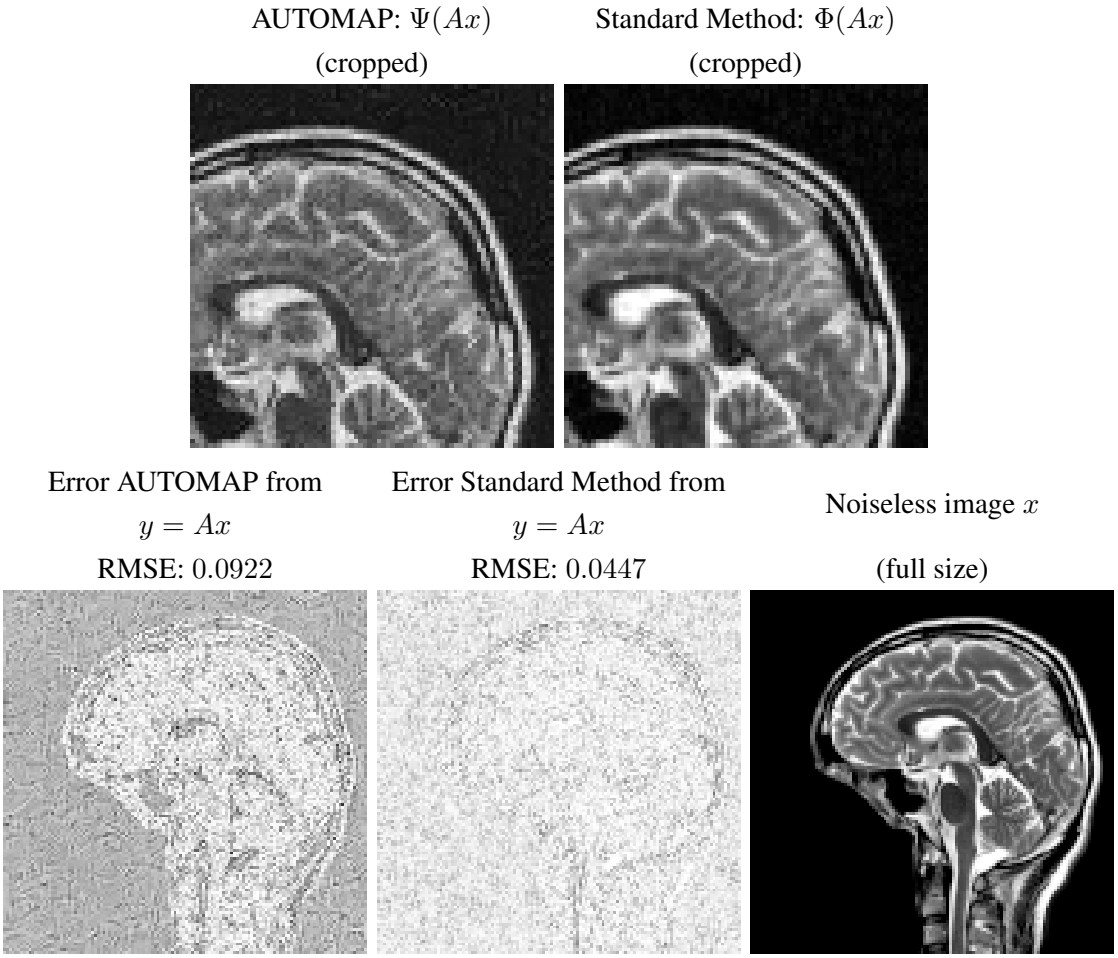


Figure 2.4: (**AUTOMAP produces serious reconstruction artefacts on images outside the training dataset**). In this experiment we have chosen an image $x$ with higher contrast and more details than the images from the training set. In the upper row we see zoomed in versions of AUTOMAP's and a standard method's reconstruction from measurements $y = Ax$. As we see from the figure, AUTOMAP shows no "*reduction in reconstruction artefacts compared with conventional handcrafted reconstruction methods*" reconstruction has striking reconstruction artefacts. The standard method in this example is based on a LASSO decoder with wavelet reconstruction. In the lower row we visualize the error map of AUTOMAP and the standard methods reconstruction, along with the true image $x$. The error maps are computed by taking the largest pixel wise absolute difference between (all) the reconstructed images and the true image $x$. Let $\iota$ denote this difference. The error maps are then normalized by the same $\iota$ for all reconstructed imaged, as $\mathbb{I} - \iota^{-1}|x - \tilde{x}|$, where $\mathbb{I} \in \mathbb{C}^N$ is a vector of ones and $\tilde{x}$ is a reconstructed image.

(A) (**Overperformance causes instability**) Let $x'' = x'$. Then, Theorem 2.2.1 states that a reconstruction map $\Psi$ that *overperforms* – it recovers two images $x$ and $x'$ well (i.e. (2.2)) from measurements that are close (i.e. (2.3)) – must be unstable (i.e. (2.4)). A stable reconstruction map can therefore recover either $x$ accurately or $x'$ accurately, but not both. Thus, there is a trade-off between how accurately $x$ and $x'$ can be recovered and how stable the reconstruction map can be. Note that the situation described can possibly occur with learning, e.g. when $x$ and $x'$ are elements of the training set and at the same time $Ax \approx Ax'$. The latter condition is not encouraged through learning, but if $A$ has a non-trivial kernel and the training set is not chosen accordingly, it can occur.

(B) (**Inconsistent performance causes instability**) Now let $x'' = \Psi(Ax')$. Theorem 2.2.1 states that a map $\Psi$ that recovers $x$ well (i.e. (2.2)), but reconstructs some $x'$ having similar measurements (i.e. (2.3)) poorly, i.e. $\|x - x''\| = \|x - \Psi(Ax')\| \gg \eta_1$, must be unstable. Hence, inconsistent performance – recovering $x$ well but some nearby $x'$ poorly – causes instability. This situation can easily occur with learning, e.g. when $x$ is in the training set and $x'$ is close to this set, but not in it. Precise numerical values are shown in Table 2.1, thus verifying that Theorem 2.2.1 predicts AUTOMAP's instability.

(C) (**Probability of instability can become arbitrarily large**) As shown in part $(2)$ of Theorem 2.2.1, the probability of instabilities occurring can get arbitrarily close to 1 while the noise level stays bounded. Recall Fig. 2.2.

**Remark 2.2.2** (The lower bound in (2.4) related to instability)**.** For $\epsilon > 0$ and $y \in \mathbb{C}^m$, we define the local $\epsilon$-Lipschitz constant of a mapping $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ as

$$L^\epsilon(\Phi, y) = \sup_{\substack{z \in \mathbb{C}^m \\ 0 < \|z-y\| \leq \epsilon}} \frac{\|\Phi(z) - \Phi(y)\|}{\|z - y\|}.$$

Given this definition, part $(1)$ can be reformulated as the local Lipschitz constant being unbounded: There is a closed non-empty ball $\mathcal{B} \subset \mathbb{C}^m$ centred at $y = Ax$ such that, for all $\epsilon \geq \eta_1$, the local $\epsilon$-Lipschitz constant at any $\tilde{y} \in \mathcal{B}$ is bounded from below:

$$L^\epsilon(\Psi, \tilde{y}) \geq \frac{1}{\eta} \left( \|x - x''\| - 2\eta_1 \right).$$

*Proof of Theorem 2.2.1.* **Statement (1)**: We start by applying the reverse triangle inequality twice and get

$$\|\Psi(Ax) - \Psi(Ax')\| \geq \|x'' - x\| - \|\Psi(Ax) - x\| - \|\Psi(Ax') - x''\|. \tag{2.5}$$

By assumptions (2.2) and (2.3) it follows that

$$\|\Psi(Ax') - \Psi(Ax)\| > \|x'' - x\| - 2\eta_1. \tag{2.6}$$

Using that the inequality in (2.6) is strict, and that $\Psi$ is continuous function, for example a neural network, we can find an $\varepsilon > 0$, such that for $\tilde{e} := Ax' - Ax$, we have that

$$\|\Psi(Ax + e) - \Psi(Ax)\| \geq \|x - x''\| - 2\eta_1, \quad \text{for all } e \in \mathcal{B}(\tilde{e}, \varepsilon). \tag{2.7}$$

Moreover, by assumption (2.3) we have $\|\tilde{e}\| \leq \eta_2$.

**Statement (2)**: Let $\{\Omega, \mathbb{P}, \mathcal{F}\}$ be a probability space and $e : \Omega \to \mathbb{C}^m$ be an absolutely continuous random vector with a strictly positive probability density function $g\colon \mathbb{C}^m \to \mathbb{R}_{>0}$. Let $\mu$ denote the Lebesgue measure on $\mathbb{C}^m$. Now denote $z := \tilde{e}$. It then follows that

$$c = \mathbb{P}(e \in \mathcal{B}(z, \varepsilon)) = \int_{\mathcal{B}(z,\varepsilon)} g \, d\mu > 0,$$

by the assumption that $g$ is strictly positive. Note that the noise level for the constant $c$ is bounded, as for $e \in \mathcal{B}(z, \varepsilon)$ we have that $\|e\| < \varepsilon + \eta_2$. Then, by (2.7),

$$\mathbb{P}\left(\|\Psi(Ax + e) - \Psi(Ax)\| \geq \|x - x''\| - 2\eta_1\right) \geq c,$$

proving the second part of the theorem.

**Statement (3)**: Let $\epsilon > 0$ be fixed and let $e\colon \Omega \to \mathbb{C}^m$ be a Gaussian vector centered at $z = Ax' - Ax$, with covariance matrix $\Sigma = \sigma^2 I \in \mathbb{C}^{2m\times 2m}$, where $\sigma > 0$ is a scalar. As the joint probability density function of a (multivariate) Gaussian distribution is a real valued function, we identify a complex vector $t \in \mathbb{C}^m$, with it's real counterpart $\widehat{t} = (\mathrm{Re}(t), \mathrm{Im}(t))^\top \in \mathbb{R}^{2m}$, where $\mathrm{Re}(t)$ and $\mathrm{Im}(t)$ denote the real and imaginary component of $t$, respectively. Similarly, we write $\widehat{z} \in \mathbb{R}^{2m}$ for $z \in \mathbb{C}^m$. We then have that

$$\begin{aligned}\mathbb{P}(e \in \mathcal{B}(z, \varepsilon)) &= \frac{1}{\sqrt{(2\pi)^{2m}\det(\Sigma)}} \int_{\mathcal{B}(z,\varepsilon)} \exp\left(-\frac{1}{2}(\widehat{t} - \widehat{z})^\top \Sigma^{-1}(\widehat{t} - \widehat{z})\right) d\mu(t) \\ &= \frac{1}{\sqrt{(2\pi\sigma^2)^{2m}}} \int_{\mathcal{B}(0,\varepsilon)} \exp\left(-\frac{1}{2\sigma^2}\|\widehat{t}\|_{\ell^2}^2\right) d\mu(t).\end{aligned}$$

This yields, again by using (2.7), that

$$\begin{aligned}c &= \mathbb{P}\left(\|\Psi(Ax + e) - \Psi(Ax)\| \geq \|x - x''\| - 2\eta_1\right) \\ &\geq \frac{1}{\sqrt{(2\pi\sigma^2)^{2m}}} \int_{\mathcal{B}(0,\varepsilon)} \exp\left(-\frac{1}{2\sigma^2}\|\widehat{t}\|_{\ell^2}^2\right) d\mu(t).\end{aligned} \tag{2.8}$$

We observe that as $\sigma \to 0$, the lower bound in (2.8) tends to 1. Thus, if we choose $\sigma > 0$ small enough we get that $c \geq 1 - \epsilon$. □

The above theorem demonstrates that stable neural networks can only be obtained by training in ways that prohibit conditions (2.2) and (2.3) from occurring for small $\eta_1$ when $x$ and $x''$ are far apart. The AUTOMAP methodology, for example, does not do this. In fact its training process – which, in particular, is completely independent of $A$ – can encourage such conditions to occur. As we show in Table 2.1, the trained AUTOMAP reconstruction map does indeed satisfy (2.2) and (2.3) for a triple $\{x, x', x''\}$, where $\|x - x''\| \gg \eta_1$, and is therefore unstable in Fig. 2.1. By contrast, "handcrafted" methods have mechanisms for balancing the performance-stability trade-off when the matrix $A$ satisfies particular conditions. Fig. 2.2 and Fig. 2.6 demonstrate the robustness of such methods, which is described in more detail in Section 2.2.3.

| Condition | Quantity | $j=1$ | $j=2$ | $j=3$ |
|---|---|---|---|---|
| | $\eta_j$ | 4.9 | 4.9 | 4.9 |
| (2) | $\|\Psi(Ax)-x\|_{\ell^2}+\|\Psi(Ax'_j)-x''_j\|_{\ell^2}$ | 4.8 | 4.8 | 4.8 |
| (3) | $\|Ax-Ax'_j\|_{\ell^2}$ | 2.0 | 3.0 | 4.2 |
| (4) | $\|\Psi(Ax+e_j)-\Psi(Ax)\|_{\ell^2}$ | 13.7 | 23.9 | 35.7 |
| | $\|x-x''\|_{\ell^2}-\eta$ | 12.5 | 22.6 | 34.3 |
| | Sharpness ratio | 1.11 | 1.06 | 1.04 |
| Other quantities | $\|x'\|_{\ell^2}$ | 25.5 | 25.1 | 24.9 |
| | $\|Ax'\|_{\ell^2}$ | 12.1 | 11.9 | 11.9 |

Table 2.1: **Verifying that Theorem 2.2.1 predicts AUTOMAP's instability**. The instabilities of AUTOMAP shown in Fig. 2.1 are explained by Theorem 2.2.1. To see this, consider the image $x$, measurement matrix $A$ and perturbations $e_j = Ar_j$, $j = 1, 2, 3$, from Fig. 2.1, and let $\Psi\colon \mathbb{C}^m \to \mathbb{C}^N$ be the AUTOMAP reconstruction map. To apply Theorem 2.2.1, we now choose suitable values for $x' = x'_j$, $x'' = x''_j$ and $\eta = \eta_j$ to verify conditions (2.2)–(2.4). We set these as follows: $x'_j = x + r_j$, $x''_j = \Psi(Ax'_j)$ and $\eta_j = \|\Psi(Ax) - x\|_{\ell^2}$. We also define Sharpness ratio $= \frac{\|\Psi(Ax+e_j)-\Psi(Ax)\|_{\ell^2}}{\|x-x''_j\|_{\ell^2}-\eta}$, and report its value in the table. Note that a sharpness ratio $\geq 1$ means that condition (2.4) holds. This table confirm conditions (2.2)–(2.4) of Theorem 2.2.1 for these choices of $x'_j$, $x''_j$ and $\eta_j$ and with $e_j$ chosen as above. Further, the sharpness ratio is close to one in all cases, meaning that the estimate instability estimate of Theorem 2.2.1, given by the right-hand side of (2.4), very precisely describes the true instability seen in Fig. 2.1.

**Theorem 2.2.1 provides conditions for instability, regardless of the imaging model**

Theorem 2.2.1 provides *sufficient conditions* for instabilities, along with various probabilistic statements quantifying how likely these instabilities are. In particular, the theorem states that if a neural network $\Psi$ (which is, by definition, a continuous map) satisfies conditions (2.2) and (2.3), and additionally, $\|x - x''\| \gg \eta_1$, then $\Psi$ must be unstable. Moreover, even if $m \geq N$ and $A$ is well-conditioned, neither of the conditions (2.2) and (2.3) necessitate that $\|x - x''\|$ be small. Thus, Theorem 2.2.1 provides sufficient conditions for instabilities regardless of the imaging modality, and, in particular, regardless of whether or not $A$ has a non-trivial null space.

On the other hand, instabilities are more likely to arise when $A$ has a non-trivial null space. Indeed, suppose that $A \in \mathbb{C}^{m\times N}$ has a non-trivial null space $\mathcal{N}(A) \subset \mathbb{C}^N$ and let $x$ be an image satisfying (2.2). Since $A$ has a non-trivial null space, there are many, in fact, infinitely-many, vectors $x'$ of the from $x' = x + v$ for which $\|Ax - Ax'\| = \|Av\| \leq \eta_2$. In other words, there are many ways in which the condition $\|Ax - Ax'\| \leq \eta_2$ can be achieved. If just one of those vectors $x'$ yields $\|\Psi(Ax') - x''\| < \eta_1$ for some $x''$ not too close to $x$ (i.e. $\|x - x''\| \gg \eta_1$), then by Theorem 2.2.1 instabilities arise. The key point is that these conditions are extremely weak ones. If there is a non-trivial null space, there are many ways in which (2.2)–(2.3) can occur. In particular, the existence of a single triple $\{x, x', x''\}$ satisfying (2.2)–(2.3) implies the existence

of infinitely-many other triples for which these conditions also hold. In other words, there are infinitely-many images for which instability occurs. In fact, item $(3)$ of Theorem 2.2.1 shows that the probability of instabilities occurring can become arbitrarily close to 1, even while the noise level of the corresponding perturbations stays bounded.

**AUTOMAP is susceptible to the instability phenomenon of Theorem 2.2.1**

As demonstrated in Table 2.1, AUTOMAP satisfies the conditions of the theorem for the example shown in Fig. 2.1. A key point here, as indicated by (2.3), is that AUTOMAP yields many 'candidate' images $x^j$ for which (2.2) may hold. However, AUTOMAP has no mechanism to simultaneously prevent (2.3) and the condition $\|x''-x\| \gg \eta_1$ from occurring. As consequence, one observes the instabilities and inconsistent generalization seen in our experiments.

### 2.2.3 Theorem 2.2.1 in the broader context – The performance stability trade-off

Theorem 2.2.1 is true for any reconstruction method that is continuous, and it is therefore relevant to all methods that construct neural network reconstruction maps (which are continuous by definition). In this section, we discuss the consequences of this theorem, focusing on item (A), the performance stability trade-off and the need for conditions on the matrix $A \in \mathbb{C}^{m\times N}$. We start by recalling this trade-off.

**Balancing the trade-off between stability and performance for undersampled acquisitions requires conditions on $A$**

Let $x' = x''$ in Theorem 2.2.1 and observe that the conditions (2.2) and (2.3) imply that $\Psi$ very accurately recovers $x$ and $x'$. If at the same time $\|x-x'\| \gg 2\eta_1$ and $\|Ax-Ax'\| < \eta_2$, it follows from (2.4) that $\Psi$ is unstable. Thus if $\Psi$ overperforms on two samples $x$ and $x'$ (potentially in the training set), it is susceptible to instabilities if these images are mapped to roughly the same measurements.

A pertinent question is under which conditions will there exist images $x$ and $x'$ causing instabilities. We shall see that this relates to the sampling matrix $A$. For simplicity we suppose throughout that $\eta = \eta_2 = 2\eta_1$, and recall that $\Psi$ is unstable whenever $\|Ax - Ax'\| \leq \eta$ and $\|x - x'\| \gg \eta$. That is, we need two images $x$ and $x'$, such that

$$\|A(x - x')\| \ll \|x - x'\|. \tag{2.9}$$

It is important to note that there are scenarios where no such pair $x$ and $x'$, satisfying (2.9), exist. Indeed, let $m \geq N$, $\|\cdot\| = \|\cdot\|_{\ell^2}$ and suppose that the singular values of $A$ are centered around 1. Then $\|A(x-x')\|_{\ell^2} \approx \|x-x'\|_{\ell^2}$ for all $x, x' \in \mathbb{C}^N$ and, as a result, no method which

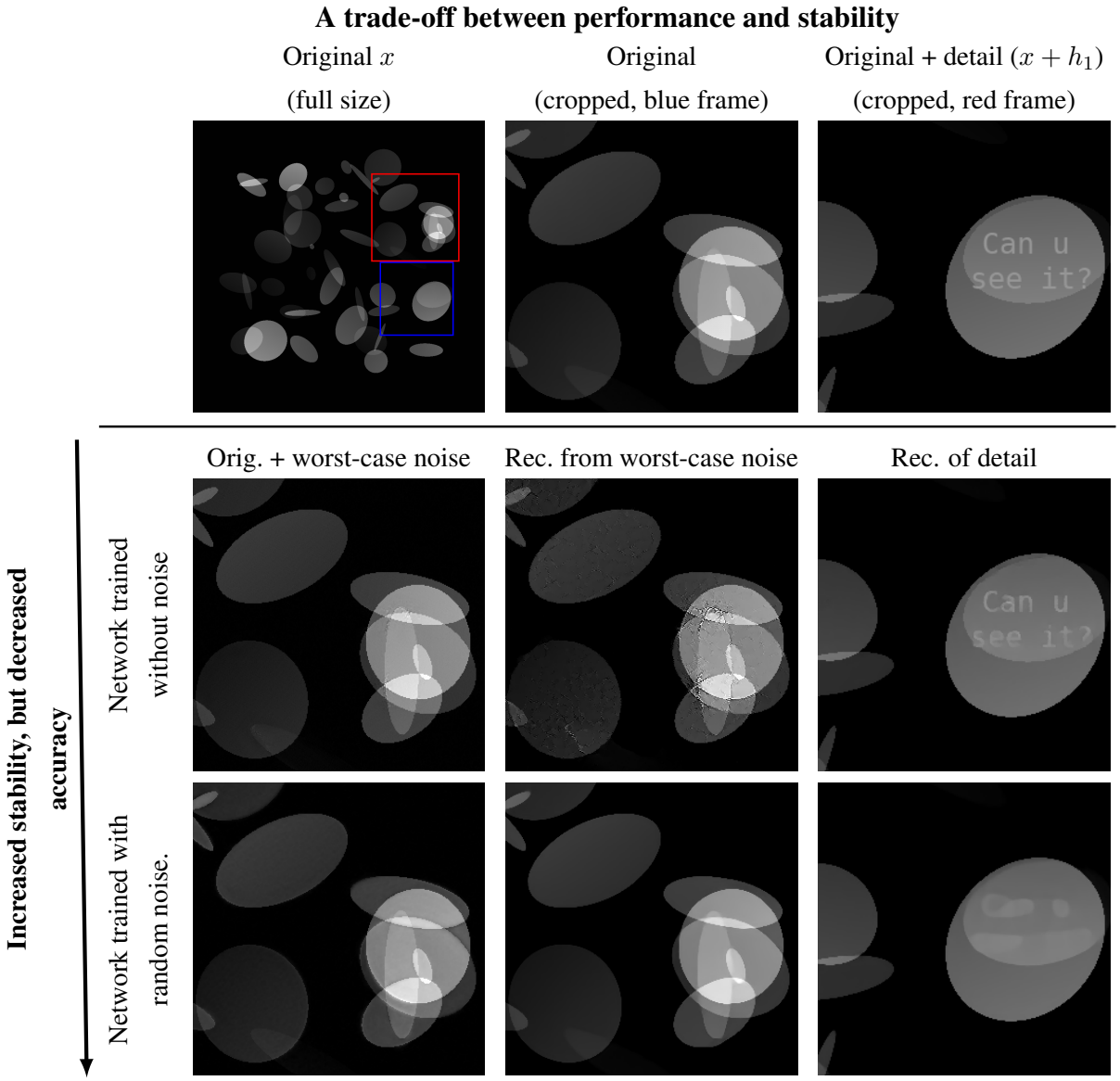


Figure 2.5: **Finding the right balance between performance and stability is delicate**. Theorem 2.2.1 sheds light on a universal trade-off between performance and stability for undersampled image acquisitions. This figure (from [46], Fig. 4) tries to elucidate this phenomenon by testing the performance and stability of two trained neural networks. The first network, whose reconstructions can be seen in the middle row, is trained on noiseless measurements of ellipses images. This makes the neural network susceptible to tiny perturbations, either in the form of a worst-case perturbation (middle column) or in the form of a tiny detail (right column) in the image. As we can see from the figure, this network can reconstruct the tiny "can u see it?" detail $h$ with high accuracy (it has high performance), but it is also unstable with respect to worst-case perturbations. The second network, whose reconstructions can be seen in the last row, is trained with Gaussian noise added to the measurements during the training process. This makes the network fairly stable with respect to worst-case perturbations, but also less accurate, as it is insensitive to the tiny detail $h$, and washes this detail out in its reconstruction (low performance). Thus, finding the right balance between performance and stability for undersampled image acquisitions is key.

accurately recover $x$ and $x'$, are susceptible to instabilities due to overperformance. This is one of the reason methods such as SENSE [151], works so well for moderate acceleration factors.

On the other hand, if $A$ is ill-conditioned, or does not have full rank, it is trivial to find two vectors $x$ and $x'$, satisfying $\|A(x-x')\| \ll \|x-x'\|$. Moreover, (as discussed previously) these scenarios must also be considered to attain the acceleration factors of standard imaging setups. Thus, to have any possibility of successful recovery when $A$ satisfies (2.9) it is necessary to restrict the set of images one wants to recover well to some class $\mathcal{M}_1 \subset \mathbb{C}^N$ for which

$$\|x - x'\| \leq C\|A(x - x')\| \quad \text{for all } x, x' \in \mathcal{M}_1 \subset \mathbb{C}^N, \tag{2.10}$$

for some constant $C > 0$ that is not too small. When this holds, one avoids the scenario (2.9) that leads to instabilities. Notice that for a not unreasonably small constant $C$, the condition (2.10), avoids the flaw in (2.9) leading to the potential unstable reconstruction of the elements $x$ and $x'$. In particular, it is also clear that for a fixed set $\mathcal{M}_1$, (2.10) cannot hold for arbitrary $A$. Therefore, conditions on $A$ are needed to ensure stable recovery.

Such conditions have been established in [21]. The lower and upper RIP directly relate to (2.10), for the lower RIP, and the upper RIP probihits (2.9) from occurring. Assuming the lower and upper RIP yields a robust and instance optimal decoder, as proven in [Theorem 9, [21]].

This conclusion stands in stark contrast to the AUTOMAP methodology. As discussed in Section 2.2.4, AUTOMAP strives to learn from "arbitrary acquisition strategies" whenever the data have "low-dimensional features". It, therefore, actively seeks to avoid any conditions on the acquisition matrix, but in doing so, becomes highly susceptible to instabilities.

### Sparse regularization achieves stability and performance through compressed sensing theory

Having shown the need for such conditions, we now illustrate through the example of sparse regularization how suitable conditions on the acquisition matrix $A$ can be successfully exploited to guarantee both performance and stability. This is achieved via the theory of compressed sensing (see [74] for an in-depth introduction).

The relevant condition on $A$ in this case is the following:

**Definition 2.2.3** (Robust Null Space Property)**.** A matrix $A \in \mathbb{C}^{m\times N}$ satisfies the *robust Null Space Property (rNSP)* of order $1 \leq s \leq N$ with constants $0 < \upsilon < 1$ and $\gamma > 0$ if

$$\|P_\Omega x\|_{\ell^2} \leq \frac{\upsilon}{\sqrt{s}}\|P_{\Omega^c} x\|_{\ell^1} + \gamma\|Ax\|_{\ell^2},$$

for all $x \in \mathbb{C}^N$ and any $\Omega \subseteq \{1, \ldots, N\}$ with $|\Omega| \leq s$. Here $\Omega^c = \{1, \ldots, N\} \setminus \Omega$.

Recall that a vector $x \in \mathbb{C}^N$ is *$s$-sparse* if it has at most $s$ nonzero components. A consequence of the rNSP is that the difference between two $s$-sparse vectors cannot be close to the null space

of $A$. Indeed, the rNSP implies that

$$\|x - x'\|_{\ell^2} \leq D\|Ax - Ax'\|_{\ell^2}, \quad \text{for all } s\text{-sparse vectors } x \text{ and } x'. \tag{2.11}$$

where the constant $D = (3 + \upsilon)\gamma/(1 - \upsilon)$ (see [74, Thm. 4.25]).

An implication of the rNSP is that compressed achieves stable and accurate recovery for vectors that are approximately $s$-sparse in some unitary (sparsifying) transform $V \in \mathbb{C}^{N\times N}$. The following is a typical result in the literature. For convenience we consider the QCBP and SR-LASSO decoders only, see Section 2.3.1. Other decoders can also be addressed (see [5, Chpt. 6]):

**Theorem 2.2.4** (rNSP implies a balance between stability and performance)**.** *Suppose that the matrix $AV^* \in \mathbb{C}^{m\times N}$ satisfies the robust null space property of order $s$, with constants $\gamma > 0$ and $0 < \upsilon < 1$. Then for any $x \in \mathbb{C}^N$ and $y = Ax + e$, with $e \in \mathbb{C}^m$ we have that*

$$c^\sharp \in \operatorname{argmin}_{z\in\mathbb{C}^N}\|z\|_{\ell^1} \quad \textit{subject to} \quad \|AV^*z - y\|_{\ell^2} \leq \delta, \tag{2.12}$$

*satisfies*

$$\|x - V^*c^\sharp\|_{\ell^2} \leq \frac{C_1}{\sqrt{s}}\sigma_s(Vx)_1 + D_1\delta,$$

*if $\|e\|_{\ell^2} \leq \delta$, and*

$$\widetilde{c} \in \operatorname{argmin}_{z\in\mathbb{C}^N}\lambda\|z\|_{\ell^1} + \|AV^*z - y\|_{\ell^2}, \tag{2.13}$$

*satisfies*

$$\|x - V^*\widetilde{c}\|_{\ell^2} \leq 2\frac{C_2}{\sqrt{s}}\sigma_s(Vx)_1 + \left(\frac{C_2}{\sqrt{s\lambda}} + D_2\right)\|e\|_{\ell^2}$$

*if $\lambda \leq C_2/(D_2\sqrt{s})$, where $\sigma_s(x)_1 = \inf\{\|x - z\|_{\ell^1} : z \text{ is } s\text{-sparse}\}$ denotes the $\ell^1$-distance to an $s$-sparse vector, and where the constants $C_1, D_1, C_2, D_2 > 0$ only depend on $\gamma$ and $\upsilon$.*

**Theorem 2.2.5** (Theorem 4.22 [74], $p = 2$)**.** *Suppose that the matrix $A \in \mathbb{C}^{m\times N}$ satisfies the robust null space property of order $s$ with constants $0 < \upsilon < 1$ and $\gamma > 0$. Then, for any $x \in \mathbb{C}^N$ a solution $c^\sharp$ to* (2.12) *with $y = Ax + e$, and $\|e\|_{\ell^2} \leq \delta$ approximates the vector $x$ with*

$$\|x - V^*c^\sharp\|_{\ell^2} \leq \frac{C_1}{\sqrt{s}}\sigma_s(Vx)_1 + D_1\delta,$$

*for some constants $C_1, D_2 > 0$ depending only on $\upsilon$ and $\gamma$.*

**Theorem 2.2.6** (Theorem 6.5 [5])**.** *Suppose that the matrix $A \in \mathbb{C}^{m\times N}$ satisfies the robust null space property of order $s$ with constants $0 < \upsilon < 1$ and $\gamma > 0$. Then, for any $x \in \mathbb{C}^N$ a solution $\widetilde{c}$ to* (2.13) *with $y = Ax + e$, and $\|e\|_{\ell^2} \leq \delta$*

$$\|x - V^*\widetilde{c}\|_{\ell^2} \leq 2\frac{C_2}{\sqrt{s}}\sigma_s(Vx)_1 + \left(\frac{C_2}{\sqrt{s\lambda}} + D_2\right)\|e\|_{\ell^2}$$

*if $\lambda \leq C_2/(D_2\sqrt{s})$, and where the constants $C_2, D_2 > 0$ only depend on $\gamma$ and $\upsilon$.*

*Proof of Theorem 2.2.4.* The statement for the minimizer $c^\sharp$ can be found as Theorem 4.22 in [74], while the statement for the minimizer $\tilde{c}$ can be found as Theorem 6.5 in [5]. □

The key point is that all vectors that are (approximately) $s$-sparse in some transform $V$ – as is the case for natural images in, say, a wavelet transform – have guaranteed reconstruction performance, as measured through the $\sigma_s(Vx)_{\ell^1}/\sqrt{s}$ term. Further, one also has stability, as indicated by the term $D_1\delta$ in the error bound. Notably, the method only yields stable and accurate recovery of vectors that are (approximately) $s$-sparse under the assumption that $A$ has the rNSP. This highlights an important example of how conditions on $A$ can be used to ensure performance.

### 2.2.4 The theoretical premise for AUTOMAP is not satisfied in undersampled acquisitions

In this section we explain that the main premise for the theoretical analysis of a robust and stable fully learned decoder according to the AUTOMAP methodology [192], is not satisfied in undersampled acquisitions. Hence, any further derivations and conclusions, for example, stability and robustness, based on this assumption do not hold. Furthermore, we explain why the training process in the AUTOMAP methodology cannot protect against these instabilities. However, well-known classical methods, such as compressed sensing methods under certain conditions on the sampling matrix, remain stable and accurate and avoid the pitfalls of Theorem 2.2.1. The essence of the AUTOMAP methodology is best summarised by a quote from [192]:

> *"AUTOMAP provides a new paradigm for image reconstruction that learns a reconstruction function for arbitrary acquisition strategies conditioned upon low-dimensional features of real-world data to improve artefact reduction and reconstruction accuracy for noisy and undersampled acquisitions."*

The two keywords here are *arbitrary acquisition strategies* and *undersampled acquisitions*. These keywords are important when examining the theoretical premise for the AUTOMAP methodology, which is the following:

> (**The theoretical premise of AUTOMAP**) "*We denote the data* $\{x_i, y_i\}_{i=1}^n$*, where for the* $i$*th observation* $y_i$ *indicates a* $n \times n$ *set of input parameters, and* $x_i$ *indicates the* $n \times n$ *real, underlying images. We assume that (1) there exists a unknown smooth and homeomorphic function*
>
> $$f : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}, \text{ such that } x = f(y), \tag{2.14}$$
>
> *and (2)* $\{y_i\}_{i=1}^n$*,* $\{x_i\}_{i=1}^n$ *lie on unknown smooth manifolds* $\mathcal{Y}$ *and* $\mathcal{X}$*, respectively.*"

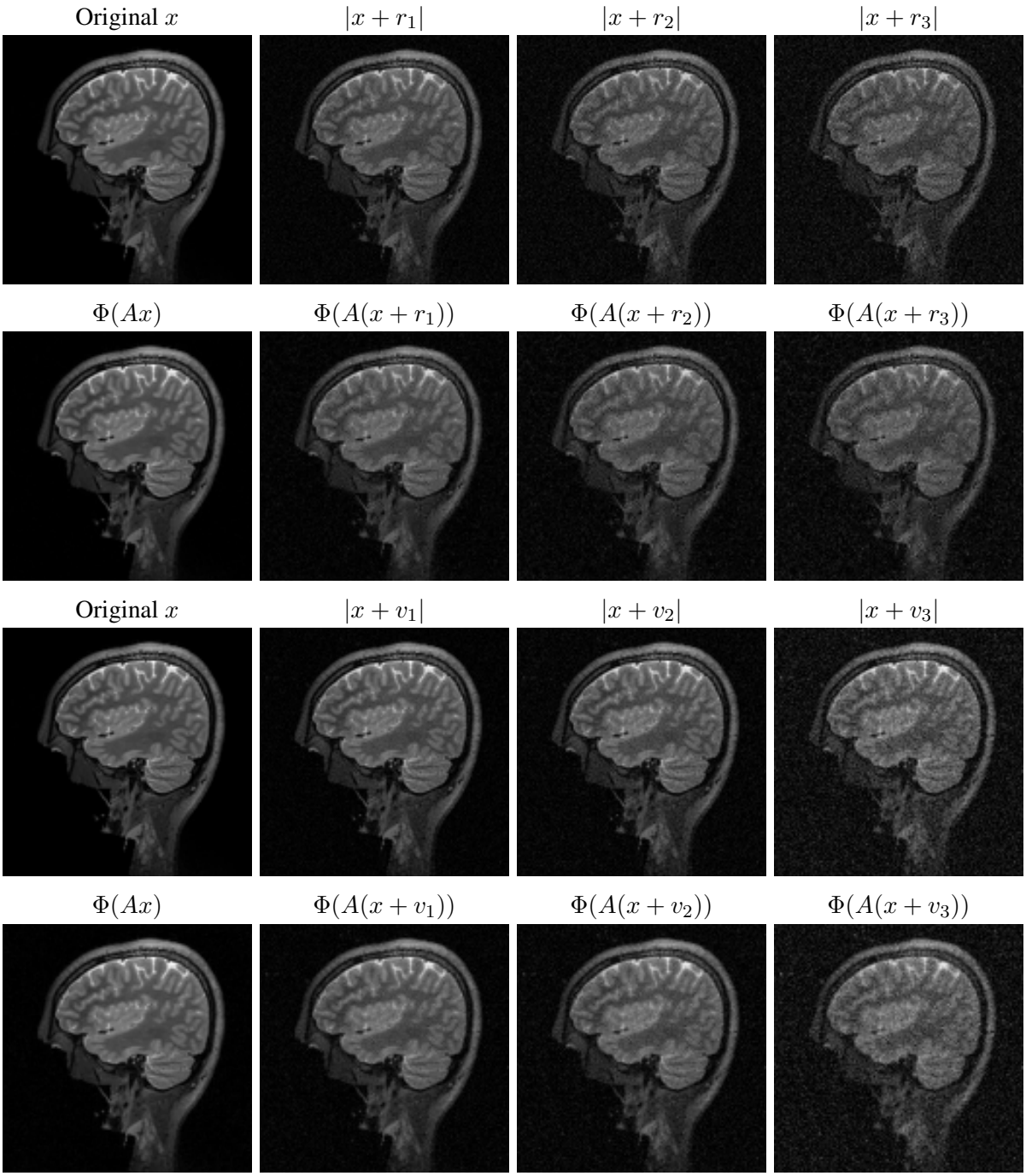


Figure 2.6: (**Standard method is stable to worst-case perturbations**). We compute worst-case perturbations for a standard method $\Phi\colon \mathbb{C}^m \to \mathbb{C}^N$ based on a LASSO decoder with wavelet reconstruction. In this experiment $A \in \mathbb{C}^{m\times N}$ is the same sampling operator as used in Fig. 2.1. Top row: The perturbations $r_1, r_2, r_3 \in \mathbb{C}^N$ used to simulate worst-case effect for AUTOMAP in Fig. 2.1, are added to the image $x \in \mathbb{C}^N$. Second row: The standard method $\Phi$'s reconstruction of the images $x + r_j$, $j = 0, 1, 2, 3$ with $r_0 = 0 \in \mathbb{C}^N$, from the data $A(x + r_j)$, $j = 0, 1, 2, 3$. Third row: We compute perturbations $v_1, v_2, v_3 \in \mathbb{C}^N$ meant to simulate worst-case effect for the standard method $\Phi$. All $v_j$'s are computed so that $\|Av_j\|_{\ell^2} > \|Ar_j\|_{\ell^2}$ for $j = 1, 2, 3$. Bottom row: The standard method $\Phi$'s reconstruction of the images $x + v_j$, $j = 0, 1, 2, 3$ with $v_0 = 0 \in \mathbb{C}^N$, from the data $A(x + v_j)$, $j = 0, 1, 2, 3$. See Section 2.3.4 for details on how the $v_j$'s are computed.

**Remark 2.2.7.** In the above quote, we have swapped the letters $x$ and $y$ to make it consistent with our notation, in which $y$ denotes measurements and $x$ denotes an image. In [192] these letters have the opposite meaning. Note also that the use of the letter $n$ in both $\mathbb{R}^{n^2}$ and the collections $\{x_i\}_{i=1}^n$, $\{y_i\}_{i=1}^n$ is not a misprint, but a clash of notation in [192].

The crucial assumption in the quote above is that the function $f$ is a homeomorphism, and this is the function one tries to learn. However, this assumption fails to hold in all undersampled acquisition scenarios. Indeed, if $f$ is a homeomorphism it has a smooth inverse function $g : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ (that is $g \circ f(y) = y$), and hence the noise-free forward acquisition process in the apparatus, for example the MRI machine, is given by $g$, since $y = g(x)$ when $x = f(y)$. The function $g$ is also bijective and it has an inverse map given by $f$. However, as we now explain, this fails to be the case in general, thus defying the whole principle of "arbitrary acquisition strategies" and "undersampled acquisitions".

Indeed, for the single-coil MRI model in the AUTOMAP paper [192], the noiseless forward model is given by

$$y = P_\Omega F x, \tag{2.15}$$

where $P_\Omega \in \mathbb{R}^{n^2 \times n^2}$ is the projection

$$(P_\Omega z)_i = \begin{cases} z_i & i \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

with $|\Omega| = m < n^2$, and $F \in \mathbb{C}^{n^2 \times n^2}$ is the two-dimensional discrete Fourier transform. Now, since $|\Omega| = m < n^2$ the matrix $P_\Omega F$ has a non-trivial null space. Hence, the map $g(x) = P_\Omega F x$ does not have an inverse and consequently the assumption above pertaining to the existence of the homeomorphism $f$ fails.

### Extension to parallel MRI

It is clear from the above quote that the dimensions of the images and measurements are $n \times n$. Hence, the intended model in the AUTOMAP methodology does not allow for parallel MRI. Indeed, in parallel MRI we can have more measurements $m$ than pixels $N$ in the image. However, we will show that even if we alter the setup in the AUTOMAP methodology to accommodate parallel MRI, the underlying assumption on the existence of the homeomorphism $f$ cannot be satisfied when the sampling matrix $A$ is rank deficient. Note that rank deficiency of $A$ is typical in accelerated acquisition, as we highlight below.

**Remark 2.2.8. The homeomorphism $f$ in the AUTOMAP framework cannot exist when $A$ is rank deficient.** Suppose first that $A \in \mathbb{R}^{m \times N}$ is a rank deficient sampling matrix and that $m > N$. Thus, the measurements are given by $y = Ax$, where $x$ denotes the image of interest. To fit this into the AUTOMAP framework we let $n^2 = m$ (we assume for convenience that $m$

is a square number). Given arbitrary data $\{y_i\}$ and $\{x_i\}$, where $y_i = Ax_i$, we interpret the $x_i$s (that originally have dimension $N < m = n^2$) as embedded in the larger space $\mathbb{R}^{n^2}$. Hence, the existence of a homeomorphism $f : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ such that $f(y_i) = x_i$ for all $i$ means that there is a bijection $g : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ such that $g(x_i) = y_i$ for all $i$. Since this holds for arbitrary data, the matrix $A$ must have full rank, and we have reached a contradiction.

With this in hand, we now demonstrate numerically that standard matrices in parallel MRI can have non-trivial numerical[1] null spaces. This means that for many of the most interesting imaging scenarios in undersampled parallel MRI, the homeomorphism $f$ will not exist.

To demonstrate that $A$ in (2.20) is rank deficient when using state-of-the-art acceleration factors, we have generated multiple matrices $A$ given by (2.20) and computed their rank. The matrices were generated using the acceleration factors and sampling patterns from the fastMRI challenge [188] and using a different number of coils $c$. For each of the matrices, the sensitivity profiles $S_i$ were generated using code from [87], where the coils are arranged in a circle with centres that are equidistant to the origin. As we can see from Table 2.2 the numerical rank of $A$ is always not full. As mentioned above, this implies that the homeomorphism $f$ will cease to exist when using standard floating-point precision.

### 2.2.5 Fully learned neural networks and AUTOMAP

Recall the standard definition of a neural network. A function $\Psi\colon \mathbb{R}^n \to \mathbb{R}^p$ taking input $u$ is said to be an $L$-layer neural network if it has the form

$$\begin{aligned} a^0 &= u \in \mathbb{R}^n \\ a^l &= \rho^l(W^l a^{l-1} + b^{l-1}) \in \mathbb{R}^{n_l}, \quad l = 1, \dots L, \\ \Psi(u) &= a^L, \end{aligned}$$

where $n = n_0$, $p = n_L$, $W^l \in \mathbb{R}^{n_l \times n_{l-1}}$, $b^l \in \mathbb{R}^{n_l}$, and $\rho^l\colon \mathbb{R} \to \mathbb{R}$ is a *continuous* function acting component-wise on a vector. Here the vectors $b^l$ are the *biases*, the $W^l$'s are called the *weights*, and $n_l$'s are called the *widths* of the networks. A specific choice of dimensions $\mathbf{n} = (n_0, \dots, n_L)$ and activation functions $\boldsymbol{\rho} = (\rho^1, \dots, \rho^L)$ is called the *architecture* of the network. We let $\mathcal{NN}_{\mathbf{n},\boldsymbol{\rho}}$ denote the set of all neural networks with a given architecture.

We note that if we let $\rho^1 = \cdots = \rho^{L-1}$ all be the same, and let $\rho^L = \mathrm{Id}$ be the identity mapping, then we recover the definition often used by mathematicians [148]. However, such a definition would exclude the AUTOMAP architecture, introduced below, and many other architectures used in practice. The continuity assumption on $\rho^l$ is used by the AUTOMAP architecture (and most other neural networks used in practice), and is necessary to get the probabilistic statement in Theorem 2.2.1. However, if this is relaxed, the neural network could be discontinuous and

[1]The word "numerical" here means that if one computes the rank of these matrices in the standard way (by counting the number of singular values above a threshold based on machine precision), they are rank-deficient.

**Computed rank of matrices used in parallel MRI for different acceleration factors**

| **Number of coil** | **4×acceleration** | | | **8×acceleration** | | |
|---|---|---|---|---|---|---|
| **sensitivity profiles** | Numerical rank | $m$ | $N$ | Numerical rank | $m$ | $N$ |
| 2 | 7533 | 8192 | 16384 | 4224 | 4352 | 16384 |
| 4 | 12942 | 16384 | 16384 | 7281 | 8704 | 16384 |
| 6 | 14746 | 24576 | 16384 | 8641 | 13056 | 16384 |
| 8 | 14892 | 32768 | 16384 | 8701 | 17408 | 16384 |
| 10 | 13552 | 40960 | 16384 | 9443 | 21760 | 16384 |
| 12 | 13793 | 49152 | 16384 | 9242 | 26112 | 16384 |
| 16 | 13543 | 65536 | 16384 | 8983 | 34816 | 16384 |
| 20 | 13471 | 81920 | 16384 | 9000 | 43520 | 16384 |

Table 2.2: In parallel MRI, the acquisition matrix $A \in \mathbb{C}^{m\times N}$ is modeled as the block matrix in (2.20). We compute the rank of this matrix in single precision for the two equidistant sampling patterns from the fastMRI challenge [188] with 4 and 8 times acquisition speed-up (see Fig. 2.10) and for a different number of receiver coils $c$. For each receiver coil, we generate the sensitivity profiles $S_i$ using code from [87]. These sensitivity profiles correspond to a scanner where the coils are arranged in a circle with centres that are equidistant to the origin. As can be seen from the table, $A$ is rank deficient even when we use as many as 20 coils. For completeness, we recall that the rank of a matrix $A \in \mathbb{C}^{m\times N}$ is upper bounded by $\operatorname{rank}(A) \leq \min\{m, N\}$, and that the dimension of $A$'s null space is given by $\dim(\mathcal{N}(A)) = N - \operatorname{rank}(A)$. For details on how the rank is computed, we refer to Section 2.3.7.

hence make sudden jumps. This would imply instability; hence the continuity assumption is crucial if one seeks stable reconstruction.

### 2.2.6 Description of AUTOMAP's architecture

The AUTOMAP methodology allows for learning image reconstructions from Radon projections, spiral non-Cartesian Fourier measurements, misaligned Fourier measurements and undersampled Fourier measurements [192]. In Fig. 2.7 we report the architecture used in the latter experiments, see also Fig. 2 in [192]. This architecture along with weights and biases for a pretrained network were obtained through communicating with the authors of [192]. Using the fact that the operations "subtract mean", "reshape", and "convolution" are all linear mappings, from Fig. 2.7 one can see that AUTOMAP's architecture is given by

$$\tilde{\mathbf{n}} = [2m, 25000, 64N^2, 64N^2, N^2] \text{ and } \tilde{\boldsymbol{\rho}} = [\tanh, \tanh, \text{ReLU}, \text{Id}].$$

Thus, all networks in the set $\mathcal{NN}_{\tilde{\mathbf{n}},\tilde{\boldsymbol{\rho}}}$ have the AUTOMAP architecture.

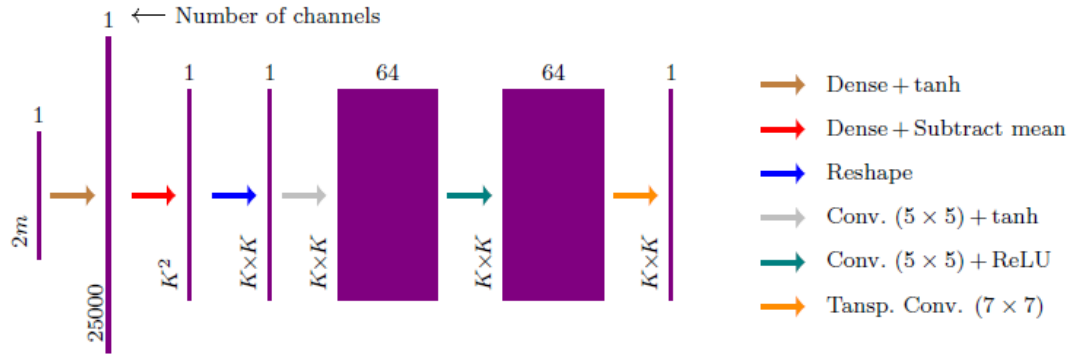


Figure 2.7: (**AUTOMAP's architecture**). The numbers written vertically indicate the dimension of the tensors in one (or two) of the dimensions, whereas the numbers on top refer to the last dimension (often called channel dimension). The arrows indicate which operations are used to map between the tensors. Here $N = K^2 = (128)^2$ and $m = 9855$. The input dimension is $2m$ as a complex vector $y \in \mathbb{C}^m$ is mapped to a vector in $\mathbb{R}^{2m}$, by splitting $y$ into its real and imaginary components. The parenthesis after the "Conv." is meant to indicate the size of the kernel. All the convolutions are zero-padded in order to not reduce the dimension.

### 2.2.7 The training process of AUTOMAP typically yields a small training error

The training process of AUTOMAP for undersampled acquisitions is as follows. Given a set of medical training images $\{x^j\}_{j=1}^T \subset \mathbb{C}^N$, and noisy measurements $y^j = Ax^j + e^j$, $j \in \{1, \dots, T\}$, the neural network obtained via domain-transform manifold learning is obtained by minimizing the following objective function

$$\widehat{\Psi} \in \underset{\widetilde{\Psi} \in \mathcal{NN}_{\tilde{\mathbf{n}},\tilde{\boldsymbol{\rho}}}}{\operatorname{argmin}} \frac{1}{T} \sum_{j=1}^{T} \frac{1}{2} \|x^j - \widetilde{\Psi}(y^j)\|_{\ell^2}^2 + \lambda' J(\widetilde{\Psi}, y^j), \tag{2.16}$$

using the RMSProp optimization algorithm. Note that $\lambda' = 0.0001$ in (2.16) and $J : \mathcal{NN}_{\mathbf{n},\boldsymbol{\rho}} \times \mathbb{C}^m \to \mathbb{R}_{\geq 0}$ computes the $\ell^1$-norm of the output from $\widetilde{\Psi}$'s ReLU activation function, when applied to the input $y^j$. This training procedure results in a neural network $\Psi$ (the AUTOMAP network) with a very small mean-squared-error (MSE) of the reconstructed images, both for the images in the training set and the validation set, (as documented in Extended Data Fig. 3c in [192]). In other words, the training procedure in AUTOMAP produces a network $\Psi$ that satisfies

$$\|\Psi(y^j) - x^j\|_{\ell^2}^2 \leq \delta_j, \qquad \delta_j \geq 0, \quad j = 1, \dots, T, \tag{2.17}$$

for small values $\delta_j$.

## 2.3 Methods

The methods section contains the information needed to state our theoretical results, explain why standard non-AI based methods are stable and document the numerical tests. In Section 2.3.1 we introduce basic concepts used in undersampled acquisition and compressive imaging and standard methods for solving these problems. In Section 2.3.3 we consider computational aspects, needed to reproduce our results.

### 2.3.1 Compressive imaging and undersampled acquisition - Modeling the measurement process

Compressive imaging pertains to the accurate and stable reconstruction of images from undersampled measurements. Typically, one models this problem as the discrete linear problem:

$$\text{Given the measurements } y = Ax + e, \text{ recover } x. \tag{2.18}$$

Here $y \in \mathbb{C}^m$ is the data acquired by the sensing device, $A \in \mathbb{C}^{m\times N}$ is a model of the measurement process (known as the *measurement matrix*), $x \in \mathbb{C}^N$ is a vectorized version of the image, and $e \in \mathbb{C}^m$ is a (unknown) noise term, accounting for inaccuracies in the measurement model, measurement noise and other artefacts.

For undersampled acquisitions, an overall goal is to not sample more data than necessary. To model the degree of undersampling one uses a mask (projection) operator $P_\Omega$ defined as follows. For a given sampling mask (set) $\Omega \subsetneq \{1,\dots,N\}$, of cardinality $|\Omega| = m < N$, we let $P_\Omega \in \mathbb{C}^{m\times N}$, denote the projection which extracts the elements of a vector, indexed by $\Omega$. That is, for $x \in \mathbb{C}^N$, we have $(P_\Omega x)_i = x_{\Omega(i)}$ for $i = 1,\dots,m$, where $\Omega(i)$, denotes the $i$'th element in $\Omega$, using the natural ordering. In Fig. 2.8, we show the sampling mask $\Omega$ used in [192].

In single-coil MRI (which is the model used in [192]) the standard model for the sensing matrix $A$ is

$$A = P_\Omega F, \tag{2.19}$$

where $F \in \mathbb{C}^{N\times N}$ is a 2-dimensional discrete Fourier transform (DFT) matrix. We highlight that in this model one always has $m < N$ measurements, since it is assumed that $\Omega$ is a strict subset of $\{1,\dots,N\}$.

In multi-coil (parallel) MRI, multiple receiver coils simultaneously acquire $k$-space data. To acquire as much information as possible, each of the receiver coils is adjusted to be sensitive to a limited spatial region [53]. For an MRI scanner with $c$ coils, one models this by introducing a diagonal matrix $S_i \in \mathbb{C}^{N\times N}$, $i = 1,\dots,c$ for each coil. This matrix is usually called the *sensitivity matrix* or *sensitivity profile* of the coil. Given these sensitivity profiles, the full model

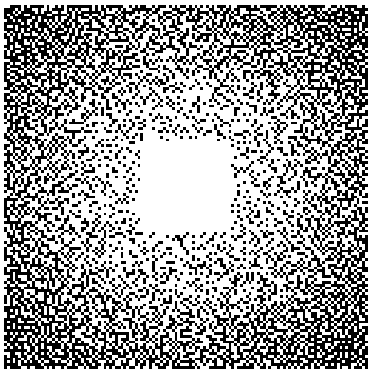

Figure 2.8: The sampling pattern $\Omega \subset \{1, \dots, N\}$, where $N = K \times K$, and $K = 128$. The white dots, corresponds to the frequencies we sample, whereas the black dots are the non-sampled frequencies.

for the acquisition matrix $A \in \mathbb{C}^{m \times N}$ is

$$A = \begin{bmatrix} P_\Omega F S_1 \\ \vdots \\ P_\Omega F S_c \end{bmatrix} \tag{2.20}$$

where $|\Omega| = m'$, $P_\Omega \in \mathbb{C}^{m' \times N}$ and $m = m'c$. Depending on the choice of $m'$ and $c$, the total number of measurements $m$ can exceed the number of pixels $N$ in the image, i.e., $m > N$. Furthermore, notice that since $|\Omega| = m' < N$, this is still a form of undersampled acquisition, as the acquisition process is involves less than full sampling in each coil.

### 2.3.2 Standard methods used for image reconstruction in this thesis

If $m \geq N$ and the matrix $A$ in (2.20) is well-conditioned, then it is natural to reconstruct the image $x$ in (2.18), using a least-squares estimator. That is, using $y \mapsto (A^*A)^{-1}A^*y$ as our reconstruction map. This approach has had tremendous empirical success and forms the basis for the classical SENSE reconstruction algorithm [151].

If, on the other hand, $A$ is ill-conditioned or has $m < N$ rows, e.g. as in (2.19), then the least-squares estimator mentioned above leads to poor performance. In such cases, it is customary to employ sparse regularization to achieve higher-quality reconstructions [71, 156]. Note that $A$ being ill-conditioned or having $m < N$ rows is often the most interesting acquisition model to consider, as it arises when one seeks to reduce the number of measurements taken, and thereby save acquisition time. Reducing acquisition times is one of the key goals of imaging research.

In this setting, standard methods for image reconstruction typically exploit the inherent sparsity of images in fixed transformed domains, such as wavelets or discrete gradients [79, 156]. These are sometimes referred to as "handcrafted" methods.

In this exposition, we focus on handcrafted methods using orthogonal wavelets, but other choices are certainly possible, see, e.g. [71, 156] for an overview. Let $V \in \mathbb{C}^{N \times N}$ denote a discrete wavelet transform with an orthogonal wavelet. Since we consider orthogonal wavelets, the matrix $V$ is unitary, i.e. its inverse $V^{-1} = V^*$ equals its adjoint. Sparse regularization

methods involve solving certain optimization problems – often termed decoders – that exploit sparsity in the given transform.

There are many ways to formulate these optimization problems, but typical examples include the quadratically-constrained basis pursuit (QCBP) problem

$$\min_{z\in\mathbb{C}^N} \|z\|_{\ell^1} \text{ subject to } \|AV^*z - y\|_{\ell^2} \leq \delta, \qquad \delta \in [0,\infty),$$

the unconstrained LASSO (U-LASSO) decoder

$$\min_{z\in\mathbb{C}^N} \|AV^*z - y\|_{\ell^2}^2 + \lambda\|x\|_{\ell^1}, \qquad \lambda \in (0,\infty),$$

the constrained LASSO (C-LASSO) decoder

$$\min_{z\in\mathbb{C}^N} \|AV^*z - y\|_{\ell^2}^2 \text{ subject to } \|z\|_{\ell^1} \leq \tau, \qquad \tau \in (0,\infty),$$

or the square-root LASSO (SR-LASSO) decoder,

$$\min_{z\in\mathbb{C}^N} \|AV^*z - y\|_{\ell^2} + \lambda\|x\|_{\ell^1}, \qquad \lambda \in (0,\infty). \tag{2.21}$$

Given a solution $\tilde{z}$ to one of these problems one computes an approximation to $x$ as $V^*\tilde{z}$.

### 2.3.3 Computational aspects

The purpose of this section is to document the many computational aspects of the paper, making this work reproducible. We document how the worst-case perturbations, seen in Fig. 2.1 and Fig. 2.6, are computed. We next show that the mean value that is used to produce the image-independent random Gaussian noise in Fig. 2.2 can be comprehensively changed, yet AUTOMAP still produces images with severe reconstruction artefacts. Next, we discuss how the standard method we use is designed and implemented. Finally, we explain how one can compute the rank of sampling matrices in imaging problems.

### 2.3.4 Computing worst-case perturbations

The worst-case perturbations shown in Fig. 2.1 and Fig. 2.6 are computed using the method developed in [6]. Given an almost everywhere differentiable function $f\colon \mathbb{C}^m \to \mathbb{C}^N$, this method is based on maximizing the objective function

$$Q(u) = \frac{1}{2}\|f(A(x+u)) - x\|_{\ell^2}^2 - \frac{\widetilde{\lambda}}{2}\|u\|_{\ell^2}^2, \quad \widetilde{\lambda} > 0, \tag{2.22}$$

where $A \in \mathbb{C}^{m\times N}$, as usual, denotes the model of the sampling process.

To maximize (2.22) we use a gradient-based method. We start by drawing the real and imaginary components of an initial perturbation $u_0 \in \mathbb{C}^N$ from a uniform distribution on $[0,1]$. We then

multiply the resulting $u_0$ by a scalar $\widetilde{\tau} > 0$, to adjust its magnitude to an appropriate size. Given the rescaled $u_0$, we perform the following iteration, known as *gradient ascent with momentum*,

$$\xi_{i+1} = \widetilde{\gamma}\xi_i + \widetilde{\eta}\nabla_u Q(u_i)$$
$$u_{i+1} = u_i + \xi_{i+1}$$

to find a $u$ maximizing (2.22). Here $\xi_0 = 0$ and the scalars $\widetilde{\gamma} > 0$ and $\widetilde{\eta} > 0$ are known as the *momentum* and *learning rate*, respectively. The number of iterations, $\bar{n} \in \mathbb{N}$, required to find a perturbation simulating worst-case effect is very dependent on the parameters and function involved, and is fine tuned for each image. In Table 2.3 we report the parameters used to produce the perturbations $r_j$ and $v_j$ in Fig. 2.1 and Fig. 2.6. These perturbations then are given by $r_j, v_j = u_{\bar{n}}$ for given $j \in \{1, 2, 3\}$. Where the norm of the perturbation increases as $j$ increases.

| Algorithm | $\widetilde{\lambda}$ | $\widetilde{\gamma}$ | $\widetilde{\eta}$ | $\widetilde{\tau}$ |
|---|---|---|---|---|
| LASSO | 0.01 | 0.9 | 0.01 | 1e-3 |
| AUTOMAP | 0.1 | 0.9 | 0.001 | 1e-5 |

Table 2.3: Parameters used to compute the perturbations $r_j$ and $v_j$, meant to illustrate worst-case effect in Fig. 2.1 and Fig. 2.6.

### 2.3.5 Computing image-independent perturbations

In Fig. 2.2, we demonstrate that the worst-case perturbations are image-independent and lies in ball with a certain magnitude. That is, we compute perturbations

$$e_j = e^1_{\text{pert},j} + e^2_{\text{pert},j}, \qquad j = 0, 1, 2, 3, 4$$

where $e^1_{\text{pert},j}$, $(j = 1, 2, 3)$ denote the worst-case perturbations computed in Fig. 2.1, and $e^2_{\text{pert},j}$ is random Gaussian noise. From the figure, we can see that these perturbations cause severe artefacts to AUTOMAP's reconstruction of the three brain images $x^{(1)}$, $x^{(2)}$ and $x^{(3)}$. In this experiment we have used the brain image from Fig. 2.1 to generate the worst-case perturbations $e^1_{\text{pert},j}$. Thus, a pressing question is, therefore, whether the perturbations $e_{\text{pert},j}$ are truly image-independent, since $e^1_{\text{pert},j}$ is generated from a brain image. To check this, we have generated a new set of perturbations

$$\hat{e}_j = \hat{e}^1_{\text{pert},j} + \hat{e}^2_{\text{pert},j}, \qquad j = 0, 1, 2, 3, 4$$

in Fig. 2.9, where $\tilde{e}^1_{\text{pert},0} = 0$ and the perturbations $\tilde{e}^1_{\text{pert},j}$ for $j = 1, 2, 3$ are worst-case perturbations computed for AUTOMAP from a knee MR image that is shown in the lower right corner of Fig. 2.9. As we can see from Fig. 2.9, AUTOMAP is also unstable with respect to the perturbations $\tilde{e}_{\text{pert},j}$, where the perturbations have been generated without any knowledge of the

image one wants to recover. We conclude from this that AUTOMAP is also highly susceptible to truly image-independent perturbations. To compute the perturbations $\tilde{e}^1_{\text{pert},j}$, $j = 1, 2, 3$, we used the algorithm described in Section 2.3.4, with the parameters for AUTOMAP as described in Table 2.3.

### 2.3.6 Details of the standard reconstruction method used for comparison

We now describe the standard method used in this work. A requirement for testing the stability of a reconstruction method with the algorithm described in Section 2.3.4 is the ability to compute the gradient of the reconstruction method efficiently. To this end, we implemented a standard reconstruction method based on sparse regularization in Tensorflow, thereby allowing efficient computation of the gradient. Our reconstruction method solves the SR-LASSO optimization problem using the Chambolle & Pock primal-dual iteration [36], with the additional shift updates introduced in [37, 96]. We describe this method in detail in the following.

We write $H = AV^*$, where $A \in \mathbb{C}^{m\times N}$ is a model of the sampling device and $V^* \in \mathbb{R}^{N\times N}$ is the inverse discrete wavelet transform corresponding to an orthogonal wavelet. In this work, we consider the DB2 wavelet. For an image $x \in \mathbb{C}^N$, with wavelet coefficients $c = Vx$, we then consider noisy measurements $y = Hc + e$, and seek to solve the optimization problem

$$\min_{z\in\mathbb{C}^N} \lambda\|z\|_{\ell^1} + \|Hz - y\|_{\ell^2}. \tag{2.23}$$

where $\lambda > 0$ is a regularization parameter.

| iter. | $\lambda$ | $\tau$ | $\sigma$ | Wavelet | DWT-levels |
|---|---|---|---|---|---|
| 1000 | 0.0001 | 3/5 | 3/5 | DB2 | 3 |

Table 2.4: Parameters for the unrolled neural network used in Fig. 2.3.

To define the primal dual iterations for (2.23), we introduce the functions $g(z) := \lambda\|z\|_{\ell^1}$ and $f(\xi) := \|\xi - y\|_{\ell^2}$ and write (2.23) as $\min_{z\in\mathbb{C}^N} g(z) + f(Hz)$. Next note that the convex conjugate (sometimes called the Fenchel dual) of $f$ is $f^*(\xi) := \chi_{B_1(0)} + \mathrm{Re}\langle\xi, y\rangle$, where $\chi_S$ is the indicator function for a set $S$, taking the value 0 on $S$ and $+\infty$ outside, and $B_1(0)$ denotes the closed unit $\ell^2$-ball centered in 0, and $\mathrm{Re}\langle\cdot,\cdot\rangle$ denotes the real part of the inner product. Given initial starting values $v^0 \in \mathbb{C}^N$ and $\xi^0 \in \mathbb{C}^m$, and parameters $\sigma, \tau \in \mathbb{R}_{>0}$, satisfying $\sigma\tau\|H\| < 1$ where $\|\cdot\|$ is the operator norm, the primal-dual iterations are then computed as

$$\begin{aligned} v^{k+1} &= \mathrm{argmin}_{v\in\mathbb{C}^N}\, g(v) + \mathrm{Re}\langle Hv, \xi^k\rangle + \frac{1}{2\tau}\|v - v^k\|^2_{\ell^2} \\ &= \mathrm{argmin}_{v\in\mathbb{C}^N}\, g(v) + \frac{1}{2\tau}\|v - (v^k - \tau H^*\xi^k)\|^2_{\ell^2}, \end{aligned}$$

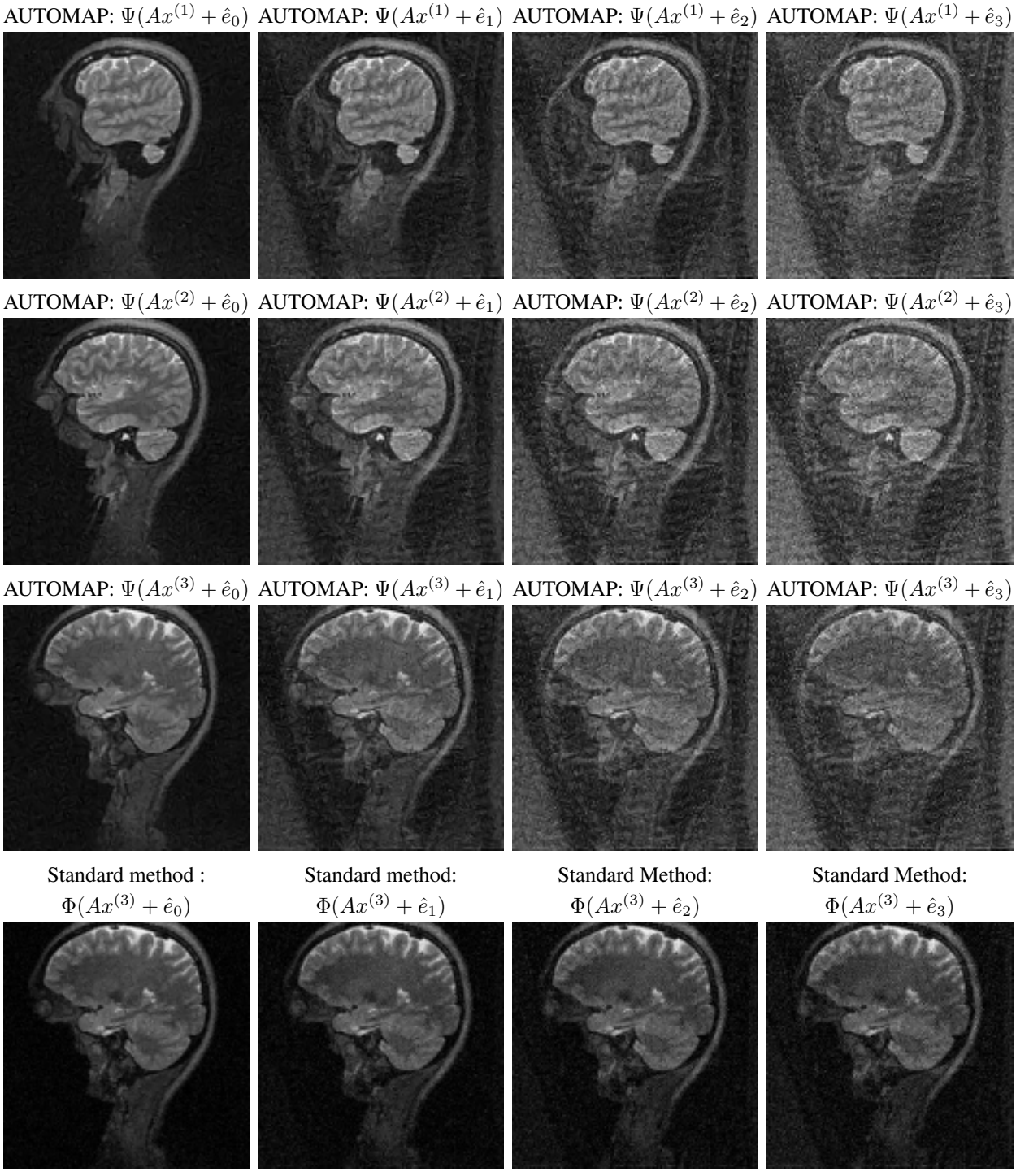


Figure 2.9: (**AUTOMAP is unstable to image-independent random Gaussian generated using a knee image**). In this experiment we consider noise of the form $\hat{e}_j = \hat{e}^1_{\text{pert},j} + \hat{e}^2_{\text{pert},j}$, where $j = 0, 1, 2, 3$. Here $\hat{e}^1_{\text{pert},j}$ is constant and $\hat{e}^2_{\text{pert},j}$ are different types of mean-zero Gaussian noise. The noise is generated as follows. Let $w_0 = 0 \in \mathbb{C}^N$. We compute wort-case perturbations $w_j \in \mathbb{C}^N$, $j = 1, 2, 3$, for AUTOMAP with respcet to the knee image seen below, and let $\hat{e}^1_{\text{pert},j} = Aw_j$, where $A \in \mathbb{C}^{m\times N}$ is the matrix from Fig. 2.1. We ensure that $\|\hat{e}^1_{\text{pert},j}\|_{\ell^2} \leq \|e^1_{\text{pert},j}\|_{\ell^2}$ for $j = 1, 2, 3$, where $e^1_{\text{pert},j}$ is the perturbation from Fig. 2.2. The remaining construction of $\hat{e}_j$ is identical to Fig. 2.2 and the reconstruction maps $\Psi$ and $\Phi$ are the same. In this experiment $\{\|\tilde{e}_j - e_j\|_{\ell^2}/\|\tilde{e}_j\|_{\ell^2}\}_{j=1}^3 = \{0.752, 0.485, 0.368\}$.

where

$$\begin{aligned}\xi^{k+1} &= \operatorname{argmin}_{\xi\in\mathbb{C}^m} f^*(\xi) - \operatorname{Re}\langle H(2v^{k+1}-v^k),\xi\rangle + \frac{1}{2\sigma}\|\xi-\xi^k\|_{\ell^2}^2 \\ &= \operatorname{argmin}_{\xi\in\mathbb{C}^m} f^*(\xi) + \frac{1}{2\sigma}\|\xi-(\xi^k-\sigma H(2v^{k+1}-v^k))\|_{\ell^2}^2 \\ &= \operatorname{argmin}_{\xi\in\mathbb{C}^m} \chi_{B_1(0)} + \frac{1}{2\sigma}\|\xi-(\xi^k+\sigma H(2v^{k+1}-v^k)-\sigma y)\|_{\ell^2}^2,\end{aligned}$$

where all the $\ell^2$ norms should be interpreted as norms over the reals, and the vectors inside the norms lying in $\mathbb{C}^N$ or $\mathbb{C}^m$ must be identified with vectors in $\mathbb{R}^{2N}$ and $\mathbb{R}^{2m}$, respectively, and likewise for the matrices $H$ and $H^*$. We recognize that the final versions of the $\operatorname{argmin}$s are proximal maps, and that the iterations can be written as

$$\begin{aligned}v^{k+1} &= \Psi_{\tau\lambda}(v^k - \tau H^*\xi^k) \\ \xi^{k+1} &= \phi(\xi^k + \sigma H(2v^{k+1}-v^k) - \sigma y)\end{aligned} \tag{2.24}$$

where $\Psi_{\tau\lambda}\colon \mathbb{C}\to\mathbb{C}$ is applied component-wise to a vector and is defined as

$$\Psi_{\tau\lambda}(z) = \frac{z}{|z|}\max\{0,|z|-\tau\lambda\}, \quad \text{for } z\in\mathbb{C}$$

and $\phi\colon \mathbb{C}^m\to\mathbb{C}^m$ is defined as $\phi(z) = z\min\{1, 1/\|z\|_{\ell^2}\}$.

We note that for the experiments in Fig. 2.3 the matrix $A = P_\Omega F$, is a subsampled Fourier transform. This implies that $\|H\| = \|P_\Omega F V^*\| = \|P_\Omega F\| = 1$, since both $F$ and $V$ are unitary matrices. Here $\|\cdot\|$ denotes the standard matrix norm induced by the $\ell^2$-norm. This gives the condition $\sigma\tau < 1$. The complete set of parameters for the experiment can be found in Table 2.4.

### 2.3.7 Computing the rank of a matrix using MATLAB

In Table 2.2 we numerically compute the rank of the matrix $A$ in (2.20), to determine the dimension of $A$'s numerical null space. In this section we elaborate on how we computed the rank $A$ in this table.

In imaging problems, such as (parallel) MRI, computing the rank can be challenging as the dimensions of the matrices under consideration easily exceeds the memory capacity of standard workstations. In particular, it is not feasible to use MATLAB's `rank` function naively for such matrices.

Though storing the matrices used in imaging is challenging, one can compute matrix-vector multiplications with such matrices efficiently using fast transforms, such as the Fast Fourier Transform (FFT). Thus by using algorithms relying solely on matrix-vector multiplications, it is still possible to determine the rank of such matrices. One way of doing this is by computing their singular values, using, for example, MATLAB's `svds` function, and then counting the number of singular values above a certain threshold. Below, we describe in detail how MATLAB chooses this threshold.

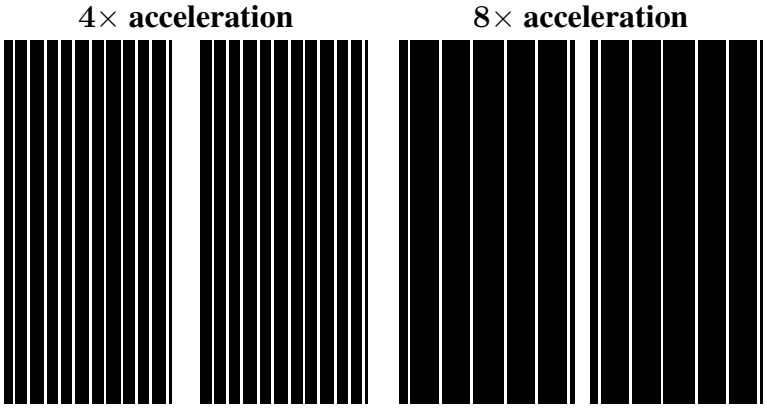


Figure 2.10: (**Sampling patterns generated using code from the fastMRI challenge [188]**) Two different sampling patterns $\Omega \subset \{1, \dots, N\}$. Each white dot in the images represents an element in $\Omega$, whereas the black pixels denote the elements in $\{1, \dots, N\} \setminus \Omega$. On the left we have $|\Omega|/N = 1/4$ and on the right we have $|\Omega|/N = 1/8$. The sampling patterns $\Omega$ simulate the equidistant sampling patterns used in the fastMRI challenge for the brain dataset [188]. Here 8% and 4% of all k-space lines are fully sampled from the centre for 4 and 8 times acceleration, respectively.

For completeness, recall that the rank of a matrix $A \in \mathbb{C}^{m \times N}$ is equal to its number of non-zero singular values. That is, if we denote the singular values of $A$ by $\sigma_i \in \mathbb{R}_{\geq 0}$, and order them by size $\sigma_1 \geq \cdots \geq \sigma_{\min\{m,N\}} \geq 0$, then the rank of $A$ is the largest $r$, such that $\sigma_r > 0$.

In practice, the singular values of a matrix $A$ can be very small without being zero. When doing computations with such matrices, it is no longer useful to consider the rank of $A$. Instead one considers the *numerical* rank of $A$. That is the largest $r$ such that $\sigma_r \geq \epsilon_{\text{tol}} > 0$, where $\epsilon_{\text{tol}} > 0$, is a threshold parameter used to discard all singular values below a certain threshold. When MATLAB computes the rank of a matrix numerically, it uses the following tolerance criterion[2]:

```
tol = max(size(A)) * eps(norm(A))
```

Here the `norm(A)` is the matrix norm induced by the $\ell^2$-vector norm and is equal to the largest singular value of $A$. The function `eps(...)` is defined as follows:

> "*[...] returns the positive distance from abs(x) to the next larger floating-point number of the same precision as x.*"
>
> — From MATLAB's documentation of the `eps` function[3].

To compute the rank of the matrices $A$ considered in Table 2.2, we computed their singular values using MATLAB's `svds` function and used the threshold rule described above in single-precision to determine their rank. Our use of single precision is based on the fact that the majority of neural networks are trained on GPUs in single precision.

[2]See `https://se.mathworks.com/help/matlab/ref/rank.html`
[3]See `https://se.mathworks.com/help/matlab/ref/eps.html`

### 2.3.8 Data availability

The raw data used for Fig. 2.1, Fig. 2.2, Fig. 2.3 and Fig. 2.6 are form the publicly available dataset [69]. The raw data used for Fig. 2.4 is provided by GE Healthcare and is available from `https://www.gehealthcare.in/products/magnetic-resonance-imaging/1-5t/optima-mr450w-with-gem`. Network weights, sampling pattern and edited data is available from `https://www.mn.uio.no/math/english/people/aca/vegarant/data/storage_automap_final.zip` . The knee image used in Fig. 2.9 is provided by GE Healthcare and is available from `https://www.gehealthcare.in/products/magnetic-resonance-imaging/1-5t/optima-mr450w-with-gem`.

Code used to generate all the figures are available at `https://github.com/vegarant/automap_not_robust`.

## 2.4 Conclusion

**The importance of AUTOMAP and future work.** The AUTOMAP methodology radically departs from current standard "*handcrafted*" reconstruction methods and differs significantly from recent DL-based methods for image reconstruction. Letting the learning process be dictated by the data only, and circumventing any connection to physical and mathematical modelling, is new. This attempt is to be welcomed. On the one hand, imprecise modelling of the physics is an ongoing topic of research in image reconstruction. But on the other hand, the ensuing instability shows the importance of balancing the reconstruction methodology between physical modelling and physics-independent learning. As shown, AUTOMAP is not robust as it has no mechanism to balance the necessary performance-stability trade-off described in Theorem 2.2.1. However, it should be emphasized that this is not exclusive to AUTOMAP, and AI generated hallucinations, instabilities and unpredictability are a serious concern regarding the new generation of AI-based reconstruction methods [6]. Naturally, there are various ways one might strive to stabilize a DL method by modifying its setup. One may, for example, perform adversarial training with perturbations added to the measurements in the training process. In Fig. 2.5 we show the effect of this and display the performance-stability trade-off demonstrated in Theorem 2.2.1 in practice. As is clear from the figure, one can increase the robustness, but at the cost of losing details in the recovered image. This work has highlighted and provided insight into these critical issues. Yet, the question of how to achieve high performance *and* robustness with AI-based methods remains open.

# Chapter 3

# The troublesome kernel: hallucinations and instabilities in deep learning for inverse problems

AI methods are fundamentally changing the sciences and scientific computing, however, the issue of trustworthiness has become a serious issue given the overwhelming empirical evidence that deep learning leads to unstable methods in a variety of applications. Recently, another phenomenon of AI generated hallucinations providing false yet realistic looking artefacts has been reported on – adding to the list of fascinating yet trust reducing AI phenomena. In this chapter, we present a comprehensive mathematical analysis explaining the many facets of AI generated hallucinations, the links to instabilities, but also how stable AI methods can hallucinate. Our results establish four crucial issues for AI methods in inverse problems that can be interpreted as 'no free-lunch' phenomena:

(1) overly accurate AI methods will wrongly transfer details from one image to another reconstructed image creating a hallucination.

(2) there is an accuracy-hallucination trade-off.

(3) there is an accuracy-stability trade-off, and optimising these trade-offs is difficult through standard training processes.

(4) hallucinations can occur due to any kind of noise model and probability distribution on the data used, and standard training typically encourages the conditions leading to AI generated hallucinations.

Finally, our results show that accurate and hallucination-free methods can only be achieved by having information about the kernel of the sampling operator encoded in the recovery algorithm. Based on this, we initiate a program for reducing hallucinations in DL in inverse problems.

This chapter is based on joint work with Vegard Antun, University of Oslo, who provided the code and figures, and Ben Adcock, Simon Fraser University, and was supervised by Anders C. Hansen, University of Cambridge.

## 3.1 Introduction

The impact that deep learning has had in recent years in machine learning (ML) applications such as image classification, speech recognition and natural language processing can hardly be overstated. Perhaps unsurprisingly, the development and use of DL for challenging problems in the computational sciences has recently become an active area of inquiry. Areas of particular notice include numerical PDEs [60, 185], discovering PDE dynamics [158], Uncertainty Quantification and high-dimensional approximation [162].

ML, especially DL, may be the future of computational imaging, as there has been an unprecedented trend of work in this direction. However, ML is also famous for fundamental 'no free-lunch' results, and thus it should not be a surprise that such phenomena may also happen for ML methods in inverse problems. Image reconstruction from measurements is an important task in a wide range of scientific, industrial and medical applications, including, but by no means limited to, electron and fluorescence microscopy, seismic imaging, nuclear magnetic resonance (NMR), magnetic resonance imaging (MRI) and X-Ray CT. The last several years have witnessed the emergence of a variety of different trained Neural Networks (NNs) for image reconstruction which claim to achieve competitive, and sometimes even superior, performance to current state-of-the-art techniques [8]. Notably, their potential has been described by *Nature* as 'transformative' [171].[1] Our results follow the tradition in ML on 'no free-lunch' theorems and establish these in connection with AI generated hallucinations and instabilities in inverse problems and image reconstruction.

In many fields, the perils and limitations of deep learning used in inverse problems have become an issue of concern. In particular, instabilities and AI generated hallucinations – as documented in Figures 3.1-3.2, across many image modalities – have become a serious issue, for example in the fastMRI challenge 2020 [138]. This problem seems to be universal across many imaging modalities.

> *"The potential lack of generalization of deep learning-based reconstruction methods as well as their innate unstable nature may cause false structures to appear in the reconstructed image that are absent in the object being imaged."* – From "On hallucinations in tomographic image reconstruction" (2020) [138].

The potential lack of generalization of deep learning-based reconstruction methods as well as their innate unstable nature may cause false structures to appear in the reconstructed image that

[1] To be specific, [171] is titled 'AI transforms image reconstruction' and features a new DL approach [192] which 'improves speed, accuracy and robustness of biomedical image reconstruction'.

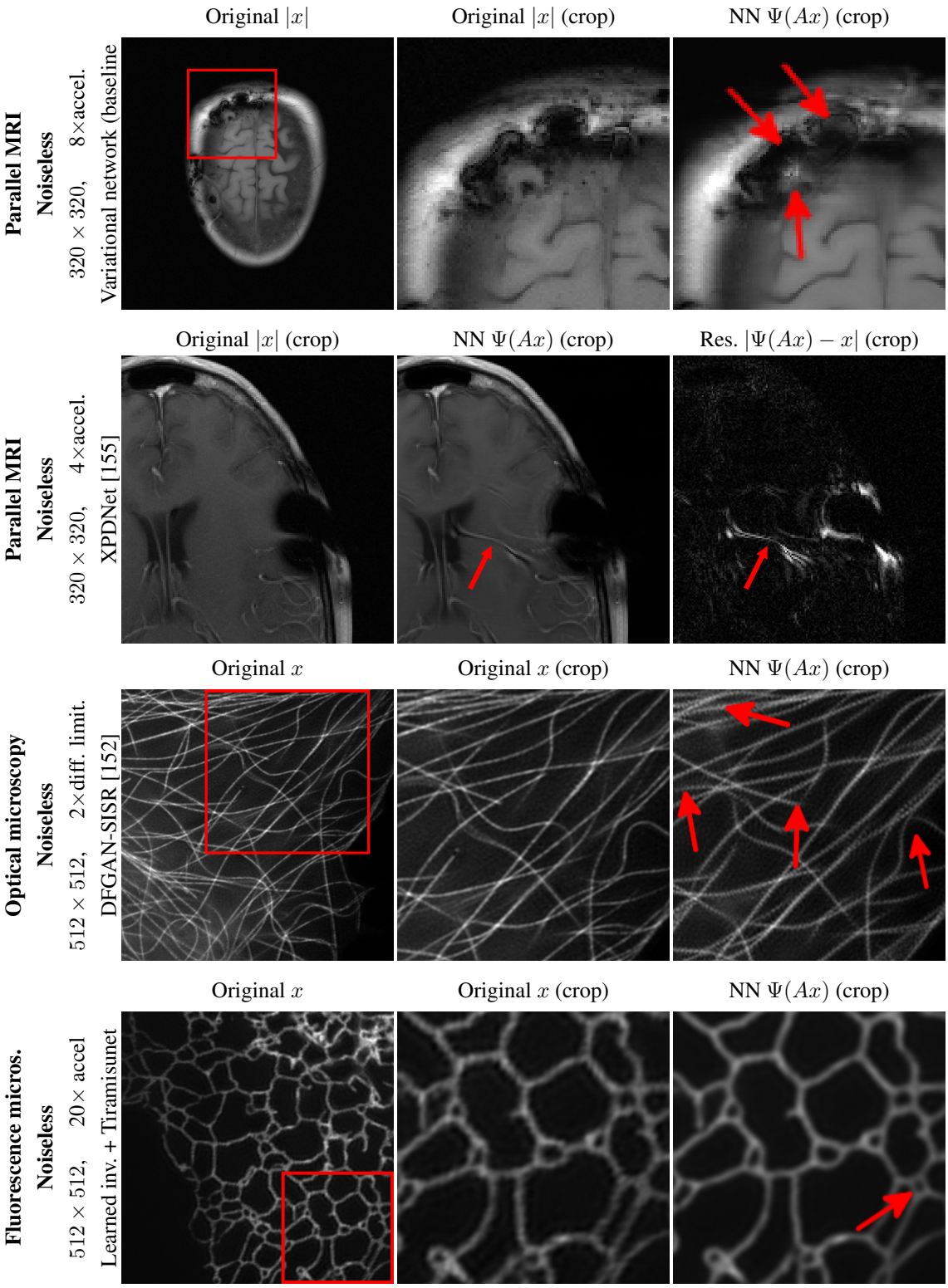


Figure 3.1: (**AI generated hallucinations in different imaging modalities**). Different neural networks $\Psi\colon \mathbb{C}^m \to \mathbb{R}^N$ are shown to hallucinate, i.e. create realistic artifacts, when evaluated on test data. Note that $m$, $N$ and $A$ vary between the experiments. In the first three rows we consider networks from the cited publications. In row four, we have trained a network $\Psi$ on data from [152]. Here the network $\Psi$ is given by $y \mapsto \phi(Ly)$ where $\phi\colon \mathbb{R}^N \to \mathbb{R}^N$ is a learnable Tiramisunet [105] and $L \in \mathbb{R}^{N\times m}$ is a learnable pseudoinverse.

are absent in the object being imaged. The Facebook fastMRI challenge focuses on medical imaging, yet AI generated hallucinations are apparent in microscopy as well.

> *"The most serious issue when applying deep learning for discovery is that of hallucination. [...] These hallucinations are deceptive artifacts that appear highly plausible in the absence of contradictory information and can be challenging, if not impossible, to detect."* – From "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction", *Nature Methods* (2019) [14].

DL in super resolution techniques, which can be viewed as an undersampled inverse problem, may be problematic as additional or removed elements in the reconstruction can occur.

> *"However, if the neural network encounters unknown specimens, or known specimens imaged with unknown microscopes, it can produce nonsensical results."* – From "The promise and peril of deep learning in microscopy", *Nature Methods* (2021) [99].

As mentioned in Chapter 1, it should be noted that non-robustness and untrustworthiness of DL systems is not just an issue in the sciences, as these issues may have legal implications. The European Commission is in the process of outlining a legal framework for the use of AI.

> *"On AI, trust is a must, not a nice to have. [...] The new AI regulation will make sure that Europeans can trust what AI has to offer. [...] High-risk AI systems will be subject to strict obligations before they can be put on the market: [requiring] High level of robustness, security and accuracy."* – Europ. Comm. outline for legal AI (April 2021) [66].

Thus, trustworthiness, stability and accuracy of neural networks used in inverse problems is a fundamental issue that must be resolved. By studying and mathematically understanding why instabilities and AI generated hallucinations can occur, we shed light on how these issues may be overcome in order to safely incorporate DL into inverse problems.

### 3.1.1 Outline

In the following sections, we present a summary of our main results in non-technical terms, Section 3.2 and a review of related work, Section 3.3. Then, we present our main theorems, Section 3.4. In Section 3.6 we discuss the main results and provide insights into their relevance in various applications. Finally, in Sections 3.4.4 and 3.4.5, we consider training and regularization in DL. Section 3.7 contains further implementation details for the numerical experiments. Code accompanying this work can be found at `https://github.com/vegarant/troub_ker`.

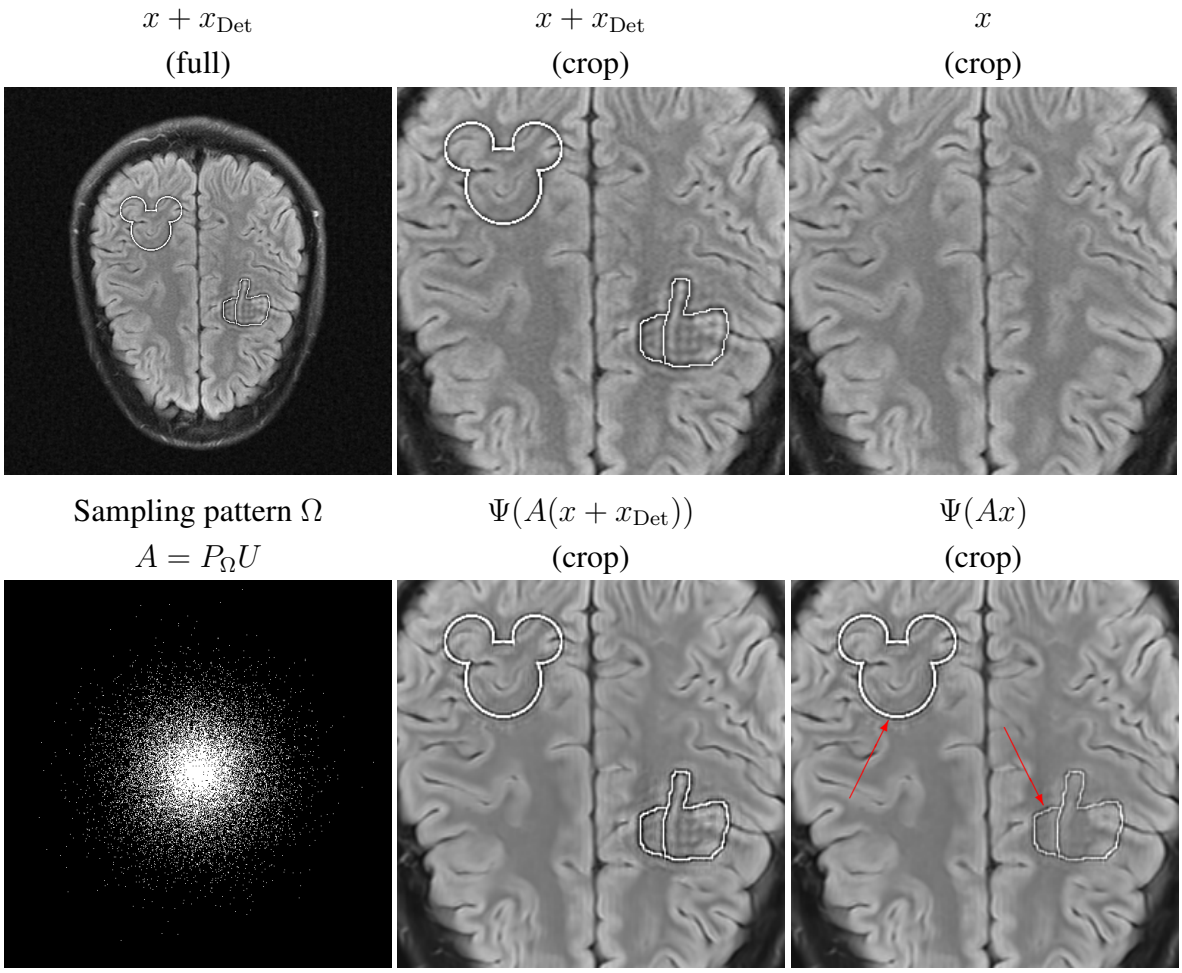


Figure 3.2: (**Detail transfer**) We consider the setup from Theorem 3.4.2, with two images $x + x_{\text{Det}}$ and $x$, seen in the top row. Here $x_{\text{Det}}$ is the Mickey and thumb detail seen in the upper left image. The detail $x_{\text{Det}}$ is constructed so Mickey is in the null space of $A$ and the thumb is very close to the null space of $A$. The neural network $\Psi$ is trained to high accuracy on the pair $(A(x + x_{\text{Det}}), x + x_{\text{Det}})$, and 1200 other images from the fastMRI dataset. As we can see from the lower right image, the neural network $\Psi$ transfers the detail $x_{\text{Det}}$ onto the image $x$. In this example $A = P_\Omega F$, where $F \in \mathbb{C}^{N\times N}$ is a two-dimensional Fourier transform, and $|\Omega|/N = 1/16$, $N = 512^2$.

### 3.1.2 Problem outline

We consider the following discrete finite-dimensional linear inverse problem:

$$\text{Given measurements } y = Ax + e, \text{ recover } x. \tag{3.1}$$

Here $A \in \mathbb{C}^{m\times N}$, with $1 \leq \operatorname{rank}(A) < N$ is the *sampling operator* (also called *measurement matrix*), $y \in \mathbb{C}^m$ is a vector of *measurements*, $e \in \mathbb{C}^m$ is noise and $x \in \mathbb{C}^N$ is the (unknown) object to recover (typically a discrete image). Moreover we denote the kernel of $A$ by

$$\mathcal{N}(A) = \{x \in \mathbb{C}^N : Ax = 0\}.$$

While seemingly simple, the model (3.1) is sufficient to model many applications, including all of those mentioned above. See Section 1.5 for more details on applications of (3.1). Moreover,

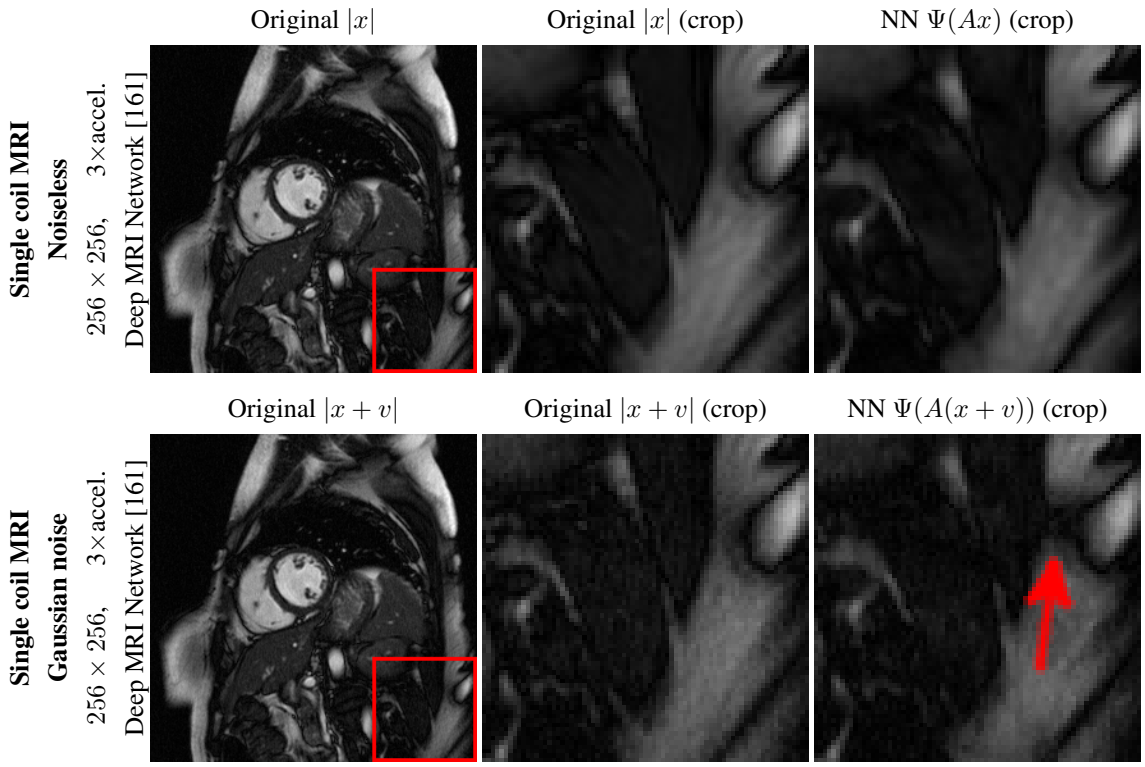


Figure 3.3: (**AI generated hallucinations due to Gaussian noise**). Different neural networks $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ are shown to hallucinate by adding Gaussian noise to measurements. The Deep MRI network (DM) $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ [161] is unstable with respect to Gaussian noise. We compute 100 Gaussian noise vectors $w_j \in \mathbb{C}^N$ and pick (using the eyeball metric) the $w_j$ for which $\Psi(A(x + w_j))$ gives the larges artefact, and label it $v_1$. Here $A \in \mathbb{C}^{m\times N}$ is a subsampled Fourier transform. Repeating the experiment with 20 (and 1) new noise vectors yields a perturbation $v_2$ (and $v_3$). $\Psi$ introduces a false dark area, indicated by the red arrows. The sparse regularization decoder (SRD) $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ is applied to the same measurements $A(x + v_j)$, $j = 1, 2, 3$.

since $\mathrm{rank}(A) < N$ in all of these applications, solving (3.1) is a challenging even in the noiseless case ($e = 0$). Since the kernel of $A$ is non-trivial, as a result, there exist infinitely many $x$'s mapping to the same measurements $y$, making (3.1) ill-posed.

Methods for recovering $x$ from measurements $y$, therefore, have to rely on an underlying assumption that $x$ belongs to some set $\mathcal{M}_1 \subset \mathbb{C}^N$, to make the problem tractable. Thus, rather than solving the problem (3.1), one solves

$$\text{Given measurements } y = Ax + e, \text{ of } x \in \mathcal{M}_1 \text{ recover } x, \tag{3.2}$$

where

$$\mathcal{M}_1 \text{ is referred to as the initial domain.}$$

In many practical scenarios the sampling matrix $A \in \mathbb{C}^{m\times N}$ is of the form $P_\Omega U$, where $U \in \mathbb{C}^{N\times N}$ is an isometry and $P_\Omega$ is the linear operator choosing rows of $U$ according to the index set $\Omega \subset \{1, \dots, N\}$ with $|\Omega| = m$. We will refer to such an $A$ as a *subsampled isometry*.

As summarised in Chapter 1, there are different methods for solving inverse problems. In this chapter we focus on the distinction of these methods into two types: *model based* and *learning based methods*.

(i) For model based methods, one makes explicit assumptions about $\mathcal{M}_1$, and designs the reconstruction method based on the choice of $\mathcal{M}_1$. Common choices for $\mathcal{M}_1$ includes $s$-sparse vectors, unions of subspaces, sparsity in wavelet, curvelets or shearlets, or sparse gradient etc. The corresponding reconstruction methods are typically based on solving a regularised least squares problem, with the predominant choice being $\ell^1$-regularization.

(ii) Learning based methods, on the other hand, make few or no explicit assumptions about $\mathcal{M}_1$. For these methods, one is typically given a large training dataset

$$\mathcal{T} = \{(Ax_1, x_1), \dots, (Ax_K, x_K)\} \subset \mathbb{C}^m \times \mathbb{C}^N,$$

where the $x_j$s often are assumed to be a subset of $\mathcal{M}_1$. Using this set $\mathcal{T}$ one learns a mapping $\Psi \colon \mathbb{C}^m \to \mathbb{C}^N$ by solving an optimization problem. A common choice is to let $\Psi$ be the minimizer of

$$\Psi \in \operatorname{argmin}_{\varphi \in \mathcal{NN}} \sum_{x \in \mathcal{T}} \|\varphi(Ax) - x\|^2 + J(\varphi)$$

for an appropriate function class $\mathcal{NN}$ of functions $\varphi \colon \mathbb{C}^m \to \mathbb{C}^N$, for example neural networks, and regularization term $J \colon \mathcal{NN} \to \mathbb{R}_{\geq 0}$. See Section 1.4.1 for a more detailed overview of DL used in inverse problems.

## 3.2 Summary of main results

Following up on the introduction, we define hallucinations in solutions to inverse problems, as *realistically looking artefacts that appear in the reconstruction, which are not present in the ground truth image.* Moreover, these artefacts would not be present in solutions obtained by most other methods given the same measurement and, hence, are only detectable with contradicting information. Various examples of such hallucinations are given in Fig. 3.1. This paper presents a comprehensive mathematical analysis explaining the many facets of the instability phenomenon and AI generated hallucinations in DL for inverse problems.

### 3.2.1 Hallucinations and instabilities

In this section we summarise – in non-technical terms – our results on how and why hallucinations and instabilities occur.

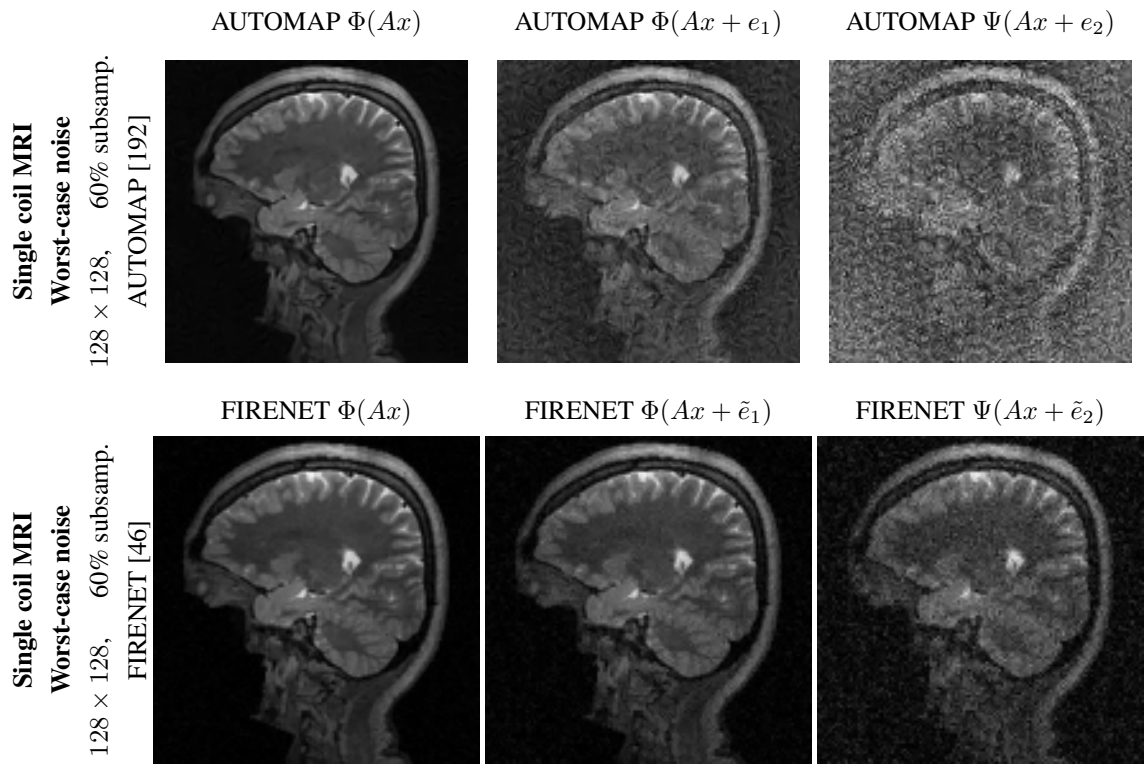


Figure 3.4: (**AI is not robust**). Different neural networks $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ are shown to be unstable with respect to worst-case noise.

(M1) **Detail transfer and hallucinations (summary of Theorem 3.4.2).** Suppose that the null space of $A$ satisfies $\mathcal{N}(A) \neq \{0\}$, then we have the following.

(1) A reconstruction mapping that recovers a detail in an image, where the detail is close to or in the kernel $\mathcal{N}(A)$ of the sampling matrix $A$, will wrongly transfer this detail to another reconstructed image yielding a hallucination created by the reconstruction mapping.

(2) There will always exist NNs that can recover elements close to or in $\mathcal{N}(A)$. Thus, trained NNs with great approximation qualities and small error on the training data may be bound to hallucinate.

Consequence: There is an accuracy-hallucination trade-off. If the reconstruction is too good on a certain image, a detail will be wrongly transferred to another reconstructed image. Figure 3.2 illustrates detail transfer in practice.

**Note:** A reconstruction mapping does not have to be unstable in order to hallucinate. Instabilities are sufficient (see Theorem 3.4.7) for the hallucination phenomenon, but not necessary. The hallucinating reconstruction map in Theorem 3.4.2 can be completely stable, in the sense that it is Lipschitz continuous.

(M2) **No free lunch I: Accuracy on certain inputs imply hallucinations on others – regardless of probability distribution (summary of Theorem 3.4.4).** Suppose that $\mathcal{N}(A) \neq \{0\}$. Then, NNs that are accurate on certain inputs are bound to hallucinate on others inputs. This can happen despite the existence of reconstruction algorithms that do not hallucinate on the same data. This happens also if the input data is drawn randomly from a probability distribution, and it happens regardless of the distribution.

Consequence: Hallucinations can only be prevented – while keeping accuracy – by adding extra information to the reconstruction algorithm that encodes information about the initial domain $\mathcal{M}_1$ and the kernel of the sampling operator $A$. In particular, the difference between the hallucinating NN and the non-hallucinating algorithm in Theorem 3.4.4 is that the non-hallucinating algorithm has encoded information about the initial domain $\mathcal{M}_1$ and the kernel of the sampling operator $A$.

(M3) **No free lunch II: Hallucinations and instabilities – Overperformance yields both (summary of Theorem 3.4.7).** Suppose that $\mathcal{N}(A) \neq \{0\}$. If a reconstruction map recovers an image well, there is a limit on how well it can perform on other inputs. In particular, there are other inputs, such that if the reconstruction mapping performs too well on these inputs we have the following.

(1) The reconstruction map becomes unstable, and the instability becomes worse the smaller the reconstruction error is on these other inputs.

(2) The reconstruction map will hallucinate and either provide false positive or negatives (add or remove details).

(3) The hallucinations do not just happen in worst-case scenarios but also for random noise.

(4) The instabilities do not just happen in worst-case scenarios but also for random noise.

Consequence: There is an accuracy-stability trade-off. If the reconstruction is too good on two particular inputs the reconstruction map becomes unstable. The instability also implies hallucinations.

**Note:** To avoid instabilities one must control the accuracy-stability trade-off and sacrifice some accuracy on certain inputs to ensure stability. The big problem is to figure out which images the accuracy – stability trade-off applies to. Current training approaches of NNs do not have a way of determining and optimising the accuracy-stability trade-off. Thus, one may end up with either poor performance and too good stability or too strong performance at the cost of severe instability. Figure 3.3 and Figure 3.6 demonstrate Theorem 3.4.7 in practice.

### 3.2.2 Optimal maps that optimise the accuracy-stability trade-off

As we pointed out above there is an accuracy-stability trade-off when constructing recovery mappings in inverse problems. In this section we provide a non-technical summary of our results on the difficulty of optimising the accuracy-stability trade-off through training.

(M1) **Training a NN that optimises the accuracy-hallucination/stability trade-off is hard (summary of Theorem 3.4.12).** Suppose that $\mathcal{N}(A) \neq \{0\}$. Then, for any $\delta \in (0, 1/5)$ there are uncountably many initial domains $\mathcal{M}_1$ and training sets such there exists a NN with training error (error on the training set) less than $\delta$, but no such NN can be an optimal mapping – regardless of training procedure.

Consequence: Optimising the accuracy-hallucination/stability trade-off with training can only be done by controlling the training error. In particular, any optimal map will satisfy a particular training error, and a smaller training error will yield a non-optimal reconstruction map.

**Note:** In specific cases one can construct NNs that are optimal maps, in the sense that they optimise the accuracy-hallucination/stability trade-off. Yet, these cases require specific knowledge about the initial domain $\mathcal{M}_1$ and the kernel of the sampling operator $A$.

(M2) **The optimal map can sometimes be trained, but training is itself unstable (summary of Theorem 3.4.15).** Given any invertible matrix $U \in \mathbb{C}^{N\times N}$. Then, there are exponentially (in $N$) many sampling patterns $\Omega$ such that when considering $A = P_\Omega U$, there are uncountably many domains $\mathcal{M}_1$ and uncountably many different training sets, such that the minimisers of basic regularised cost functions will yield an optimal reconstruction map. However, all such training procedures are unstable in the following way: By adding an element to the training set, the minimisers of the cost function are not optimal anymore.

Consequence: Optimising the accuracy-hallucination/stability trade-off is hard to do in practice as slight changes in the training data may change the optimality of the trained NN.

## 3.3 Related work

Our focus in this work is on the *underdetermined* setting where $A \in \mathbb{C}^{m\times N}$ is rank deficient, as is now common in practice. In many applications, the number of measurements is severely limited, due to time, cost, power or other constraints, hence one commonly faces the situation where $m \ll N$. An overview of different applications can be found in Section

**List of networks use in the numerical experiments**

| NN name | Authors | Title | Journal | Year |
|---|---|---|---|---|
| AUTOMAP [192] | B. Zhu et al. | Image reconstruction by domain-transform manifold learning | Nature | 2018 |
| DFGAN-SISR [152] | C. Qiao et al. | Evaluation and development of deep neural networks for image super-resolution in optical microscopy | Nature Methods | 2021 |
| Baseline FastMRI [138] | M. J. Muckley et al. | Results of the 2020 fastMRI challenge for machine learning MR image reconstruction | IEEE Trans. Med. Imaging | 2021 |
| FBPConvNet [106] | K. H. Jin et al. | Deep convolutional neural network for inverse problems in imaging | IEEE Trans. Image Process. | 2017 |
| Deep MRI [161] | J. Schlemper et al. | A deep cascade of convolutional neural networks for MR image reconstruction | Int. Conf. Inf. Proc. Med. Imaging | 2017 |
| XPDNet [155] | Z. Ramzi et al. | XPDNet for MRI reconstruction: An application to the 2020 fastMRI challenge | ISMRM annual meeting | 2021 |
| FIRENET [46] | M. Colbrook et al. | Can stable and accurate neural networks be computed? – On the barriers of deep learning and Smale's 18th problem | PNAS | 2022 |

Table 3.1: List of networks use in the numerical experiments

1.5. To design a good reconstruction mapping, one generally requires additional information on the images to be reconstructed. An example is sparse regularization, which rose to prominence in the imaging community with the introduction of compressed sensing in the 2000's by Candès, Romberg & Tao [29] and Donoho [57]. Sparse regularization techniques, as in [2, 3, 13, 29, 31, 44, 57, 74, 122, 150], are typically *untrained*: they exploit the inherent *sparsity* of natural images in fixed transform domains (e.g. wavelets or discrete gradient). Note that conditions such as the *Restricted Isometry Property* [31] and *robust Null Space Property* [74], which ensure stable and accurate recovery with compressed sensing, are precisely statements about the sampling operators null space. Over the last fifteen years, such techniques, supported by the theory of *compressed sensing*, have become the state-of-the-art for many different image reconstruction tasks. On the other hand, more recent DL techniques are *data-driven*: they seek to implicitly learn a suitable image structure from a training database of existing images.

Instabilities in DL for classification problems were first discovered in [174]. A significant development was the *DeepFool* package of Moosavi–Dezfooli, Fawzi & Frossard [135], which was followed by the construction of so-called *universal adversarial perturbations* [134]. The construction of and mitigation against *adversarial attacks* is an active area of research. To the best of our knowledge [6] and [102] were the first works to demonstrate the instability phenomenon for inverse problems in imaging. An example of this is the inability of convolutional neural networks to provide a stable and accurate reconstruction for CT inverse problems [166]. Generally, instabilities and especially AI generated hallucinations, have become a serious issue, in the fastMRI challenge [68, 138, 188] and in other

applications [14, 16, 99]. The necessity for further research for DL used in MRI is outlined in [41]. Yet, to date, there is little theoretical assessment of why and when additional or removed elements in the reconstruction and AI generated hallucinations occur in DL used for inverse problems. An approach to theoretically describe AI generated hallucinations is given in [166].

The work of Jin, McCann, Froustey & Unser [106] was influential in highlighting the promise of DL for inverse problems in imaging. This is now a rapidly evolving area of research, which we will not attempt to summarise. See [4, 8, 131] for overviews of current techniques. Note that sparse regularization has been used as the basis for some DL techniques, e.g. by using DL to recover the parts of an image that sparse regularization cannot, such as in [24], or by designing NN architectures through the process of *unrolling* an optimization algorithm (see e.g. [8]). Finally, let us note that this work considers deterministic approaches to inverse problems. For an in-depth treatment of Bayesian approaches, see, for instance, the work of Stuart [172] and Dashti & Stuart [50], as well as references therein and [8]. AI generated hallucinations also occur in such applications [45]. It has been highlighted that statistical methods in DL also suffer from the above outlined pitfalls [99, 129].

## 3.4 Main results

Our objective in this work is to elucidate the many reasons for instabilities and hallucinations in linear inverse problems. We do so by providing a number of sufficient conditions for the two phenomena to occur. Only by establishing such sufficient conditions (and eventually also necessary conditions), can we establish the methodological barriers associated with these phenomena, and guide the design of safe and secure reconstruction methods for (3.2).

**Notation**: Given a set $\mathcal{M}_1 \subset \mathbb{C}^N$ and a matrix $A \in \mathbb{C}^{m\times N}$, we let

$$\mathcal{M}_2 = A\mathcal{M}_1 = \{Ax : x \in \mathcal{M}_1\}$$

denote the range of $A$ with domain $\mathcal{M}_1$. We assume throughout that the rank of $A$ is bounded by $1 \leq \operatorname{rank}(A) < N$, and we denote null space of $A$ by $\mathcal{N}(A) \subset \mathbb{C}^N$. For a set $\Omega \subset \{1, \dots, N\}$, $P_\Omega \in \mathbb{C}^{N\times N}$ is the projection onto the canonical basis index by $\Omega$, i.e., for $x \in \mathbb{C}^N$,

$$(P_\Omega x)_i = x_i$$

if $i \in \Omega$ and 0 otherwise. We sometimes abuse notation slightly and assume $P_\Omega \in \mathbb{C}^{m\times N}$, where $m = |\Omega|$, by ignoring the zero entries. Similarly for a subspace $\mathcal{V} \subset \mathbb{C}^N$, we let $P_\mathcal{V}$ denote the projection onto $\mathcal{V}$. Throughout we let $\|\cdot\|$ denote a norm on $\mathbb{C}^N$ and $\|\cdot\|$ denote a norm on $\mathbb{C}^m$. We let

$$\mathcal{B}(x, r) = \{z \in \mathbb{C}^N : \|x - z\| \leq r\}$$

denote the closed ball centered at $x \in \mathbb{C}^N$ with radius $r > 0$. If $x \in \mathbb{C}^m$, then $\mathcal{B}(x, r)$ denotes a ball with the norm $\|\cdot\|$. For a set $\mathcal{M}_1 \subset \mathbb{C}^N$, we use the notation

$$\mathcal{M}_1^\nu = \{z \in \mathbb{C}^N : \exists x \in \mathcal{M}_1 \text{ such that } \|z - x\| < \nu\}$$

to denote the $\nu$-neighborhood of $\mathcal{M}_1$ in $\mathbb{C}^N$. Both of these notations are extended in the natural way for balls and sets in $\mathbb{C}^m$. Finally, whenever use the phrase "algorithm", this is always meant it in the sense of a Blum-Shub-Smale (BSS) machine [19], see [18] for an introduction. In particular, this means that the algorithm can perform arithmetic operations with real numbers, and check if two real numbers are equal.

In the following, consider the vanilla case of *feedforward* neural networks, although many of the results also apply to more exotic setups. Following remark 1.4.1, we identify $y \in \mathbb{C}^m$ with $\tilde{y} \in \mathbb{C}^{m'}$, where $m' = 2m$. An $L$-layer feedforward neural network is a function $\Psi : \mathbb{R}^{m'} \to \mathbb{R}^{N'}$ of the form

$$\Psi(y) = V_L(\rho(V_{L-1}(\rho(\dots\rho(V_1(y)))))), \quad y \in \mathbb{R}^{m'},$$

where each $V_j : \mathbb{R}^{n_{j-1}} \to \mathbb{R}^{n_j}$ is an affine map

$$V_j(y^{j-1}) = W_j y^{j-1} + b_j, \qquad W_j \in \mathbb{R}^{n_j \times n_{j-1}}, \quad b_j \in \mathbb{R}^{n_j},$$

and

$$y^j = \rho(V_j(y^{j-1})) \in \mathbb{R}^{n_j},$$

and $\rho : \mathbb{R} \to \mathbb{R}$ is a non-linear function, $\rho(y) = (\rho(y_i))$ for $y = (y_i)$, and $n_0 = m'$, $n_L = N'$. The $W_j$'s are referred to as *weights* and the $b_j$'s as *biases*. The number $L$ is the *depth* of the network, and $n_l$ is the *width* of its $l$th layer. The function $\rho$ is the *activation function*. Typical choices for $\rho$ are the *Rectified Linear Unit (ReLU)*, defined by $\rho(y) = \max\{0, y\}$, or the *sigmoid*, defined by $\rho(y) = \frac{1}{1+e^{-y}}$. The *architecture* of a neural network refers to choice of the depth $L$, widths $n_1, \dots, n_{L-1}$ and activation function $\rho$. Write $n = (n_0, n_1, \dots, n_L)$. The class of neural networks with a given architecture is denoted as $\mathcal{NN}$ or $\mathcal{NN}_{\mathbf{n}}$. Going back to the complex case, using Remark 1.4.1, we call a map $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ that depending on the input $y \in \mathbb{C}^m$ yields a neural network $\Psi' : \mathbb{C}^m \to \mathbb{C}^N$, an *adaptive* neural network.

In *inverse problems*, the goal is to construct a mapping that takes the noisy measurements $y = Ax + e$ of some unknown image $x$ as input and returns $x$ (or some approximation to it) as output. In DL for inverse problems (see, e.g. [8, 106, 131]) this is achieved using a *training set*

$$\mathcal{T} = \{(y^j, x^j) : y^j = Ax^j + e^j,\ j = 1, \dots, K\},$$

consisting of pairs of the form $(Ax + e, x)$, where $x$ is a training image and $y = Ax + e$ are its noisy measurements. For more details on DL for inverse problems, see Section 1.4.1.

### 3.4.1 AI generated hallucinations – detail transfer

One cause of AI generated hallucinations in DL for solving inverse problems is *detail transfer*. As shown in Fig. 3.2 a detail from one image in the training set can easily be transferred to another image in the training or test set. In the following we use the formal definition of a detail.

**Definition 3.4.1** (Detail)**.** Let $A \in \mathbb{C}^{m\times N}$ with $1 \leq \operatorname{rank}(A) < N$, $H \in \mathbb{C}^{N\times N}$ be unitary, $\delta > 0$ and $x \in \mathbb{C}^N$. Then, $x_{\mathrm{Det}}\mathbb{C}^N$ is a detail relative to the vector $x$ and the sampling matrix $A$, if

$$\|Ax_{\mathrm{Det}}\| \leq \delta,$$

and $\|x\| >> \|x_{\mathrm{Det}}\|$. Moreover, the detail is localised and satisfies,

$$|\{i \in \{1,...N\} : (Hx_{\mathrm{Det}})_i \neq 0\}| << |\{i \in \{1,...N\} : (Hx)_i \neq 0\}|.$$

An underlying mathematical mechanism for this behaviour is explained in the following theorem.

**Theorem 3.4.2** (AI Hallucinations - Detail Transfer)**.** *Let $A \in \mathbb{C}^{m\times N}$ with $1 \leq \operatorname{rank}(A) < N$ and $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be Lipschitz continuous with constant $L$. Let $\delta > 0$ and $x, x_{\mathrm{Det}} \in \mathbb{C}^N$, where $x_{\mathrm{Det}}$ is a detail that satisfies $\|Ax_{\mathrm{Det}}\| \leq \delta$ and $\|x_{\mathrm{Det}}\| >> (1+2L)\delta$.*

*(1) Suppose that*

$$\|\Psi\big(A(x+x_{\mathrm{Det}})\big) - (x+x_{\mathrm{Det}})\| \leq \delta. \tag{3.3}$$

*Then for every $e \in \mathcal{B}(0,\delta)$, there exists a $z \in \mathbb{C}^m$ with $\|z\| \leq (1+2L)\delta$, such that*

$$\Psi(Ax+e) = x + x_{\mathrm{Det}} + z.$$

*(2) If the activation function $\rho$ is Lipschitz continuous and not a polynomial, then there exists a neural network $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ such that that* (3.3) *is satisfied.*

The above theorem states that for an image $x$, with a detail $x_{\mathrm{Det}}$ lying close to the kernel of $A$, a Lipschitz continuous decoder hallucinates in the following sense. If the decoder reconstructs $x + x_{\mathrm{Det}}$ to a certain precision $\delta$, then it will hallucinate as it transfers the detail $x_{\mathrm{Det}}$ to any noisy measurements of $x$. Note that if $\|Ax_{\mathrm{Det}}\| \leq \delta$, the condition $\|x_{\mathrm{Det}}\| >> (1+2L)\delta$ can easily be satisfied at the same time if $x_{\mathrm{Det}}$ lies close to or in the kernel of the sampling matrix $A$. This is exemplified in Figure 3.2. This is problematic as such hallucinations, for instance in medical applications, may not be recognizable as unrealistic artefacts to the practitioner's eye.

In order to prove part $(2)$ of Theorem 3.4.2, we will use a classical and well-known result of Pinkus [148], concerning the uniform approximation of continuous functions by neural

networks. Denote the set of functions implementable by a network with $n \in \mathbb{N}$ hidden units and one hidden layer and one output unit by,

$$\mathcal{NN}_1^\rho = \{\phi : \mathbb{R}^m \to \mathbb{R} : \ \phi(x) = \sum_{i=1}^{n} \beta^i \rho(\alpha^i x - \theta^i), \text{ for some } n \in \mathbb{N}\},$$

where $\beta^i \in \mathbb{R}$, $\alpha^i \in \mathbb{R}^m$, $\theta^i \in \mathbb{R}$ for $i = 1, ...n$ and $\alpha^i x = \sum_{j=1}^m \alpha_j^i x_j$ and, moreover, $\rho : \mathbb{R} \to \mathbb{R}$ denotes a common activation function. With this definition at hand we can state Pinkus' result.

**Theorem 3.4.3** (Theorem 3.1. [148])**.** *Let $\rho \in C(\mathbb{R})$. Then, $\mathcal{NN}_1^\rho$ is dense in $C(X)$ with respect to the uniform convergence on compact sets if and only if $\rho$ is not a polynomial.*

*Proof of Theorem 3.4.2.* **Part (1)**; let $e \in \mathcal{B}(0, \delta)$. We have that

$$\Psi(Ax + e) = \Psi(A(x + x_{\text{Det}}) + (e - Ax_{\text{Det}})) = \Psi(A(x + x_{\text{Det}})) + \hat{z},$$

where

$$\begin{aligned} \|\hat{z}\| &= \|\Psi(A(x + x_{\text{Det}}) + (e - Ax_{\text{Det}})) - \Psi(A(x + x_{\text{Det}}))\| \\ &\leq L\|e - Ax_{\text{Det}}\| \leq L2\delta. \end{aligned}$$

Furthermore, by assumption we have that

$$\Psi(A(x + x_{\text{Det}})) = x + x_{\text{Det}} + \tilde{z},$$

where

$$\|\tilde{z}\| = \|\Psi(A(x + x_{\text{Det}})) - (x + x_{\text{Det}})\| \leq \delta.$$

Combining the above, we get that

$$\Psi(Ax + e) = x + x_{\text{Det}} + z_e,$$

where $\|z\| = \|z_e\| = \|\hat{z} + \tilde{z}\| \leq (1 + 2L)\delta$.

**Part (2)** is an application of Theorem 3.1 [148] using Remark 1.4.1, for the activation function $\rho$ being Lipschitz continuous and not a polynomial. □

### 3.4.2 Inevitable hallucinations – Despite existence of non-hallucinating algorithm

As the previous theorem is very general in its assumptions, the question arises how this is related to DL used for solving inverse problems. The following Theorem 3.4.4 puts this into a precise relation. In the following theorem the set $\mathcal{T}_1$ can be considered to be the projection onto the second component of the training set $\Pi_2(\mathcal{T}) = \mathcal{T}_1$.

**Theorem 3.4.4** (No free-lunch in inverse problems)**.** *Let $A \in \mathbb{C}^{m\times N}$, with $1 \leq \operatorname{rank}(A) < N$ and $\mathcal{T}_1 \subset \mathbb{C}^N$ be non-empty. Suppose that there is a $\delta > 0$ and a Lipshitz continuous $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ such that*

$$\max_{x\in\mathcal{T}_1} \|\Psi(Ax) - x\| \leq \delta.$$

*Then, for any $\epsilon, C > 0$ there is a family $\mathcal{C}$ of sets $\mathcal{M}_1 \subset \mathbb{C}^N$ with $\mathcal{T}_1 \subset \mathcal{M}_1$ and $A\mathcal{M}_1 \subset (A\mathcal{T}_1)^\epsilon$. Moreover, for each $\mathcal{M}_1 \in \mathcal{C}$ the following happens simultaneously.*

*(1)* (In-distribution hallucination). *For any probability distribution $\mathcal{D}$ on $\mathcal{M}_1$, with the property that $\mathbb{P}(X \in \mathcal{T}_1) \leq q$ for $X$ drawn according to $\mathcal{D}$, it holds that*

$$\mathbb{P}\Big(\Psi(AX) = X + z \in \mathcal{M}_1^{2\delta} \text{ for some } z \text{ with } \|z\| \geq C\delta, \Big) \geq 1 - q.$$

*(2)* ($\Psi$ will hallucinate, yet there exists an algorithm producing an accurate NN). *Assume that $\Psi$ is a neural network. Then, there exists an algorithm $\Gamma$ taking inputs in $A\mathcal{M}_1$ such that for each $y = Ax$ we have $\Gamma(y) = \Phi_y$ is an adaptive neural network $\Phi_y : \mathbb{C}^m \to \mathbb{C}^N$ such that*

$$\|\Phi_y(y) - x\| \leq \delta, \text{ for all } x \in \mathcal{M}_1, \text{ with } y = Ax.$$

**Remark 3.4.5.** A possible consequence of Theorem 3.4.4 is that in certain settings hallucinations may not be rare events. In part $(1)$, a hallucination $z$ is present in the reconstruction of $X$, drawn according to $\mathcal{D}$, from measurements $AX$ with a high probability. Specifically, this probability is $1$ minus the probability that $X \in \mathcal{T}_1$ and, hence, is close to $1$ when the probability that $X$ is in the training set $\mathcal{T}_1$ is very small. There may exist algorithms that prevent hallucinations, but they will have built in extra information about the problem. Typically, minimising a cost function in the training process of the NN will not be sufficient to prevent hallucinations.

The above theorem is an extension of Theorem 3.4.2 to DL in inverse problems. Hallucinations in DL decoders occur due to conditions easily established with standard training. We are given a training set and Lipschitz continuous decoder that satisfies $\max_{x\in\mathcal{T}_1} \|\Psi(Ax) - x\| \leq \delta$. To be specific, on the reals this decoder can without loss of generality be assumed to be a neural network by Theorem 3.1 [148]. Yet for Theorem 3.4.4 to hold we merely need to assume that the decoder is Lipschitz continuous. Then, there are uncountably many test sets $\mathcal{M}_1$ on which the decoder will hallucinate. Moreover, the test sets will have similar measurements to those of the original set depending on the noise level $A\mathcal{M}_1 \subset (A\mathcal{T}_1)^\epsilon$ and, thus, be sufficiently realistic looking measurements. The hallucination $z$ is added by the decoder to the measured vector $X$ drawn from a test set $\mathcal{M}_1$. The hallucination can become arbitrarily large, as the theorem holds with $\|z\| \geq C\delta$ for any $C > 0$. However, the reconstruction satisfies $\Psi(AX) \in \mathcal{M}_1^{2\delta}$ and, thus, may look sufficiently realistic as is still close to the original set $\mathcal{M}_1$. Moreover, this occurs with probability greater than $1 - q$ for any $X$ drawn according to any probability distribution $\mathcal{D}$. Where the probability distribution

has to satisfy $\mathbb{P}(X \in \mathcal{T}_1) \leq q$ for $X$ drawn according to $\mathcal{D}$ and, hence, relates this to the likelihood $q$ that $X$ is in the training set. Moreover, under the same conditions there exists an algorithm producing an accurate neural network on each test set.

**Remark 3.4.6.** Part $(1)$ of Theorem 3.4.4 can easily be extended to consider noisy measurements. However, this is omitted in order to emphasize the existence of an accurate neural network in the case that the trained network hallucinates.

*Proof of Theorem 3.4.4.* **Part (1)**; Let $L > 0$ denote the Lipschitz constant of $\Psi$. We start by constructing $\mathcal{M}_1$. Since $\operatorname{rank}(A) \geq 1$, there is at least one Cartesian coordinate vector $v$ such that $Av \neq 0$. Now, let $\eta$ be a multiple of $v$ such that $\|A\eta\| = \hat{\epsilon}$, where $\hat{\epsilon} < \min\{\epsilon, \delta/L\}$. Let $\xi \in \mathcal{N}(A)$ be such that $\|\eta + \xi\| \geq (C+2)\delta$. Fix $x' \in \mathcal{T}_1$ and let

$$\mathcal{M}_1 = \mathcal{T}_1 \cup \{x' + e^{ni}(\eta + \xi)\}_{n=1}^{\infty}.$$

Observe that $A\mathcal{M}_1 \subset (A\mathcal{T}_1)^{\epsilon}$, as $\mathcal{T}_1 \subseteq \mathcal{M}_1$ and $\|Ax' - A(x' + e^{ni}(\eta+\xi))\| = \|A\eta\| \leq \epsilon$ for any $n \in \mathbb{N}$. To create an uncountable family $\mathcal{C}$, pick any $\hat{\epsilon} \in (0, \min\{\epsilon, \delta/L\}]$ and repeat the construction.

Next let $n \in \mathbb{N}$ and notice that

$$\Psi(A(x' + e^{ni}(\eta + \xi))) = \Psi(A(x' + e^{ni}\eta)) = \Psi(Ax') + \tilde{z},$$

where $\|\tilde{z}\| \leq \delta$. Indeed,

$$\|\tilde{z}\| = \|\Psi(A(x' + e^{ni}\eta) - \Psi(A(x')\| \leq L\|A(x' + e^{ni}\eta) - Ax')\| \leq \delta.$$

Note that, since – by assumption – $\max_{x \in \mathcal{T}_1} \|\Psi(Ax) - x\| \leq \delta$, we have that $\Psi(Ax') = x' + \hat{z}$, with $\|\hat{z}\| \leq \delta$. Hence,

$$\begin{aligned}\Psi(A(x' + e^{ni}(\eta + \xi))) &= \Psi(Ax') + \tilde{z} = x' + \hat{z} + \tilde{z} \\ &= x' + e^{ni}(\eta + \xi) + \hat{z} + \tilde{z} - e^{ni}(\eta + \xi).\end{aligned}$$

We have that

$$\|\hat{z} + \tilde{z} - e^{ni}(\eta + \xi)\| \geq \|\eta + \xi\| - \|\hat{z} + \tilde{z}\| \geq C\delta.$$

Moreover, we have that $\|\Psi(A(x' + e^{ni}(\eta + \xi))) - x'\| \leq 2\delta$, which implies that $\Psi(Ax) \in \mathcal{M}_1^{2\delta}$ for any $x \in \mathcal{M}_1$.

Moreover, by assumption, we have that $\mathbb{P}(X \in \mathcal{T}_1) \leq q$ for $X$ drawn according to $\mathcal{D}$. It follows, by the definition of $\tilde{M}_1$, that

$$\mathbb{P}\big(X = x' + e^{ni}(\eta + \xi) \text{ for some } n \in \mathbb{N}\big) \geq 1 - q.$$

By the observations above,

$$\mathbb{P}\Big(\Psi(AX) = X + z \in \mathcal{M}_1^{2\delta} \text{ for some } z \text{ with } \|z\| \geq C\delta, \Big) \geq 1 - q,$$

follows.

**Part (2)**; will be shown for our choice of $\mathcal{M}_1 \in \mathcal{C}$. Recall that by "algorithm" we mean a BSS machine. In particular, this means that the algorithm can perform arithmetic operations with real numbers, and check if two real numbers are equal. We describe the algorithm informally as follows:

The algorithm takes input $y \in A\mathcal{M}_1$ and starts by checking if $y = Ax$ for some $x \in \mathcal{T}_1$. If it is the case, it outputs a neural network $\Gamma(y) = \Phi$ whose weights and biases are all zero, except the bias in the final layer, which is $x$. Otherwise, it continues as follows.

Let $t = (Ax')_1$, $s = (A\eta)_1$ and $y_1$ denote the first entries of $Ax'$, $A\eta$ and $y$, respectively (if $s$ is zero, consider the algorithm which chooses a different index). The algorithm computes $r = (y_1 - t)/s$ which equals $e^{\mathrm{i}n}$ for some $n \in \mathbb{N}$. Next the algorithm computes $\tilde{x} = x' + r(\eta + \xi)$, and outputs a neural network $\Phi_y \colon \mathbb{C}^m \to \mathbb{C}^N$, whose weights and biases are all zero, except the bias in the final layer, which equals $\tilde{x}$. Moreover, the neural network satisfies,

$$\|\Phi_y(Ax) - x\| \leq \delta, \text{ for all } x \in \mathcal{M}_1, \text{ with } y = Ax.$$

Then, the network we obtain is given by,

$$\begin{aligned} \Phi : \mathbb{C}^m \to & \mathbb{C}^N \\ y \mapsto & \Phi(y) = \Psi(y) \qquad \text{for} \quad y \in A\mathcal{T}_2 \\ & \Phi(y) = P^{\perp}_{\mathcal{N}(A)} x_j + \exp(\mathrm{i}n(l))(\eta + \xi) \qquad \text{else.} \end{aligned}$$

□

The above algorithm can be phrased in pseudo-code as follows,

(1) For $i \in \{1, ...k\}$: if $y = Ax_i$, then $\Phi_y = \Psi$, else continue.

(2) Set $l = 1$ and $c(l) = 0$. While $c(l) = 0$, if $(A\eta)(l) \neq 0$, then

$$c(l) = (y(l) - (Ax')(l)/(A\eta)(l),$$

else $l = l + 1$.

(3) If $\nexists i \in \{1, ...k\}$ with $y = Ax_i$. Set $n(l) = 1$. While $c(l) \neq \exp(\mathrm{i}n(l))$, set $n(l) = n(l) + 1$. Output $n(l) \in \mathbb{N}$.

(4) Let $\Phi$ be a neural network such that $\Phi(y) = \tilde{x} = x' + \exp(\mathrm{i}n(l))(\eta + \xi)$.

The neural network $\Phi_y \colon \mathbb{C}^m \to \mathbb{C}^N$ is such that its weights and biases are all zero, except the bias in the final layer, which equals $\tilde{x}$. Note that the choice of $n(l) \in \mathbb{N}$ is unique, as the sequence $(e^{\mathrm{i}n})_{n=1}^{\infty}$ has distinct values. This can be proven by a simple trigonometric argument.

### 3.4.3 Instabilities and AI generated hallucinations - additional or removed elements in the reconstruction

The following Theorem 3.4.7, is an application and extension of Theorem 2.2.1, Chapter 2. It states that any neural network that recovers images in the training set whose difference lies either in or close to $\mathcal{N}(A)$, admits a lower bound on its noisy measurements around a certain vector that is reconstructed with small error. Under certain conditions, which are easily satisfied in the case of DL using standard training, this lower bound can become arbitrarily large. Hence, a small perturbation can cause large artefacts in the reconstructed image. In addition, Theorem 3.4.7 also guarantees the existence of both additional or removed elements in the reconstruction under the same conditions. Thus, it also explains the vulnerability of DL to small structural changes in the input. However, the general result below sheds additional light on how it may be challenging to protect even against nearly mean zero Gaussian noise, a noise model that is very common in imaging apparatuses across the sciences, for example in MRI.

**Theorem 3.4.7** (Hallucinations and instabilities due to training)**.** *Let $A \in \mathbb{C}^{m\times N}$ and $\eta_1, \eta_2 > 0$. Further, let $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be a neural network and $\mathcal{T} := \{(y^k, x^k) : y^k = Ax^k + e^k, x^k \in \mathcal{M}_1, e^k \in \mathcal{B}(0,\epsilon), k = 1,\dots,K/\{j\}\}\bigcup\{(y^j, \tilde{x}^j)\} \subset \mathbb{C}^m \times \mathbb{C}^N$ be a training set. Suppose that there is $i \in \{1,...K\}/\{j\}$, with $\|x^i - \tilde{x}^j\| >> 2\eta_1$, such that*

$$\|\Psi(y^i) - x^i\| < \eta_1, \qquad \|\Psi(y^j) - \tilde{x}^j\| < \eta_1, \tag{3.4}$$

*and that*

$$\|y^i - y^j\| \leq \eta_2. \tag{3.5}$$

*Then the following hold:*

*(1)* (Instability). *There is a $v \in \mathbb{C}^m$ with $\|v\| \leq \eta_2$ and a closed non-empty ball $\mathcal{B}_v \subset \mathbb{C}^m$ centered at $v$, such that,*

$$\|\Psi(y^i + e) - \Psi(y^i)\| \geq \|x^i - \tilde{x}^j\| - 2\eta_1, \quad \forall\, e \in \mathcal{B}_v. \tag{3.6}$$

*(2)* (AI hallucinations – additional or removed elements in the reconstruction). *There exists $z \in \mathbb{C}^N$ and $v \in \mathbb{C}^m$, with $\|z\| \geq \|x^i - \tilde{x}^j\|$ and $\|v\| \leq \eta_2$, and closed non-empty balls $\mathcal{B}_x$, $\mathcal{B}_z$ and $\mathcal{B}_v$ centred at $x^i$, $z$ and $v$, respectively, such that*

$$\|\Psi(Au + e) - (u + \zeta)\| \leq \eta_1, \quad \forall\, u \in \mathcal{B}_{x^i},\, \zeta \in \mathcal{B}_z,\, e \in \mathcal{B}_v. \tag{3.7}$$

*(3)* (AI hallucinations are not rare events). *If $E = \{E_1, ..., E_m\}$ is an absolutely continuous random vector, with a strictly positive probability density function, then there is a $c > 0$ and $z \in \mathbb{C}^N$ such that*

$$\mathbb{P}\big(\|\Psi(Au + E) - (u + \zeta)\| \leq \eta_1\big) \geq c, \tag{3.8}$$

*for all $u \in \mathcal{B}_{x^i}$, $\zeta \in \mathcal{B}_z$. Moreover, for any $0 < \delta < 1$, there is a Gaussian distribution on $E$, such that* (3.8) *holds with $c = 1 - \delta$.*

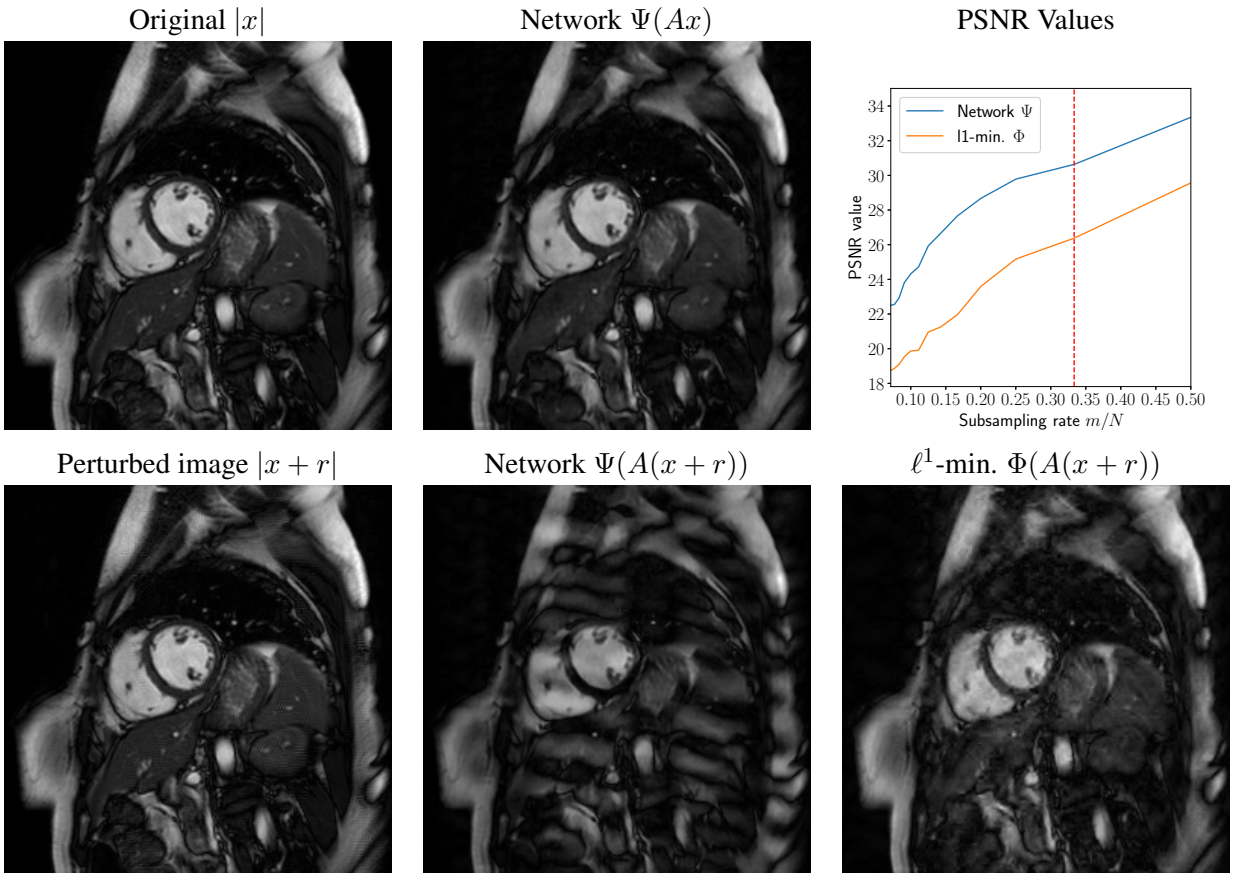


Figure 3.5: (**Instability through overperformance.**) In all experiments $A \in \mathbb{C}^{m\times N}$ is a subsampled Fourier transform. Left: (top) Original image $|x|$ and (bottom) perturbed image $|x+r|$. Middle: Reconstruction of the original and perturbed images, $x$ and $x+r$ from measurements $Ax$ (top) and $A(x+r)$ (bottom) using the deep MRI network $\Psi\colon \mathbb{C}^m \to \mathbb{C}^N$ with $m/N = 0.33$. Right: (bottom) reconstruction from $A(x+r)$ with $m/N = 0.33$ using the reconstruction map $\Phi\colon \mathbb{C}^m \to \mathbb{C}^N$ obtained using standard $\ell^1$-minimization with wavelets. In the top right image we plot PSNR-values against subsampling rate for the two reconstruction maps $\Psi$ and $\Phi$. We choose different sampling rates $m/N$ in the interval $\left[\frac{1}{14}, \frac{1}{2}\right]$ and preform reconstruction of the image $x$ from measurements $Ax$. The red dashed line indicates at which sampling rate the network is trained.

*(4)* (Instabilities are not rare events). *If $E = \{E_1, ..., E_m\}$ is an absolutely continuous random vector, with a strictly positive probability density function, then there is a $c > 0$ such that*

$$\mathbb{P}\big(\|\Psi(y^i + E) - \Psi(y^i)\| \geq \|x^i - \tilde{x}^j\| - 2\eta_1\big) \geq c. \tag{3.9}$$

*Moreover, for any $0 < \delta < 1$, there is a Gaussian distribution on $E$ such that* (3.9) *holds with $c = 1 - \delta$.*

Theorem 3.4.7 states that a continuous decoder which reconstructs two elements whose measurements lie close sufficiently well, will become unstable and produce false negatives and positives. The conditions in the theorem are very general, yet Proposition 3.4.8 states that there exists a neural network, defined on closed and bounded subsets of $\mathbb{R}^m$, with bounded width and depth satisfying these conditions.

**Proposition 3.4.8** ( [163], Theorem 1)**.** *Let $A \in \mathbb{R}^{m\times N}$, with $1 \leq \operatorname{rank}(A) < N$. Let $\mathcal{M}_1 \subseteq \mathbb{R}^N$ be closed and bounded, $\mathcal{M}_2^{\mathcal{E}} = (A\mathcal{M}_1)^{\epsilon} \subset \mathbb{R}^N$. Let $\Psi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ be continuous. Then, for any $\delta > 0$ there exists a neural network $\phi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ with $L = 11$ layers and width $W = N36m(2m + 1)$ that satisfies,*

$$\sup_{\substack{x\in\mathcal{M}_1 \\ e\in\mathcal{B}(0,\epsilon)}} \|\Psi(Ax + e) - \phi(Ax + e)\| \leq \delta.$$

Furthermore, training typically encourages a small reconstruction error, as in (3.4) regardless of (3.5). Precisely, if the training set has at least two elements $(y^i, x^i)$ and $(y^j, x^j)$ for which $\|y^i - y^j\|$ is small and that are reconstructed sufficiently well, then instabilities necessarily occur. This effect is exemplified in Fig. 3.5. Further, note that as $x^j \neq \tilde{x}^j$ or $x^j = \tilde{x}^j$, the conditions in the theorem include two cases. Firstly, if $x^j \neq \tilde{x}^j$ the neural network $\Psi$ already has a hallucination, as in not reconstructing $x^j$ but some other element $\tilde{x}^j$ and, then, this leads to further hallucinations and instabilities. Secondly, if $x^j = \tilde{x}^j$, $\Psi$ reconstructs two elements well and, then, this leads to hallucinations and instabilities. In both cases $\Psi$ is necessarily unstable (3.6), both to worst-case perturbations and random noise. Furthermore, the latter occurs for a large class of different noise models, such as Gaussian noise, as these perturbations can be in a ball $\mathcal{B}_v$. The hallucinations are due to false positives $\tilde{z} \in \mathcal{B}_z$ in the reconstruction that look sufficiently realistic. In general, the additional or removed elements in the reconstruction for measurements of $\tilde{x} \in \mathcal{B}_{x^i}$ with respect to random noise $E$ can occur with a high probability (3.8). These effects are exemplified in Figs. 3.6 and 3.3.

**Remark 3.4.9.** Note that Theorem 3.4.7 is an applictaion and extension of Theorem 2.2.1, Chapter 2. However, we state the proofs for the sake of completeness.

*Proof of Theorem 3.4.7.* The proof of Theorem 3.4.7 follows directly from the proof of Theorem 2.2.1. □

The results of Theorem 3.4.7 are exemplified in Fig. 3.6. Further, Fig. 3.8 shows that recovering elements close to the null space of $A$ causes hallucinations. Here the null space of the sampling operator $A$ is larger for higher subsampling rates and worse hallucinations are occurring with larger subsampling rates.

In order to demonstrate the ease with which a method can become unstable, Theorem 3.4.7 is deliberately formulated with the very weak conditions (3.4) and (3.5). Although the instabilities always happen in a ball, Theorem 3.4.7 does not indicate the size of these balls. Here one needs slightly stronger assumptions on the decoder $\Psi$ and the following definition. For $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ be a reconstruction map, we define the local $\epsilon$-Lipschitz constant of $\Phi$ at $y \in \mathbb{C}^m$ as

$$L^{\epsilon}(\Phi, y) = \sup_{0<\|z-y\|\leq\epsilon} \frac{\|\Phi(z) - \Phi(y)\|}{\|z - y\|}.$$

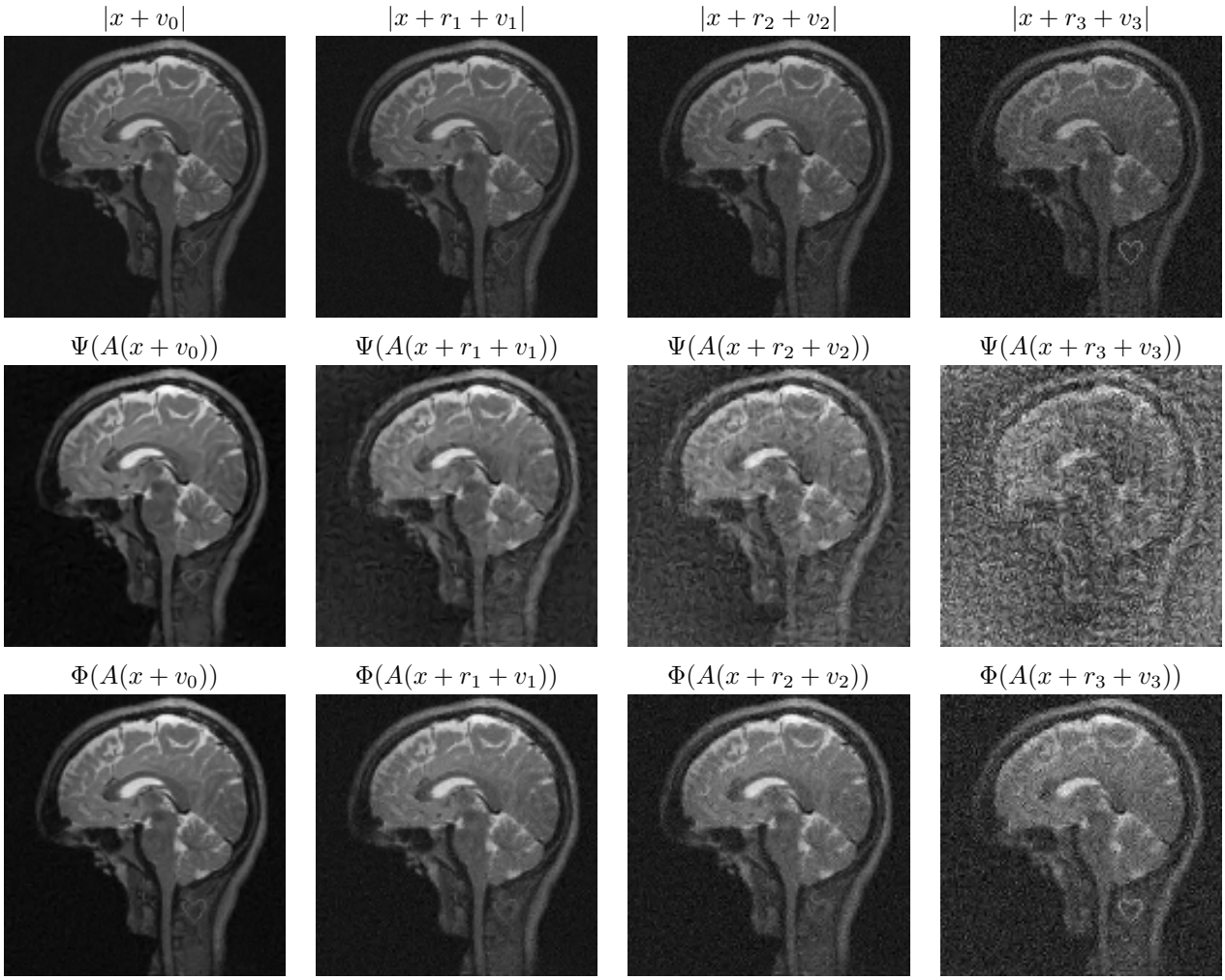


Figure 3.6: (**The instabilities are "stable"**). Adding a random perturbation to a 'bad' perturbation yields a 'bad' perturbation. In [6] the effect of small worst-case perturbations $r_i$, $i = 1, 2, 3$, on the AUTOMAP network [192] for recovering an image $x$ from its measurements $y = Ax$, is studied. Here $A \in \mathbb{C}^{m \times N}$ is a subsampled discrete Fourier transform, which is the standard mathematical model for MRI. Here perturbations $\|v_0\|_2 = \|v_1\|_2 < \|v_2\|_2 < \|v_3\|_2$ are constructed with magnitude $\|v_i\|_2/\|r_i\|_2 = 1/2$ for $i = 1, 2, 3$. The $v_i s$ are constructed as $v_i = A^* e_i$ where the real and imaginary components of $e_i \in \mathbb{C}^m$ are drawn independently from a normal distribution $\mathcal{N}(0, 1)$ and then rescaled to get the desired norm. Top row: The magnitude of the sampled images $|x + r_i + v_i|$, for $i = 0, 1, 2, 3$, with $r_0 = 0$. Middle row: The reconstruction obtained by the AUTOMAP network $\Psi : \mathbb{C}^m \to \mathbb{C}^N$, given measurements $A(x + r_i + v_i)$, $i = 0, 1, 2, 3$. Third row: Reconstruction from $A(x + r_i + v_i)$, $i = 0, 1, 2, 3$ using a sparse regularization decoder $\Phi \colon \mathbb{C}^m \to \mathbb{C}^N$.

With this definition, we can obtain a Corollary of Theorem 3.4.7 that provides a potential explanation of instabilities in larger sets.

**Corollary 3.4.10 (Instabilities in larger sets).** *Let $A \in \mathbb{C}^{m\times N}$ be full rank, $m \leq N$, $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be continuous and let $\eta > 0$. Consider $\mathbb{C}^m$ and $\mathbb{C}^N$ equipped with their respective $\ell^2$-norms, denoted $\|\cdot\|_{\ell^2}$. Suppose that there exists $x \in \mathcal{N}(A)^\perp$, $x' \in \mathbb{C}^N$ and $r_1 > 0$ such that*

$$\|x' - \Psi(Ax')\|_{\ell^2} < \eta, \quad \|z - \Psi(Az)\|_{\ell^2} < \eta, \quad \|A(x' - z)\|_{\ell^2} \leq \eta, \quad \forall z \in \mathcal{B}_x,$$

*where $\mathcal{B}_x = B_{\|\cdot\|_{\ell^2}}(x, r_1)$ is the open ball of radius $r_1$ centred at $x$. Then there exists a closed ball $\mathcal{B}_y = \mathcal{B}_{\|\cdot\|_{\ell^2}}(y, r_2) \subset \mathbb{C}^m$ of radius $r_2 \geq \sigma_{\min}(A)r_1$ centred at $y = Ax$ such that the following holds. For every $\tilde{y} \in \mathcal{B}_y$, the local $\varepsilon$-Lipschitz constant satisfies*

$$L^\epsilon(\Psi, \tilde{y}) \geq \frac{1}{\varepsilon}\left(\operatorname{dist}(x', \mathcal{B}_x) - 2\eta\right), \quad \epsilon \geq \eta,$$

*where $\operatorname{dist}(x', \mathcal{B}_x) = \inf\{\|x' - z\|_{\ell^2} : z \in \mathcal{B}_x\}$ and $\sigma_{\min}(A)$ is the smallest singular value of $A$.*

*Proof of Corollary 3.4.10.* Let $\tilde{y}$ with $\|y - \tilde{y}\|_{\ell^2} \leq \sigma_{\min}(A)r_1$ (this gives the closed ball $\mathcal{B}_y$) and write $\tilde{x} = A^\dagger\tilde{y}$, where $\dagger$ denotes the pseudoinverse. Since $x \in \mathcal{N}(A)^\perp$ and $A$ is full rank, we have $x = A^\dagger y$. In particular, $\|\tilde{x} - x\|_{\ell^2} \leq 1/\sigma_{\min}(A)\|\tilde{y} - y\|_{\ell^2} \leq r_1$. Hence $\tilde{x} \in \mathcal{B}_x$. Set $e = A(x' - \tilde{x})$ and observe that $\|e\|_{\ell^2} \leq \eta \leq \varepsilon$. Then, since $A\tilde{x} = \tilde{y}$,

$$\begin{aligned} L_\epsilon(\Psi, \tilde{y}) &\geq \frac{\|\Psi(A\tilde{x} + e) - \Psi(A\tilde{x})\|_{\ell^2}}{\|e\|_{\ell^2}} \\ &\geq \frac{\|\Psi(Ax') - \Psi(A\tilde{x})\|_{\ell^2}}{\varepsilon} \\ &\geq \frac{\|x' - \tilde{x}\|_{\ell^2} - \|x' - \Psi(Ax')\|_{\ell^2} - \|\tilde{x} - \Psi(A\tilde{x})\|_{\ell^2}}{\varepsilon} \\ &\geq \frac{\operatorname{dist}(x', \mathcal{B}_x) - 2\eta}{\varepsilon}. \end{aligned}$$

This completes the proof. □

### 3.4.4 Optimal maps are hard to train

Now we proceed to defining optimal decoders for (3.2). Recall that a multivalued mapping is traditionally noted with double arrows as

$$\varphi : \mathcal{M}_2 \rightrightarrows \mathbb{C}^N,$$

where in the following cases $(\mathcal{M}_2, d_2)$ and $(\mathbb{C}^N, d_1)$ are metric spaces. Note that in the following results the metrics $d_1, d_2$ are induced by norms. Yet, in Chapter 4 the same setting is considered and we provide results for metric spaces. We assume that the set $\varphi(x)$

is bounded for all $x \in \mathcal{M}_2$. Thus, to measure the distance between $X, Z \subset \mathbb{C}^N$ we use the Hausdorff metric $d_1^H$ on bounded subsets of $\mathbb{C}^N$. For bounded sets $X, Z \subset \mathbb{C}^N$, the Hausdorff metric is defined by

$$d_1^H(Z, X) = \max\{ \sup_{x \in X} \inf_{z \in Z} d_1(z, x), \sup_{z \in Z} \inf_{x \in X} d_1(z, x) \}.$$

With slight misuse of notation we will denote a singleton $\{x\} \subset \mathbb{C}^N$ by $x$, and we notice that $d_1^H(Z, x)$ is an upper bound on the largest possible distance between $x$ and any point in $Z$.

In this setting, the set of all possible noisy measurements $y$ is

$$\mathcal{M}_2^\epsilon := \{y \in \mathbb{C}^m : \exists y' \in \mathcal{M}_2, d_2(y, y') \leq \epsilon\}, \tag{3.10}$$

where $\mathcal{M}_2 = A\mathcal{M}_1$ and $\mathcal{M}_1 \subseteq \mathbb{C}^N$ is the initial domain. The above definition can be generalized to any subset $X \subseteq \mathcal{X}$ of a metric space $(\mathcal{X}, d)$ to obtain its $\epsilon$-neighbourhood $X^\epsilon \subseteq \mathcal{X}$.

A reconstruction mapping that optimises the accuracy-stability trade-off should satisfy the following definition. Moreover, this definition is consistent with the tradition in approximation theory and the seminal work by Cohen, Dahmen and DeVore [44].

**Definition 3.4.11** (Optimal map)**.** Let $A : \mathbb{C}^N \to \mathbb{C}^m$ be linear, $\mathcal{M}_1 \subseteq \mathbb{C}^N$ and $\mathbb{C}^m$ be equipped with metrics $d_1$, respectively $d_2$, $\mathcal{M}_2 = A(\mathcal{M}_1)$ and $\epsilon > 0$. Define the optimality constant

$$c_{\text{opt}}^\epsilon(A, \mathcal{M}_1) = \inf_{\varphi : \mathcal{M}_2^\epsilon \rightrightarrows \mathbb{C}^N} \sup_{x \in \mathcal{M}_1, e \in \mathcal{B}_\epsilon} d_1^H(\varphi(Ax + e), x). \tag{3.11}$$

Since the infimum may not be attained we define an approximate optimal map as follows. We say that $\varphi_\delta : \mathcal{M}_2^\epsilon \rightrightarrows \mathbb{C}^N$, $\delta \in (0, 1]$ is *a family of approximate optimal maps for* $(A, \mathcal{M}_1)$ if

$$\sup_{x \in \mathcal{M}_1, e \in \mathcal{B}_\epsilon} d_1^H(\varphi_\delta(Ax + e), x) \leq c_{\text{opt}}^\epsilon(A, \mathcal{M}_1) + \delta, \tag{3.12}$$

and that $\varphi_0$ is an *optimal map* if $\varphi_0$ satisfies (3.12) with $\delta = 0$.

With this definition in hand, the question is now whether or not DL provides a reconstruction mapping that is either optimal or approximately optimal.

**Theorem 3.4.12.** *Let $A \in \mathbb{C}^{m \times N}$ with $\operatorname{rank}(A) \geq 1$, where $m < N$, $K \in \{2, \ldots, \infty\}$, $\delta \leq 1/5$ and $\mathcal{B} \subset \mathbb{C}^N$ be the closed unit ball. Let $d_1, d_2$ be induced by norms on $\mathbb{C}^N$, respectively $\mathbb{C}^m$, denoted $\|\cdot\|$ and let $\epsilon > 0$ be sufficiently small. Then, there exist uncountably many $\mathcal{M}_1 \subset \mathcal{B}$, such that for each $\mathcal{M}_1$ there exist uncountably many sets $\mathcal{T} \subset \mathcal{M}_2^\epsilon \times \mathcal{M}_1$ with $|\mathcal{T}| = K$, where $\mathcal{M}_2^\epsilon = (A\mathcal{M}_1)^\epsilon$, satisfying the following. There always exist a neural network $\Psi$ such that*

$$\|\Psi(y) - x\| \leq \delta, \qquad \forall (y, x) \in \mathcal{T}. \tag{3.13}$$

*However, any training process yielding a map $\Psi$ with training error as in* (3.13) *fails to produce an optimal map as $\Psi$ satisfying* (3.13) *cannot be optimal (even if it is multivalued with the norm replaced by the Hausdorff metric). Moreover, the collection of such mappings satisfying* (3.13) *does not contain a family of approximate optimal maps. If $K$ is finite, one can choose $|\mathcal{M}_1| = K + 1$.*

We stress that Theorem 3.4.12 is *not* a statement about overfitting. Overfitting occurs when a network performs well on the training set, but poorly on the test set. This phenomenon is caused by the fact that the architecture of the network is fixed, and hence its ability to fit data is limited (it can fit the training set, but not the test set). Indeed, it is a classical result in approximation theory that any set of data points (e.g. the union of the training and test sets) can be interpolated by a neural network of sufficient size [148]. Similarly, any continuous function can be approximated by a large enough network. So even if the trained network would suffer from overfitting, and hence lack performance on the test set, there will exist another neural network that interpolates all data points in the training set as well as the test set. What Theorem 3.4.12 describes is a phenomenon that happens *for all* mappings. Thus, no restriction on the network architecture is used. Moreover, according to Theorem 3.4.12, the size of $\mathcal{M}_1$ can be $K + 1$ and therefore, finite. Hence the phenomenon arises regardless of the interpolation power guaranteed by classical results. More directly, one could simply let $\mathcal{T}$ contain both the training sets and test sets in $(1)$. Theorem 3.4.12 then says that one can have excellent performance on both the training and test sets, but still be a suboptimal mapping.

**Remark 3.4.13.** The condition that $\delta \leq 1/5$ is related to the assumption that $\mathcal{M}_1$ is a subset of the unit ball $\mathcal{B}$. The theorem holds for arbitrary $\delta > 0$, provided this ball is suitably enlarged.

*Proof of Theorem 3.4.12.* First, let $\{x_1, \ldots, x_K\}$ be $K$ distinct elements in $\mathcal{N}(A))^\perp$ such that $\|x_1\| = 1/2$ and $0 < \|x_j\| \leq 1$ and such that $\|Ax_j - Ax_i\| \geq 2\epsilon$ for all $i \neq j \in \{1, \ldots, K\}$, which is possible for $\epsilon \geq 0$ chosen to be sufficiently small. Thus, define $\mathcal{M}_2^\epsilon := \{Ax_i + e : e \in \mathcal{B}(0, \epsilon), i = 1, \ldots, K\}$, where $\mathcal{B}(0, \epsilon) \subseteq \mathbb{C}^m$ is the open ball around $0$ with radius $\epsilon > 0$ and all $y, y' \in \mathcal{M}_2^\epsilon$ are such that $y \neq y'$. Note that we can do this since, by assumption, $\operatorname{rank}(A) \geq 1$. Choose, by the assumption that $m < N$, $z_1 \in \mathcal{N}(A)$ with $\|z_1\| = 1/2$. Let $\mathcal{M}_1 = \{x_1 + z_1, x_1, \ldots, x_K\}$ and observe that $\mathcal{M}_1 \subset \mathcal{B}$. We argue by contradiction and suppose that there exists a map $\Psi : \mathcal{M}_2^\epsilon \to \mathbb{C}^N$ with

$$\|\Psi(Ax + e) - x\| \leq \delta, \qquad \forall x \in \mathcal{M}_1, \forall e \in \mathcal{B}(0, \epsilon). \tag{3.14}$$

In particular for $e_1 \in \mathcal{B}(0, \epsilon)$, $\|\Psi(Ax_1 + e_1) - x_1\| \leq \delta$. However, since $z_1 \in \mathcal{N}(A)$, we have $\Psi(A(x_1 + z_1) + e_1) = \Psi(Ax_1 + e_1)$ and therefore $\|\Psi(A(x_1 + z_1) + e_1) - (x_1 + z_1)\| \geq \|z_1\| - \delta \geq 1/2 - 1/5 > \delta$, which contradicts (3.14). Since $K$ is arbitrary, we get the result. Moreover the same argument works for multivalued maps using the Hausdorff

distance. In order to get uncountably many different $\mathcal{M}_1$'s, as mentioned in the statement of the theorem, one can simply multiply the original choice of $\mathcal{M}_1$ by complex numbers of modulus 1.

Let $\mathcal{M}_1$ be as defined previously, and set $\mathcal{T} = \{x_1, \dots, x_K\}$. Define the map $\psi_0 : \mathcal{M}_2^\epsilon \to \mathbb{C}^N$ by

$$\psi_0(y) = \begin{cases} x_1 + \frac{1}{2}z_1 & \text{if } y = Ax_1 + e \quad \forall e \in \mathcal{B}_{d_2}(0,\epsilon) \\ x_1 + \frac{1}{2}z_1 & \text{if } y = A(x_1 + z_1) + e \quad \forall e \in \mathcal{B}_{d_2}(0,\epsilon) \\ x_j & \text{otherwise} \end{cases}. \tag{3.15}$$

This is well defined by the assumption that $\|Ax_j - Ax_i\| \geq 2\epsilon$ for all $i \neq j \in \{1, \dots, K\}$. Then, by (3.15),

$$\begin{aligned} c_{\text{opt}}^\epsilon(A, \mathcal{M}_1) &= \inf_{\varphi : \mathcal{M}_2^\epsilon \rightrightarrows \mathbb{C}^N} \sup_{x \in \mathcal{M}_1} \sup_{e \in \mathcal{B}(0,\epsilon)} d_1^H(\varphi(Ax+e), x) \leq \sup_{x \in \mathcal{M}_1} \sup_{e \in \mathcal{B}(0,\epsilon)} \|\psi_0(Ax+e) - x\| \\ &\leq \sup_{j \geq 1} \sup_{e \in \mathcal{B}(0,\epsilon)} \|\psi_0(Ax_j + e) - x_j\| \vee \|\psi_0(A(x_1+z_1)+e) - (x_1+z_1)\| = \frac{1}{4}, \end{aligned} \tag{3.16}$$

where $a \vee b$ denotes the standard maximum of real numbers $a, b$. However, for any mapping $\Psi : \mathcal{M}_2^\epsilon \rightrightarrows \mathbb{C}^N$ with

$$d_1^H(\Psi(Ax_j + e), x_j) \leq \delta, \quad \forall j = 1, \dots, K, \tag{3.17}$$

for all $e \in \mathcal{B}(0,\epsilon)$, we have that for $e_1 \in \mathcal{B}(0,\epsilon)$

$$\begin{aligned} &\sup_{x \in \mathcal{M}_1} \sup_{e \in \mathcal{B}(0,\epsilon)} d_1^H(\Psi(Ax+e), x) \geq d_1^H(\Psi(A(x_1+z_1)), x_1+z_1) \\ &\quad = d_1^H(\Psi(Ax_1+e_1), x_1+z_1) \geq \|z_1\| - d_1^H(\Psi(Ax_1+e_1), x_1) \geq \frac{1}{2} - \frac{1}{5} = \frac{3}{10} > \frac{1}{4}. \end{aligned}$$

Thus, by (3.16), it follows that $\Psi$ is not an optimal map. Furthermore, it is clear that no family of maps satisfying (3.17) can be approximately optimal. □

### 3.4.5 Stability versus performance: Setting the regularization parameter is challenging

Creating a stable neural network for inverse problems is not hard. Indeed, one may simply consider the zero network, a highly stable, yet not entirely useful network. The challenge is combining stability with performance. Note that Theorem 3.4.7 reveals that it is the current training process that causes the instabilities. Indeed, solving, for example,

$$\Psi \in \operatorname*{argmin}_{\tilde{\Psi} \in \mathcal{NN}} \frac{1}{K} \sum_{j=1}^{K} \text{Cost}(\tilde{\Psi}, y^j, x^j), \tag{3.18}$$

where Cost is an appropriate cost function, for example

$$\mathrm{Cost}(\tilde{\Psi}, y^i, x^i) = \|x^i - \tilde{\Psi}\left(y^i\right)\|_{\ell^2}^2,$$

produces a network with small training error

$$\|\Psi(y) - x\|_{\ell^2}^2 \leq \delta, \qquad \forall (y, x) \in \mathcal{T}, \tag{3.19}$$

for some $\delta > 0$, where $\mathcal{T}$ denotes the training set. One may attempt to overcome overfitting or instabilities in a similar way by adding a regularization term to the objective function. Specifically, one solves

$$\Psi \in \underset{\hat{\Psi} \in \mathcal{NN}}{\operatorname{argmin}} \frac{1}{|\mathcal{T}|} \sum_{(y,x)\in\mathcal{T}} \frac{1}{2}\|x - \hat{\Psi}\left(y\right)\|_{\ell^2}^2 + \lambda J(\hat{\Psi}) \tag{3.20}$$

where $\lambda \in \mathbb{R}_+ := [0, \infty)$, and $J : \mathcal{NN} \to \mathbb{R}$ is a function from a class of neural networks to the reals. This raises the question: how does one set the regularisation parameter $\lambda$ in (3.20)? In this section, we investigate. To do so, we need the following:

**Definition 3.4.14** (Optimal $\lambda$)**.** Given a pair $(A, \mathcal{M}_1)$, a class $\mathcal{NN}$ of neural networks and a training set $\mathcal{T}$ we say that $\lambda \in \mathbb{R}_+$ is optimal for $\{(A, \mathcal{M}_1), \mathcal{NN}, \mathcal{T}\}$, if there is a minimiser of (3.20) that is an optimal map for $(A, \mathcal{M}_1)$.

As Theorem 3.4.15 reveals setting the right $\lambda$ is a highly delicate, and ironically, highly unstable problem:

**Theorem 3.4.15** (Setting $\lambda$ is delicate)**.** *Let $U \in \mathbb{C}^{N\times N}$ be an invertible matrix with $N \geq 4$. Let $d_1, d_2$ be induced by by the $\ell^2$-norm on $\mathbb{C}^N$, respectively $\mathbb{C}^m$, denoted $\|\cdot\|_{\ell^2}$. There exist $2^N - 2N - 2$ sampling patterns $\tilde{\Omega}$ and for each of them a sampling pattern $\Omega$, that is not full sampling, such that if $A = P_\Omega U$ and $\tilde{A} = P_{\tilde{\Omega}} U$ we have the following. Let $K \in \{2, \ldots, \infty\}$, $\mathcal{NN} = \mathcal{NN}_n$ be any class of ReLU neural networks with at least one hidden layer, where $n = (m, n_1, \ldots, n_{L-1}, N)$, $m = \max\{|\Omega|, |\tilde{\Omega}|\}$ (see Remark 3.4.16), $n_j \geq 2m$, ($n_j \geq 4m$, if we consider it as a real-valued network, see Remark 1.4.1) and $J : \mathcal{NN} \to \mathbb{R}$. Then, there is a $\lambda_{\mathrm{opt}} \in \mathbb{R}_+$ and uncountably many domains $\mathcal{M}_1$ of size $K$ such that for each $\mathcal{M}_1$ there are uncountably many training sets $\mathcal{T} \subset A(\mathcal{M}_1) \times \mathcal{M}_1$, $\tilde{\mathcal{T}} \subset \tilde{A}(\mathcal{M}_1) \times \mathcal{M}_1$, such that*

$$\lambda_{\mathrm{opt}} \text{ is optimal for } \{(\tilde{A}, \mathcal{M}_1), \mathcal{NN}, \tilde{\mathcal{T}}\} \text{ and } \{(A, \mathcal{M}_1), \mathcal{NN}, \mathcal{T}\}.$$

*However, there exists an uncountable collection $\mathcal{S} \subseteq \mathbb{C}^N$ such that the following holds:*

*(1)* (The sampling pattern $\tilde{\Omega}$ makes $\lambda_{\mathrm{opt}}$ unstable with respect to $\mathcal{S}$). *If $(\tilde{A}x, x), x \in \mathcal{S}$, is either added to the training set $\tilde{\mathcal{T}}$, or replaces a specific element in $\tilde{\mathcal{T}}$, then either there is no element of*

$$\underset{\hat{\Psi} \in \mathcal{NN}}{\operatorname{argmin}} \frac{1}{|\tilde{\mathcal{T}}|} \sum_{(\tilde{y},\tilde{x})\in\tilde{\mathcal{T}}} \frac{1}{2}\|\tilde{x} - \hat{\Psi}\left(\tilde{y}\right)\|_{\ell^2}^2 + \lambda J(\hat{\Psi}), \tag{3.21}$$

*for any $\lambda \in \mathbb{R}_+$ that is an optimal map for $(\tilde{A}, \mathcal{M}_1)$, or there is another $\tilde{\lambda}_{\mathrm{opt}} \neq \lambda_{\mathrm{opt}}$ that is optimal for $\{(\tilde{A}, \mathcal{M}_1), \mathcal{NN}, \tilde{\mathcal{T}}\}$ whereas $\lambda_{\mathrm{opt}}$ is not.*

*(2)* (The sampling pattern $\Omega$ makes $\lambda_{\text{opt}}$ stable with respect to $\mathcal{S}$). *Yet, given any subset* $\mathcal{V} \subset \mathcal{S}$ *such that if* $\{(Ax, x) \,|\, x \in \mathcal{V}\}$ *is either added to* $\mathcal{T}$ *or replaces elements in* $\mathcal{T}$*, then* $\lambda_{\text{opt}}$ *is still optimal for* $\{(A, \mathcal{M}_1), \mathcal{NN}, \mathcal{T}\}$.

**Remark 3.4.16.** Note that in the case $|\Omega| \neq |\tilde{\Omega}|$ and $m = \max\{|\Omega|, |\tilde{\Omega}|\}$ we interpret the action of $\Psi \in \mathcal{NN}_n$ on $y$ denoted, with slight abuse of notation, as $\Psi(y)$ as $\Psi$ acting on $y \oplus 0$ if $y$ has dimension less than $m$.

Theorem 3.4.15 says that given an invertible matrix $U$, there is an abundance of sampling patterns $\Omega$ and $\tilde{\Omega}$, as well as training sets $\mathcal{T}$ and $\tilde{\mathcal{T}}$ plus a large set $\mathcal{S}$, such that setting the optimal parameter $\lambda$ is highly unstable with respect to changes in the training set $\tilde{\mathcal{T}}$ from elements in $\mathcal{S}$ when considering the sampling pattern $\tilde{\Omega}$. However, at the same time, setting the optimal parameter $\lambda$ is highly stable with respect to changes in the training set $\mathcal{T}$ from the same elements in $\mathcal{S}$ when considering the sampling pattern $\Omega$. Hence, changes in the training sets from the same collection $\mathcal{S}$ can give vastly different results. The conclusion is therefore that unless one has prior information about the training data, or a potential way of learning this information, setting the $\lambda$ parameter is a delicate affair. Ironically, one ends up with a potentially unstable problem in order to cure the instability issue in the original problem.

*Proof Theorem 3.4.15.* Let $\tilde{\Omega} \subset \{1, \dots, N\}$ such that $2 \leq |\tilde{\Omega}| \leq N - 2$. Note that there are $2^N - 2N - 2$ different choices of $\tilde{\Omega}$.

Moreover, let $\Omega = \tilde{\Omega} \cup \{j\}$ where $j \notin \tilde{\Omega}$, so $\Omega \neq \{1, \dots, N\}$ by the fact that $|\tilde{\Omega}| \leq N - 2$. Choose $\mathcal{M}_1 \subset \mathcal{N}(\tilde{A})^\perp$ of size $K$. If $K$ is infinite, choose $\mathcal{M}_1$ to be countable. By multiplying $\mathcal{M}_1$ by any real number we clearly get uncountably many different choices of $\mathcal{M}_1$. For simplicity of notation let $\mathcal{M}_2 = A(\mathcal{M}_1)$ and $\tilde{\mathcal{M}}_2 = \tilde{A}(\mathcal{M}_1)$.

Choose $i \in \tilde{\Omega}$. Let $\tilde{\mathcal{T}} \subset \tilde{\mathcal{M}}_2 \times \mathcal{M}_1$ be any finite non-zero collection such that there is exactly one pair $(\hat{y} = \tilde{A}\hat{x}, \hat{x}) \in \tilde{\mathcal{T}}$ such that

$$\tilde{\mathcal{T}} \setminus \{(\hat{y}, \hat{x})\} \subset \{(\tilde{y}, \tilde{x}) \,|\, \tilde{y} = \tilde{A}\tilde{x}, P_i\tilde{y} = 0\}, \quad P_i\hat{y} \neq 0, \quad P_{\tilde{\Omega}\setminus\{i\}}\hat{y} = 0, \tag{3.22}$$

where $P_i$ denotes the projection onto the $i$-th coordinate. Note that such a choice is possible since $2 \leq |\tilde{\Omega}|$. Choose any non-empty $\mathcal{T} \subset \mathcal{M}_2 \times \mathcal{M}_1$. As both $\mathcal{T}$ and $\tilde{\mathcal{T}}$ can be multiplied by any real number that would not change any of the properties outlined above, we clearly have uncountably many different choices of $\mathcal{T}$ and $\tilde{\mathcal{T}}$. Note that

$$\tilde{A}^\dagger y = x \text{ if } y = \tilde{A}x, \quad x \in \mathcal{M}_1, \tag{3.23}$$

where $\tilde{A}^\dagger$ denotes the pseudoinverse. This fact will be crucial later in the argument.

Next we write all networks as complex valued for clarity, yet it should implicitly be understood that they can be written as real-valued, by doubling all dimensions. Consider the $L$-layer ReLU neural network $\tilde{\Psi} : \mathbb{C}^{|\tilde{\Omega}|} \to \mathbb{C}^N$ defined by

$$\tilde{\Psi}(x) = \tilde{A}^\dagger W_2 \rho(\dots \rho(W_1 W_2 \rho(W_1 W_2 \rho(W_1 x)))),$$

$$W_1 = [1,-1]^T \otimes I_{|\tilde{\Omega}|}, \quad W_2 = [1,-1] \otimes I_{|\tilde{\Omega}|}, \tag{3.24}$$

where we use the notation $I_d$ for the $d$-dimensional identity matrix and $\otimes$ denotes the Kronecker product.

Observe that for any pair $(\tilde{y} = \tilde{A}\tilde{x}, \tilde{x}) \in \tilde{\mathcal{M}}_2 \times \mathcal{M}_1$ we have

$$\begin{aligned}\tilde{\Psi}(\tilde{y}) &= \tilde{A}^\dagger W_2 \rho(\dots \rho(W_1 W_2 \rho(W_1 W_2 \rho(W_1 \tilde{y}))))\\ &= \tilde{A}^\dagger W_2 \rho(\dots \rho(W_1 W_2 \rho(W_1([1,-1] \otimes I_{|\tilde{\Omega}|})\rho(([1,-1]^T \otimes I_{|\tilde{\Omega}|})\tilde{y}))))\\ &= \tilde{A}^\dagger W_2 \rho(\dots \rho(W_1 W_2 \rho(W_1([1,-1] \otimes I_{|\tilde{\Omega}|})[\rho(\tilde{y}), \rho(-\tilde{y})]^T)))\\ &= \tilde{A}^\dagger W_2 \rho(\dots \rho(W_1 W_2 \rho(W_1(\rho(\tilde{y}) - \rho(-\tilde{y}))))) = \tilde{A}^\dagger \tilde{y} = \tilde{x},\end{aligned} \tag{3.25}$$

which follows from (3.24), (3.23) and the easy observation that, since $\rho$ is the ReLU function, $\rho(\tilde{y}) - \rho(-\tilde{y}) = \tilde{y}$. Also, recalling that $n = (m, n_1, \dots, n_{L-1}, N)$, $m = \max\{|\Omega|, |\tilde{\Omega}|\}$ and the assumption that $n_j \geq 2m$, it is clear that by replacing $W_j$ with $W_j \oplus 0$ we can without loss of generality assume that $\tilde{\Psi} \in \mathcal{NN}_1$. Hence, by setting $\lambda_{\text{opt}}$ to zero, and using (3.25) we have that

$$\sum_{(\tilde{y},\tilde{x}) \in \tilde{\mathcal{T}}} \frac{1}{2} \|\tilde{x} - \tilde{\Psi}(\tilde{y})\|_{\ell^2}^2 + \lambda_{\text{opt}} J(\hat{\Psi}) = 0. \tag{3.26}$$

We will continue with this choice of $\lambda_{\text{opt}}$ throughout the argument. We claim that $\lambda_{\text{opt}}$ is optimal for $\{(\tilde{A}, \mathcal{M}_1), \mathcal{NN}, \tilde{\mathcal{T}}\}$. Indeed, (3.26) implies that $\tilde{\Psi}$ is a minimiser of (3.21) for $\lambda = \lambda_{\text{opt}}$. Thus, we only need to show that $\tilde{\Psi}$ is an optimal mapping. To see this we observe that by (3.25) it follows that

$$\begin{aligned}c_{\text{opt}}(\tilde{A}, \mathcal{M}_1) &= \inf_{\varphi : \tilde{\mathcal{M}}_2 \rightrightarrows \mathbb{C}^N} \sup_{x \in \mathcal{M}_1} d_1^H(\varphi(\tilde{A}x), x)\\ &\leq \sup_{x \in \mathcal{M}_1} d_1^H(\tilde{\Psi}(\tilde{A}x), x) = 0,\end{aligned} \tag{3.27}$$

and hence our claim that $\lambda_{\text{opt}}$ is optimal for $\{(\tilde{A}, \mathcal{M}_1), \mathcal{NN}, \tilde{\mathcal{T}}\}$ is true. That $\lambda_{\text{opt}}$ also is optimal for $\{(A, \mathcal{M}_1), \mathcal{NN}, \mathcal{T}\}$, will be shown when we consider part $(2)$.

We will now establish the collection $\mathcal{S}$ as described in the statement of the theorem. Note that $\mathcal{N}(\tilde{A}) \cap \mathcal{N}(A)^\perp$, is non-zero by our choice of $\Omega$. We choose a non-zero $x \in \mathcal{N}(\tilde{A}) \cap \mathcal{N}(A)^\perp$ and let $z = \hat{x} + x$ where $\hat{x}$ is from (3.22), and let $\mathcal{S}$ denote any non-zero uncountable collection of multiples of $z$.

To show $(1)$, we note that it is enough to show that $\lambda_{\text{opt}}$ is no longer optimal if we add or replace a specific element of $\tilde{\mathcal{T}}$ with an element from $\mathcal{S}$. First, suppose we replace $(\hat{y}, \hat{x}) \in \tilde{\mathcal{T}}$ by $(\hat{y}, \hat{x} + x)$, and define the neural network $\Phi \in \mathcal{NN}$ by

$$\Phi(x) = T\rho(\dots \rho(W_1 W_2 \rho(W_1 W_2 \rho(W_1 x)))), \tag{3.28}$$

where $W_1$ and $W_2$ are defined in (3.24) and

$$T = C([1,-1] \otimes I_{|\tilde{\Omega}|}), \qquad C = \tilde{A}^\dagger P_{\tilde{\Omega} \setminus \{i\}} + (\frac{1}{\hat{y}^i}(\hat{x} + x) \otimes e_i^T).$$

We observe that for $(\tilde{y}, \tilde{x}) \in \tilde{\mathcal{T}}$ with $P_i\tilde{y} = 0$ we have $\Phi(\tilde{y}) = \tilde{\Psi}(\tilde{y}) = \tilde{x}$, and for $(\hat{y}, \hat{x} + x) \in \tilde{\mathcal{T}}$ we have by (3.22) that

$$\Phi(\hat{y}) = \left(\tilde{A}^\dagger P_{\tilde{\Omega}\setminus\{i\}} + \left(\frac{1}{\hat{y}^i}(\hat{x} + x) \otimes e_i^T\right)\right)\hat{y} = \hat{x} + x.$$

Hence, the objective function in (3.21) is zero at $\Phi$ whenever $\lambda = \lambda_{\text{opt}}$. Thus, any

$$\hat{\Phi} \in \operatorname*{argmin}_{\hat{\Psi}\in\mathcal{NN}} \frac{1}{|\tilde{\mathcal{T}}|} \sum_{(\tilde{y},\tilde{x})\in\tilde{\mathcal{T}}} \frac{1}{2}\|\tilde{x} - \hat{\Psi}(\tilde{y})\|_{\ell^2}^2 + \lambda_{\text{opt}} J(\hat{\Psi}) \tag{3.29}$$

will satisfy $\hat{\Phi}(\hat{y}) = \hat{x} + x$, which means that $\sup_{x\in\mathcal{M}_1} d_1^H(\hat{\Phi}(\tilde{A}x), x) \neq 0$ which by (3.27) means that $\hat{\Phi}$ is not an optimal map. Hence $\lambda_{\text{opt}}$ is no longer optimal for $\{(\tilde{A}, \mathcal{M}_1, \mathcal{NN}, \tilde{\mathcal{T}})\}$.

Now let us consider the case where $z$ is added to $\tilde{\mathcal{T}}$. Consider the neural net $\tilde{\Phi}$ defined by (3.28) with

$$T = D([1, -1] \otimes I_{|\tilde{\Omega}|}), \qquad D = \tilde{A}^\dagger P_{\tilde{\Omega}\setminus\{i\}} + (\frac{1}{\hat{y}^i}(\hat{x} + \frac{1}{2}x) \otimes e_i^T).$$

For $(\tilde{y}, \tilde{x}) \in \tilde{\mathcal{T}}$ with $P_i\tilde{y} = 0$ we have $\tilde{\Phi}(\tilde{y}) = \tilde{x}$, and for $(\hat{y}, \hat{x} + x), (\hat{y}, \hat{x}) \in \tilde{\mathcal{T}}$ we have by (3.22) that

$$\tilde{\Phi}(\hat{y}) = \left(\tilde{A}^\dagger P_{\tilde{\Omega}\setminus\{i\}} + (\frac{1}{\hat{y}^i}(\hat{x} + \frac{1}{2}x) \otimes e_i^T)\right)\hat{y} = \hat{x} + \frac{1}{2}x.$$

Hence, $\frac{1}{|\tilde{\mathcal{T}}|}\sum_{(\tilde{y},\tilde{x})\in\tilde{\mathcal{T}}} \frac{1}{2}\|\tilde{x} - \tilde{\Phi}(\tilde{y})\|_{\ell^2}^2 + \lambda_{\text{opt}} J(\tilde{\Phi}) = \frac{1}{|\tilde{\mathcal{T}}|}\|\frac{1}{2}x\|_{\ell^2}^2$, and therefore any minimiser $\hat{\Psi}$ of (3.29) will satisfy

$$\frac{1}{|\tilde{\mathcal{T}}|} \sum_{(\tilde{y},\tilde{x})\in\tilde{\mathcal{T}}} \frac{1}{2}\|\tilde{x} - \hat{\Psi}(\tilde{y})\|_{\ell^2}^2 + \lambda_{\text{opt}} J(\hat{\Psi}) \leq \frac{1}{4|\tilde{\mathcal{T}}|}\|x\|_{\ell^2}^2. \tag{3.30}$$

However, by (3.27), any optimal map $\varphi$ for $(\tilde{A}, \mathcal{M}_1)$ will satisfy

$$\sup_{x\in\mathcal{M}_1} d_1^H(\varphi(\tilde{A}x), x) = 0.$$

Thus, $\frac{1}{|\tilde{\mathcal{T}}|}\sum_{(\tilde{y},\tilde{x})\in\tilde{\mathcal{T}}} \frac{1}{2}\|\tilde{x} - \varphi(\tilde{y})\|_{\ell^2}^2 = \frac{1}{2|\tilde{\mathcal{T}}|}\|x\|_{\ell^2}^2$, and therefore, by (3.30), no minimiser $\hat{\Psi}$ of (3.29) can be an optimal map for $(\tilde{A}, \mathcal{M}_1)$. Hence $\lambda_{\text{opt}}$ is not optimal for $\{(\tilde{A}, \mathcal{M}_1, \mathcal{NN}, \tilde{\mathcal{T}})\}$.

Consider part (2). Let $\mathcal{D} \subset \mathcal{M}_1 \cup \mathcal{S}$ be any finite non-empty set, and let $\mathcal{T} = \{(Ax, x) : x \in \mathcal{D}\}$. We shall prove that $\lambda_{\text{opt}}$ is optimal for $\{(A, \mathcal{N}(A)^\perp), \mathcal{NN}, \mathcal{T}\}$ for any such $\mathcal{T}$. Note that this is a stronger statement than in the theorem, as $\mathcal{M}_1 \cup \mathcal{S} \subset \mathcal{N}(A)^\perp$. From (3.27) it is clear that $y \mapsto A^\dagger y$ is an optimal map for $(A, \mathcal{N}(A)^\perp)$. Using the network $\tilde{\Psi}$ from (3.25), where $\tilde{A}^\dagger$ is replaced by $A^\dagger$ in the last layer, it is clear that $\tilde{\Psi}$ is an optimal map, and a minimiser of (3.29) for $\lambda_{\text{opt}}$ when we sum over $\mathcal{T}$. Thus $\lambda_{\text{opt}}$ is optimal for $\{(A, \mathcal{N}(A)^\perp), \mathcal{NN}, \mathcal{T}\}$. □

## 3.5 Outlook and potential remedies for AI generated hallucinations

If as in current research, one applies methods in DL for solving (3.1) in different methods, summarised in Section 1.5 and [14, 68, 99, 102, 166], the conditions for our theoretical results are met. All of the modalities, in Section 1.5, can be written as (3.1) and as the corresponding inverse problem is generally undetermined, our theoretical results apply. Thus our main results provide insight on the difficulties of protecting against hallucinations and optimising the accuracy-stability trade-off for DL used to solve many different inverse problems. Below follows a discussion and examples of the challenges of training NNs that avoid hallucinations while keeping performance and stability.

### 3.5.1 Remedies for causes of AI generated hallucinations and instabilities

In the following subsections possible remedies for causes of AI generated hallucinations and instabilities are indicated. When discussing possible remedies it is necessary for assessment to determine how these should be evaluated. NN algorithms for solving inverse problems can be examined through the traditional pillars of numerical analysis, namely, *accuracy* and *stability*. Naturally, these issues are tied to the conditioning of the sampling procedure and for ill-conditioned problems there exist theoretical limitations. There is a growing awareness that such techniques have not yet been subject to the same rigorous standards as more well-established methods in scientific computing [11]. Moreover, there is evidence that such techniques, in their current guise at least, do not yet meet these standards. For instance in image reconstruction, recently it has been demonstrated that existing DL algorithms, despite offering purportedly 'superior immunity to noise' [171], are often highly *unstable* [6, 102] and often suffer from *AI generated hallucinations* [14, 16, 68, 99, 138, 188]. The instability phenomenon in DL for inverse problems is on the one hand similar, but on the other hand different, to the better-known phenomenon of *adversarial attacks* in DL for classification problems [35]. In both cases such susceptibilities potentially have serious consequences. Firstly, areas where DL techniques are designed to perform tasks hereto performed by humans. For example, automatic diagnosis in medicine, as has already been approved for commercial use in April 2018 by the US Food and Drug Administration (FDA) [70]. Similarly, a recent publication in *Science* [73], for example, warns about the potentially severe consequences in insurance fraud. Secondly, as noted above, areas where DL techniques replace well-established algorithms in the computational sciences. Instabilities therein may lead to incorrect scientific predictions, with serious downstream consequences. A serious concern in medical imaging are AI generated hallucinations yielding artefacts that are not recognizable as such and, hence, hard to detect in applica-

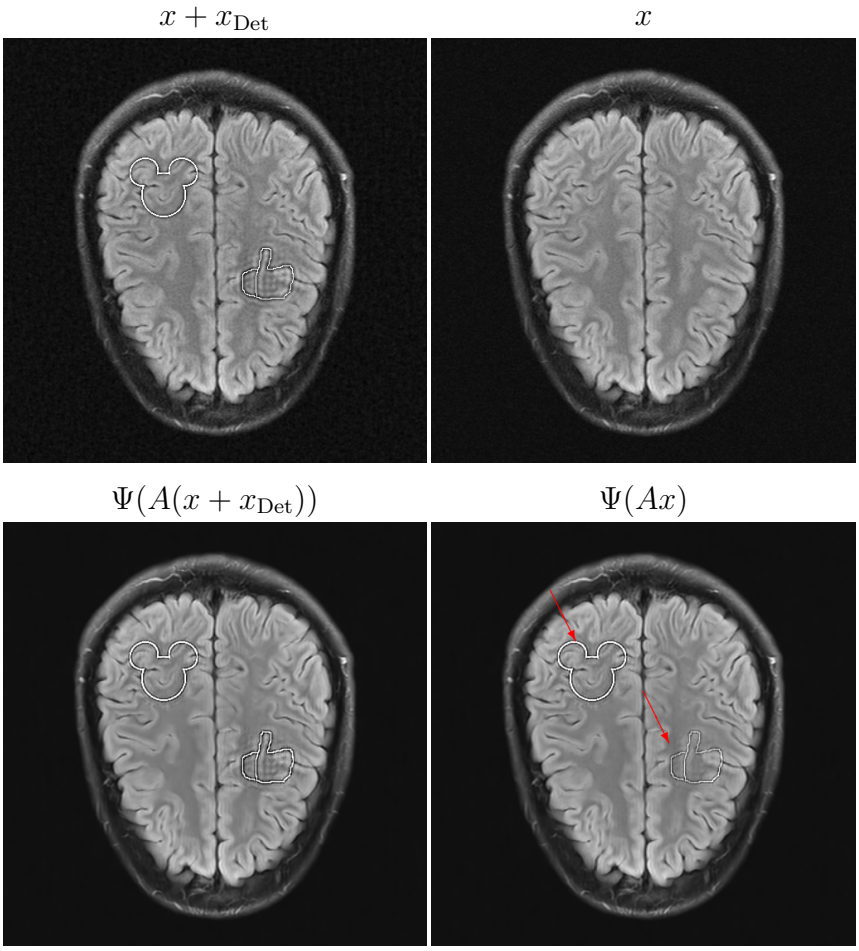


Figure 3.7: (**Detail transfer**) We consider the setup from Theorem 3.4.2, with two images $x + x_{\mathrm{Det}}$ and $x$, seen in the top row. Here $x_{\mathrm{Det}}$ is the Mickey and thumb detail seen in the upper left image. The detail $x_{\mathrm{Det}}$ is constructed so that $\|P_{\mathcal{N}(A)^\perp} x_{\mathrm{Det}}\|$ is small, and the neural network $\Psi$ is trained to high accuracy on the pair $(A(x + x_{\mathrm{Det}}), x + x_{\mathrm{Det}})$, and 1200 other images from the fastMRI dataset. As we can see from the lower right image, the neural network $\Psi$ transfers the detail $x_{\mathrm{Det}}$ onto the image $x$. In this example $A = P_\Omega F$, where $F \in \mathbb{C}^{N\times N}$ is a two-dimensional Fourier transform, and $|\Omega|/N = 1/16$, $N = 512$.

tions [14, 16, 138, 166]. This establishes the need to assess the accuracy and stability of NN algorithms for solving inverse problems. However, this leads to the following question.

**Remark 3.5.1. Why are AI generated hallucinations an issue if DL has shown to be more accurate than state of the art methods in image reconstruction?**
This question arises, for example in Fig. 3.7, the neural network $\Psi$ reconstructs the Mickey-Mouse feature and the Facebook thumb from $x$, yet the CS reconstruction washes out the Mickey-Mouse feature in the reconstruction. The key point here is, that for state of the art methods, such as CS, when a detail is present in the reconstruction it usually exists. However, details may be washed out due to subsampling. With a given time and cost budget, for example in MRI, one uses subsampling from a higher resolution as it encaptures more details than full sampling from a lower resolution. Even though full sampling would

give all the information of the data one may still obtain artefacts in the reconstruction. A well-known example is the Gibbs ringing phenomenon. Hence, state of the art methods when using subsampling from a higher resolution may miss a detail. Yet, except for known artefacts, which are also quantifiable due to error bounds on the reconstruction accuracy, these methods do not add seemingly realistic details. An example of such quantifiable error bounds is given by Theorem 2.2.4, in Chapter 2. On the contrary, for DL methods there may be details apparent in the reconstruction that do not exist- thus the network may hallucinate. Yet, the DL reconstruction may be more accurate in certain cases and reconstruct details that are washed out by state of the art methods. As Fig. 3.7 exemplifies and Theorems 3.4.4 and 3.4.2 show, this comes at the cost of hallucinations. Hence, there is an *accuracy hallucination trade-off*.

Concerning possible remedies for AI generated hallucinations, note that **stable recovery methods can still hallucinate through detail transfer**. There has been a wide variety of attempts to make NNs in inverse problems more stable. However, this does not mean that one can automatically cure the detail-transfer hallucination phenomenon. Indeed, stabilisation techniques during training may help improve the stability-accuracy trade-off but may still provide hallucinating NNs. In Figure 3.7 we demonstrate how a stabilisation technique, known as "Jittering" and presented in [80], still yields hallucinations through detail transfer. Note that stabilising neural nets through training on noisy data has also been investigated for the classification problem [167].

If stability does not protect against AI hallucinations, the question arises how these hallucinations are caused and if the cause can be protected against. Figs. 3.3 and 3.2 give examples of such hallucinatory features, which may be caused by a combination of the above outlined explanations: a detail transferred from another image in the training set or an imperceptible change in the image causes an artefact in the reconstruction that cannot easily be dismissed as unphysical. These features occur with non-zero probability, for the first case due to detail transfer with respect to any kind of noise and for the latter, due to mean-zero, random perturbations of the measurements. These effects, the NN decoder becoming unstable and producing hallucinations, are not related to a particular network architecture. As shown in Table 3.1 these effects occur for various architectures trained using various methods. Furthermore, this is undermined by our theoretical results: the conditions do not specify any architecture. The only conditions on the decoders are continuity or Lipshitz continuity and sufficiently good reconstruction of specific elements in the training set. As outlined training typically encourages these conditions. Yet, for example, in sparse regularization methods conditions on the sampling operator protect from these. See Section 2.2.3 in Chapter 2. Our results indicate that a reason for instabilities and hallucinations is of lack of knowledge of the null-space of the sampling operator $A$ built into the decoder $\Psi$ and the data $\mathcal{M}_1$. This automatically also sheds light on possible remedies. For instance, the sampling operator, this also includes the sampling pattern, the training data and all test sets should be chosen to prohibit the above outlined conditions from occurring. This could be improved by choos-

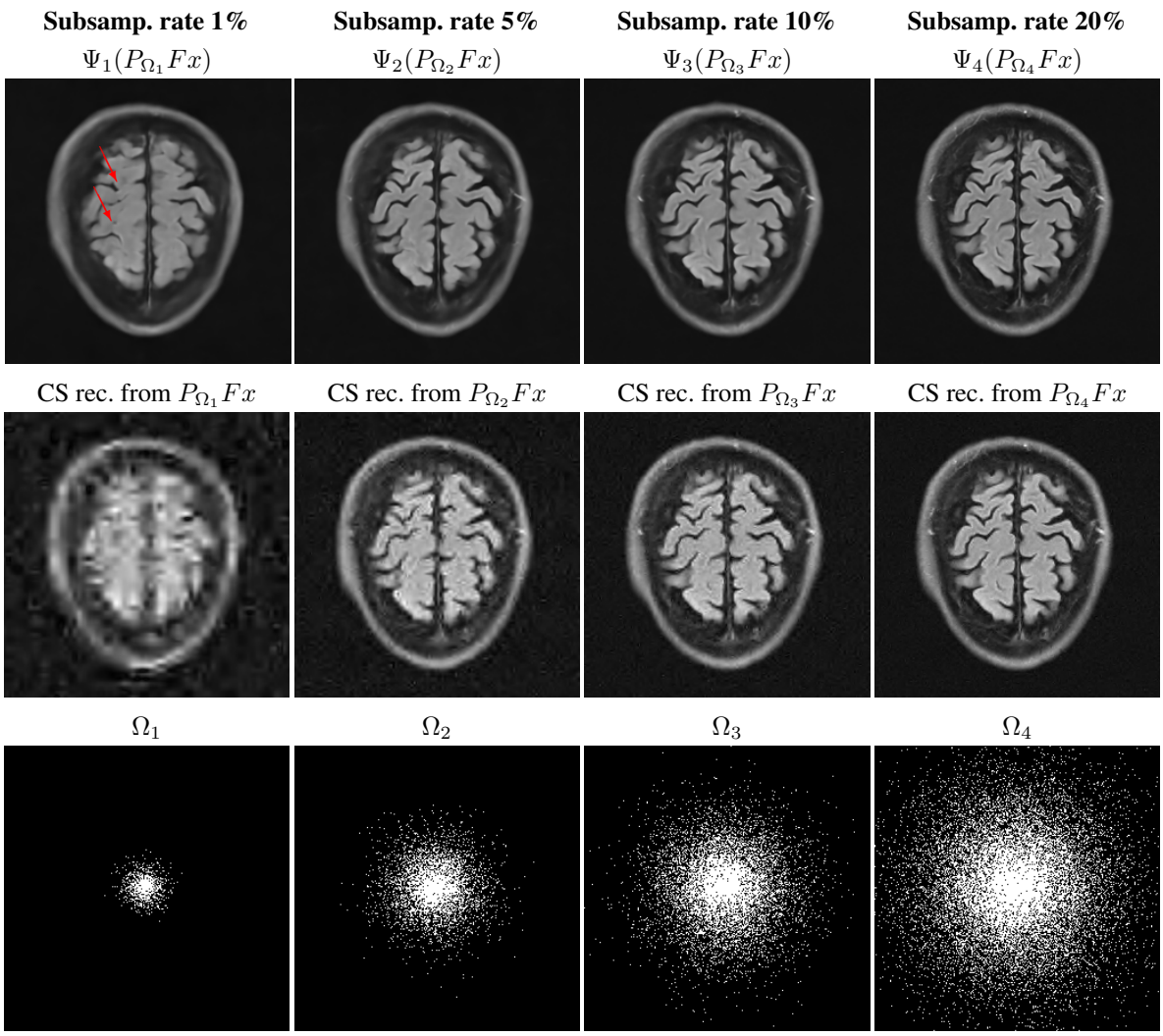


Figure 3.8: (**Trained NNs produce realistic looking images irrespectively of the sampling rate**). We train four neural networks $\Psi_j$ on 1200 images of size $N = 256 \times 256$ from the fastMRI challenge using Fourier sampling with the sampling patterns $\Omega_j \subset \{1, \dots, N\}$ seen in the last row. Here each $\Psi_j$ is trained on samples from the matrix $P_{\Omega_j}F$, where $F \in \mathbb{C}^{N\times N}$ is the two-dimensional Fourier matrix. We can see that as the sampling rate decreases the CS reconstruction produces more and more artifacts and for 1% subsampling, the image quality of the reconstructed image is too poor to provide any details about the underlying brain. The trained neural network, on the other hand, reconstructs realistic looking images at all sampling rates, however, the reconstructed image is not necessarily an accurate representation of the underlying brain, as the upper left image shows. In this experiment we used the neural network architecture $y \mapsto \phi_j(F^* P^*_{\Omega_j} y)$, where $\phi_j \colon \mathbb{C}^N \to \mathbb{C}^N$ is a U-net.

ing an appropriate regularization that enhances awareness of the method of the null space $\mathcal{N}(A)$ and the data $\mathcal{M}_1$. The precise implementation of these potential remedies is highly dependent on the application considered and, hence, a topic for future research. The upper and lower bounds of reconstruction accuracy based on the null space $\mathcal{N}(A)$ and the data $\mathcal{M}_1$, are examined in more detail in Chapter 4.

## 3.6 Discussion of existing remedies against instability

We provide an overview how despite standard attempts to protect against instabilities, AI generated hallucinations and instabilities still occur. This includes methods such as enforcing data consistency, training with random sampling patterns, adversarial training, augmenting the training set and adding random noise - also referred to as jittering.

### 3.6.1 Do bad perturbations occur in practice?

In the following, we discuss different noise models which may be relevant in practice. Based on Section 2.2.1, Chapter 2, recall the following. In practice, perturbations often arise as random noise on the measurements, i.e. realizations of a mean-zero random variable $e_{\text{pert}} : \Omega \to E$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $E = \mathbb{C}^m$ equipped with a norm $\|\cdot\|$ and the Borel measure. The fact that (3.6) and (3.7) in Theorem 3.4.7 hold for all perturbations within a ball implies lower bounds on the probability of having 'bad' perturbations realized by $e_{\text{pert}}$, as shown in the last two parts of Theorem 3.4.7. In the following sections we consider $\eta = \eta_1 = \eta_2$. This discussion pertains to *generic* noise, i.e. realizations of mean-zero, random variables. Yet in many applications, the measurement are corrupted not only by generic noise, but also by other phenomena. This is the case for instance in MRI, where factors such as small patient motion or small anatomic differences cause specific corruptions in the measurements. In such settings, a more suitable model of the random perturbation is

$$e_{\text{pert}} = e^1_{\text{pert}} + e^2_{\text{pert}} : \Omega \to E,$$

where $e^1_{\text{pert}}$ is a random variable that accounts for the non-generic part of the perturbation and $e^2_{\text{pert}}$ is a mean-zero random variable accounting for the generic part. While it is typically straightforward in applications to identify a reasonable model for $e^2_{\text{pert}}$ (e.g. Gaussian, Poisson, etc), it is usually much less straightforward to model $e^1_{\text{pert}}$. This, motivates establishing stability guarantees that address worst-case perturbations. If a realization of $e^1_{\text{pert}}$ results in a damaging perturbation, then we obtain

$$\mathbb{P}\left(\|\Psi(y + e_{\text{pert}}) - (x + z)\| \leq \eta \mid e^1_{\text{pert}} = e\right) \geq 1 - \epsilon. \tag{3.31}$$

Since $\epsilon$ will typically be small, this represents a high probability event: generic noise added to a damaging perturbation cannot counteract the damage. Such a phenomenon is shown

empirically in Fig. 3.6. This argument also demonstrates how difficult it may be to guarantee robustness to physical perturbations. In view of (3.31), doing so would involve ensuring that $e^1_{\text{pert}}$ rarely gives damaging perturbations and that $1-\epsilon$ is small.

### 3.6.2 The instability phenomenon is not easy to remedy

Having established the presence of instabilities, the next question to ask is: how might one make DL more robust? There are many strategies, yet proceeding in an ad-hoc fashion is both time- and resource-consuming. By establishing processes that lead to instabilities and AI generated hallucinations, Theorems 3.4.7, 3.4.4 and 3.4.2 are useful tools for excluding approaches that are unlikely to succeed in preventing instabilities. We now highlight three such strategies. The key idea is that any remedy which does not enforce awareness of the null space $\mathcal{N}(A)$ and the data $\mathcal{M}_1$, will remain susceptible to instabilities and AI hallucinations.

#### Enforcing consistency

Consistency of the reconstruction with the measured data is often desirable in practice, and many emerging DL strategies for inverse problems seek to enforce this property [89, 101]. However, this does not prohibit instabilities. Indeed, let $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be an arbitrary reconstruction map ($\Psi$ need not be a NN). Then, consistency of $\Psi$, i.e.

$$A\Psi(y) = y, \quad \forall y = Ax + e,\, x \in \mathcal{M}_1, \quad e \in \mathcal{B}(0,\epsilon). \tag{3.32}$$

does nothing to help one avoid the conclusions of Theorems 3.4.7 and 3.4.2. Indeed, the corresponding conditions

$$\|\Psi(Ax+e) - x\| < \eta, \quad \|\Psi(Ax'+e') - x'\| < \eta, \quad \|Ax + e - (Ax' + e')\| \leq \eta, \tag{3.33}$$

pertain to the quality of $\Psi$ as an approximation, and are unrelated to its consistency. In fact, if $\Psi(Ax) = x$ and $\Psi(Ax') = x'$, where $e = e' = 0$, i.e. $\Psi$ recovers $x$ and $x'$ perfectly, then clearly $\Psi$ is also consistent for $x$ and $x'$. A rather general approach to approximate consistency, as suggested in [89, 101], is to consider a set $\mathcal{S} \subset \mathbb{C}^N$ which either contains the images of interest, or approximates these images well. Then, for the $\ell^2$-norm one defines the reconstruction mapping $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ as

$$y \mapsto \Phi(y) \in \operatorname*{argmin}_{\tilde{x} \in \mathcal{S}} \frac{1}{2}\|A\tilde{x} - y\|^2_{\ell^2}. \tag{3.34}$$

However, if $x$ and $x'$ satisfy $\|Ax - Ax'\|^2_{\ell^2} \leq \eta$ and additionally, $x, x' \in \mathcal{S}$, then (3.33) still holds, and, thus, instabilities occur.

**Training with random sampling patterns**

As noted, in applications such as MRI, the measurement matrix takes the form $A = P_\Omega F$, where $\Omega \subseteq \{1, \dots, N\}$, $|\Omega| = m$ is the set of frequencies sampled and $P_\Omega$ is the projection onto these indices. Another approach to improve the robustness of DL, suggested in [170], is to train with many different sampling patterns at once. Specifically, one now considers the decoder $\Psi$ as a map from $\mathbb{C}^N \to \mathbb{C}^N$ and the measurement vector $y$ as an element of $\mathbb{C}^N$, where the components in $y$ that correspond to the unsampled indices are set to zero. The mapping $\Psi$ is then found by training on the data $\{(y^{ji}, x^j) : j = 1, \dots, K, i = 1, \dots, L\}$, where $y^{ji} = P_{\Omega_i} x_j$ and $\Omega_i$ is the $i$th sampling pattern. For instance, one may define

$$\Psi \in \underset{\tilde{\Psi} \in \mathcal{NN}}{\operatorname{argmin}} \frac{1}{KL} \sum_{j=1,i=1}^{K,L} \frac{1}{2} \| x^i - \tilde{\Psi}\left(P_{\Omega_j} F x^i\right) \|_{\ell^2}^2. \tag{3.35}$$

Once $\Psi$ is trained, it is used to reconstruct an image $x$ from measurements $y = P_\Omega F$ acquired from a given sampling pattern $\Omega$ of size $|\Omega| = m$ (which may or may not be equal to $\Omega_i$ for some $i$). In particular, even though it is trained using $\Omega_1, \dots, \Omega_L$, when used as a reconstruction map one only has access to data from one sampling pattern of size $m$, and not all $L$ sampling patterns used in the training. This type of training does nothing to obviate instabilities, and in fact, may make them more likely. Training on more data, as in (3.35), is likely to improve the quality of the reconstruction map $\Psi$, making it easier to achieve the conditions

$$\|\Psi(P_\Omega F x) - x\|_{\ell^2}^2 < \eta, \quad \|\Psi(P_\Omega F x') - x'\|_{\ell^2}^2 < \eta, \quad \|x - x'\|_{\ell^2}^2 \gg \eta.$$

Yet, the amount and variety of the training data is completely unrelated to the null space of $P_\Omega F$, thus it does nothing to mitigate against the condition

$$\|P_\Omega F x - P_\Omega F x'\|_{\ell^2}^2 \leq \eta.$$

Note that the network used in Figs. 3.3 and 3.5 is trained with random sampling as discussed above, yet it is highly unstable. Moreover, concerning AI hallucinations Fig. 3.8 indicates a rather cumbersome finding. Namely, trained NNs produce realistic looking images irrespectively of the sampling rate. Usually, the performance of standard methods tends to decline with lowes sampling rates. Yet, as shown in Fig. 3.8 this is not necessarily the case for learned methods. This may be an indication that such methods are more prone to satisfying the conditions for additional or removed elements in the reconstruction arising in Theorem 3.4.7.

**Adversarial training/augmenting the training set**

In image classification, a common strategy to enhance robustness to perturbations is to perform adversarial training [82, 135, 174]. One may view this as a way of increasing the

size of the training set. Other strategies for example include data augmentation. There is no reason why an increase in the amount of training data will mitigate against instabilities and hallucinations in inverse problems. As long as the class $\mathcal{NN}$ of neural networks is rich enough to ensure a small training error, then the trained network $\Psi$ will satisfy the following. Firstly, the conditions of Theorem 3.4.7 are encouraged,

$$\|\Psi(Ax+e)-x\|<\eta, \quad \|\Psi(Ax'+e')-x'\|<\eta, \tag{3.36}$$

for some small $\eta>0$ and all $(Ax+e,x)$, $(Ax'+e',x')$ in the training set. The size of the training set is irrelevant. If

$$\|Ax+e-(Ax'+e')\|\leq\eta, \quad \|x-x'\|\gg\eta, \tag{3.37}$$

for any two such pairs $(Ax+e,x)$, $(Ax'+e',x')$, then Theorem 3.4.7 applies. Secondly, the conditions of Theorem 3.4.4 are encouraged as well. We have that

$$\|\Psi(A(x+x_{\mathrm{Det}}))-(x+x_{\mathrm{Det}})\|\leq\eta \tag{3.38}$$

and if at the same time this element satisfies

$$\|Ax_{\mathrm{Det}}\|\leq\eta, \tag{3.39}$$

then Theorem 3.4.4 applies. Thus, $\Psi$ will hallucinate on measurements of $x$ uncountably many other objects with any kind of noise added. Moreover, (3.37) and (3.38) may even be encouraged by more training data. Firstly, since there are simply more pairs of $(Ax+e,x)$, $(Ax'+e',x')$ available that satisfy (3.36) and, secondly, as there are more elements in the training set that satisfy (3.39).

#### Adding random noise

In [8, p.138] and [80], the prospect of adding additional random noise to the measurements has been raised as a potential way to combat instabilities. This is also referred to as *jittering* [80]. This is a tempting idea, and it would have succeeded had the collection of 'bad' perturbations belonged to a set of measure zero. However, as Theorem 3.4.7 reveals this is not the case: the 'bad' perturbations have balls around them containing further 'bad' perturbations. Recall also the discussion in §3.6.1 on probabilistic aspects of instabilities. This phenomenon is illustrated in Fig. 3.6, where small random noise is added to the perturbation without having any mitigating effect. Moreover, in the case of hallucinations due to detail transfer, Theorem 3.4.4 states that these occur regardless of the noise added, one could also add no noise, $e=0$. Thus, jittering is no remedy for hallucinations due to detail transfer.

## 3.7 Methods

In this section, we describe the methods used to generate the various numerical results.

**Parameters for** (3.40)

| | Wavelet | $\eta$ |
|---|---|---|
| Fig. 3.6. | DB2 | 0.1 |
| Fig. 3.5. | DB4 | 0.001 |

**Parameters for** (3.41)

| | $n$ | $\alpha_1$ | $\alpha_2$ |
|---|---|---|---|
| Fig. 3.3. | 500 | 1 | 1 |

Table 3.2: Parameters for the sparse regularization decoders.

### 3.7.1 Sparse regularization decoders

The sparse regularization decoder used in Figure 3.5, is

$$\underset{z\in\mathbb{C}^N}{\text{minimize}} \|z\|_{\ell^1} \text{ subject to } \|AH^\top z - y\|_{\ell^2}^2 \leq \eta, \tag{3.40}$$

where $\eta \geq 0$ is a noise parameter and $H \in \mathbb{R}^{N\times N}$ is a discrete wavelet transform with a Daubechies orthonormal wavelet. The sampling operator $A$ is a subsampled discrete Fourier transform in all of these experiments. To search for a minimiser of the above optimization problem, we used the SPGL1 [181] software package. The chosen parameters in each figure can be found in Table 3.2.

For Fig. 3.3, we used the sparse regularization decoder introduced in [128]. This is a more advanced decoder, supporting different types of regularizes and also non-Fourier sampling operators, such as Radon sampling.

The decoder proposed in [128] tries to iteratively solve the optimization problem

$$\underset{z\in\mathbb{C}}{\text{minimize}} \sum_{j=1}^{J} \lambda_j \|W_j H_j z\|_{\ell^1} + \text{TGV}_\alpha^2(z) \text{ subject to } Az = y \tag{3.41}$$

using $n$ iterations. Here the $W_j$'s are diagonal weighting matrices, $\lambda_j \in \mathbb{R}_+$ are weighting parameters, and $H_j$ is the $j$'th subband in a shearlet transform. The weights $W_j$ and $\lambda_j$ are updated iteratively between each iteration. The $\text{TGV}_\alpha^2(z)$ term is a second order Total Generalised Variation operator depending on two parameters $\alpha = (\alpha_1, \alpha_2)$, where the first order term (TV) is weighted by $\alpha_1$ and a second order (generalised) term weighted by $\alpha_2$.

In all experiments we used shearlets with 4 scales and directional parameters $[0, 0, 1, 1]$. The complete set of parameters can be found in Table 3.2.

### 3.7.2 Creating Gaussian noise in $\mathcal{N}(A)^\perp$

In Fig. 3.3 and 3.6 we construct Gaussian vectors $v \in \mathbb{C}^N$ of a fixed magnitude, all lying in $\mathcal{N}(A)^\perp$. This is done as follows. We draw the real and imaginary components of a vector $e \in \mathbb{C}^m$ form a Normal distribution $\mathcal{N}(0, 10)$ and compute $v$ as $\alpha A^* e = v$, where $\alpha \in \mathbb{R}_+$ is a scalar. The $\alpha$ is chosen so that $v$ gets the desired norm. A Gaussian random variable

is still Gaussian after a linear map, so the vector $v$ is Gaussian. Furthermore, since since $AA^* = I$ for the special case where $A = P_\Omega F$ is a subsampled discrete Fourier transform, it follows that $v \in \mathcal{N}(A)^\perp$.

## 3.8 Conclusion

The purpose of this paper is to initiate a programme into the rigorous foundations of NNs and DL from the dual pillars of numerical analysis, *accuracy* and *stability*. While much of the focus on DL in the machine learning community has been on discrete problems such as classification, this paper aims to highlight both the challenges and potential when applying DL to continuous problems in computational mathematics. Due to both their ubiquity in the computational sciences and the recent activity on data-driven approaches for them, we have chosen to focus on inverse problems. For inverse problems, the conclusions from our findings are decidedly mixed: current approaches to training cannot ensure stable methods; even if they do, the resulting methods may not offer state-of-the-art performance; regularization strategies may not fix these issues. Furthermore, instabilities and AI generated hallucinations are not rare events, able to be dismissed by all bar a small group of theoreticians (recall Fig. 3.3). Should one therefore give up on the DL approach to inverse problems? Of course not. The rich approximation theory – dating back to the classical *Universal Approximation Theorem* (see, e.g. [148]) but including many recent advances such as [162, 186] – says that NNs have the potential to give rise to powerful methods for inverse problems in imaging. Our hope is that these findings, in particular the crucial role of *awareness* of the reconstruction method of the null space $\mathcal{N}(A)$ and the data $\mathcal{M}_1$, spur new research into devising better ways to design and train stable and accurate DL algorithms. Moreover, as we have shown there is an accuracy-stability and an accuracy-hallucination trade-off. Yet, the results also provide insights into increasing stability of DL for solving inverse problems and avoiding hallucinations by preventing the conditions in our main results from occurring.

# Chapter 4

# On existence, accuracy, stability and learning of approximate decoders for ill-posed inverse problems

The following chapter is concerned with universal accuracy bounds for ill-posed inverse problems and the existence, stability and robustness of decoders for such problems. Moreover, it aims to provide insight into the application and possibilities of DL applied to solve ill-posed inverse problems. This chapter is based on joint work with Paolo Campodonico, University of Cambridge, who contributed with discussions and comments, Vegard Antun, University of Oslo, who proofread this chapter and was supervised Anders C. Hansen, University of Cambridge.

## 4.1 Introduction

Solving ill-posed inverse problems is an ongoing area of active research. In this chapter, we investigate the existence, stability and learning of solutions for ill-posed, possibly non-linear, inverse problems.

In the previous chapters we considered to problem of recovering a vector $x \in \mathbb{C}^N$ given a measurement $y \in \mathbb{C}^m$ of the form

$$y = Ax + e \tag{4.1}$$

where $A : \mathbb{C}^N \to \mathbb{C}^m$ is a linear operator, called a *sampling operator*, and $e \in \mathbb{C}^m$ is additive noise.

The above model, (4.1), is standard for most medical imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT) [65] and compressive fluorescence microscopy [187]. Details on applications are given in Section 1.5. In this chapter,

we will consider the case in which the system in (4.1) is ill-posed, possibly non-linear, and, thus, obtaining a stable, accurate and robust decoder is challenging. For instance, this is the case in compressive imaging, which pertains to accurate and stable reconstruction of images from undersampled measurements.

Comapred to (4.1) there are, however, many other possible models for inverse problems: for example, the operator $A$ might be non-linear, and the noise $e$ might be multiplicative instead of additive, or a combination of both. These inverse problems can be represented as

$$\begin{aligned} y &= \mathcal{A}(x) + e && \text{(nonlinear operator, additive noise)} \\ y &= \mathcal{A}(x) \odot e && \text{(multiplicative noise)} \\ y &= e_1 \odot \mathcal{A}(x) + e_2 && \text{(mixture of additive and multiplicative noise)} \end{aligned}$$

where $\mathcal{A} : \mathbb{C}^N \to \mathbb{C}^m$ is non-linear, and $\odot$ represents the entrywise multiplication of vectors.

When formalising an inverse problem, it is fundamental to specify a noise model, and similarly the class of vectors $x$ that we wish to reconstruct. This is a crucial point that will be highlighted throughout this chapter and was introduced in the previous chapters. In particular, we assume that the noise $e$ belongs to a set $\mathcal{E} \subseteq \mathbb{C}^m$. In the case of a mixture of additive and multiplicative noise one can assume that $(e_1, e_2)$ belongs to a set $\mathcal{E} \subseteq \mathbb{C}^m \times \mathbb{C}^m$. We will use the first case for notational ease in the following. We assume the unknown $x$ belongs to a class $\mathcal{M}_1 \subseteq \mathbb{C}^N$, which is sometimes referred to as a 'manifold'.

In the following, we consider a general version of inverse problems that encompasses the representations presented above. Namely, a general measurement model of the form

$$y = F(x, e), \quad x \in \mathcal{M}_1,\ e \in \mathcal{E} \tag{4.2}$$

where $F : \mathcal{M}_1 \times \mathcal{E} \to \mathbb{C}^m$, $\mathcal{M}_1 \subseteq \mathbb{C}^N$, and $\mathcal{E} \subseteq \mathbb{C}^m$. The set of all possible noisy measurements $y$ is

$$\begin{aligned} \mathcal{M}_2^{\mathcal{E}} &:= \{y \in \mathbb{C}^m : \exists x \in \mathcal{M}_1, \exists e \in \mathcal{E}, y = F(x, e)\} \\ &= \mathrm{Im}(F) = F(\mathcal{M}_1 \times \mathcal{E}) \end{aligned}$$

which is the image of the measurement model $F$. By definition $F$ is surjective onto $\mathcal{M}_2^{\mathcal{E}}$.

The aim of solving the inverse problem (4.2) is to obtain a solution map, more commonly referred to as a decoder,

$$\varphi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{C}^N$$

that is *accurate*, *stable*, and *robust*.

If (4.2) is ill-posed, in the sense that the set $\pi_1(F^{-1}(y)) = \{x \in \mathcal{M}_1 : \exists e \in \mathcal{E}, y = F(x, e)\}$ given $y \in \mathcal{M}_2^{\mathcal{E}}$ (where $\pi_1$ denotes projection on the first component) contains more than one element or is unbounded [8], then it may not be possible to achieve exact reconstruction. Thus one needs to choose a specific element from $\pi_1(F^{-1}(y))$ given $y \in \mathcal{M}_2^{\mathcal{E}}$ in

order to obtain a solution to (4.2). We measure the quality of the reconstruction and of the measurements by equipping $\mathbb{C}^N$ and $\mathbb{C}^m$ with metrics $d_1$ and $d_2$ respectively. Possibilities for obtaining a solution include minimising the reconstruction error $d_1(\varphi(y), x)$ for all $x \in \pi_1(F^{-1}(y))$ or the average reconstruction error when considering a probabilistic model.

The decoder $\varphi$ is *accurate*, if it faithfully reconstructs $x$ given a measurement $y = F(x, 0)$. In particular this means, that the distance $d_1(\varphi(F(x, 0)), x)$ is small. The decoder is *stable*, if it reconstructs $x$ sufficiently well given a measurement $y = F(\tilde{x}, 0)$ where $\tilde{x}$ is a perturbed version of $x$. This means that the distance $d_1(\varphi(F(\tilde{x}, 0)), x)$ is small, for $d_1(\tilde{x}, x)$ small. This definition is also used in [21,74]. The decoder is *robust*, if it reconstructs $x$ sufficiently well given a noisy measurement $y = F(x, e)$. In particular, the distance $d_1(\varphi(F(x, e)), x)$ is small. As the presence of noise and the perturbation of $x$ are closely related, robustness and stability are related. For example, the modulus of continuity determines the decoders stability, as in [166]. Overall, an accurate, stable and robust decoder will be able to recover a vector $x$ given a potentially noisy measurement $y = F(\tilde{x}, e)$ for $\tilde{x} \approx x$.

A key point in finding a stable, accurate and robust decoder is choosing the set $\mathcal{M}_1 \subset \mathbb{C}^N$ of vectors to reconstruct and determine conditions on the noisy measurement model $F$. In particular, there also have to be conditions on the noise model $\mathcal{E}$. The notation $\mathcal{M}_1 \subseteq \mathbb{C}^N$ is guided by the current trend in research to describe the set which is to be reconstructed as a 'manifold'. Since in many applications $\mathcal{M}_1$ is not a manifold in the usual mathematical definition, we will simply assume $\mathcal{M}_1$ to be a subset of $\mathbb{C}^N$. This also encompasses previous methods for solving inverse problems. Traditionally these methods are based on a particular structure of the set $\mathcal{M}_1$, such as unions of linear subspaces [20] (and in particular sparse vectors, which are central in compressed sensing [74]), but also point clouds [1, 109] and smooth manifolds [12, 61] have been considered.

In order to encompass probabilistic models, in which we aim to minimise the average reconstruction error, we will equip the sets $\mathcal{M}_1$ and $\mathcal{E}$ with measures $\mu_1$ and $\nu$. Then, (4.2) can model many possible settings for inverse problems:

(1) For the set $\mathcal{M}_1 \subseteq \mathbb{C}^N$, this includes: $\mathcal{M}_1 = \Sigma_s = \{x \in \mathbb{C}^N : |\{i : x_i \neq 0\}| \leq s\}$, the set of $s$-sparse vectors, $\mathcal{M}_1$ being a union of subspaces, a point cloud, a manifold or a general set, equipped with a probability measure $\mu_1$. In the atter case $\mu_1$ can represent a prior, in the Bayesian setting, on $\mathcal{M}_1$.

(2) For the noise model $\mathcal{E} \subseteq \mathbb{C}^m$ this includes: $\mathcal{E} = B_{d_2}(0, \varepsilon)$ for some $\varepsilon \geq 0$ or $\mathcal{E} = \mathbb{R}^m$, $\nu \in \{$ Gaussian distribution, Poisson distribution,... $\}$.

Opposed to this a-priori approach, in which the structure of $\mathcal{M}_1$ is assumed in advance, more recent data-driven methods are trying to learn the decoder on the set $\mathcal{M}_1$ by using a finite set of data $\{(y_1, x_1), \ldots, (y_n, x_n)\} \subset \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$, for some $n \in \mathbb{N}$. We aim at providing a theoretical basis for understanding the performance limitations of both state of

the art methods and recent data-driven approaches in underdetermined and ill-posed inverse problems.

### 4.1.1 Problem outline and related work

As an ongoing topic of research ill-posed inverse problems have been studied in different areas. They have been studied from a statistical perspective [145], using iterative deep neural networks [4], by using regularization [48] and even in fields, such as radio tomography of the ionosphere [77], and also as a textbook topic [62]. Despite this extensive amount of research, there is little to be found on fundamental accuracy bounds of approximate solutions to ill-posed inverse problems. Considering the Bayesian approach to inverse problems, there exist some a posteriori accuracy estimation bounds under specific conditions on normed spaces [119, 120, 184]. Accuracy and error bounds for non-linear ill-posed inverse problems under certain restrictions have been studied in [112]. Non-linear and possibly ill-posed inverse problems have a wide range of theoretical and industrial applications [63]. In some iterative reconstruction approaches, trade-offs between convergence speed and reconstruction accuracy have been established [81]. Moreover, most mentioned approaches consider additive noise, however often multiplicative noise models are of interest for studying inverse problems [10, 103, 165, 191]. To the best of our knowledge there are no fundamental accuracy bounds for ill-posed inverse problems with multiplicative noise. Yet, our framework also encompasses the additive noise model [15, 100, 110].

Compared to standard methods, data-driven approaches using deep learning for solving inverse problems (4.2) have reported superior accuracy in different applications [14,171,192]. As established in the previous chapters, this can potentially lead to instabilities, which is also highlighted in [6,83]. In fact, there is a variety of research that has established that artificial intelligence techniques based on deep learning are universally unstable, in image classification [67, 116, 135, 142, 174], and later in applications ranging from audio and speech recognition [32, 33, 190] to natural language processing [123] and automatic diagnosis in medicine [73]. Instabilities, such as false positives, false negatives and especially AI hallucinations, have also been an issue in the fastMRI challenge [138] and in microscopy [14,99]. An example of this is the inability of convolutional neural networks to provide a stable and accurate reconstruction for CT inverse problems [166]. As highlighted in the previous chapters, stability and robustness of neural networks in inverse problems are now an active area of machine learning research. There exist numerous empirical studies with a wide variety of results [80,138,166], on additional or removed elements in the reconstruction and AI hallucinations in DL used for inverse problems. However, to the best of our knowledge there are no fundamental performance and accuracy limits for data-driven approaches using deep learning for solving inverse problems. Yet, for solving underdetermined and ill-posed inverse problems, there exist several approaches involving neural networks. For example, invertible neural networks aim at learning the inverse process implicitly and use so-called

additional latent output variables in order to *„capture the information otherwise lost“* [7]. Another approach is distributional learning, which aims to circrumvent these problems by sampling enough data. As mentioned in Chapter 1 in fully-learned Bayes estimation, due to the lack of training data, this *„is inapplicable to cases when data are acquired using novel instrumentation“* [8]. Another novel approach, coined AUTOMAP [192], claims to achieve superior immunity to noise and high performance, was highlighted in [171] and does not employ any knowledge of $(F, \mathcal{M}_1)$. There exist numerous extensions of this framework, yet in [8] it is argued that these fully learned generic approaches are infeasible as they would involve learning a practical not attainable large amount of weights from supervised data. Moreover, as established in Chapter 2 AUTOMAP and other fully learned approaches suffer from instabilities and lack of robustness. For a detailed overview of deep learning in inverse problems and stability of robustness for deep learning we refer to [8, 126, 131]. Furthermore, it is not clear if these learned approaches solve the problem at hand. As for ill-posed inverse problems the set $\{x \in \mathcal{M}_1 : \exists e \in \mathcal{E}, y = F(x, e)\}$ given $y \in \mathcal{M}_2^{\mathcal{E}}$ has a strictly positive diameter, there is a fundamental accuracy limit. Moreover, conditions on $\mathcal{M}_1$ and the measurement model $F$ have to be satisfied in order for a stable and robust reconstruction to exist.

In general, when sampling a large amount of data, if the underlying set $\mathcal{M}_1 \subseteq \mathbb{C}^N$ that one wants to reconstruct from data obtained through the measurement model $F$, is not known, then conditions are needed in order to guarantee estimates on the reconstruction error. An example of reconstruction errors, are the average error on the data, referred to as *empirical error* and the average error on $\mathcal{M}_1$, referred to as *generalization error*, in statistical learning [133]. However, the conditions on $\mathcal{M}_1$ are often too restrictive for severely ill-posed settings to guarantee accurate and robust reconstruction. Some examples of such choices are to be found in [38, 51, 58, 72, 85]. In the case that the measurement model is given by a linear $A$ and additive noise, the condition $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$ is assumed to guarantee accurate and robust reconstruction, as in [21, 44, 178]. This establishes the need for conditions on $(A, \mathcal{M}_1)$ and the measurement model $F$ in order to provide theoretical guarantees for stability, robustness and accuracy in ill-posed settings. Ideally, then, these conditions should be applicable to deep learning in inverse problems and cases, where the assumptions for state of the art methods are not satisfied anymore.

### 4.1.2 Contribution

In this chapter, our main contribution is to establish an extensive framework that highlights the importance of knowledge of the measurement model $F$ and $\mathcal{M}_1$ when attempting to solve (4.2). This framework provides accuracy bounds for solutions of ill-posed inverse problems in terms of $\mathcal{M}_1$ and the measurement model $F$. In particular, in the case of a linear $A$ these bounds can be stated in terms of its kernel $\mathcal{N}(A)$. These concepts are established by extending assumptions and concepts previously used in standard methods

for inverse problems and applied to prove our main Theorem 4.2.9. In the following, we summarise our main results.

(M1) **Upper and lower bounds on reconstruction accuracy for ill-posed inverse problems (Summary Theorem 4.2.9).** Theorem 4.2.9 contributes to assessing ill-posed inverse problems, by establishing a *universal optimality constant*, that includes the best worst-case noise, the average and the statistical reconstruction error for the reconstruction of (4.2). Under general assumptions, Theorem 4.2.9 *provides upper and lower bounds on the optimality constant*. Thereby it yields theoretical limits of a decoder's accuracy in all above mentioned settings. Moreover, in parts $(2), (3)$ explicit optimization problems yielding the decoders that attain the optimality constant are derived.

Key points: A key point in the proof is the use of the Measurable Maximum Theorem [40], to prove that the optimal map is measurable. Moreover, a disintegration of measure is used [39], as assumption on $(F, \mathcal{M}_1)$ in order to obtain upper and lower bounds on the optimality constant. These bounds then, hold for any data distribution or measure $\mu_1$ on $\mathcal{M}_1$ and general noise models $(\mathcal{E}, \nu)$. Another, key point is the compactness assumption on $\mathcal{M}_1$ to obtain a bounded set $\{x \in \mathcal{M}_1 : \exists e \in \mathcal{E}, y = F(x, e)\}$ given $y \in \mathcal{M}_2^{\mathcal{E}}$. This assumption is needed for ill-posed inverse problems to obtain a well-defined optimal decoder.

(M2) **Best worst-case noise reconstruction error (Summary Theorem 4.2.3).** In the special case of the best worst-case noise reconstruction error, we can show that the upper and lower bounds on the optimality constant relate to well-known concepts if $A$ is linear. Namely, to the condition $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$, which is commonly necessary for exact recovery or also directly implied by assuming the robust null space property. This condition is obviated and the diameter of the intersecting sets determines the best worst-case reconstruction error. In this setting, we define the decoder, which obtains the smallest possible worst-case reconstruction error that can appear in (4.2). We call this decoder *the optimal map with worst-case noise*.

Key points: The optimal map with worst-case noise is the mapping that optimises the accuracy-hallucination trade-off, as established in Chapter 3.

(M3) **Approximability by neural networks (Summary Theorem 4.2.21).** We show that under certain conditions the optimal decoder for (4.1) can be approximated by a neural network and that it is accurate and robust given general conditions on $(A, \mathcal{M}_1)$. Furthermore, in this ill-posed setting, we identify sufficient and necessary conditions on $(A, \mathcal{M}_1)$ such that an optimal decoder can be approximated by a neural network.

Key points: Key points in the proof are the use of set-valued analysis and the $\ell_2$ norm and that $\mathcal{M}_1$ satisfies regularity properties. In a general setting,

we provide sufficient and necessary conditions for continuity of optimal map with worst-case noise, and hence, its robustness. Furthermore, as the optimal decoder may be set-valued we provide error bounds on the reconstruction as shown in Theorem 4.2.30. Together with the universal instability theorem [5, 83], which provides a theoretical explanation for additional or removed elements in the reconstruction, Theorem 4.2.30 possibly gives insight into why AI hallucinations occur. Moreover, the stability result of Theorem 4.2.30 and accuracy bounds of Theorem 4.2.9, as well as the general conditions established on general $(A, \mathcal{M}_1)$ are a generalization of the theory established by [21] and [44]. Thus, these conditions provide valuable means of theoretically assessing existence and stability of approximate decoders for underdetermined systems.

(M4) **Learnability of optimal maps (Summary Corollary 4.2.28).** Lastly, we relate these results to learned methods for reconstruction and asssess whether training may yield an optimal map with worst-case noise for (4.1) in Corollary 4.2.28 and Figure 4.1.

Key points: Our results are closely related to foundational aspects in approximation theory and the theory of fundamental decoders [21], which is a generalisation of the sparsity based approach. In particular, the well-known framework for fundamental decoders as presented in [21] is extended. Here the robust instance optimal (rIOP) decoder is related to the optimal map with worst-case noise. Thus, the optimal map framework allows for assessing stability and accuracy of approximate decoders for underdetermined inverse problems on arbitrary bounded sets $\mathcal{M}_1$. This can be applied to a large variety of applications, including deep learning, for inverse problems. This is shown in Corollary 4.2.28 and Figure 4.1.

### 4.1.3 Outline

In Section 4.2, we present a framework - *the optimality constant* - that provides upper and lower bounds on reconstruction accuracy for a wide range of considered errors, including statistical, average and worst-case reconstruction errors for (4.2). Moreover, necessary and sufficient conditions are presented that allow approximating the decoder that achieves the best worst-case reconstruction error for (4.1) given a set $\mathcal{M}_1$ and a sampling operator $A$ by neural networks. This is followed by a detailed discussion of our main results related to the use of DL for solving (4.2) and (4.1), in Section 4.2.5. Moreover, stability results based on this approach are presented. This framework is compared to fundamental results in approximation theory, Section 4.3.1, and in Section 4.4, to the fundamental decoder framework, as introduced in [21].

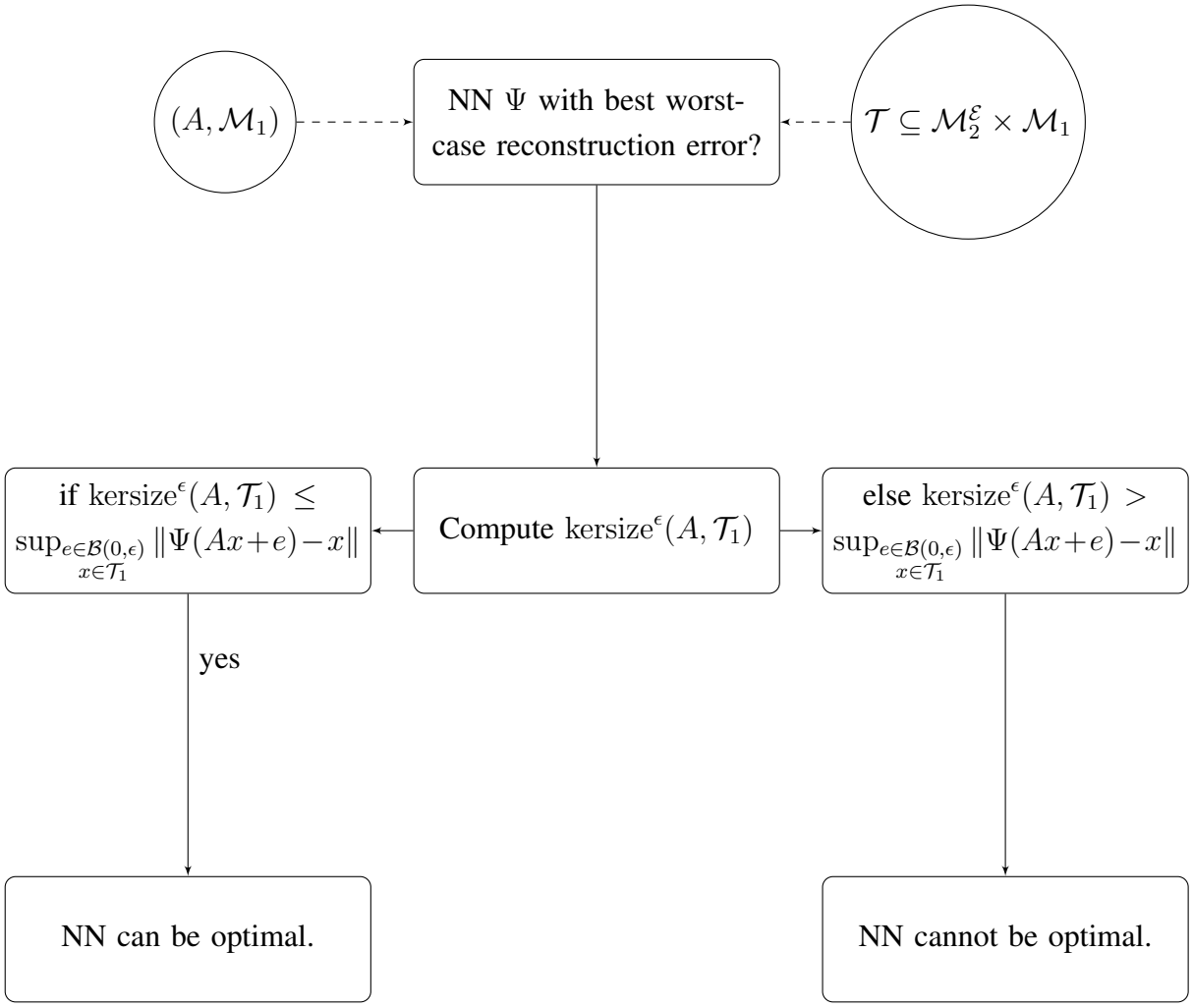


Figure 4.1: **(Illustration of Corollary 4.2.28)** The figure illustrates Corollary 4.2.28, where $\epsilon \geq 0$ is the noise level and the sampling operator $A \in \mathbb{C}^{m \times N}$ is linear. The projection onto the second component of the training set is denoted $\pi_2(\mathcal{T}) = \mathcal{T}_1 \subseteq \mathcal{M}_1$. Moreover, we consider the optimality constant with worst-case noise for (4.1) and the kernel size, as introduced in Defnition 4.2.8, is abbreviated by $\mathrm{kersize}(A, \mathcal{T}_1, \mathcal{B}_{d_2}(0, \epsilon), \infty) = \mathrm{kersize}^{\epsilon}(A, \mathcal{T}_1)$.

## 4.2 Main results

In the following, we provide the necessary preliminaries for the main results.

### 4.2.1 Notation

Given a set $\mathcal{M}_1 \subset \mathbb{C}^N$. For $\mathcal{E} \subseteq \mathbb{C}^m$, the set of noisy measurements is denoted by

$$\mathcal{M}_2^{\mathcal{E}} = \{y \in \mathbb{C}^m : \exists x \in \mathcal{M}_1, \exists e \in \mathcal{E}, y = F(x, e)\}.$$

In the case that $A \in \mathbb{C}^{m \times N}$ is linear, we assume that the rank of $A$ is bounded by $1 \leq \mathrm{rank}(A) < N$, and denote null space of $A$ by $\mathcal{N}(A) \subset \mathbb{C}^N$. Moreover, in this case we let

$$\mathcal{M}_2 = A\mathcal{M}_1 = \{Ax : x \in \mathcal{M}_1\}$$

denote the range of $A$ with domain $\mathcal{M}_1$. For a subspace $\mathcal{V} \subset \mathbb{C}^m \times \mathbb{C}^N$, we let $\pi_{\mathcal{V}}$ denote the projection onto $\mathcal{V}$ and the projection onto the first component of the product space is abbreviated by $\pi_1$. $d_1$ denotes a metric on $\mathbb{C}^N$ and $d_2$ denote a metric on $\mathbb{C}^m$, unless specified otherwise. We let

$$\mathcal{B}_{d_1}(x, r) = \{z \in \mathbb{C}^N : d_1(x, z) \leq r\}$$

denote the closed ball centered at $x \in \mathbb{C}^N$ with radius $r > 0$. If $x \in \mathbb{C}^m$, then $\mathcal{B}_{d_2}(x, r)$ denotes a ball with respect to $d_2$. Moreover, recall the definition of the $\epsilon$-neighbourhood $X^\epsilon \subseteq \mathcal{X}$, for a subset $X \subseteq \mathcal{X}$ of a metric space $(\mathcal{X}, d)$ in (3.10):

$$X^\epsilon := \{x \in \mathcal{X} : \exists x' \in X, d(x, x') \leq \epsilon\}.$$

Generally, decoders for an inverse problem as in (4.2) are obtained by an optimization problem. Usually, non-convex or convex optimization problems without sufficient constraints do not have a unique solution and, thus, this may yield a set-valued decoder. Hence, we consider multivalued maps denoted by $\phi : \mathbb{C}^m \rightrightarrows \mathbb{C}^N$, where $\phi(y) \subseteq \mathbb{C}^N$ is non-empty and bounded for $y \in \mathbb{C}^m$. In this case, it is common to consider the Hausdorff metric between bounded subsets of $\mathbb{C}^N$. For two bounded subsets $Z, X \subset \mathbb{C}^N$ we denote Hausdorff distance by

$$d_1^H(Z, X) = \max\{ \sup_{x \in X} \inf_{z \in Z} d_1(z, x), \sup_{z \in Z} \inf_{x \in X} d_1(z, x) \}.$$

With slight misuse of notation we denote a singleton $\{x\} \subset \mathcal{M}_1$ by $x$. Note that $d_1^H(Z, x)$ is an upper bound on the largest possible distance between $x$ and any point in $Z$, i.e.

$$d_1^H(Z, x) = \sup_{z \in Z} d_1(z, x).$$

### 4.2.2 Optimality bounds with worst-case noise

In order to present our main results, we will consider the following definitions. Recall the definition of the optimal map 3.4.11, from Chapter 3. This definition is extended to the concept of the optimality constant for arbitrary measure spaces and noise models. The optimality constant can be applied to a wide range of settings, for example, to determine bounds for the mean reconstruction error of (4.2).

Given an inverse problem of the form (4.2), one important question is the following: what is the smallest reconstruction error that I can get? The answer is the *optimality constant*, defined below. Note that this topic was discussed in a less general setting in Chapter 3, Definition 3.4.11. As stated before, this definition aligns with the tradition in approximation theory and the seminal work by Cohen, Dahmen and DeVore [44].

**Definition 4.2.1.** Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ and $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^e \subseteq \mathbb{C}^m$ be surjective. Define the *optimality constant with worst-case noise* of the problem (4.2) as

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty) = \inf_{\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N} \sup_{x \in \mathcal{M}_1} \sup_{e \in \mathcal{E}} d_1^H(x, \varphi(F(x, e))). \tag{4.3}$$

A function $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ that attains such an infimum is called an *optimal map with worst-case noise*.

In other words, the optimality constant with worst-case noise gives the best worst-case reconstruction error.

Note the similarity of this definition and the definition of a *compressive* $m$*-width* in Foucart & Rauhut's book Definition 10.2 [74]. Here, in particular norms are used rather than the metric $d_1$, and the infimum is merely taken over linear maps from $\mathbb{C}^N$ to $\mathbb{C}^m$. A more comprenhensive comparison is given in section 4.3.1. The above Definition 4.2.1 constitutes an error estimate or loss functional for a conservative reconstruction method, as suggested in [4].

In order to analyse ill-posed inverse problems, we now define the notion of the *kernel size*. This is a generalization of the well-known condition $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$, which is often assumed when solving ill-posed inverse problems. This relation is elaborated on in Section 4.4, precisely in Remark 4.4.9.

**Definition 4.2.2** (Kernel size with worst-case noise)**.** Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ and $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}} \subseteq \mathbb{C}^m$ be surjective. Define the *kernel size* of the problem (4.2) as

$$\text{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty) = \sup_{\substack{x, x' \in \mathcal{M}_1, \exists e, e' \in \mathcal{E} \text{ s.t.} \\ F(x,e) = F(x', e')}} d_1(x, x'). \tag{4.4}$$

The kernel size gives the maximum distance between any two points $x, x' \in \mathcal{M}_1$ which can lead to the same measurement $y = F(x, e) = F(x', e')$ for $e, e' \in \mathcal{E}$ .

Our initial result gives an upper and lower bound on the optimality constant in terms of the kernel size. Moreover, it provides a variational expression for the optimal map with noise.

**Theorem 4.2.3** (Optimality bounds with worst-case noise)**.** *Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ and $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}} \subseteq \mathbb{C}^m$ be surjective. Assume that $\mathcal{M}_1$ is compact and that all closed balls in $\mathbb{C}^N$ with respect to $d_1$ are compact.*

*(1)* ***(Worst and best case reconstruction error bounds)****. We have the bounds*

$$\text{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty)/2 \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq \text{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty). \tag{4.5}$$

*(2)* ***(Explicit form of the optimal map with worst-case noise)****. Moreover, the optimal map with the noise is given by*

$$\Psi(y) = \text{argmin}_{z \in \mathbb{C}^N} \sup_{(x,e) \in F^{-1}(y)} d_1(x, z) \tag{4.6}$$

*and has non-empty values.*

The proof of Theorem 4.2.3 follows a similar structure to the more involved proof of our main result, Theorem 4.2.9.

*Proof of Theorem 4.2.3.* **Part (1)**, we first prove the lower bound

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq 2c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

Let $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$. Let $y \in \mathcal{M}_2^{\mathcal{E}}$. Let $(x, e), (x', e') \in \mathcal{M}_1 \times \mathcal{E}$ such that $F(x, e) = F(x', e') = y$, then by the triangle inequality

$$d_1(x, x') \leq d_1^H(x, \varphi(F(x, e))) + d_1^H(x', \varphi(F(x', e'))).$$

Taking the supremum,

$$\sup_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x, x') \leq \sup_{(x,e)\in F^{-1}(y)} d_1^H(x, \varphi(F(x, e))) + \sup_{(x',e')\in F^{-1}(y)} d_1^H(x', \varphi(F(x', e'))).$$

Taking the supremum with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$ and the infimum over $\varphi \in \mathcal{C}$ gives

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq 2c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

**Part (2)**, and the upper bound. $\Psi$ has non-empty values, since the minimum in (4.13) is attained. Fix $y \in \mathcal{M}_2^{\mathcal{E}}$. For $z \in \mathbb{C}^N$, let

$$f_y(z) = \sup_{(x,e)\in F^{-1}(y)} d_1(x, z).$$

$f_y$ is continuous. Since $f_y : (\mathbb{C}^N, d_1) \to [0, \infty)$ is a function between metric spaces, continuity is equivalent to showing that for every sequence $z_n \to z$ there exists a subsequence $(z_{n_k})_k$ of $(z_n)_n$ such that $f_y(z_{n_k}) \to f_y(z)$. Let $z_n \to z$ wrt. $d_1$. Then, for $(x, e) \in F^{-1}(y)$, since the function $d(x, \cdot)$ is continuous, there exists a subsequence $(z_{n_k})_{k\in\mathbb{N}}$ such that

$$d_1(x, z) - 1/k \leq d_1(x, z_{n_k}) \leq d_1(x, z) + 1/k.$$

The supremum with respect to $(x, e) \in F^{-1}(y)$ satisfies

$$\sup_{(x,e)\in F^{-1}(y)} d_1(x, z) - 1/k \leq \sup_{(x,e)\in F^{-1}(y)} d_1(x, z_{n_k}) \leq \sup_{(x,e)\in F^{-1}(y)} d_1(x, z) + 1/k.$$

Taking the limit $k \to \infty$ yields, $f_y(z_{n_k}) \to f_y(z)$. Thus, $f_y$ is continuous. Now define

$$r_y = \sup_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x, x'),$$

which satisfies $r_y < \infty$ as $\mathcal{M}_1$ is compact. For $(x, e) \in F^{-1}(y)$,

$$\Psi(y) = \operatorname{argmin}_{z\in\mathbb{C}^N} \sup_{(x',e')\in F^{-1}(y)} d_1(z, x') = \operatorname{argmin}_{z\in B_{d_1}(x,r_y)} f_y(z). \tag{4.7}$$

In fact, if $z \in \mathbb{C}^N \setminus B_{d_1}(x, r_y)$, then

$$f_y(z) = \sup_{(x'e') \in F^{-1}(y)} d(z, x') \geq d(x, z) > r_y \geq f_y(x).$$

Therefore, $\inf_{z \in \mathbb{C}^N} f_y(z) \geq r_y \geq \inf_{z \in B_{d_1}(x, r_y)} f_y(z)$, which proves (4.7). By assumption $B_{d_1}(x, r_y) \subset \mathbb{C}^N$ is compact, since it is a closed ball with respect to the metric $d_1$. Therefore, the minimum in (4.7) can be attained by the Extreme Value Theorem. This shows that the argmin is non-empty, and hence that $\Psi$ has non-empty values on $\mathcal{M}_2^{\mathcal{E}}$.

**Claim 1:** $\Psi$ is an optimal map. Let $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$. Fix $y \in \mathcal{M}_2^{\mathcal{E}}$. By the definition of $\Psi$,

$$d_1^H(\Psi(y), x) \leq d_1(z, x) \quad \text{for every } (x, e) \in F^{-1}(y)$$

for every $z \in \varphi(y)$. In particular, taking the supremum with respect to $z \in \varphi(y)$, which coincides with the Hausdorff distance

$$\begin{aligned} d_1^H(\Psi(y), x) &\leq \sup_{z \in \varphi(y)} d_1(z, x) \\ &= d_1^H(\varphi(y), x) \quad \text{for every } (x, e) \in F^{-1}(y). \end{aligned}$$

Hence,

$$\operatorname*{essup}_{(x,e) \in F^{-1}(y)} d_1^H(\Psi(y)), x) \leq \operatorname*{essup}_{(x,e) \in F^{-1}(y)} d_1^H(\varphi(y), x).$$

By considering the supremum with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$, we obtain

$$\operatorname*{essup}_{y \in \mathcal{M}_2^{\mathcal{E}}} \sup_{(x,e) \in F^{-1}(y)} d_1^H(\Psi(y)), x) \leq \sup_{y \in \mathcal{M}_2^{\mathcal{E}}} \sup_{(x,e) \in F^{-1}(y)} d_1^H(\varphi(y), x).$$

This can be rewritten as

$$\sup_{(x,e) \in \mathcal{M}_1 \times \mathcal{E}} d_1^H(\Psi(F(x, e))), x) \leq \sup_{(x,e) \in \mathcal{M}_1 \times \mathcal{E}} d_1^H(\varphi(F(x, e))), x).$$

Now, as $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ was arbitrary, we obtain

$$\sup_{(x,e) \in \mathcal{M}_1 \times \mathcal{E}} d_1^H(\Psi(F(x, e))), x) \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

The opposite inequality holds trivially since $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$. Therefore, $\Psi$ is an optimal map. **Claim 2**; finally, we prove the upper bound. By the minimisation property of $\Psi$, for every $(x', e') \in F^{-1}(y)$,

$$\sup_{(x,e) \in F^{-1}(y)} d_1^H(x, \Psi(y)) \leq \sup_{(x,e) \in F^{-1}(y)} d_1(x, x').$$

Thus, taking the supremum with respect to $(x', e') \in F^{-1}(y)$ and with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$, yields

$$\begin{aligned} \sup_{(x,e) \in \mathcal{M}_1 \times \mathcal{E}} d_1^H(x, \Psi(F(x, e))) &\leq \sup_{y \in \mathcal{M}_2^{\mathcal{E}}} \sup_{\substack{(x,e) \in F^{-1}(y) \\ (x',e') \in F^{-1}(y)}} d_1(x, x') \\ &= \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty). \end{aligned}$$

As $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$, we deduce that

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq \text{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

□

The previous theorem gives upper and lower bounds when considering the best worst-case reconstruction error. In many applications, however, it is interesting to evaluate the *average* reconstruction error. Since "average" is simply another word for "integral", in the following we will consider $\mu_1$ and $\nu$ to be finite measures on $\mathcal{M}_1$ and $\mathcal{E}$ respectively.

### 4.2.3 General optimality bounds

In order to present our main result, we will consider the following definitions which generalise the optimality constant and kernel size to the case when $\mathcal{M}_1$ and $\mathcal{E}$ are equipped with measures $\mu_1$ and $\nu$. Before, we introduce the definition of a *disintegration of a measure*.

**Definition 4.2.4** (Disintegration of measure)**.** Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ be equipped with finite positive measures $\mu_1$, respectively $\nu$. Let $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}}$ be surjective and $\rho = F_*(\mu_1 \otimes \nu)$ be the pushforward measure defined on $\mathcal{M}_2^{\mathcal{E}}$. We will say that $\mu_1 \otimes \nu$ *admits a disintegration of measure with respect to* $F$ if there exists a $\rho$-almost everywhere uniquely determined family of probability measures $\{(\mu_1 \otimes \nu)^y\}_{y \in \mathcal{M}_2^{\mathcal{E}}}$ on $\mathcal{M}_1 \times \mathcal{E}$, such that

(a) The function $\mathcal{M}_2^{\mathcal{E}} \ni y \mapsto (\mu_1 \otimes \nu)^y(B) \in \mathbb{R}$ is Borel measurable for every Borel measurable subset $B \subseteq X$.

(b) $(\mu_1 \otimes \nu)^y$ is concentrated on $F^{-1}(y)$ for $\rho$-almost every $y \in \mathcal{M}_2^{\mathcal{E}}$, i.e. for every measurable $B \subseteq \mathcal{M}_1 \times \mathcal{E}$

$$(\mu_1 \otimes \nu)^y(B) = (\mu_1 \otimes \nu)^y(B \cap F^{-1}(y)) \quad \text{for } \rho\text{-almost every } y \in \mathcal{M}_2^{\mathcal{E}}.$$

(c) for every positive Borel-measurable function $f : \mathcal{M}_1 \times \mathcal{E} \to [0, \infty]$

$$\int_{\mathcal{M}_1 \times \mathcal{E}} f(x,e) d(\mu_1 \otimes \nu)(x,e) = \int_{\mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} f(x,e) d(\mu_1 \otimes \nu)^y(x,e)\, d\rho(y).$$

Note that, as the definition of a disintegration of measure requires Borel measurable sets, in the following $\mathcal{M}_1$ and $\mathcal{E}$ are equipped with the corresponding Borel sigma algebras.

**Remark 4.2.5.** We call the sets $F^{-1}(y)$ *feasible sets*. Moreover, Theorem 2 [39], states that the measures $(\mu_1 \otimes \nu)^y$ are probability measures if and only if $F$ is surjective. In this case $(\mu_1 \otimes \nu)^y$ can be directly related to the conditional expectation of $(x, e)$ given measurements $y$, see [39] for further details.

Having introduced measures on $\mathcal{M}_1$ and $\mathcal{E}$ and a disintegration of measure, we can now generalise the optimality constant with worst-case noise.

**Definition 4.2.6** (Optimality constant)**.** Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ and $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}} \subseteq \mathbb{C}^m$ be surjective. Let $\mu_1$ and $\nu$ be finite measures on $\mathcal{M}_1$ and $\mathcal{E}$ respectively, and assume that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$. Let

$$\mathcal{C} := \{\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N : (x, e) \mapsto d_1^H(x, \varphi(F(x, e)) \text{ is measurable}\}.$$

The optimality constant is defined, for $p \in [1, \infty)$ , as

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p) = \inf_{\varphi \in \mathcal{C}} \left( \int_{\mathcal{M}_1 \times \mathcal{E}} d_1^H(x, \varphi(F(x, e)))^p \, d(\mu_1 \otimes \nu)(x, e) \right)^{\frac{1}{p}}, \tag{4.8}$$

and for $p = \infty$,

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty) = \inf_{\varphi \in \mathcal{C}} \operatorname*{essup}_{(x,e) \in \mathcal{M}_1 \times \mathcal{E}} d_1^H(x, \varphi(F(x, e))) \tag{4.9}$$

where the essential supremum is taken with respect to $\mu_1 \otimes \nu$. A map $\Psi \in \mathcal{C}$ attaining the infimum in (4.8) or (4.9) is called an *optimal map*.

**Remark 4.2.7.** The optimality constant can be related to many well-known settings. For $p = 2$, the optimality constant is the minimal squared reconstruction error. For $p = 1$ and $\mu_1, \nu$ probability measures, the optimality constant is the (mean) statistical reconstruction error and for $p = 2$ the mean squared error (MSE). Moreover, for $p = \infty$, if the essential supremum coincides with the supremum, the optimality constant is the best worst-case noise reconstruction error defined in Definition 4.2.2.

**Definition 4.2.8** (Kernel size)**.** Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$, $\mathcal{E} \subseteq \mathbb{C}^m$ and $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}} \subseteq \mathbb{C}^m$ be surjective. Let $\mu_1$ and $\nu$ be finite measures on $\mathcal{M}_1$ and $\mathcal{E}$ respectively, and assume that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$. The kernel size of the inverse problem (4.2) is defined, for $p \in [1, \infty)$, as

$$\begin{aligned} &\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p) \\ &\quad = \left( \int_{\mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} \int_{F^{-1}(y)} d(x, x')^p \, d(\mu_1 \otimes \nu)^y(x, e) \, d(\mu_1 \otimes \nu)^y(x', e') \, d\rho(y) \right)^{\frac{1}{p}} \end{aligned}$$

and for $p = \infty$ as

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty) = \operatorname*{essup}_{y \in \mathcal{M}_2^{\mathcal{E}}} \operatorname*{essup}_{\substack{(x,e) \in F^{-1}(y) \\ (x',e') \in F^{-1}(y)}} d_1(x, x'),$$

where the essential suprema are taken with respect to $(\mu_1 \otimes \nu)^y$.

Intuitively, the kernel size gives the average distance between the $x$-components of $(x, e), (x', e') \in \mathcal{M}_1 \times \mathcal{E}$ that have the same measurement $F(x, e) = F(x', e') = y$ in the case of $p \in [1, \infty)$, and the maximum of such distances when $p = \infty$.

Our first result states existence and explicit form of the decoders obtaining the optimality constant and the kernel size gives upper and lower bounds on the optimality constant.

**Theorem 4.2.9** (Optimality bounds)**.** *Assume that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$. Let $\mathcal{M}_1$ be compact and $d_1$ be such that all closed balls in $\mathbb{C}^N$ are compact. Then, the following holds for every $p \in [1, \infty]$,*

*(1)* ***(Worst and best case reconstruction error bounds)****. We have the bounds*

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p)/2 \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p) \leq \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p).$$

*(2)* ***(Explicit form of the optimal map with a.e. worst-case noise)****.*
*The minimizer $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ for $c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty)$ in* (4.9) *exists and is given by*

$$\Psi(y) = \operatorname{argmin}_{z \in \mathbb{C}^N} \operatorname*{essup}_{(x,e) \in F^{-1}(y)} d_1(x, z). \tag{4.10}$$

*Moreover, $\Psi$ has compact values.*

*(3)* ***(Explicit form of the optimal map with any noise model)****. For $p \in [1, \infty)$ the minimizer $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ for $c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p)$ in* (4.8) *exists and is given by*

$$\Psi(y) = \operatorname{argmin}_{z \in \mathbb{C}^N} \int_{F^{-1}(y)} d_1(x, z)^p \, d(\mu_1 \otimes \nu)^y(x, e). \tag{4.11}$$

*Moreover, $\Psi$ has compact values.*

**Remark 4.2.10.** Note that Theorem 4.2.3 could be seen as a special case of Theorem 4.2.9 in the case when the measures $\mu_1$, $\nu$ and the function $F$ are such that the essential supremum (with respect to $\mu_1 \otimes \nu$ and with respect to $(\mu_1 \otimes \nu)^y$for every $y \in \mathcal{M}_2^{\mathcal{E}}$) coincides with the supremum. We decided to keep the two definitions distinct, because it may be hard to ensure such a condition, while Theorem 4.2.3 still holds without requiring any assumptions on the measures.

Unfortunately if $\mathcal{M}_1$ is not bounded, the kernel size can be infinity and, hence, is not useful anymore. However, the compactness assumption is only needed in order to obtain an explicit form of the optimal map and prove its existence. Moreover, the condition that $d_1$ is such that all closed balls in $\mathbb{C}^N$ are compact, is need in order for the argmin to exist by the Extreme Value Theorem. Thus, for the lower and upper bounds compactness is not necessary. The sparse case can be treated but only for bounded sparse vectors.

In order to prove Theorem 4.2.9, we will make use of Theorem 18.19 [40].

**Theorem 4.2.11** (Measurable Maximum Theorem, Theorem 18.19 [40])**.** *Let $X$ be a separable metrisable space and $(S, \Sigma)$ a measurable space. Let $\varphi : S \rightrightarrows X$ be a weakly measurable correspondence with non-empty compact values, and suppose $f\colon S \times X \to \mathbb{R}$ is a Caratheodory function. Define the value function $m : S \to \mathbb{R}$ by*

$$m(s) = \max_{x \in \varphi(s)} f(s, x)$$

*and the correspondence $\mu : S \rightrightarrows X$ of maximisers by*

$$\mu(s) = \{x \in \varphi(s) \ : \ f(s, x) = m(s)\}.$$

*Then, $m$ is measurable, $\mu$ has non-empty and compact values, and $\mu$ is measurable and admits a measurable selector.*

Moreover, the following Propositions, 4.2.13 and 4.2.15, and Corollary, 4.2.14, are needed for the proof of Theorem 4.2.9.

**Theorem 4.2.12** ((iii), Theorem 2 [39])**.** *Let $\mu$ have a $(F, \rho)$-disintegration $\{\mu^y\}$, with $\mu$ and $\rho$ each sigma-finite. Then, the measures $\{\mu^y\}$ are probabilities for $\rho$-almost all $y \in Y$ if and only if $\rho = F_*\mu$.*

**Proposition 4.2.13.** *Let $X$ be a Radon space and $\mu_X$ be a probability measure. Let $F : X \to Y$ be a surjective, Borel measurable function. Let $\mu_Y := F_*\mu_X$ be the pushforward measure of $\mu_X$ onto $Y$. Assume $\mu_X$ admits a disintegration with respect to $F$. Then, for $A \subseteq X$ measurable,*

$$\mu_X(A) = 0 \iff \mu_X^y(A) = 0 \text{ for } \mu_Y\text{-almost every } y \in Y$$

*Proof of Proposition 4.2.13.* As there exists a disintegration of $\mu_X$ with respect to $F$, this is a $\mu_Y$-almost everywhere uniquely determined family of probability measures $\{\mu_X^y\}_{y \in Y}$ on $\mu_X$ by Theorem 4.2.12. Now taking $f = \mathbb{1}_A$ we have

$$\mu_X(A) = 0 \iff 0 = \int_X \mathbb{1}_A \, d\mu_X = \int_Y \int_{F^{-1}(y)} \mathbb{1}_A \, d\mu_X^y \, d\mu_Y.$$

The integral of a positive function with respect to a positive measure is zero if and only if the integrand function is zero almost-everywhere, so we continue:

$$\iff \int_{F^{-1}(y)} \mathbb{1}_A \, d\mu_X^y = 0 \quad \text{for } \mu_Y\text{-almost every } y \in Y.$$

And since $d\mu_X^y$ is supported on $F^{-1}(y)$ for $\mu_Y$-almost every $y \in Y$, we continue

$$\begin{aligned} &\iff \int_Y \mathbb{1}_A \, d\mu_X^y = 0 \quad \text{for } \mu_Y\text{-almost every } y \in Y \\ &\iff \mu_X^y(A) = 0 \text{ for } \mu_Y\text{-almost every } y \in Y. \end{aligned}$$

□

Under the same assumptions as in Proposition 4.2.13 as a corollary, we can split the essential supremum with respect to $\mu_X$ in a "double" essential supremum with respect to $\{\mu_X^y\}_{y\in Y}$ and $\mu_Y$. In particular, for any measurable function $f : X \to [0, +\infty]$, define $f_y(x) := f(x)\mathbb{1}_{F^{-1}(y)}(x)$. We will denote by $\|\cdot\|_\infty$ the essential supremum of a function.

**Corollary 4.2.14.** *Let $X$ be a Radon space and $\mu_X$ be a probability measure. Let $F : X \to Y$ be a surjective, Borel measurable function. Let $\mu_Y := F_*\mu_X$ be the pushforward measure of $\mu_X$ onto $Y$. Assume $\mu_X$ admits a disintegration with respect to $F$. Let $f : X \to [0, +\infty]$ be a measurable function. Then*

$$\|f\|_\infty = \operatorname*{essup}_{y\in Y} \|f_y\|_\infty.$$

*Proof of Corollary 4.2.14.* For $\geq$: By definition of essential supremum, $|f(x)| \leq \|f\|_\infty$ for $\mu_X$-almost every $x \in X$. This means that $\mu_X(\{x \in X : |f(x)| > \|f\|_\infty\}) = 0$. By Proposition 4.2.13 and using the fact that $\mu_X^y$ is concentrated on $F^{-1}y$, we deduce that for $\mu_Y$-almost every $y \in Y$ we have

$$\begin{aligned} 0 &= \mu_X^y(\{x \in X : |f(x)| > \|f\|_\infty\}) \\ &= \mu_X^y(\{x \in X : |f(x)| > \|f\|_\infty, F(x) = y\}) \\ &= \mu_X^y(\{x \in X : |f_y(x)| > \|f\|_\infty\}). \end{aligned}$$

Hence, we have $\|f_y\|_\infty \leq \|f\|_\infty$ for $\mu_Y$-almost every $y \in Y$. This means that $\operatorname{essup}_{y\in Y} \|f_y\|_\infty \leq \|f\|_\infty$. For $\leq$: By definition of essential supremum of the function $y \mapsto \|f_y\|$, we know that for $\mu_Y$-every $y \in Y$ we have $\|f_y\|_\infty \leq \operatorname{essup}_{y\in Y} \|f_y\|_\infty$. This means that for $\mu_Y$-every $y \in Y$, for $\mu_X^y$-every $x \in X$ we have $|f_y(x)| \leq \operatorname{essup}_{y\in Y} \|f_y\|_\infty$. Equivalently,

$$\begin{aligned} &\mu_X^y(\{x \in X : |f_y(x)| > \operatorname*{essup}_{y\in Y} \|f_y\|_\infty\}) = 0 \quad \text{for } \mu_Y\text{-almost every } y \in Y \iff \\ &\mu_X^y(\{x \in X : |f(x)| > \operatorname*{essup}_{y\in Y} \|f_y\|_\infty, F(x) = y\}) = 0 \quad \text{for } \mu_Y\text{-almost every } y \in Y \iff \\ &\mu_X^y(\{x \in X : |f(x)| > \operatorname*{essup}_{y\in Y} \|f_y\|_\infty\}) = 0 \quad \text{for } \mu_Y\text{-almost every } y \in Y \iff \\ &\mu_X(\{x \in X : |f(x)| > \operatorname*{essup}_{y\in Y} \|f_y\|_\infty\}) = 0, \end{aligned}$$

where in the last step we have used Proposition 4.2.13. This means that $|f(x)| \leq \operatorname{essup}_{y\in Y} \|f_y\|_\infty$ for $\mu_X$-almost every $x \in X$. Equivalently, $\|f\|_\infty \leq \operatorname{essup}_{y\in Y} \|f_y\|_\infty$. □

**Proposition 4.2.15.** *Let $X$ be a Radon space and $\mu_X$ be a probability measure. Let $F : X \to Y$ be a surjective, Borel measurable function. Let $\mu_Y := F_*\mu_X$ be the pushforward measure of $\mu_X$ onto $Y$. Assume $\mu_X$ admits a disintegration with respect to $F$. Let $g : X \to \mathbb{R}$ be Borel measurable. Then the function*

$$m\colon Y \to \mathbb{R}, \quad m(y) = \operatorname*{essup}_{x\in F^{-1}(y)} g(x)$$

*where each essential supremum is taken with respect to $\mu^y$, is Borel measurable.*

*Proof of Proposition 4.2.15.* Let $a \in \mathbb{R}$. Then showing following set is measurable

$$\begin{aligned} m^{-1}((a,+\infty)) &= \{y \in Y : m(y) > a\} \\ &= \{y \in Y : \mu^y(\{x \in X : g(x) > a\}) > 0\} \\ &= \{y \in Y : \int_X \mathbb{1}_{g^{-1}((a,\infty))}\, d\mu^y > 0\} \\ &= \{y \in Y : h_a(y) > 0\} \\ &= h_a^{-1}((0,+\infty)), \end{aligned}$$

where $h_a(y) = \int_X \mathbb{1}_{g^{-1}((a,\infty))}\, d\mu^y$, is equivalent to showing that $m$ is measurable. In particular, since $g$ is Borel measurable, $g^{-1}((a,\infty))$ is a measurable subset of $X$, and $\mathbb{1}_{g^{-1}((a,\infty))}$ is a measurable function on $X$. By the disintegration theorem, Theorem 1 [39], the function $y \mapsto \int_X f\, d\mu^y$ is measurable for every measurable $f$. Hence, $h_a(y) = \int_X \mathbb{1}_{g^{-1}((a,\infty))}\, d\mu^y$ is a measurable function and so $h_a^{-1}((0,+\infty)) = m^{-1}((a,+\infty))$ is a measurable subset of $Y$. As $a \in \mathbb{R}$ was arbitrary, this proves that $m$ is Borel-measurable. □

Now in order to prove Theorem 4.2.9, we split it up into two theorems, Theorem 4.2.16 and Theorem 4.2.17. In order to prove these theorems, we will use the above propositions and corollary. Moreover, as mentioned above we will make use of Theorem 18.19 [40].

**Theorem 4.2.16 (Lower bound of part $(1)$, Theorem 4.2.9).** *Assume that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$ and that $\mathcal{M}_1$ is compact. Then, the following holds for every $p \in [1,\infty]$:*

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p) \leq 2c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p).$$

*Proof of Theorem 4.2.16.* Let $\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$. Let $y \in \mathcal{M}_2^{\mathcal{E}}$. Let $(x,e),(x',e') \in \mathcal{M}_1 \times \mathcal{E}$ such that $F(x,e) = F(x',e') = y$, then by the triangle inequality

$$d_1(x,x') \leq d_1^H(x, \varphi(F(x,e))) + d_1^H(x', \varphi(F(x',e'))). \tag{4.12}$$

**Part (1)**; for the case $p = \infty$, the essential supremum satisfies

$$\operatorname*{essup}_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x,x') \leq \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(x, \varphi(F(x,e))) + \operatorname*{essup}_{(x',e')\in F^{-1}(y)} d_1^H(x', \varphi(F(x',e'))).$$

Applying Proposition 4.2.13, taking the essential supremum with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$ and the infimum over $\varphi \in \mathcal{C}$ gives

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq 2c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

**Part (2)**; for the case $p \in [1,\infty)$, integrating (4.12) twice with respect to $(\mu_1 \otimes \nu)^y$ and using that $(a+b)^p \leq 2^p(a^p+b^p)$ for $a, b \geq 0$, we obtain

$$\begin{aligned}
&\int_{F^{-1}(y)} \int_{F^{-1}(y)} d_1(x,x')^p \, d(\mu_1 \otimes \nu)^y(x,e) \, d(\mu_1 \otimes \nu)^y(x',e') \leq \\
&\int_{F^{-1}(y)} \int_{F^{-1}(y)} (d_1^H(x, \varphi(F(x,e))) + d_1^H(x', \varphi(F(x',e'))))^p \, d(\mu_1 \otimes \nu)^y(x,e) d(\mu_1 \otimes \nu)^y(x',e') \\
&= 2^p \int_{F^{-1}(y)} d_1^H(x, \phi(F(x,e)))^p \, d(\mu_1 \otimes \nu)^y(x,e),
\end{aligned}$$

where in the last step we used the fact that $(\mu_1 \otimes \nu)^y$ is a probability measure. Now, integrating the above inequality with respect to $\rho$ on $\mathcal{M}_2^{\mathcal{E}}$ and raising to the power $\frac{1}{p}$ gives that

$$\begin{aligned}
\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p) &\leq \left( 2^p \int_{\mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} d_1^H(x, \varphi(F(x,e)))^p \, d(\mu_1 \otimes \nu)^y(x,e) \, d\rho(y) \right)^{\frac{1}{p}} \\
&= 2 \left( \int_{\mathcal{M}_1 \times \mathcal{E}} d_1^H(x, \varphi(F(x,e))) \, d(\mu_1 \otimes \nu)(x,e) \right)^{\frac{1}{p}}.
\end{aligned}$$

Since $\varphi \in \mathcal{C}$ was arbitrary, by taking the infimum over $\varphi \in \mathcal{C}$ we obtain that

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p) \leq 2 c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p).$$

□

**Theorem 4.2.17 (Upper bound of part $(1)$, Theorem 4.2.9 and parts $(2), (3)$).** *Assume that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$, that $\mathcal{M}_1$ is compact, and that with respect to $d_1$ closed balls are compact. Then, the following holds for every $p \in [1,\infty]$:*

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p) \leq \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p)$$

*and the optimal map $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ is given by,*

$$\Psi(y) = \operatorname{argmin}_{z \in \mathbb{C}^N} \operatorname*{essup}_{(x,e) \in F^{-1}(y)} d_1(x,z) \qquad (p = \infty) \tag{4.13}$$

$$\Psi(y) = \operatorname{argmin}_{z \in \mathbb{C}^N} \int_{F^{-1}(y)} d_1(x,z)^p \, d(\mu_1 \otimes \nu)^y(x,e) \qquad (p \in [1,\infty)) \tag{4.14}$$

*for every $y \in \mathcal{M}_2^{\mathcal{E}}$. Moreover, the optimal map has non-empty compact values.*

*Proof of Theorem 4.2.17.* We distinguish the cases $p = \infty$ and $p \in [1,\infty)$. First, we consider the case $p = \infty$.

**Part (1)**; $\Psi$ has non-empty values, since the minimum in (4.13) is attained. Fix $y \in \mathcal{M}_2^{\mathcal{E}}$. For $z \in \mathbb{C}^N$, let

$$f_y(z) = \operatorname*{essup}_{(x,e) \in F^{-1}(y)} d_1(x,z).$$

$f_y$ is continuous. Since $f_y : (\mathbb{C}^N, d_1) \to [0, \infty)$ is a function between metric spaces, continuity is equivalent to showing that for every sequence $z_n \to z$ there exists a subsequence $(z_{n_k})_k$ of $(z_n)_n$ such that $f_y(z_{n_k}) \to f_y(z)$. Let $z_n \to z$ wrt. $d_1$. Then, for $(x, e) \in F^{-1}(y)$, since the function $d(x, \cdot)$ is continuous, there exists a subsequence $(z_{n_k})_{k\in\mathbb{N}}$ such that

$$d_1(x, z) - 1/k \leq d_1(x, z_{n_k}) \leq d_1(x, z) + 1/k.$$

The essential supremum with respect to $(x, e) \in F^{-1}(y)$ satisfies

$$\operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x, z) - 1/k \leq \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x, z_{n_k}) \leq \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x, z) + 1/k.$$

Taking the limit $k \to \infty$ yields, $f_y(z_{n_k}) \to f_y(z)$. Thus, $f_y$ is continuous. Now define

$$r_y = \operatorname*{essup}_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x, x').$$

In particular, this implies that $d(x, x') \leq r_y$ for $(\mu_1 \otimes \nu)^y$-almost every $(x', e') \in F^{-1}(y)$. More precisely, define for $(x, e) \in F^{-1}(y)$ the set

$$E_{x,e} = \{(x', e') \in F^{-1}(y) : d(x, x') > r_y\}.$$

Then, $(\mu_1 \otimes \nu)^y(\{(x, e) \in F^{-1}(y) : (\mu_1 \otimes \nu)(E_{x,e}) \neq 0\}) = 0$. Define

$$G_y := \{(x, e) \in F^{-1}(y) : (\mu_1 \otimes \nu)^y(E_{x,e}) \neq 0\}.$$

If $(x, e) \in G_y$, then the $x$-projection of the feasible set $\pi_1(F^{-1}(y))$ is contained within the ball $B_{d_1}(x, r_y) = \{z \in \mathbb{C}^N : d_1(z, x) \leq r_y\}$, up to a measure zero set. More precisely, $(\mu_1 \otimes \nu)^y\big(F^{-1}(y) \cap (\mathbb{C}^N \setminus (B_{d_1}(x, r_y) \times \mathcal{E}))\big) = 0$. This is immediate, since $F^{-1}(y) \cap (\mathbb{C}^N \setminus (B_{d_1}(x, r_y) \times \mathcal{E})) \subseteq E_{(x,e)}$. For $(x, e) \in G_y$,

$$\Psi(y) = \operatorname{argmin}_{z\in\mathbb{C}^N} \operatorname*{essup}_{(x',e')\in F^{-1}(y)} d_1(z, x') = \operatorname{argmin}_{z\in B_{d_1}(x,r_y)} f_y(z). \tag{4.15}$$

In fact, if $z \in \mathbb{C}^N \setminus B_{d_1}(x, r_y)$, then

$$f_y(z) = \operatorname*{essup}_{(x'e')\in F^{-1}(y)} d_1(z, x') \geq d_1(x, z) > r_y \geq f_y(x).$$

Therefore $\inf_{z\in\mathbb{C}^N} f_y(z) \geq r_y \geq \inf_{z\in B_{d_1}(x,r_y)} f_y(z)$, which proves (4.15). By assumption $B_{d_1}(x, r_y) \subset \mathbb{C}^N$ is compact, since it is a closed ball with respect to the metric $d_1$. Therefore, the minimum in (4.15) can be attained by the Extreme Value Theorem. This shows that the argmin is non-empty, and hence that $\Psi$ has non-empty values on $\mathcal{M}_2^{\mathcal{E}}$.

**Part (2)**; let us show that $\Psi$ is measurable and that is has non-empty, compact values. We apply the maximum measurable theorem. Note that this theorem works with minimsers

instead of maximisers, since $\min f = -\max(-f)$ for any function $f$. Theorem 4.2.11 is applied with $S = \mathcal{M}_2^{\mathcal{E}}$, $X = \mathbb{C}^N$,

$$\begin{aligned} &\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N, \qquad \varphi(y) = \mathcal{M}_1 + \mathcal{B}_{d_1}(0, 2\text{diam}(\mathcal{M}_1)) \\ &f : \mathcal{M}_2^{\mathcal{E}} \times \mathbb{C}^N \to \mathbb{R}, \; f(y,z) = f_y(z) = \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x,z). \end{aligned}$$

In order to apply the theorem, we need to prove the following claims:

**Claim 1:** $\varphi$ is weakly-measurable with non-empty compact values. $\varphi$ is weakly measurable and has non-empty values since it is constant. Moreover, $\mathcal{M}_1 + \mathcal{B}_{d_1}(0, 2\text{diam}(\mathcal{M}_1))$ is compact, as closed bounded balls with respect to the metric $d_1$ are compact, $\mathcal{M}_1$ is compact, and the sum of two compact sets is compact. **Claim 2:** $f$ is Caratheodory. In particular, $f(y,\cdot) = f_y$ is continuous for every fixed $y \in \mathcal{M}_2^{\mathcal{E}}$ and $f(\cdot, z)$ is measurable for every fixed $z \in \mathcal{M}_1$. For every fixed $y \in \mathcal{M}_2^{\mathcal{E}}$, the function $f_y$ is continuous on $\mathbb{C}^N$ as proven above. For every fixed $z$, the function $f(\cdot, z) : y \mapsto \operatorname{essup}_{(x,e)\in F^{-1}(y)} d_1(x,z)$ is Borel measurable thanks to the disintegration theorem by Proposition 4.2.15.

Then, by Theorem 4.2.11, the possibly multivalued function $\mu : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$,

$$\mu(y) = \operatorname{argmin}_{z\in\mathcal{M}_1+\mathcal{B}_{d_1}(0,2\text{diam}(\mathcal{M}_1))} \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x,z) \tag{4.16}$$

is measurable. As for any $(x,e) \in F^{-1}(y)$, $\mathcal{B}_{d_1}(x, r_y) \subseteq \mathcal{M}_1 + \mathcal{B}_{d_1}(0, 2\text{diam}(\mathcal{M}_1))$ and by (4.15), we get that $\Psi = \mu$. Hence, $\Psi$ is measurable and has non-empty, compact values.

**Part (3)**; let us prove that $\Psi \in \mathcal{C}$, i.e. that the function

$$r_\Psi : \mathcal{M}_1 \times \mathcal{E} \to \mathbb{R}, \quad r_\Psi(x,e) = d^H(x, \Psi(F(x,e))) = \sup_{z\in\Psi(F(x,e))} d_1(x,z)$$

is measurable. This claim follows from the Measurable Maximum Theorem 4.2.11 with $S = \mathcal{M}_1 \times \mathcal{E}$, $X = \mathbb{C}^N$, and

$$\begin{aligned} &\varphi = \Psi \circ F : \mathcal{M}_1 \times \mathcal{E} \rightrightarrows \mathbb{C}^N, && \varphi(x,e) = \Psi(F(x,e)) \\ &f = d \circ (\pi_1, \text{id}) : (\mathcal{M}_1 \times \mathcal{E}) \times \mathbb{C}^N \to \mathbb{R}, && f((x,e),z) = d_1(x,z). \end{aligned}$$

We can apply the theorem since the function $\varphi$ is weakly measurable (since it is the composition of the continuous function $F$ and the measurable function $\Psi$), and has compact values as proven above. The function $f$ is continuous and hence Caratheodory. Then, Theorem 4.2.11 implies that

$$m(y) = \max_{z\in\Psi(F(x,e))} d_1(x,z)$$

is measurable. It is immediate to notice that $m = r_\Psi$, so we conclude that $r_\Psi$ is measurable. Hence $\Psi \in \mathcal{C}$.

**Part (4)**; $\Psi$ is an optimal map. Let $\varphi \in \mathcal{C}$. Fix $y \in \mathcal{M}_2^{\mathcal{E}}$. By the definition of $\Psi$,

$$d_1^H(\Psi(y), x) \le d_1(z,x) \quad \text{for } (\mu_1 \otimes \nu)^y\text{-almost every } (x,e) \in F^{-1}(y)$$

for every $z \in \varphi(y)$. In particular, taking the supremum with respect to $z \in \varphi(y)$, which coincides with the Hausdorff distance

$$\begin{aligned} d_1^H(\Psi(y), x) &\leq \sup_{z\in\varphi(y)} d_1(z, x) \\ &= d_1^H(\varphi(y), x) \quad \text{for } (\mu_1 \otimes \nu)^y\text{-almost every } (x, e) \in F^{-1}(y). \end{aligned}$$

Hence,

$$\operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(\Psi(y)), x) \leq \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(\varphi(y), x).$$

By considering the essential supremum with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$, we obtain

$$\operatorname*{essup}_{y\in\mathcal{M}_2^{\mathcal{E}}} \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(\Psi(y)), x) \leq \operatorname*{essup}_{y\in\mathcal{M}_2^{\mathcal{E}}} \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(\varphi(y), x).$$

Due to Corollary 4.2.14, this can be rewritten as

$$\operatorname*{essup}_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\Psi(F(x, e))), x) \leq \operatorname*{essup}_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\varphi(F(x, e))), x).$$

Now, as $\varphi \in \mathcal{C}$ was arbitrary, we obtain

$$\operatorname*{essup}_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\Psi(F(x, e))), x) \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

The opposite inequality holds trivially since $\Psi \in \mathcal{C}$. Therefore, $\Psi$ is an optimal map.

**Part (5)**; by the minimisation property of $\mu$, thus of $\Psi$, for every $(x', e') \in F^{-1}(y)$,

$$\operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1^H(x, \Psi(y)) \leq \operatorname*{essup}_{(x,e)\in F^{-1}(y)} d_1(x, x').$$

Thus, taking the essential supremum with respect to $(x', e') \in F^{-1}(y)$ and with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$, yields

$$\begin{aligned} \operatorname*{essup}_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(x, \Psi(F(x, e))) &\leq \operatorname*{essup}_{y\in\mathcal{M}_2^{\mathcal{E}}} \operatorname*{essup}_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x, x') \\ &= \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty). \end{aligned}$$

As $\Psi \in \mathcal{C}$, we deduce that

$$c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty) \leq \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

For $p \in [1, \infty)$.

**Part (1)**; $\Psi$ in (4.11) has non-empty values. Fix $y \in \mathcal{M}_2^{\mathcal{E}}$. Define $f_y : \mathbb{C}^N \to [0, \infty]$,

$$f_y(z) = \int_{F^{-1}(y)} d_1(x, z)^p \, d(\mu_1 \otimes \nu)^y(x, e)$$

for $z \in \mathbb{C}^N$. $f_y$ is continuous. Let $z_n \to z$ with respect to $d_1$. Then for $(x,e) \in F^{-1}(y)$, since the function $d_1(x,\cdot)^p$ is continuous, there exists a subsequence $(z_{n_k})_{k\in\mathbb{N}}$ in $\mathbb{C}^N$ such that for every $k \in \mathbb{N}$,

$$d_1(x,z)^p - 1/k \le d_1(x,z_{n_k}) \le d_1(x,z)^p + 1/k.$$

Integrating over $(x,e) \in F^{-1}(y)$ and using that $(\mu_1 \otimes \nu)^y$ is a probability measure, we get

$$\begin{aligned}\int_{F^{-1}(y)} d_1(x,z)^p \, d(\mu_1\otimes\nu)^y(x,e) - 1/k \le \int_{F^{-1}(y)} d_1(x,z_{n_k})^p \, d(\mu_1\otimes\nu)^y(x,e) \\ \le \int_{F^{-1}(y)} d_1(x,z)^p \, d(\mu_1\otimes\nu)^y(x,e) + 1/k.\end{aligned}$$

Taking the limit $k \to \infty$ yields $f(z_{n_k}) \to f(z)$. Since $f : \mathbb{C}^N \to [0,\infty)$ is a function between metric spaces, $f$ is continuous. Define,

$$\begin{aligned} r_y &= \operatorname*{essup}_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_1(x,x') \\ E_{x,e} &= \{(x',e') \in F^{-1}(y) : d_1(x,x') > r_y\} \\ G_y &= \{(x,e) \in F^{-1}(y) : (\mu_1\otimes\nu)^y(E_{x,e}) \neq 0\}. \end{aligned}$$

For $(x,e) \in G_y$,

$$\Psi(y) = \operatorname{argmin}_{z\in\mathbb{C}^N} \int_{F^{-1}(y)} d_1(x',z)^p \, d(\mu_1\otimes\nu)^y(x',e') = \operatorname{argmin}_{z\in B_{d_1}(x,2r_y)} f_y(z). \tag{4.17}$$

In fact, if $z \in \mathbb{C}^N \setminus B_{d_1}(x,2r_y)$, then for $(\mu_1\otimes\nu)^y$-almost every $(x',e') \in F^{-1}y$,

$$d_1(z,x') \ge d_1(z,x) - d_1(x,x') > 2r_y - r_y = r_y.$$

Thus,

$$\begin{aligned} f_y(z) &= \int_{F^{-1}(y)} d_1(z,x')^p \, d(\mu_1\otimes\nu)^y(x',e') \\ &> \int_{F^{-1}(y)} r_y^p / d(\mu_1\otimes\nu)^y(x',e') = r_y^p. \end{aligned}$$

The previous inequality holds for any $z \in \mathbb{C}^N \setminus B_{d_1}(x,2r_y)$. On the other hand,

$$\begin{aligned} f_y(x) &= \int_{F^{-1}(y)} d_1(x,x')^p \, d(\mu_1\otimes\nu)^y(x',e') \\ &\le \int_{F^{-1}(y)} r_y^p \, d(\mu_1\otimes\nu)^y(x',e') = r_y^p. \end{aligned}$$

Therefore $\inf_{z\in\mathbb{C}^N} f_y(z) \ge r_y^p \ge \inf_{z\in B_{d_1}(x,2r_y)} f_y(z)$, which proves (4.17). By assumption the set $B_{d_1}(x,2r_y) \subset \mathbb{C}^N$ is compact, since it is a closed ball with respect to the metric

$d_1$. Therefore, the minimum in (4.15) can be attained by the Extreme Value Theorem. This shows that the argmin is non-empty, and hence that $\Psi$ has non-empty values on $\mathcal{M}_2^{\mathcal{E}}$.

**Part (2)**; $\Psi \in \mathcal{C}$. Again it is sufficient to prove that $\Psi$ is measurable. We will apply the maximum measurable theorem, 4.2.11, with $S = \mathcal{M}_2^{\mathcal{E}}$, $X = \mathbb{C}^N$,

$$\varphi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N, \qquad \varphi(y) = \mathcal{M}_1 + \mathcal{B}_{d_1}(0, 2\text{diam}(\mathcal{M}_1))$$
$$f : \mathcal{M}_2^{\mathcal{E}} \times \mathbb{C}^N \to \mathbb{R},\ f(y,z) = f_y(z) = \int_{F^{-1}(y)} d_1(x,z)^p \, d(\mu \otimes \nu)^y(x,e).$$

In order to apply the theorem, we need to prove the following claims: **Claim 1:** $\varphi$ is weakly-measurable with non-empty compact values. This was proved for $p = \infty$. **Claim 2:** $f$ is Caratheodory, i.e. measurable in the first argument and continuous in the second argument. We need to show that $f(y,\cdot) = f_y$ is continuous for every fixed $y \in \mathcal{M}_2^{\mathcal{E}}$ and that $f(\cdot, z)$ is measurable for every fixed $z \in \mathcal{M}_1$. For every fixed $y \in \mathcal{M}_2^{\mathcal{E}}$, the function $f_y$ is continuous on $\mathbb{C}^N$ as proven above. For every fixed $z$, the function $f(\cdot, z) : y \mapsto \int_{F^{-1}(y)} d_1(x,z)^p \, d(\mu \otimes \nu)^y(x,e)$ is Borel measurable due to the properties of the disintegrations of measure. Then, by Theorem 4.2.11, the possibly multivalued function $\mu : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathcal{M}_1$,

$$\mu(y) = \text{argmin}_{\substack{z \in \mathcal{M}_1: \\ \exists e' \in \mathcal{E} : F(z,e') = y}} \int_{F^{-1}y} d_1(x,z)^p \, d(\mu \otimes \nu)^y(x,e)$$

is measurable. Moreover, by (4.17) and as for $(x,e) \in G_y$, $B_{d_1}(x, 2r_y) \subseteq \mathcal{M}_1 + \mathcal{B}_{d_1}(0, 2\text{diam}(\mathcal{M}_1))$, $\mu = \Psi$. Hence, $\Psi$ is measurable.

**Part (3)**; $\Psi$ is an optimal map. Let $\varphi \in \mathcal{C}$. By the definition of $\Psi$,

$$\int_{F^{-1}(y)} d_1^H(\Psi(y), x)^p \, d(\mu_1 \otimes \nu)^y(x,e) \leq \int_{F^{-1}(y)} d_1(z,x)^p \, d(\mu_1 \otimes \nu)^y(x,e)$$

for every $z \in \varphi(y)$. In particular, taking the supremum with respect to $z \in \varphi(y)$, which coincides with considering the Hausdorff distance, and using Fatou's Lemma yields

$$\begin{aligned}\int_{F^{-1}(y)} d_1(\Psi(y), x)^p \, d(\mu_1 \otimes \nu)^y(x,e) &\leq \sup_{z \in \varphi(y)} \int_{F^{-1}(y)} d_1(z,x)^p \, d(\mu_1 \otimes \nu)^y(x,e) \\ &\leq \int_{F^{-1}(y)} \sup_{z \in \varphi(y)} d_1(z,x)^p \, d(\mu_1 \otimes \nu)^y(x,e) \\ &\leq \int_{F^{-1}(y)} d_1^H(\varphi(y), x)^p \, d(\mu_1 \otimes \nu)^y(x,e).\end{aligned}$$

By integrating with respect to $y \in \mathcal{M}_2^{\mathcal{E}}$, we obtain

$$\begin{aligned}&\int_{y \in \mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} d_1^H(\Psi(y), x)^p \, d(\mu_1 \otimes \nu)^y(x,e) \, d\rho(y) \\ &\leq \int_{y \in \mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} d_1^H(\varphi(y), x)^p \, d(\mu_1 \otimes \nu)^y(x,e) \, d\rho(y).\end{aligned}$$

Due to the disintegration of measure, this can be rewritten as

$$\int_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\Psi(F(x,e)),x)\; d(\mu_1\otimes\nu)(x,e)$$
$$\leq \int_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\varphi(F(x,e)),x)\; d(\mu_1\otimes\nu)(x,e).$$

Now, as $\varphi\in\mathcal{C}$ was arbitray and by raising both sides to the power $\frac{1}{p}$, we obtain

$$\left(\int_{(x,e)\in\mathcal{M}_1\times\mathcal{E}} d_1^H(\Psi(F(x,e)),x)^p\; d(\mu_1\otimes\nu)(x,e)\right)^{\frac{1}{p}} \leq c_{\text{opt}}(F,\mathcal{M}_1,\mathcal{E},p).$$

The opposite inequality holds trivially, as $\Psi\in\mathcal{C}$. Therefore, $\Psi$ is an optimal map.

**Part (4)**; by the minimisation property of $\mu$, hence also of $\Psi$, for every $(x',e')\in F^{-1}(y)$:

$$\int_{F^{-1}y} d_1^H(x,\Psi(y))^p\; d(\mu\otimes\nu)^y(x,e) \leq \int_{F^{-1}y} d_1(x,x')^p\; d(\mu\otimes\nu)^y(x,e). \tag{4.18}$$

Integrating (4.18) with respect to $(\mu_1\otimes\nu)^y$ yields,

$$\int_{F^{-1}(y)} d_1^H(x,\Psi(y))^p\; d(\mu_1\otimes\nu)^y(x,e)$$
$$\leq \int_{F^{-1}(y)}\int_{F^{-1}(y)} d_1(x,x')^p\; d(\mu_1\otimes\nu)^y(x,e)\; d(\mu_1\otimes\nu)^y(x',e'), \tag{4.19}$$

where we used that $(\mu_1\otimes\nu)^y$ is a probability measure. Integrating both sides over $y\in\mathcal{M}_2^{\mathcal{E}}$ with respect to $\rho$ we obtain

$$\int_{\mathcal{M}_2^{\mathcal{E}}}\int_{F^{-1}(y)} d_1^H(x,\Psi(y))^p\; d(\mu_1\otimes\nu)^y(x,e)\; d\rho(y)$$
$$\leq \int_{\mathcal{M}_2^{\mathcal{E}}}\int_{F^{-1}(y)}\int_{F^{-1}(y)} d_1(x,x')^p\; d(\mu_1\otimes\nu)^y(x,e)\; d(\mu_1\otimes\nu)^y(x',e')\; d\rho(y).$$

Using the disintegration of the measure $\mu_1\otimes\nu$ on the left hand side of the above inequality

$$\int_{\mathcal{M}_1\times\mathcal{E}} d_1^H(x,\Psi(F(x,e)))^p\; d(\mu_1\otimes\nu)(x,e) \leq \text{kersize}(F,\mathcal{M}_1,\mathcal{E},p)^p.$$

Then, raising both sides to the $\frac{1}{p}$-th power, gives

$$\left(\int_{\mathcal{M}_1\times\mathcal{E}} d_1^H(x,\Psi(F(x,e)))^p\; d(\mu_1\otimes\nu)(x,e)\right)^{\frac{1}{p}} \leq \text{kersize}(F,\mathcal{M}_1,\mathcal{E},p).$$

And finally, since $\Psi\in\mathcal{C}$, we conclude

$$c_{\text{opt}}(F,\mathcal{M}_1,\mathcal{E},p) \leq \text{kersize}(F,\mathcal{M}_1,\mathcal{E},p).$$

□

The following proposition states sufficient conditions for a disintegration of measure to exist.

**Proposition 4.2.18.** *Let $\mathcal{M}_1 \subseteq \mathbb{C}^N$ be equipped with the Borel $\sigma$-algebra and a finite positive measure $\mu_1$. Let $\mathcal{E} \subseteq \mathbb{C}^m$ be equipped with a finite positive measure $\nu$ on $\mathcal{E}$, and such that the completion of the product measure $\mu_1 \otimes \nu$ is a complete Radon measure. Let $d_1, d_2$ be metrics on $\mathcal{M}_1$, respectively $\mathcal{M}_2^{\mathcal{E}}$, such that $(\mathcal{M}_2^{\mathcal{E}}, d_2)$ is a separable metric space. Assume that $F$ is surjective and continuous. Let $\rho = F_*(\mu_1 \otimes \nu)$ be the pushforward measure defined on $\mathcal{M}_2^{\mathcal{E}}$. Then, there exists disintegration of the measure $\mu_1 \otimes \nu$. Moreover, for every $y \in \mathcal{M}_2^{\mathcal{E}}$ the measure $(\mu_1 \otimes \nu)^y$ is a probability measure.*

In order to prove Proposition 4.2.18, we make use of the following theorem.

**Theorem 4.2.19** (Theorem 1 [39])**.** *Let $\mu$ be a sigma-finite Radon measure on a metric space $(X, d_X)$ and let $F : X \to Y$ be a measurable map. Let $\rho$ be a sigma-finite measure on $Y$ that dominates the pushforeward measure $F_*\mu$. If $Y$ is countably generated and contains all the singleton sets $\{y\}$, then $\mu$ has a $(F, \rho)$-disintegration. The $\mu^y$ measures are uniquely determined up to an almost sure equivalence: if $\mu_0^y$ is another $(F, \rho)$-disintegration then $\rho(\{y \in Y : \mu_0^y = \mu^y\}) = 0$.*

Given the setting in Theorem, 4.2.19, one can obtain a slightly stronger result than the one of Theorem 4.2.12.

*Proof of Proposition 4.2.18.* As $(\mathcal{M}_2^{\mathcal{E}}, d_2)$ is a seperable metric space, the Borel sigma algrebra is countably generated. Moreover, the mapping $F$ is also measurable, as it is continuous. Then, by Theorem 1 and Theorem 2 [39], there exists disintegration of the measure $\mu_1$:

$$\int_{\mathcal{M}_1 \times \mathcal{E}} f(x) d(\mu_1 \otimes \nu)(x, e) = \int_{\mathcal{M}_2^{\mathcal{E}}} \int_{F^{-1}(y)} f(x) d(\mu_1 \otimes \nu)^y(x, e)\, d\rho(y)$$

for any positive measurable function $f$. This means that for every $y \in \mathcal{M}_2^{\mathcal{E}}$ we have a measure $(\mu_1 \otimes \nu)^y$ on the feasible set $F^{-1}(y) = \{(x, e) \in \mathcal{M}_1 \times \mathcal{E} \mid F(x, e) = y\}$. Moreover, for every $y \in \mathcal{M}_2^{\mathcal{E}}$ the measure $(\mu_1 \otimes \nu)^y$ is a probability measure. □

Note that the assumption that $F$ is continuous, is the case in most inverse problems where the sampling operator $A$ is assumed to be continuous.

### 4.2.4 Approximability of optimal maps by neural networks

As our main aim is to obtain bound for the reconstruction quality of neural networks used as decoders for (2.18), we define neural networks in the following way.

**Definition 4.2.20** (Neural network)**.** Let $L \in \mathbb{N}$. $\phi : \mathbb{R}^m \to \mathbb{R}^N$ is called a neural network, if it is of the form $\phi(x) = \mathcal{L}_L \circ \sigma \circ \mathcal{L}_{L-1} \circ \cdots \circ \mathcal{L}_1 \circ \sigma \circ \mathcal{L}_0(x)$, for $x \in \mathbb{R}^m$ and $\mathcal{L}_i$ being affine linear transformations for $i \in \{1, \cdots L\}$ and $\sigma' : \mathbb{R} \to \mathbb{R}$ is an activation function, which is a continuous function that is not a polynomial. For any $k \in \mathbb{N} \setminus \{0\}$ for $y \in \mathbb{R}^k$, $\sigma(y) = [\sigma'(y_1), \cdots \sigma'(y_k)]^\top$ is applied pointwise. Moreover, $\mathcal{NN}$ denotes a family of neural networks specified by the affine linear transformations and an activation function.

However, note that the family of neural networks $\mathcal{NN}$ can be obtained by various different architectures. See Section 1.4 for an overview of DL and DL for solving inverse problems.

In the following we will consider specific cases of Definition 4.2.20, as used in Theorem 3.1 [148] and [163]. Our next results state that the optimal map with worst-case noise can be uniformly approximated on arbitrary balls. Definition 4.2.20 with the activation function presented in [163], yields Theorem 4.2.21. Note that in the following two theorems, we consider additive noise and a linear measurement model. In particular, $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}}$, is given by $F(y) = Ax + e$ for $y \in \mathcal{M}_2^{\mathcal{E}}$ and fixed $A \in \mathbb{C}^{m \times N}$.

**Theorem 4.2.21 (Existence of neural network approximating the best worst-case decoder).** *Let $m < N$, $\epsilon \geq 0$, $\rho > 0$, $A \in \mathbb{R}^{m \times N}$ be linear and bounded, let $z \in \mathbb{R}^N$ and $\mathcal{M}_1 = \mathcal{B}_{d_1}(z, \rho) \subseteq \mathbb{R}^N$ and let $\mathcal{E} = B_{d_2}(0, \varepsilon)$, $\mathcal{M}_2^{\mathcal{E}} = \mathcal{M}_2 + \mathcal{E}$. Let $d_1, d_2$ be induced by the $\ell^2$-norm and $\Psi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ be the optimal map with worst-case noise, which minimises* (4.2.1).

*Then:*

*(1) $\Psi$ is continuous. Moreover, for any $\delta > 0$, there exists a neural network $\phi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ with $L = 11$ layers and width $W = N36m(2m + 1)$ that satisfies,*

$$\sup_{y \in \mathcal{M}_2^{\mathcal{E}}} \|\Psi(y) - \phi(y)\|_{\ell^2} \leq \delta.$$

*(2) For $\mathcal{E} = \{0\}$, $\Psi$ is Lipschitz continuous.*

The following auxiliary results are used in order to prove Theorem 4.2.21. Note that Proposition 4.2.25, is an extension of Proposition 3.4.8, in Chapter 3, for the noisy case. Moreover, Lemma 4.2.23 and 4.2.24, as well as Corollary 4.2.22 from [136] used in the proof are presented below. The results presented in the following use the theory of convergence of convex sets. As we only use the notion of pointwise convergence, we direct the reader to [136] for the precise definition of such a convergence of convex sets.

**Corollary 4.2.22.** *( [136], Corollary of Theorem A, with $T = Id$) Let $K$ and $(K_n)_N$ be nonepmty closed convex subsets of $X$ such that $K_n \to K$. Then there exists for each $n \in \mathbb{N}$ one and only one point $u_n$ such that*

$$\langle u_n,\ x - u_n \rangle \geq 0 \quad (x \in K_n) \tag{4.20}$$

*and there exists a unique solution $u$ of inequality*

$$\langle u,\ x-u\rangle \geq 0 \quad (x \in K). \tag{4.21}$$

*Moreover, $u_n \to u$.*

**Lemma 4.2.23** ( [136], Lemma 1.4)**.** *Let $X$ be a reflexive Banach space, and let $K$ be a closed convex subset of $X$, whose interior is non-empty, and $(S_n)_n$ is a sequence of closed convex subsets of $X$, such that $S_n \to S$ is $X$. Then $K \cap S_n \to K \cap S$ in $X$.*

**Lemma 4.2.24** ( [136], Lemma 1.6)**.** *Let $X$ be a reflexive Banach space. Let $(S_n)_n$ be a sequence of subsets of $X$ such that $S_n \to S$, $(v_n)_n$ a sequence of vectors of $X$ such that $v_n \to v$ in $X$. Then $S_n + v_n \to S + v$ in $X$.*

For the proofs of the above results, see [136].

**Proposition 4.2.25** ( [163], Theorem 1)**.** *Let $A \in \mathbb{R}^{m\times N}$, with $1 \leq \operatorname{rank}(A) < N$. Let $\mathcal{M}_1 \subseteq \mathbb{R}^N$ be closed and bounded, $\mathcal{M}_2^{\mathcal{E}} = (A\mathcal{M}_1)^{\epsilon} \subset \mathbb{R}^N$. Let $\Psi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ be continuous. Then, for any $\delta > 0$ there exists a neural network $\phi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ with $L = 11$ layers and width $W = N36m(2m+1)$ that satisfies,*

$$\sup_{\substack{x\in\mathcal{M}_1 \\ e\in\mathcal{B}_{\|\cdot\|_{\ell^2}}(0,\epsilon)}} \|\Psi(Ax+e) - \phi(Ax+e)\|_{\ell^2} \leq \delta.$$

Given the above results, we can now prove Theorem 4.2.21.

*Proof of Theorem 4.2.21.* **Part (1)**; First note that for any closed and bounded $Z \subseteq \mathbb{R}^N$ and any $z \in \mathbb{R}^N$, we have that $\tilde{z} \in \mathbb{R}^N$ is the center of the smallest enclosing ball around $Z$ if and only if $\tilde{z} + z$ is the center of the smallest enclosing ball around $Z + z$. Thus, in the following we will restrict $\mathcal{M}_1$ to the ball around $0 \in \mathbb{R}^N$ with radius $\rho = 1$. Let us now prove that for $y \in \mathcal{M}_2^{\mathcal{E}}$ the optimal map with worst-case noise, given in (4.6), satisfies

$$\begin{aligned}\Psi(y) &= \operatorname{argmin}_{z\in\mathbb{R}^N} d^H(z, A^{-1}(B_{\|\cdot\|_{\ell^2}}(y,\epsilon)) \cap \mathcal{M}_1) \\ &= \operatorname{argmin}\{\|x\|_{\ell^2} : x \in A^{-1}(B_{\|\cdot\|_{\ell^2}}(y,\epsilon)) \cap \mathcal{M}_1\}.\end{aligned} \tag{4.22}$$

To see this, notice that equation (4.22) gives the projection of $0 \in \mathbb{R}^N$ onto the feasible set

$$\pi_1(F^{-1}(y)) = A^{-1}(B_{\|\cdot\|_{\ell^2}}(y,\epsilon)) \cap \mathcal{M}_1,$$

which is convex because both $\mathcal{M}_1$ and $B_{\|\cdot\|_{\ell^2}}(y,\epsilon)$ are convex. By the classic theorem of projection onto convex sets in Hilbert spaces (see [23], Theorem 5.2), we have that the projection is unique. We will call such point $\operatorname{proj}_{\pi_1(F^{-1}(y))}(0)$. Moreover, it satisfies

$$\langle 0 - \operatorname{proj}_{\pi_1(F^{-1}(y))}(0),\ x - \operatorname{proj}_{\pi_1(F^{-1}(y))}(0)\rangle \leq 0 \tag{4.23}$$

for every $x \in A^{-1}(B_{\|\cdot\|_{\ell^2}}(y,\epsilon)) \cap \mathcal{M}_1$. We now claim that $\Psi(y) = \mathrm{proj}_{\pi_1(F^{-1}(y))}(0)$. In order to do so, we need to prove that $\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)$ is the center of the smallest ball containing $\pi_1(F^{-1}(y))$. We claim that

$$\pi_1(F^{-1}(y)) \subseteq B_{\|\cdot\|_{\ell^2}}\Big(\mathrm{proj}_{\pi_1(F^{-1}(y))}(0), \sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2}\Big) \tag{4.24}$$

and that this inclusion is minimal. Firstly, note that $\mathrm{proj}_{\pi_1(F^{-1}(y))}(0) \in \mathcal{N}(A)^\perp$ by the minimality of its norm.

Let us prove minimality. We will show that, if a ball contains the feasible set, i.e. $B_{\|\cdot\|_{\ell^2}}(p,r) \supseteq \pi_1(F^{-1}(y))$, then necessarily its radius satisfies

$$r \geq \sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2}.$$

Let $v \in \mathcal{N}(A)$, $\|v\|_{\ell^2} = 1$. Consider the points

$$x_1 = \mathrm{proj}_{\pi_1(F^{-1}(y))}(0) + v\sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2},$$
$$x_2 = \mathrm{proj}_{\pi_1(F^{-1}(y))}(0) - v\sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2}.$$

Then, $x_1, x_2 \in \pi_1(F^{-1}(y))$, since $Ax_1 = Ax_2 = A\,\mathrm{proj}_{\pi_1(F^{-1}(y))}(0) \in B_{\|\cdot\|_{\ell^2}}(y,\epsilon)$ and $\|x_1\|_{\ell^2} = \|x_2\|_{\ell^2} = 1$. We now have

$$2\sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2} = d_1(x_1,x_2) \leq d_1(x_1,p) + d_1(p,x_2) \leq 2r. \tag{4.25}$$

Hence, $r \geq \sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2}$ as claimed.

Let us prove the inclusion

$$\pi_1(F^{-1}(y)) \subseteq B_{\|\cdot\|_{\ell^2}}\Big(\mathrm{proj}_{\pi_1(F^{-1}(y))}(0), \sqrt{1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2}\Big).$$

Let $x \in \pi_1(F^{-1}(y))$. Then

$$\begin{aligned}
&\|x - \mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2 \\
&= \|x\|_{\ell^2}^2 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2 + 2\langle x - \mathrm{proj}_{\pi_1(F^{-1}(y))}(0), -\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\rangle \\
&\leq 1 - \|\mathrm{proj}_{\pi_1(F^{-1}(y))}(0)\|_{\ell^2}^2.
\end{aligned}$$

Hence, the inclusion is established.

Thus, $\Psi(y) = \mathrm{proj}_{\pi_1(F^{-1}(y))}(0)$, since this is the center of the smallest ball containing $\pi_1(F^{-1}(y))$.

Let us now prove that $\Psi$ is continuous. The main results that we will use are Lemma 4.2.23 and 4.2.24.

Moreover, we will apply Corollary of Theorem A from [136] with $T = Id$ being the identity operator, which is bounded, hemicontinuous and satisfies the required properties. In this special case, we can use Corollary 4.2.22.

Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_2^{\mathcal{E}}$ such that $y_n \to y$ for some $y \in \mathbb{C}^m$. Note that $y \in \mathcal{M}_2^{\mathcal{E}}$ since $\mathcal{M}_2$ is closed. Let us prove that $\Psi(y_n) \to \Psi(y)$. Let us rewrite

$$\begin{aligned}\pi_1(F^{-1}(y)) &= \mathcal{M}_1 \cap A^{-1}(B_{\|\cdot\|_{\ell^2}}(y,\epsilon)) = \mathcal{M}_1 \cap A^{-1}(y + B_{\|\cdot\|_{\ell^2}}(0,\epsilon)) \\ &= \mathcal{M}_1 \cap (A^\dagger y + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon))\end{aligned}$$

where $A^\dagger$ denotes the Moore-Penrose inverse. Analogously,

$$\pi_1(F^{-1}(y_n)) = \mathcal{M}_1 \cap (A^\dagger y_n + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon)).$$

Since $A$ is bounded, then $A^\dagger$ is continuous. In particular, $A^\dagger y_n \to A^\dagger y$. Applying Lemma 4.2.24 with $v_n = A^\dagger y_n$, $v = A^\dagger y$, $S_n = S = A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon)$ for every $n \in \mathbb{N}$, we obtain that $(A^\dagger y_n + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon)) \to (A^\dagger y + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon))$. Now, since $\mathcal{M}_1 = B_{\|\cdot\|_{\ell^2}}(0,1) \subseteq \mathbb{R}^N$ is convex, closed and has a non-empty interior, applying Lemma 4.2.23 with $K = \mathcal{M}_1$ we obtain that

$$\begin{aligned}\pi_1(F^{-1}(y_n)) &= \mathcal{M}_1 \cap (A^\dagger y_n + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon)) \\ &\to \mathcal{M}_1 \cap (A^\dagger y + A^{-1}B_{\|\cdot\|_{\ell^2}}(0,\epsilon)) = \pi_1(F^{-1}(y)).\end{aligned}$$

Now, on the one hand, by the characterization of projections into Hilbert spaces, we know that the points $\Psi(y_n)$ are characterized by being the unique solutions to

$$\langle \Psi(y_n),\ x - \Psi(y_n)\rangle \geq 0 \quad \text{for} \quad x \in F^{-1}(y_n),\ n \in \mathbb{N}, \tag{4.26}$$

and similarly $\Psi(y)$ is the unique point satisfying

$$\langle \Psi(y),\ x - \Psi(y)\rangle \geq 0 \quad \text{for} \quad x \in \pi_1(F^{-1}(y)). \tag{4.27}$$

On the other hand, applying Corollary 4.2.22 with $K_n = F^{-1}(y_n), K = \pi_1(F^{-1}(y))$, there exist unique points $(u_n)_n$ and $u$ satisfying equations (4.26) and (4.27) such that $u_n \to u$. By uniqueness of $u_n$ and $\Psi(y_n)$, and uniqueness of $u$ and $\Psi(y)$, we deduce that $u_n = \Psi(y_n)$ for every $n \in \mathbb{N}$ and $u = \Psi(y)$. Thus, we can conclude that $\Psi(y_n) \to \Psi(y)$. This proves that $\Psi$ is continuous.

In order to obtain a neural network that approximates the optimal map to arbitrary, precision we apply the version following theorem from [163]. As the theorem was originally stated for closed cubes, we extend $\Psi$ continuously by Theorem 4.1 [59], from the convex set $\mathcal{M}_2^{\mathcal{E}}$ to a closed cube. Then, we apply Theorem 1 [163], stated in Proposition 4.2.25, and then restrict the neural network to $\mathcal{M}_2^{\mathcal{E}}$ again.

To prove (2), set $\epsilon = 0$ and let $y, y' \in \mathcal{M}_2$, note that $\Psi(y) \in \mathcal{N}(A)^\perp$. To prove that $\Psi = A^\dagger_{|\mathcal{M}_2}$, we show that $\Psi$ satisfies the properties of the Moore-Penrose inverse. Let $x \in \mathcal{M}_1$, then

$$A\Psi(Ax) = y = Ax.$$

Then, we have that

$$\|\Psi(y) - \Psi(y')\|_{\ell^2} = \|A^\dagger A(\Psi(y) - \Psi(y'))\|_{\ell^2} \leq \|A^\dagger\|\|A\Psi(y) - A\Psi(y')\|_{\ell^2} = K\|y - y'\|_{\ell^2}.$$

Thus, $\Psi$ is Lipshitz contiuous with $K = \|A^\dagger\|$. □

The following theorem gives necessary and sufficient conditions such that the optimal map with worst-case noise can be approximated by a neural network. Definition 4.2.20 with the activation function, which continuous and not a polynomial, from Theorem 3.1 [148], where $L = 1$, yields the next theorem.

**Theorem 4.2.26** (**Necessary and sufficent conditions for approximability by a neural network of the optimal map**). *Let $A \in \mathbb{R}^{m\times N}$ be linear and bounded, $\mathcal{M}_1 \subseteq \mathbb{R}^N$ be closed and bounded, $\epsilon > 0$, $\mathcal{E} = B_{d_2}(0, \varepsilon)$, $\mathcal{M}_2^{\mathcal{E}} = (A\mathcal{M}_1) + \mathcal{E}$ and $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{R}^N$ be the optimal map with worst-case noise, as in* (4.2.1)*. Moreover, let $d_1, d_2$ be induced by the $\ell^p$-norm for $p \in (1, \infty)$.*

*Then, the following are equivalent:*

*(1) $\Psi$ is continuous.*

*(2) For all $\delta > 0$ there exists a neural network $\phi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ such that*

$$\sup_{y\in\mathcal{M}_2^{\mathcal{E}}} \|\Psi(y) - \phi(y)\|_{\ell^p} \leq \delta.$$

*(3) For all $(y_n)_n \subseteq \mathcal{M}_2^{\mathcal{E}}$ such that $y_n \to y$ in $(\mathcal{M}_2^{\mathcal{E}}, d_2)$,*

$$z_n \to z$$

*in $\mathbb{R}^N, d_1$, where $z_n, z$ are the centers of the smallest enclosing balls of the feasible sets $\pi_1(F^{-1}(y_n))$, respectively $\pi_1(F^{-1}(y))$.*

*Proof of Theorem 4.2.26.* Firstly, we prove that the optimal map is single-valued. For $p \in (1, \infty)$, $\|\cdot\|_{\ell^p}$ is strictly convex and let $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{R}^N$ be the optimal map with worst-case noise. Let $y \in \mathcal{M}_2^{\mathcal{E}}$ and assume for contradiction that $z \neq z' \in \Psi(y)$. Then, for $\lambda \in (0, 1)$, $\sup_{x'\in\pi_1(F^{-1}(y))} \|x' - (\lambda z + (1-\lambda)z')\|_{\ell^p} < \lambda \sup_{x'\in\pi_1(F^{-1}(y))} \|x' - z\|_{\ell^p} + (1 - \lambda) \sup_{x'\in\pi_1(F^{-1}(y))} \|x' - z\|_{\ell^p} = \sup_{x'\in\pi_1(F^{-1}y)} \|x' - z\|_{\ell^p}$. Thus, $(\lambda z + (1-\lambda)z') \in \Psi(y)$ and $z, z' \notin \Psi(y)$ which is a contradiction. Thus, $\Psi$ is single-valued.

Secondly, to show $(2) \to (1)$, assume that there exists a neural network that uniformly approximates $\Psi$. Then, by the uniform limit theorem we have that the optimal map with worst-case noise is continuous. Thirdly, to show $(3) \to (2)$ and $(1) \to (2)$, assume that for all $(y_n)_n \subseteq \mathcal{M}_2^{\mathcal{E}}$ such that $y_n \to y$ in $(\mathcal{M}_2^{\mathcal{E}}, \|\cdot\|_{\ell^p})$,

$$z_n \to z$$

in $\mathbb{R}^N, d_1$, where $z_n, z$ are the centers of the smallest enclosing balls of the feasible sets $\pi_1(F^{-1}(y_n))$, respectively $\pi_1(F^{-1}(y))$. By Theorem, 4.2.30, we have any $z \in \Psi(y)$ let $r := \sup_{x'\in\pi_1(F^{-1}(y))} \|x' - z\|_{\ell^p}$. Then

$$\Psi(y) = \{z \in \mathbb{C}^N : \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_{\ell^p}}(z, r)\}.$$

Thus, as $\Psi$ is single-valued, we have that $\Psi$ is sequentially continuous. By Theorem 3.4.3, Theorem 3.1 [148], as $\mathcal{M}_2^{\mathcal{E}}$ and as $d_1$ is induced by the $\ell^p$-norm, for all $\delta > 0$ there exists a neural network $\phi : \mathcal{M}_2^{\mathcal{E}} \to \mathbb{R}^N$ such that

$$\sup_{y \in \mathcal{M}_2^{\mathcal{E}}} \|\Psi(y) - \phi(y)\|_{\ell^p} \leq \delta.$$

□

### 4.2.5 Discussion of main results and relation to deep learning

The next section provides another brief summary of deep learning methods for solving inverse problems based on the notation and topics considered in this chapter. For a more extensive summary see Chapter 1. Thereafter, the main results of this chapter are discussed. Moreover, the accuracy and stability of DL methods for solving ill-posed inverse problems can be assessed with our main results, as discussed in the following sections.

### 4.2.6 Summary of deep learning in inverse problems

As mentioned, in typical deep learning approaches to inverse problems, a neural network $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ is trained in order to reconstruct points from an unknown set $\mathcal{M}_1$ from measurements obtained as in (4.2). The network belongs to a class $\mathcal{NN}$ of neural networks, for example, specified by a particular architecture. If $\mathcal{M}_1$ is unknown, one only has access to a finite training set $\mathcal{T} \subseteq \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1 \subseteq \mathbb{C}^m \times \mathbb{C}^N$, with $|\mathcal{T}| = n \in \mathbb{N}$. Here, assume that the inputs in the training set $\mathcal{T}$ are exact, i.e. $\pi_2\mathcal{T} \subseteq \mathcal{M}_1$ (where $\pi_2 : \mathbb{C}^m \times \mathbb{C}^N \to \mathbb{C}^N$ is the projection on the second coordinate). This is usually not the case in practice, since most inputs will be perturbed by some approximation error. However, even in the ideal setting with exact inputs, we will show that it is impossible to assess the quality of the reconstruction.

The training procedure, which usually contains a regularisation term, can be formalized in the following way: the goal is to find

$$\Psi \in \operatorname*{argmin}_{\hat{\Psi} \in \mathcal{NN}} \frac{1}{|\mathcal{T}|} \sum_{(y,x) \in \mathcal{T}} \mathcal{L}(x, \hat{\Psi}(y)) + \lambda J(\hat{\Psi}, \mathcal{T}), \tag{4.28}$$

where $\lambda \geq 0$ is a fixed regularisation hyperparameter, $\mathcal{L} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}^+ \cup \{\infty\}$ is a loss function, and $J : \mathcal{NN} \times \mathcal{G} \to \mathbb{R}$ is a regularisation function, where $\mathcal{G} \subseteq \mathbb{C}^m \times \mathbb{C}^N$ is a family of possible training sets. The loss function $\mathcal{L}$ is often chosen to have some structure. A coomon choice is that it is proper, lower semi-continuous and convex in its second argument [179]. An example is the loss function given by the distance induced by the $\ell^2$-norm.

Considering the example loss function induced by the $\ell^2$-norm, training with regularization typically yields a small training error. In particular, this means that $\|x - \Psi(y)\|_{\ell^2}$ is small for all $(y, x) \in \mathcal{T}$. To illustrate this, consider the following. Let $n := |\mathcal{T}| \in \mathbb{N}$ and consider a set of vectorized training images $\{x^j\}_{j=1}^n \subset \mathbb{C}^N$, and noisy measurements $y^j \in \mathbb{C}^m$, $j \in \{1, \dots, n\}$. Then, as mentioned in Chapter 1, the learned neural network is obtained by minimizing the following objective function

$$\Psi \in \underset{\widetilde{\Psi} \in \mathcal{NN}}{\operatorname{argmin}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} \|x^j - \widetilde{\Psi}(y^j)\|_{\ell^2}^2 + \lambda J(\widetilde{\Psi}, y^j), \tag{4.29}$$

using an appropriate optimization algorithm, as in [130, 131]. Usually $\lambda > 0$ is chosen to be small and, thus, (4.29) yields a very small mean-squared-error on $\mathcal{T}$. To verify that the training procedure was successful, the aim is that the MSE is small not only for the images in the training, but also for those belonging to a validation set $\mathcal{V} \subseteq \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$. In other words, the training procedure typically gives a network $\Psi$ that satisfies

$$\|\Psi(y^j) - x^j\|_{\ell^2}^2 \leq \delta_j, \qquad \delta_j \geq 0, \quad j = 1, \dots, n, \tag{4.30}$$

for very small $\delta_j$. If the set $\mathcal{M}_1 \subset \mathbb{C}^N$ is not known, there is no guarantee that training yields an overall small MSE on $\mathcal{M}_1$. Moreover, the worst-case reconstruction error on a validation set may be very large, as the following example shows.

**Remark 4.2.27.** Let $d_1$ and $d_2$ Euclidean metrics on $\mathbb{R}^2$ and $\mathbb{R}$ respectively. Consider a linear inverse problem with, $\epsilon = 0$, $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $\mathcal{M}_1 = \{(\alpha, 0) : \alpha \in [0, 1]\} \bigcup \{(1, \lambda) : \lambda \in [0, 100]\}$ and $\mathcal{T} = \{(1/k, 0) : k \in \{1, ...n\}\} \subseteq \mathcal{M}_1$. Then, training on $\mathcal{T}$ would easily give $\delta_j \approxeq 0$ for all $j \in \{1, ...k\}$. Analytically we can just choose $\phi(A(\alpha, 0)) = (\alpha, 0)$ for all $\alpha \in [0, 1]$. However, $\phi(A(1, \lambda)) = (1, 0)$ and thus $\|\phi(A(1, \lambda)) - (1, \lambda)\|_{\ell^2}^2 = \lambda^2$ for any $\lambda \in [0, 100]$. Hence, on $\mathcal{M}_1$ the worst-case reconstruction error can be $\delta = 10000$.

### 4.2.7 Optimality depends on which reconstruction error is minimised

Given the short summary of deep learning methods for solving inverse problems, the following section aims at relating the accuracy and stability of the decoders obtained from DL methods with our main results, Section 4.2. We *roughly and formally* relate the error obtained on the training set to the general error on $\mathcal{M}_1$. Note that this relation is still an active area of research and we do not aim at providing comprehensive statements. As stated above, the aim of training according to (4.28), is that modulo the regularisation function the *empirical error* $I_n(\Psi_n) = \frac{1}{|\mathcal{T}|} \sum_{(y,x) \in \mathcal{T}} \mathcal{L}(x, \hat{\Psi}(y))$ is minimized. This can be compared to the optimality constant in the following way. Assuming that the training set is sampled according to a distribution $\rho$, and defining the *generalization error* as

$$I(\Psi) = \int_{(y,x) \in \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1} \mathcal{L}(x, \hat{\Psi}(y)) d\mu(y, x),$$

where $\mu$ is some measure, which may depend in various ways on the distribution $\rho$. The empirical and generalization error can be related in statistical learning theory. Specifically, an algorthim is said to generalize if $I_n(\Psi_n) \to I(\Psi)$, as $n \to \infty$. For this to be true in general there must be strong conditions met, some specific examples can be found in [55]. This convergence has been established for the classification problem, see [133]. For inverse problems [140] have established sufficient and necessary contiditions such that $I_n(\Psi_n) \to I(\Psi)$, as $n \to \infty$. Now let's assume that the empirical error converges to the generalization error. In the limit of using a training set $\mathcal{T}$ that approximates $\mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$ in some sense, the question remains what properties the decoder obtaining the minimal generalization error actually has? Does it exist and can it even be approximated by a neural network? Given the definition of the optimality constant, (4.8), it is evident that this is just the minimal generalization error for the loss-function induced by the Hausdorff metric $d_1^H$. In this case, the generalization error is the average accuracy of the reconstruction from (4.2). Thus, the generalization error, and also the reconstruction accuracy, can be bounded from above and below by $(1)$, Theorem 4.2.9. Moreover, the decoder obtaining the minimal generalization error is the optimal map given in (4.11), and, hence has bounded and compact values, by $(3)$, Theorem 4.2.9. The technical issue here is that the optimal map may be set-valued and in order for the integrals to be well-defined, it needs to be measurable. Thus, one has to obtain a mechanism in order to obtain a for example measurable selection. In the proof of Theorem 4.2.9 we applied the Measurable Maximum Theorem, Theorem 18.19 [40], in order to obtain such a measurable selection.

**The key implications of Theorem** 4.2.9 are for once provding fundamental performance and accuracy bounds on the minimal reconstruction error in various seetings, in terms of the kernel size in part $(1)$:

$$\operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p)/2 \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, p) \leq \operatorname{kersize}(F, \mathcal{M}_1, \mathcal{E}, p).$$

For $p = 2$, the optimality constant is the minimal squared reconstruction error. For $p = 1$ and $\mu_1, \nu$ probability measures, the optimality constant is the mean statistical reconstruction error and for $p = 2$ the mean squared error (MSE). Moreover, for $p = \infty$, if the essential supremum coincides with the supremum, the optimality constant is the best worst-case noise reconstruction error. The upper and lower bounds are applicable for any noise model, where $e \in \mathcal{E}$ is distributed according to $\nu$. The only restriction is that $\mu_1 \otimes \nu$ admits a disintegration of measure with respect to $F$; sufficient conditions are stated in Proposition 4.2.18. The second key implication of Theorem 4.2.9, parts $(2), (3)$ is the existence of the optimal map. The optimal map can be interpreted as returning, for every $y \in \mathcal{M}_2^{\mathcal{E}}$:

(1) $p = \infty$**:** $\Psi(y)$, in (4.10), is the centre of the smallest enclosing ball of $\pi_1(F^{-1}(y))$, ignoring $(\mu_1 \otimes \nu)$-null sets,

(2) $p \in [1, \infty)$**:** $\Psi(y)$, in (4.11), is the $p$-th centroid of a random variable $x$ distributed according to $\pi_{1*}(\mu_1 \otimes \nu)^y$.

For example, for $p = 2$, the optimal map yields the average of a random variable distributed according to $\pi_{1*}(\mu_1 \otimes \nu)^y$, as in this case in (4.11) the variance is minimized. For $p = 1$ and $N = 1$, the optimal map yields the median of $\pi_{1*}(\mu_1 \otimes \nu)^y$. If $\mu_1$ and $\nu$ are probability measures, $y$, $x$ and $e$ are random variables and Theorem 4.2.9 can also be applied to the statistical setting for inverse problems (4.2), see [8].

The best worst-case noise error can provide fundamental accuracy bounds for learning, as shown in Corollary 4.2.28. Note that in the following Corollary, $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}}$, is given by $F(y) = Ax + e$ for $y \in \mathcal{M}_2^{\mathcal{E}}$ and fixed $A \in \mathbb{C}^{m\times N}$.

**Corollary 4.2.28.** *Let $A \in \mathbb{C}^{m\times N}$, $\mathcal{M}_1 \subseteq \mathbb{C}^N$ be bounded, let $\epsilon \geq 0$, $\mathcal{E} = \mathcal{B}_{\ell_2}(0, \epsilon)$, $\mathcal{M}_2^{\mathcal{E}} = \mathcal{M}_2 + \mathcal{E}$, $\mathcal{T} \subseteq \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$ such that $|\mathcal{T}| = n \in \mathbb{N}$. Let $\delta \geq 0$ and $\Psi : \mathbb{C}^m \to \mathbb{C}^m$ such that*

$$\|\Psi(y) - x\|_{\ell_2}^2 \leq \delta, \quad \textit{for all} \quad (y, x) \in \mathcal{T}.$$

*Then, if $\delta \leq \operatorname{kersize}(A, \mathcal{M}_1, \mathcal{E}, \infty)/2$,*

$$\sup_{(y,x)\in\mathcal{M}_2^{\mathcal{E}}\times\mathcal{M}_1} \|\Psi(y) - x\|_{\ell_2}^2 \geq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

The above Corollary 4.2.28 is a simple application of Theorem 4.2.9, $(1)$. It states that obtaining a small training error on a finite training set $\mathcal{T}$, yields a lower bound on the worst-case reconstruction error. Concerning the approximability of the optimal map by neural networks, the main difficulty is that it may be set-valued and neural networks usually are continuous single-valued functions. In order to approximate a function by a neural network one can choose a single-valued selection or as in the following theorems, use assumptions such that the optimal map with worst-case noise is single-valued. For instance, Theorem 4.2.21 states that if $d_1, d_2$ are induced by the $\ell^2$-norm and $\mathcal{M}_1 = \mathcal{B}_{d_2}(z, \rho) \subseteq \mathbb{R}^N$ for any $z \in \mathbb{R}^N$, $\rho > 0$, there exists a neural network with bounded width and depth that uniformly approximates optimal map with worst-case noise. Moreover, in the noiseless case the optimal map is Lipshitz continuous. Theorem 4.2.26 states that for any $\ell^p$-norm with $p \in (1, \infty)$, the optimal map with worst-case noise can be uniformly approximated by a neural network if and only if the centers of the smallest enclosing balls of the feasible sets converge for converging measurements.

### 4.2.8 Obstacles to training the optimal map with worst-case noise

As shown in parts $(2), (3)$ of Theorem 4.2.9, the optimal map with worst-case noise may be set-valued. The following theorem establishes an upper bound on the diameter of the values of the optimal map with worst-case noise. Further, in Theorems 4.2.21 and 4.2.26 it is shown that in certain settings the optimal map with worst-case noise can be approximated by a neural network. However, the conditions in these theorems were sufficient for yielding a single-valued optimal map with worst-case noise. The following theorem implies that the

optimal map is not necessarily single-valued. This can lead to instabilities when trying to approximate it by a neural network, which is a single-valued and continuous function. This is the case as one can choose each value of the optimal map with worst-case noise from a set that has a diameter depending on $(F, \mathcal{M}_1)$.

**Remark 4.2.29.** Let $A \in \mathbb{R}^{m\times N}$ be linear and bounded and $\mathcal{M}_1 \subseteq \mathbb{R}^N$ bounded and equipped with the $\ell^\infty$-norm. Let $\epsilon \geq 0$, $\mathcal{E} = \mathcal{B}_{\ell^\infty}(0,\epsilon)$, $\mathcal{M}_2^{\mathcal{E}} = \mathcal{M}_2 + \mathcal{E}$. Let $y \in \mathcal{M}_2^{\mathcal{E}}$. For $k = 1, \dots, N$ let

$$a_k = \min_{x\in\pi_1(F^{-1}(y))} x_k \,, \quad b_k = \max_{x\in\pi_1(F^{-1}(y))} x_k$$

where $x_k$ denotes the $k$-th coordinate of the vector $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$. Thus, the feasible set $\pi_1(F^{-1}(y))$ is bounded by $\pi_1(F^{-1}(y)) \subseteq [a_1, b_1] \times ... \times [a_N, b_N]$, where such hyper-rectangle is the smallest possible that encloses $\pi_1(F^{-1}(y))$. Define

$$d_1 = \max_{k=1,\dots,N}(b_k - a_k), \quad l_1 = \operatorname*{argmax}_{k=1,\dots,N}(b_k - a_k)$$

$$d_{i+1} = \max_{k\in\{1,\dots N\}/\{l_1,\dots l_i\}}(b_k - a_k), \quad l_{i+1} = \operatorname*{argmax}_{k\in\{1,\dots N\}/\{l_1,\dots l_i\}}(b_k - a_k)$$

Then it is immediate to notice that we have $d_N \leq ... \leq d_2 \leq d_1 = \mathrm{diam}(\pi_1(F^{-1}(y)))$.

**Theorem 4.2.30** (Stability of the optimal map with worst-case noise)**.** *Let $A \in \mathbb{C}^{m\times N}$ be linear and bounded and $\mathcal{M}_1 \subseteq \mathbb{C}^N$ be bounded and equipped with a metric $d_1$. Let $\epsilon \geq 0$, $\mathcal{E} = \mathcal{B}_{d_2}(0,\epsilon)$, $\mathcal{M}_2^{\mathcal{E}} = \mathcal{M}_2 + \mathcal{E}$ be equipped with a metric $d_2$. Let $\Psi : \mathcal{M}_2^{\mathcal{E}} \rightrightarrows \mathbb{C}^N$ be the optimal map with worst-case noise. Then, for all $y \in \mathcal{M}_2^{\mathcal{E}}$,*

*(1) For any $z \in \Psi(y)$ with $r_y := \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z)$,*

$$\Psi(y) = \{z \in \mathbb{C}^N : \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{d_1}(z, r_y)\}.$$

*In particular, we have*

$$\mathrm{diam}(\Psi(y)) = \sup_{\substack{z,z'\in\mathbb{C}^N \\ \pi_1(F^{-1}(y))\subseteq\mathcal{B}_{d_1}(z,r_y) \\ \pi_1(F^{-1}(y))\subseteq\mathcal{B}_{d_1}(z',r_y)}} d_1(z, z'),$$

*and*

$$\mathrm{diam}(\Psi(y)) \leq 3\mathrm{diam}(\pi_1(F^{-1}(y))). \tag{4.31}$$

*(2) Furthermore, we have that*

$$1/2\mathrm{diam}(\pi_1(F^{-1}(y))) \leq r_y \leq \mathrm{diam}(\pi_1(F^{-1}(y))).$$

*(3) For $N > 1$, $\mathcal{M}_1 \subseteq \mathbb{R}^N$ be bounded, $A \in \mathbb{R}^{m\times N}$ and $d_1, d_2$ induced by $\ell^\infty$ norm.*

$$\mathrm{diam}(\Psi(y)) = \mathrm{diam}(\pi_1(F^{-1}(y))) - d_N. \tag{4.32}$$

**Remark 4.2.31.** Theorem 4.2.30 relates to the results of [137], where the number of farthest points from a set is investigated. A specific example of part $(1)$, Theorem 4.2.30, is the following. Consider $A \in \mathbb{R}^{1\times 2}$, $A(x_1, x_2) = x_1 + x_2$, $\mathcal{M}_1 = \{(0,1),(1,0)\}$ and equipp $\mathbb{R}^2$ with the $\ell^1$ metric, i.e. $d_1 = \ell^1$. Let $y = 1 \in \mathbb{R}$, $\epsilon > 0$, then $\pi_1(F^{-1}(y)) = \mathcal{M}_1$ and $\Psi(y) = \{(x_1, x_2) : x_1 = x_2, x_1 \in [0,1]\}$, since these are the centers of the smallest balls that enclose $\pi_1(F^{-1}(y))$ (which are balls of radius 1 taken with respect to the $\ell_1$ norm). In this case, $\operatorname{diam}(\Psi(y)) = \operatorname{diam}(\pi_1(F^{-1}(y))) = \operatorname{diam}(\mathcal{M}_1) = 2$.

Theorem 4.2.30 relates to the stability of the optimal map in the following way. The optimal map $\Psi(y)$ for given $y \in \mathcal{M}_2^{\mathcal{E}}$, implies that for $x, x' \in \pi_1(F^{-1}(y))$

$$d_1^H(\Psi(y), x) \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty), \quad d_1^H(\Psi(y), x') \leq c_{\text{opt}}(F, \mathcal{M}_1, \mathcal{E}, \infty).$$

Here, the reconstruction that $\Psi$ gives for $y$ can yield an approximation to different $x, x' \in \pi_1(F^{-1}(y))$ with a similar error. This can be related to the following formal definition of AI hallucinations, as given in Chapter 3: AI hallucinations are *realistically looking artefacts that appear in the reconstruction, which are not present in the ground truth image.*

Based on Chapter 3, formally one could say that decoder $\phi : \mathbb{C}^m \rightrightarrows \mathbb{C}^N$ *hallucinates*, if for given $\delta \geq 0$, $d_1(\phi(y), x') \leq \delta$ where $x' \in \pi_1(F^{-1}(y))$ and $x \in \pi_1(F^{-1}(y))$ is the actual ground truth and $x'$ contains a realistic looking artefact. Related to the optimal map $\Psi$, this can be interpreted as the following. The larger the diameter of $\Psi(y)$ is for given $y$, the more different vectors can be approximated by $\Psi(y)$. By Theorem 4.2.30, this relates to that the larger the feasible set $\pi_1(F^{-1}(y))$ is, $\Psi(y)$ approximates more different vectors with given accuracy.

The above theorem states that even when trying to obtain the best worst-case reconstruction error, depending on $(F, \mathcal{M}_1)$, there may be a large set of values for the set-valued reconstruction which obtains the error in (4.2.1). Thus, as illustrated in the introduction when training a neural network to achieve a small training error for solving an underdetermined system (4.2), the network which is single-valued may yield a variety of outputs given one input. This could possibly explain part of the issue of AI hallucinations occurring in imaging, as mentioned for example in [68,99,138,166]. These hallucinations occur, when there is a very small error in the reconstructed image and there are features apparent that resemble or are similar to real life scenarios, yet standard methods do not contain these features in the reconstructed image. This what Theorem 4.2.30 highlights. The approximate decoder, here the optimal map with worst-case noise, may achieve the best worst-case noise reconstruction error. However, as the underlying problem (4.2) is ill-posed, the optimal map with worst-case noise will be set-valued. Hence, in the case of inverse problems in imaging, there is a wide range of outputs determined by Theorem 4.2.30 and bound from above by (4.31) and these do not necessarily all correspond to realistic images. Yet, if the diameter of the feasible sets $\pi_1(F^{-1}(y))$ is small, the range of possible outputs and, also, hallucinations decreases as well.

Now in order to prove Theorem 4.2.30, we need the following Lemma.

**Lemma 4.2.32.** *Let $(M, d)$ be a metric space, $X, Y \subseteq M$ be bounded and non-empty. Then,*

$$|\textit{diam}(X) - \textit{diam}(Y)| \leq 2d^H(X, Y),$$

*where $d^H$ denotes the Hausdorff distance.*

*Proof of Lemma 4.2.32.* Let $d^H(X, Y) = r$. The $\epsilon$-version of the Hausdorff distance, which is equivalent to the definition 4.2.1, is given by

$$d^H(X, Y) = \inf\{\epsilon \geq 0 : X \subseteq Y^\epsilon \quad \text{and} \quad Y \subseteq X^\epsilon\}.$$

Recall, that $X^\epsilon$ is the $\epsilon$-noisy set in (3.10). Thus, we have that for all $\delta > 0$, for $y, y' \in Y$, $\exists x, x' \in X$ such that,

$$d(y, x) < r + \delta,$$

and

$$d(y', x') < r + \delta.$$

With the above, we get that

$$d(y, y') \leq d(y, x) + d(x, y') \leq d(y, x) + d(x, x') + d(x', y') \leq \text{diam}(X) + 2r + 2\delta.$$

Taking the supremum over all $y, y' \in Y$ yields,

$$\text{diam}(Y) \leq \text{diam}(X) + 2r + 2\delta.$$

Letting $\delta \to 0$ yields the first inequality. The second one is obtained analogously. □

With Lemma 4.2.32 we can now prove Theorem 4.2.30.

*Proof of Theorem 4.2.30.* **Part (1)**; first of all, let us note that $r_y$ is well defined. Indeed, from the definition of $\Psi(y)$ in (4.6), if $z, z' \in \Psi(y)$ are both minimisers, then

$$\sup_{x' \in \pi_1(F^{-1}(y))} d_1(x', z) = \sup_{x' \in \pi_1(F^{-1}(y))} d_1(x', z').$$

Hence $r_y$ is well defined. Now, for $z \in \Psi(y)$, we have for all $x' \in \pi_1(F^{-1}(y))$ that $d_1(x', z) \leq \sup_{x' \in \pi_1(F^{-1}(y))} d_1(x', z) = r_y$. Thus, $\pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{d_1}(z, r_y)$. For the reverse, let $z \in \mathbb{C}^N$ such that $\pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{d_1}(z, r_y)$. Then, we have that for any $x' \in \pi_1(F^{-1}(y))$,

$$d_1(x', z) \leq r_y = \sup_{x' \in \pi_1(F^{-1}(y))} d_1(x', z').$$

Taking the supremum over all $x' \in \pi_1(F^{-1}(y))$ and by definition of $r_y$, we get that $z \in \Psi(y)$. The optimal map with worst-case noise is given by,

$$\Psi(y) = \hat{X} = \text{argmin}_{z\in\mathbb{C}^N} \sup_{x'\in\pi_1(F^{-1}(y))} d_1(z, x'),$$

where $y = Ax + e$ for $x \in \mathcal{M}_1$ and $e \in \mathbb{C}^m$ with $d_2(Ax + e, Ax) \leq \epsilon$ and $\pi_1(F^{-1}(y)) = \{x' \in \mathcal{M}_1 : \exists e' \in \mathbb{C}^m, d_2(Ax' + e', Ax') \leq \epsilon, Ax' + e' = y\}$.For $z \in \Psi(y)$ and for all $z' \in \mathbb{C}^N$, we have that

$$\sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z) \leq \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z').$$

This gives for all $x' \in \pi_1(F^{-1}(y))$ and $z \in \Psi(y)$, that

$$d_1(x', z) \leq \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z') \leq \sup_{x',z'\in\pi_1(F^{-1}(y))} d_1(x', z') = \text{diam}(\pi_1(F^{-1}(y))).$$

Taking the supremum and infimum in alternating orders over $x' \in \pi_1(F^{-1}(y))$ and $z \in \Psi(y)$ yields,

$$d_1^H(\pi_1(F^{-1}(y)), \Psi(y)) \leq \text{diam}(\pi_1(F^{-1}(y))).$$

Then, with the property of the Hausdorff metric, Lemma 4.2.32, we get that

$$\frac{1}{2}|\text{diam}(\pi_1(F^{-1}(y))) - \text{diam}(\Psi(y))| \leq \text{diam}_{d_1}(\pi_1(F^{-1}(y))).$$

This gives the upper bound in (4.31),

$$\text{diam}(\Psi(y)) \leq 3\text{diam}(\pi_1(F^{-1}(y))). \tag{4.33}$$

**Part (2)**; let $z \in \Psi(y)$, then by the triangle inequality, we have

$$\sup_{x',x''\in\pi_1(F^{-1}(y))} d_1(x', x'') \leq \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z) + \sup_{x''\in\pi_1(F^{-1}(y))} d_1(x'', z) = 2r.$$

Thus, we get that $1/2\text{diam}(\pi_1(F^{-1}(y))) \leq r$. For the upper bound let $z \in \Psi(y)$ and note that for all $q \in \mathbb{C}^N$,

$$r_y = \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z) \leq \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', q),$$

from the definition (4.6) of $\Psi(y)$. Thus taking the supremum over all $q \in \pi_1(F^{-1}(y))$ yields,

$$r_y = \sup_{x'\in\pi_1(F^{-1}(y))} d_1(x', z) \leq \text{diam}(\pi_1(F^{-1}(y))).$$

**Part (3)**; note that, by definition of $d_i$ we have that

$$d_1 \geq d_2 \geq ... \geq d_N.$$

By $(1)$, Theorem 4.2.30, we have that $\operatorname{diam}(\Psi(y)) = \sup_{\substack{z,z' \in \mathbb{C}^N \\ \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_\infty}(z, r_y) \\ \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_\infty}(z', r_y)}} \|z - z'\|_\infty$.

We have that $r_y = \operatorname{diam}(\pi_1(F^{-1}(y)))/2$ and, thus,

$$r_y = \operatorname{diam}(\pi_1(F^{-1}(y)))/2 = (b_{l_1} - a_{l_1})/2 = d_1/2.$$

Denote the closure of $\pi_1(F^{-1}(y))$ with repsect to the $\ell^\infty$-norm by $\bar{\pi}_1(F^{-1}(y))$. Now, we have as $\pi_1(F^{-1}(y))$ is bounded that for every $i = \{2, ...N\}$ there exist $x^i, x'^i \in \bar{\pi}_1(F^{-1}(y))$ such that

$$d_i = \sup_{x,x' \in \pi_1(F^{-1}(y))} \sup_{k \in \{1,...N\}/\{l_1,...l_{i-1}\}} |x_k - x'_k| = b_{l_i} - a_{l_i}.$$

Notice that $b_{l_i} \geq a_{l_i}$ for all $i \in \{1, ...N\}$. Then, for all $z \in \mathbb{R}^N$ such that

$$z_{l_1} = 1/2(a_{l_1} + b_{l_1})$$

and for all $i \in \{2, ...N\}$,

$$z_{l_i} \in [b_{l_i} - d_1/2, a_{l_i} + d_1/2],$$

we claim that

$$\pi_1(F^{-1}(y)) \subseteq B_{\|\cdot\|_\infty}(z, r_y).$$

Furthermore, these are the smallest enclosing balls of $\pi_1(F^{-1}(y))$ by the lower bound in $(2)$, Theorem 4.2.30. Let $x \in \pi_1(F^{-1}(y))$, then we have that

$$\|x - z\|_\infty \leq r_y.$$

Let $k \in \{1, ...N\}$ then we have for $k = l_1$ that,

$$|x_{l_1} - z_{l_1}| = |x_{l_1} - 1/2(b_l + a_l)| \leq \max_{x_{l_1} \in [a_l, b_l]} |x_{l_1} - 1/2(b_{l_1} + a_{l_1})| = 1/2|b_{l_1} - a_{l_1}| = r.$$

For $k \in \{1, ...N\}$ such that $k = l_i$ with $i \in \{2, ..N\}$, we have that

$$|x_{l_i} - z_{l_i}| \leq \max_{x_{l_i} \in [a_{l_i}, b_{l_i}]} \max_{z_{l_i} \in [b_{l_i} - d_1/2, a_{l_i} + d_1/2]} |x_{l_i} - z_{l_i}| \leq r.$$

As for all $z_{l_i} \in [b_{l_i} - d_1/2, a_{l_i} + d_1/2]$ we have that for all $x_{l_i} \in [a_{l_i}, b_{l_i}]$ that,

$$-d_1/2 = b_{l_i} - d_1/2 - b_{l_i} \leq z_{l_i} - x_{l_i} \leq (a_{l_i} - b_{l_i}) + d_1/2 \leq d_1/2,$$

as $a_{l_i} - b_{l_i} \leq 0$. Thus, we get that

$$\pi_1(F^{-1}(y)) \subseteq B_{\|\cdot\|_\infty}(z, r_y).$$

With $(1)$, Theorem 4.2.30, we have that

$$\begin{aligned} \operatorname{diam}(\Psi(y)) &= \sup_{\substack{z,z' \in \mathbb{C}^N \\ \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_\infty}(z, r_y) \\ \pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_\infty}(z', r_y)}} \|z - z'\|_\infty \\ &= \operatorname{diam}(\pi_1(F^{-1}(y))) - d_N. \end{aligned}$$

This follows by the definition of $z$ such that $\pi_1(F^{-1}(y)) \subseteq \mathcal{B}_{\|\cdot\|_\infty}(z, r_y)$ and as $d_1 \geq d_2 \geq ... \geq d_N$. □

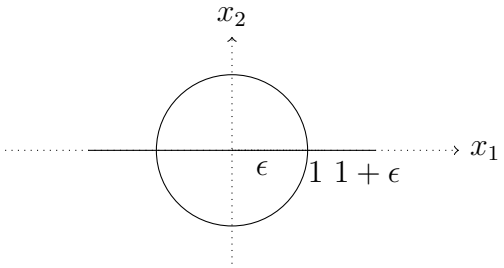


Figure 4.2: [**An example setting.**] For clarity, we consider $\mathbb{R}^N$ and $\mathbb{R}^m$ instead of their complex counterparts. For $N = 2$, $m = 1$, $d_1$ and $d_2$ Euclidean metrics on $\mathbb{R}^2$ and $\mathbb{R}$ respectively, $\epsilon \geq 0$ and $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Consider $\mathcal{M}_1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$, the ball around zero (full lines). Identifying $\mathbb{R}^1 = \mathbb{R} \times \{0\} \subseteq \mathbb{R}^2$, we have $\mathcal{M}_2 = A\mathcal{M}_1 = [-1, 1] \times \{0\}$ and $\mathcal{M}_2^{\mathcal{E}} = [-1 - \epsilon, 1 + \epsilon] \times \{0\}$ (represented in $\mathbb{R}^2$ by the full line). As a direct consequence of Theorem 4.2.9, the optimal map with worst-case noise is $\phi(y_1) = (y_1 - \epsilon, 0)$ for $y_1 > \epsilon$ and $\phi(y_1) = (y_1 + \epsilon, 0)$ for $y_1 < -\epsilon$, $\phi(y_1) = (0, 0)$ for $-\epsilon \leq y_1 \leq \epsilon$ and $c_{\text{opt}}(A, \mathcal{M}_1, \mathcal{B}_{d_2}(0, \epsilon), \infty) = 1$.

## 4.3 Theoretical background of the optimality constant

In the following sections we consider the optimal map with worst-case noise. Hence, the underlying assumption is that the essential supremum is equivalent to the supremum. Moreover, the discussion is reduced to linear sampling operators $A \in \mathbb{C}^{m \times N}$ and additive bounded noise $e \in \mathcal{B}_{d_2}(0, \epsilon)$, whenever noise is considered. In particular, $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2^{\mathcal{E}}$, is given by $F(y) = Ax + e$ for $y \in \mathcal{M}_2^{\mathcal{E}}$ and, hence, the problem (4.1) is considered. Moreover, if $\mathcal{E} = \{0\}$ the optimality constant will be denoted by $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty)$, as in this case $F = A$.

### 4.3.1 Comparison of the optimality constant to approximation theory and $n$-widths

In the noiseless case, the optimality constant, $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty)$, can be related to well-known quantities from approximation theory. As most results concerning $n$-widths are obtained for $\mathbb{R}$, in the following only real vector spaces are regarded. Firstly, under specific conditions we can bound the optimality constant from below by quantities in non-linear approximation theory, such as the continuous non-linear $m$-width and the $m$-Bernstein width. Recall the definition of the best approximation error as given in [44].

**Definition 4.3.1** (Best approximation error)**.** Let $K \subset \mathbb{R}^N$, where $\mathbb{R}^N$ is equipped with a norm $\|\cdot\|$, $A \in \mathbb{R}^{m \times N}$, $\varphi : \mathbb{R}^m \to \mathbb{R}^N$ be continuous. Then, we have the best worst-case approximation error for $K$ depending on a non-linear approximation $\varphi$,

$$E(K, A, \varphi) = \sup_{x \in K} \|x - \varphi(Ax)\|. \tag{4.34}$$

The best worst-case approximation error is the worst-case reconstruction error that is obtained by a given continuous reconstruction map $\varphi$, given linear sampling operator $A$ and a set $K$. For $K \subset X$, the best possible linear encoder and continuous decoder pair can achieve the following worst-case error bound, also referred to as continuous non-linear $m$-width.

**Definition 4.3.2** ($m$-widths)**.** Let $K \subset \mathbb{R}^N$, where $\mathbb{R}^N$ is equipped with a norm $\|\cdot\|$.

(1) The best non-linear approximation error or also the continuous non-linear $m$-width, is given by

$$d_m(K)_X = \inf_{\varphi, A} E(K, A, \varphi)_X,$$

where the infimum is taken over all $A \in \mathbb{R}^{m\times N}$ and continuous $\varphi : \mathbb{R}^m \to \mathbb{R}^N$.

(2) Define $U(K) = \{x \in K : \|x\| \leq 1\}$ to be the unit ball in $K$. The $m$-Bernstein width is given by,

$$b_m(K)_X = \sup_{\substack{X_{m+1}\subseteq X \\ \operatorname{codim}(X_{m+1})=m+1}} \sup_{\substack{\rho>0 \\ \rho U(X_{m+1})\subset K}} \rho,$$

where the second supremum is taken over all $(m+1)$-dimensional linear subspaces of $X$.

Secondly, under specific choices of $\mathcal{M}_1$ and $A \in \mathbb{R}^{m\times N}$, we can prove that the well-known Gelfand width is a lower bound for the optimality constant $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty)$. Recall the definition of the Gelfand width

**Definition 4.3.3** (Gelfand width)**.** For a normed linear space $X$ and a subset $K \subset X$, the Gelfand width is defined as

$$d^m(K)_X := \inf_{\substack{Y\subset X \\ \operatorname{codim}(Y)\leq m}} \sup\{\|x\| : x \in K \cap Y\},$$

where the infimum is taken over all subspaces $Y$ of $X$ of codimension less or equal to $m$.

Adapting the proof of Lemma 2.1 [44], under specific assumptions we can prove that the Gelfand width and the Bernstein and continuous non-linear $m$-width are lower bounds for $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty)$, (4.2.1).

**Lemma 4.3.4.** *Let $m \leq N, m, N \in \mathbb{N}$, $A \in \mathbb{R}^{m\times N}$ and $\mathcal{N}(A)$ denote the null space of $A$, $\mathcal{M}_1 \subset X = \mathbb{R}^N$ be bounded and $\mathcal{M}_2 = A(\mathcal{M}_1)$.*

*(1) If $\mathcal{M}_1 = -\mathcal{M}_1$, we have*

$$c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) \geq d^m(\mathcal{M}_1)_{\mathbb{R}^N}.$$

*(2) If $\mathcal{C}$ is restricted to continuous maps $\varphi : \mathcal{M}_2 \to \mathbb{C}^N$, then*

$$c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) \geq d_m(\mathcal{M}_1)_{\mathbb{R}^N} \geq b_m(\mathcal{M}_1)_{\mathbb{R}^N}.$$

*Proof of Lemma 4.3.4.* **Part (1)**; note that the null space $Y = \mathcal{N}(A)$ of $A$ is of codimension less than or equal to $m$. Conversely, given any space $Y \subseteq \mathbb{R}^N$ of codimension $m$, we can associate its orthogonal complement $Y^\perp$, which is of dimension $m$, and the $m \times N$ matrix $A$ whose rows are formed by any basis for $Y^\perp$. Through this identification, we see that the Gelfand width is given by

$$d^m(\mathcal{M}_1) = \inf_{A \in \mathbb{R}^{m \times N}} \sup_{\eta \in \mathcal{N}(A) \cap \mathcal{M}_1} \|\eta\|.$$

Now, if $(A, \phi)$ is any encoder-decoder pair, i.e. $A \in \mathbb{R}^{m \times N}$ and $\phi : \mathcal{M}_2 \rightrightarrows \mathbb{R}^N$, and let $z \in \phi(0)$. For any $\eta \in \mathcal{N}(A)$ we also have $-\eta \in \mathcal{N}(A)$. With this it follows that either $\|\eta - z\| \geq \|\eta\|$ or $\| - \eta - z\| \geq \|\eta\|$. Since $\mathcal{M}_1 = -\mathcal{M}_1$, we conclude that

$$d^m(\mathcal{M}_1) \leq \sup_{\eta \in \mathcal{N}(A) \cap \mathcal{M}_1} \|\eta - \phi(A\eta)\|.$$

Now this can be bounded from above,

$$\sup_{\eta \in \mathcal{N}(A) \cap \mathcal{M}_1} \|\eta - \phi(A\eta)\| \leq \sup_{\eta \in \mathcal{M}_1} \|\eta - \phi(A\eta)\|.$$

Taking an infimum over all $\phi : \mathbb{R}^m \rightrightarrows \mathbb{R}^N$, we obtain

$$d^m(\mathcal{M}_1) \leq c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty).$$

**Part (2)**; in the case that $\mathcal{M}_2 \subset \mathbb{R}^m$, the optimal map constant, restricted to continuous maps $\varphi : \mathcal{M}_2 \to \mathbb{R}^N$, can be bounded from below by the continuous non-linear $m$-width. Furthermore, by Theorem 3.1 [56], it can be bounded from below by the Bernstein $m$-width.

$$\begin{aligned} c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) &= \inf_{\varphi : \mathcal{M}_2 \to \mathbb{R}^N} \sup_{x \in \mathcal{M}_1} \|\varphi(Ax) - x\| \\ &\geq \inf_{\varphi : \mathbb{R}^m \to \mathbb{R}^N, A \in \mathbb{R}^{m \times N}} \sup_{x \in \mathcal{M}_1} \|\varphi(Ax) - x\| = d_m(\mathcal{M}_1). \end{aligned}$$

Then, we use the following theorem,

**Theorem 4.3.5** (Theorem 3.1 [56])**.** *For a normed space $X$ and $K \subset X$ we have*

$$d_m(K) \geq b_m(K).$$

This directly gives the second lower bound in $(2)$. □

**Remark 4.3.6** (The optimal map constant is not necessarily zero)**.** In the following example, let $r > 0$ and we have that

$$c_{\mathrm{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = r/2 > 0.$$

This is illustrated in Fig. 2. Let $\|\cdot\|_{\ell^2}$ denote the $\ell^2$-norm on $\mathbb{C}^N$, $\mathbb{C}^m$ respectively and let $d_1$ be induced by the $\ell^2$-norm. Let $r > 0$,

$$\mathcal{M}_1 := B_r^N(0) = \{x \in \mathbb{C}^N : \|x\|_{\ell^2} \leq r\} \subset \mathbb{C}^N,$$

$$\mathcal{M}_2 := B_r^m(0) \times \{(0, ..0)\} = \{x \in \mathbb{C}^m : \|x\|_{\ell^2} \leq r\} \times \{(0, ..0)\} \subset \mathbb{C}^N,$$

and $A : \mathcal{M}_1 \to \mathcal{M}_2$ such that $A = P_m U$ where $P_m : \mathbb{C}^N \to \mathbb{C}^m$ projects onto the first $m$ components and $U$ is unitary. Then, define $A^{-1} : \mathcal{M}_2 \to \mathcal{M}_1$ by the relevant restriction of the Moore Penrose inverse of $A$. Then, we have that

$$c_{\mathrm{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = \inf_{\varphi : \mathcal{M}_2 \overset{c}{\rightrightarrows} \mathbb{C}^N} \sup_{x \in \mathcal{M}_1} d_1^H(\varphi(Ax), x) = r,$$

as in 4.2.2. Furthermore, the infimum is attained by the set-valued map $\varphi^0 = A^{-1}$. In this example, the upper and lower bounds in part $(1)$, Theorem 4.2.9 are illustrated. We have that $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{(0, ...0)\} \times \{x \in \mathbb{C}^{N-m} : \|x\|_{\ell^2} \leq 2r\} \subset \mathbb{C}^N$, and that $\mathrm{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = \frac{1}{2}\mathrm{diam}((\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A)) = 2r$ by Proposition 4.4.7.

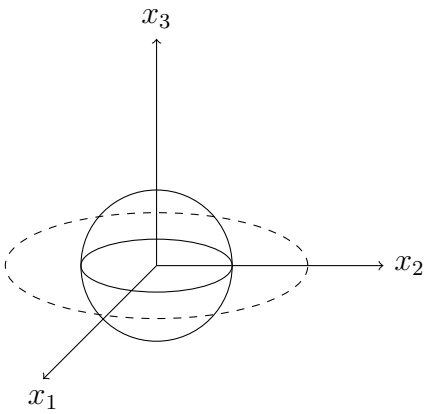


Figure 4.3: Let $N = 3$, $m = 1$ and $A = P_m U$ with $U$ the identity matrix in $\mathbb{R}^3$ and $P_m$ being the projection onto $x_3$. Drawing of the sphere $\mathcal{M}_1 = \mathcal{B}_{\|\cdot\|_{\ell^2}}(0, 1) \subseteq \mathbb{R}^3$ and $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 2\}$, dashed line, in $\mathbb{R}^3$.

## 4.4 Conditions for accurate and stable recovery

The optimality constant framework can be related to the work from Bourrier et al. [21]. Bourrier et al. introduced general conditions for accurate and stable recovery of decoders in [21]. Note that remarkably in [112] similar conditions have been established for non-linear

ill-posed problems. We will summarise the main results from [21] in the following and show how these are closely related to the optimality constant framework. Firstly, recall two definitions from [21]. The aim is to measure the performance of a decoder, $\Phi_\delta \colon \mathbb{C}^m \to \mathbb{C}^N$, by the deviation from the idealized model set $\mathcal{M}_1$. This is measured with the distance of a point $x \in \mathbb{C}^N$ to a set $X \subseteq \mathbb{C}^N$ defined by, $d_1(x, X) = \inf_{x' \in X} d_1(x, x')$. The same definition can also be applied to $\mathbb{C}^m$ with $d_2$ instead of $d_1$. The same definition can be applied to metrics induced by norms or pseudo-norms.

**Definition 4.4.1 (robust instance optimality (rIOP)).** Let $A \in \mathbb{C}^{m\times N}$, $\mathcal{M}_1 \subset \mathbb{C}^N$ and let $\delta \geq 0$. Let $\|\cdot\|_1$ and $\|\cdot\|_3$ be pseudo-norms on $\mathbb{C}^N$ and we let $\|\cdot\|_2$ be a pseudo-norm on $\mathbb{C}^m$. Let $d_i$ denote the pseudo-metric induced by $\|\cdot\|_i$ for $i = 1, 2, 3$. We call a decoder $\Phi_\delta \colon \mathbb{C}^m \to \mathbb{C}^N$, robustly instance optimal, if it fulfills the following property:

$$\|x - \Phi_\delta(Ax+e)\|_1 \leq C_1 d_3(x, \mathcal{M}_1) + C_2\|e\|_2 + \delta, \quad \text{for all } x \in \mathbb{C}^N \text{ and all } e \in \mathbb{C}^m. \tag{4.35}$$

for some constants $C_1, C_2 > 0$.

Assuming that a decoder $\Phi_\delta$ satisfies (4.35) means, that if $x \in \mathcal{M}_1$ or $x$ "lies close" to $\mathcal{M}_1$, then the decoder $\Phi_\delta$ will accurately recover $x$ depending on the noise level. The extra term $\delta \geq 0$, is present, as there are cases when there does not exist $z \in \{x\} + \mathcal{N}(A) = \{x + w : w \in \mathcal{N}(A)\}$, such that $d_3(x, z) = d_3(x, \mathcal{M}_1)$. Following the tradition from [21,44] we call a pair $(A, \Phi_\delta)$ satisfying (4.35) *robustly instance optimal* (rIOP) with constants $C_1, C_2 > 0$ and $\delta \geq 0$. A necessary, and almost sufficient, condition for the existence of a robust instance optimal decoder $\Phi_\delta$, is that $A$ satisfies the so called *robust null space property*.

**Remark 4.4.2.** In the seminal work by Cohen, Dahmen and DeVore [44], they introduced the notion of instance optimality for $\mathcal{M}_1 = \{x \in \mathbb{C}^N : x \text{ is } s\text{-sparse}\}$ being the set of $s$-sparse vectors. They derived necessary conditions for accurate recovery of $s$-sparse vectors. Later this was extended to the case where $\mathcal{M}_1$ is a union of subspaces in [146]. Finally, in [21] Bourrier et al. extended this to general sets $\mathcal{M}_1$, and pseudo-norms.

**Definition 4.4.3 (Robust null space property (rNSP)).** Let $A \in \mathbb{C}^{m\times N}$, $\mathcal{M}_1 \subset \mathbb{C}^N$. We say that the matrix $A$, satisfies the *robust null space property* (rNSP) with constants $D_1 > 0$ and $D_2 > 0$, if

$$\|x\|_1 \leq D_1 d_3(x, \mathcal{M}_1 - \mathcal{M}_1) + D_2\|Ax\|_2, \quad \text{for all } x \in \mathbb{C}^N.$$

**Remark 4.4.4** (Relation to the null space property in sparse regularization)**.** Notice that if $x \in \mathcal{N}(A)$, this reduces to the usual null space property as in [21]. For example, when $\mathcal{M}_1 = \Sigma_s = \{x \in \mathbb{C}^N : x \text{ is } s\text{-sparse}\}$, $C = 2$, and $\|\cdot\|_1 = \|\cdot\|_3$ is the $\ell^1$ norm, this condition reduces to

$$\|h\|_{\ell^1} \leq 2d_1(h, \mathcal{M}_1 - \mathcal{M}_1) := 2\sigma_{2s}(h)_1 \quad \text{for all } h \in \mathcal{N}(A).$$

In particular, since $\sigma_{2s}(x)_1 \leq \sigma_s(x)_1$, this implies the standard null space property we know from sparse regularization with an $\ell^1$-decoder.

In [21] the rIOP and rNSP are proven to be almost equivalent. Theorem 3 [21] shows that the robust null space property is a necessary condition for robust instance optimality. The converse is shown in Theorem 4 [21], namely that the robust null space property implies the existence of a robust instance optimal decoder, but with different constants.

**Theorem 4.4.5** (Theorems 3 and 4 [21])**.** *Let $A \in \mathbb{C}^{m\times N}$ and $\mathcal{M}_1 \subset \mathbb{C}^N$.*

*(1) Suppose that for all $\delta > 0$ there exists a decoder $\Psi_\delta \colon \mathbb{C}^m \to \mathbb{C}^N$ satisfying the rIOP, then $A$ stratifies the rNSP with constants $D_1 = C_1$ and $D_2 = C_2$.*

*(2) Suppose $A$ stratifies the rNSP, then for all $\delta > 0$ there exists a decoder $\Psi_\delta \colon \mathbb{C}^m \to \mathbb{C}^N$ satisfying the rIOP with constants $C_1 = 2D_1$ and $C_2 = 2D_2$.*

**Remark 4.4.6.** Initially the decoder defined in the proof of Theorem 3 [21] is set valued and, then, by the axiom of choice the authors choose a single valued decoder $\Phi_\delta$ satisfying the same properties.

### 4.4.1 Zero kernel size is a necessary condition for the rNSP

The kernel size of $(A, \mathcal{M}_1)$, is a generalization of a condition commonly applied in inverse problems, as (4.1), to guarantee exact reconstruction.

Recall, that for $\mathcal{E} = \{0\}$ the kernel size is given by,

$$\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = \sup_{\substack{x,x'\in\mathcal{M}_1 \\ Ax=Ax'}} d_1(x, x').$$

The kernel size for $\mathcal{E} = \{0\}$ has the following properties.

**Proposition 4.4.7 (Properties of the kernel size).** *Let $A \in \mathbb{C}^{m\times N}$, $\mathcal{M}_1 \subseteq (\mathbb{C}^N, d_1)$.*

*(1) $A$ is injective on $\mathcal{M}_1$ if and only if* $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0$.

*(2) If $d_1$ is translation-invariant, then*

$$\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = d_1^H(0, \mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1))$$

*(3) If $d_1$ is translation-invariant and satisfies $d_1(2x, 0) = 2d_1(x, 0)$ for every $x \in \mathbb{C}^N$, then*

$$\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = \frac{1}{2} \operatorname{diam}\big(\mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1)\big)$$

*Proof of proposition 4.4.7.* (1)

$$\begin{aligned} &\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0 \\ \iff &\sup\{d_1(x, x') : x, x' \in \mathcal{M}_1, Ax = Ax'\} = 0 \\ \iff &d_1(x, x') = 0 \ \forall x, x' \in \mathcal{M}_1 \end{aligned}$$

s.t. $Ax = Ax' \iff x = x' \ \forall x, x' \in \mathcal{M}_1$ s.t. $Ax = Ax' \iff A|_{\mathcal{M}_1}$ is injective.

(2) $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = \sup\{d_1(x, x') : x, x \in \mathcal{M}_1, Ax = Ax'\} = \sup\{d_1(0, x' - x) : x, x' \in \mathcal{M}_1, A(x' - x) = 0\} = d^H(0, \mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1))$.

(3) The set $\mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1)$ is symmetrical around the origin. For any $S \subseteq \mathbb{C}^N$ that is symmetrical, then $d_1^H(0, S) = \frac{1}{2}\operatorname{diam}(S)$, or equivalently $\operatorname{diam}(S) = 2d^H(0, S)$. In fact

(i) $s, s' \in S \implies d_1(s, s') \leq d_1(s, 0) + d_1(0, s') \leq 2d_1^H(0, S)$, take the supremum over $s, s' \in S$ to obtain $\operatorname{diam}(S) \leq 2d_1^H(0, S)$;

(ii) If $s \in S$, then $\operatorname{diam}(S) \geq d_1(s, -s) = d_1(0, 2s) = 2d_1(0, s)$. Take the supremum over $s \in S$ to obtain $\operatorname{diam}(S) \geq 2d_1^H(0, S)$.

□

For example, as introduced in [44]: Let $\Sigma_{2s} \subseteq \mathbb{C}^N$ be the set of $2s$-sparse vectors and $A \in \mathbb{R}^{m \times N}$. The condition $\Sigma_{2s} \cap \mathcal{N}(A) = \{0\}$ is equivalent to guaranteeing that there exists a decoder such that for all $x \in \Sigma_s$,

$$\phi(Ax) = x,$$

see Lemma 3.1 [44]. Note that this condition is used in many theoretical results in order to obtain recovery guarantees, such as in [38]. The condition $\Sigma_{2s} \cap \mathcal{N}(A) = \{0\}$ is equivalent to the condition $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$ with $\mathcal{M}_1 = \Sigma_s$, since $\Sigma_s - \Sigma_s = \Sigma_{2s}$. By Proposition 4.4.7, this is equivalent to $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0$. Furthermore, even when not aiming for exact recovery, but a stable and robust decoder, a necessary and sufficient condition is the robust null space property. However, this implies that $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$, as shown in Remark 4.4.9. Yet, as we do not aim for exact recovery, but the best recovery possible, we choose an arbitrary set $\mathcal{M}_1 \subseteq \mathbb{C}^N$ as opposed to sparse vectors. This is possible by generalizing this condition to the condition of the kernel size of $(A, \mathcal{M}_1)$.

**Definition 4.4.8** (Lower Lipshitz continuity)**.** Let $\mathcal{M}_1 \subset \mathbb{C}^N$ and $A \in \mathbb{C}^{m \times N}$. $(A|_{\mathcal{M}_1})^{-1}$ is lower Lipshitz continuous, if there exists $K \geq 0$ such that

$$d_1(x, x') \leq K d_2(Ax, Ax'), \tag{4.36}$$

for every $x, x' \in \mathcal{M}_1$. The smallest parameter $K$ for which (4.36) holds is called the lower Lipshitz constant of $(A|_{\mathcal{M}_1})^{-1}$.

Now, in order to show that lower Lipshitz continuity and the kernel size are more general than the rNSP and compare these, recall the definition of the rNSP generalized to metrics from [21].

**Remark 4.4.9** (Necessary conditions for the rNSP)**.** We have that $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0$ and lower Lipshitz continuity are necessary conditions for the rNSP. In particular, the

rNSP implies that $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0$. To see this assume that the rNSP holds. Then, by the definition of the kernel size and using the rNSP we get $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = \sup_{x \in \mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1)} \|x\|_1 \leq \sup_{x \in \mathcal{N}(A) \cap (\mathcal{M}_1 - \mathcal{M}_1)} D_1 d_3(x, \mathcal{M}_1 - \mathcal{M}_1) + D_2 \|Ax\|_2 = 0$. Thus, we have that $\operatorname{kersize}(A, \mathcal{M}_1, \{0\}, \infty) = 0$. Let $x, x' \in \mathcal{M}_1$, then, by the rNSP we have a constant $D_2 > 0$ such that $\|x - x'\|_1 \leq D_2 \|Ax - Ax'\|_2$. Thus, the $(A|_{\mathcal{M}_1})^{-1}$ is lower Lipshitz continuous with constant $D_2 > 0$.

Opposed to the RIP or the rNSP the kernel size allows for $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) \neq \{0\}$. The latter assumption is indeed the case in many practical settings, especially in undersampled acquisitions and/or when considering large sets $\mathcal{M}_1$. Furthermore, note that $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) \neq \{0\}$ can also easily occur, if the set $\mathcal{M}_1$ is not known. When applying the optimal map framework to deep learning yielding an approximate decoder for (4.2), a key point is to account for the assumption that $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) \neq \{0\}$. This is because, usually a neural network is trained to achieve a minimal training error on some training set $\mathcal{T} \subseteq \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$, i.e. it can be seen as trying to approximate an optimal map with noise. Yet, in practice it is hard to guarantee that the condition $(\mathcal{T}_1 - \mathcal{T}_1) \cap \mathcal{N}(A) = \{0\}$ or even $(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\}$ is satisfied. Here $\mathcal{T}_1$ is the projection onto the second component of the training set $\mathcal{T} \subset \mathcal{M}_2^{\mathcal{E}} \times \mathcal{M}_1$. For example, this can be the case if the kernel of the sampling operator $A$ or the set $\mathcal{M}_1$ is not explicitly known.

**Proposition 4.4.10** (Optimal map with the lower Lipshitz continuity)**.** *Let $A \in \mathbb{C}^{m \times N}$ and $\mathcal{M}_1 \subseteq \mathbb{C}^N$ be bounded. Let $d_2$ be translation invariant. Then, the following holds: $(A, \mathcal{M}_1)$ is lower Lipshitz continuous with constant $K > 0$ if and only if $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = 0$ and the optimal map is Lipshitz continuous with constant $K > 0$.*

The proof of Proposition 4.4.10, is an application of part $(1)$, Theorem 4.2.3, and the definition of lower Lipshitz continuity. It relates the necessary conditions, zero kernel size and lower Lipshitz continuity, of the rNSP to the optimal map. An extension of this Proposition to the optimal map with worst-case noise is provided by Theorem 4.4.15.

### 4.4.2 On the relation between robust instance optimality and the optimal map

The rIOP (4.35) is a property that depends on the parameter $\delta$. In the previous section, we assumed that the rIOP holds *for every* $\delta > 0$ and this is equivalent to the rNSP. What can be said if the rIOP holds only for *one particular value* of $\delta$? In this case, one can prove that an approximation of the rNSP holds, which we will call $\delta$-'Approximate Robust Null Space Property'.

**Definition 4.4.11** ($\delta$-Approximate Robust Null Space Property)**.** Let $\delta \geq 0$ and $d_i$ for $i = 1, 2, 3$ denote metrics incuded by pseudo-norms. We say that $A \in \mathbb{C}^{m \times N}$ satisfies the

$\delta$-Approximate Robust Null Space Property with respect to $\mathcal{M}_1$ and $\mathbb{C}^N$ with constants $K_1, K_2 > 0$ if

$$d_1(x,z) \leq K_1 d_3(x-z, \mathcal{M}_1 - \mathcal{M}_1) + K_2 d_2(Ax, Az) + \delta, \tag{4.37}$$

for all $x, z \in \mathbb{C}^N$.

The following theorem shows that the rIOP $\delta$ and the Approximate rNSP are almost equivalent.

**Theorem 4.4.12.** *(1) If there exist a reconstruction map $\Phi : \mathbb{C}^m \to \mathbb{C}^N$ that satisfies the $\delta$-rIOP for some $\delta \geq 0$, then $A$ satisfies the $2\delta$-Approximate rNSP with constants $K_1 = C_1, K_2 = C_2$.*

*(2) If $A$ satisfies the $\delta$-Approximate rNSP , then $\forall \alpha > 0$ there exists a reconstruction map $\Phi_\alpha : \mathbb{C}^m \to \mathbb{C}^N$ that satisfies the $(\delta + \alpha)$-rIOP with constants $C_1 = 2K_1, C_2 = 2K_2$.*

We separately prove the two statements in the Theorem 4.4.12.

**Theorem 4.4.13** ((1), Theorem 4.4.12)**.** *Suppose there exists $\delta \geq 0$ such that for all $x \in \mathbb{C}^N, y \in \mathbb{C}^m$:*

$$d_1^H\big(x, \Phi(y)\big) \leq C_1 d_3(x, \mathcal{M}_1) + C_2 d_2(Ax, y) + \delta \tag{4.38}$$

*Then $A$ satisfies*

$$d_1(x,z) \leq K_1 d_3(x-z, \mathcal{M}_1 - \mathcal{M}_1) + K_2 d_2(Ax, Az) + 2\delta \quad \forall x, z \in \mathbb{C}^N$$

*with $K_1 = C_1, K_2 = C_2$.*

*Proof of* (1)*, Theorem 4.4.12.* Let $x, z \in \mathbb{C}^N$ and $h = x - z$. Let $m \in \mathcal{M}_1$. Apply (4.38) on $m + h \in \mathbb{C}^N$ and $Am \in \mathbb{C}^m$, so that

$$\begin{aligned} d_1^H\big(m+h, \Phi(Am)\big) &\leq C_1 d_3(m+h, \mathcal{M}_1) + C_2 d_2(Am + Ah, Am) + \delta \\ &\leq C_1 d_3(m+h, \mathcal{M}_1) + C_2 d_2(Ah, 0) + \delta \end{aligned}$$

Using the triangle inequality and the remark, we obtain

$$\begin{aligned} d_3\big(h,0\big) = d_3\big(m+h, m\big) &\leq d_3^H\big(m+h, \Phi(Am)\big) + d_3^H\big(\Phi(Am), m\big) \\ &\leq C_1 d_3(m+h, \mathcal{M}_1) + C_2 d_2(Ah, 0) + 2\delta \end{aligned}$$

Since $m \in \mathcal{M}_1$ is arbitrary

$$\begin{aligned} d_3\big(h,0\big) &\leq \inf_{m \in \mathcal{M}_1} C_1 d_3(m+h, \mathcal{M}_1) + C_2 d_2(Ah, 0) + 2\delta \\ &= \inf_{m \in \mathcal{M}_1} C_1 d_3(h, \mathcal{M}_1 - m) + C_2 d_2(Ah, 0) + 2\delta \\ &= \inf_{m \in \mathcal{M}_1} \inf_{m' \in \mathcal{M}_1} C_1 d_3(h, m' - m) + C_2 d_2(Ah, 0) + 2\delta \\ &= C_1 d_3(h, \mathcal{M}_1 - \mathcal{M}_1) + C_2 d_2(Ah, 0) + 2\delta \end{aligned}$$

The claim follows immediately from $h = x - z$. □

**Theorem 4.4.14** ((2), Theorem 4.4.12)**.** *If $A$ satisfies*

$$d_1(x,z) \le K_1 d_3(x-z, \mathcal{M}_1 - \mathcal{M}_1) + K_2 d_2(Ax, Az) + \delta \quad \forall x, z \in \mathbb{C}^N \tag{4.39}$$

*then $\forall \alpha > 0$ there is $\Phi_\alpha : \mathbb{C}^m \rightrightarrows \mathbb{C}^N$ such that for all $x \in \mathbb{C}^N, y \in \mathbb{C}^m$*

$$d_1^H\big(x, \Phi_\varepsilon(y)\big) \le C_1 d_3(x, \mathcal{M}_1) + C_2 d_2(Ax, y) + \delta + \alpha$$

*with $C_1 = 2K_1, C_2 = 2K_2$.*

*Proof of (2), Theorem 4.4.12.* Fix $\alpha > 0$ and define

$$\Phi_\alpha(y) := \Big\{ \hat{x} \in \mathbb{C}^N : K_1 d_3(\hat{x}, \mathcal{M}_1) + K_2 d_2(A\hat{x}, y) \le \inf_{z \in \mathbb{C}^N} \Big[ K_1 d_3(z, \mathcal{M}_1) + K_2 d_2(Az, y) \Big] + \alpha \Big\}$$

Let $x \in \mathbb{C}^N, y \in \mathbb{C}^m$. Apply (4.39) with $z = \hat{x} \in \Phi_\alpha(y)$ so that

$$\begin{aligned} d_1(x, \hat{x}) &\le K_1 d_3(x - \hat{x}, \mathcal{M}_1 - \mathcal{M}_1) + K_2 d_2(Ax, A\hat{x}) + \delta \\ &\le K_1 d_3(x, \mathcal{M}_1) + K_1 d_3(\hat{x}, \mathcal{M}_1) + K_2 d_2(Ax, y) + K_2 d_2(y, A\hat{x}) + \delta \end{aligned}$$

Using that $d(a-b, c-d) = d(a, c-d+b) \le d(a,c) + d(c, c-d+b) = d(a,c) + d(0, b-d) = d(a,c) + d(b,d)$, for a metric $d$ induced by a pseudo-norm.

Consider the second and fourth terms together. From the definition of $\Phi_\alpha(y)$, by taking $z = x$ we get

$$d_1(x, \hat{x}) \le 2K_1 d_3(x, \mathcal{M}_1) + 2K_2 d_2(Ax, y) + \delta + \alpha$$

The claim follows by taking the supremum over $\hat{x} \in \Phi_\alpha(y)$. □

In order to compare the rIOP and the optimal map framework, it is necessary to assume that $d_1$ and $d_2$ are translation invariant. The next theorem states that any instance optimal map is an optimal map without noise and assuming the existence of an rIOP map bounds the optimal map constant with noise from above. Moreover, any robust instance optimal map is an optimal map with worst-case noise.

**Theorem 4.4.15 ((r)IOP implies optimal map (with noise)).** *Let $A \in \mathbb{C}^{m \times N}$, $\mathcal{M}_1 \subset \mathbb{C}^N$ and let $\|\cdot\|_i$, $i = 1, 2, 3$, be non-homogeneous (pseudo) norms. Then we have the following.*

*(1) Suppose $\Phi_0 \colon \mathbb{C}^m \to \mathbb{C}^N$ is an IOP decoder with $\delta = 0$, then $\Phi_0$ is a single valued optimal map and $c_{\mathrm{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = 0$.*

*(2) If the rIOP holds for all $\delta > 0$ with constants $C_1, C_2 > 0$, then*

$$c_{\mathrm{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = 0.$$

*(3) Assume that the noise level is bounded by $\epsilon > 0$ and that rIOP holds for fixed $\delta > 0$ with constants $C_1, C_2 > 0$, then*

$$c_{\mathrm{opt}}(A, \mathcal{M}_1, \mathcal{B}_{d_2}(0, \epsilon), \infty) \le C_2 \epsilon + \delta.$$

**Remark 4.4.16.** Part $(1)$ and $(2)$ of the above theorem follow directly from [21]. If as in Theorem 4.4.5 one assumes the rNSP, then there is a decoder satisfying the rIOP for all $\delta > 0$. This means that in the noisy case one assumes $c_{\text{opt}}(A, \mathcal{M}_1, \mathcal{B}_{d_2}(0, \epsilon), \infty) \leq C_2\epsilon$ by $(3)$, Theorem 4.4.15.

*Proof of Theorem 4.4.15.* **Part (1)**; by assumption we have that $\|x - \Phi_0(Ax)\|_1 \leq d_3(x, \mathcal{M}_1)$. We see that the right hand side is zero for all $x \in \mathcal{M}_1$, which proves the claim.

**Part (2)**; if the rIOP holds for some fixed $\delta > 0$ then by taking the supremum over all $x \in \mathcal{M}_1$ and letting $e = 0$ in (4.35) gives,

$$c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = \sup_{x \in \mathcal{M}_1} \|x - \Phi_\delta(Ax)\|_1 \leq \delta.$$

As the above inequality is satisfied for all $\delta > 0$, this $c_{\text{opt}}(A, \mathcal{M}_1, \{0\}, \infty) = 0$.

**Part (3)**; assume that the rIOP holds for some fixed $\delta > 0$ then by taking the supremum over all $x \in \mathcal{M}_1$ and all $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \epsilon$ in (4.35) gives

$$c_{\text{opt}}(A, \mathcal{M}_1, \mathcal{B}_{d_2}(0, \epsilon), \infty) = \sup_{x \in \mathcal{M}_1} \sup_{\substack{e \in \mathbb{C}^m \\ \|e\|_2 \leq \epsilon}} \|x - \Phi_\delta(Ax + e)\|_1 \leq C_2\epsilon + \delta.$$

□

## 4.5 Conclusion

In this work, we established a new theoretical framework for determining accuracy bounds of and solving ill-posed inverse problems. In special cases, we established specific conditions of the sampling operator $A$ and the underlying set to be reconstructed $\mathcal{M}_1$, which are fundamental when trying to obtain an accurate, stable and robust decoder. The concept of an optimal map with worst-case noise was introduced, which is the decoder obtaining the best possible performance. In particular, this decoder obtains the smallest worst-case reconstruction error among all vectors in $\mathcal{M}_1$, granted knowledge of the pair $(F, \mathcal{M}_1)$. Upper and lower bounds for the optimal reconstruction error are established. In the case of linear sampling operators $A$ these, are related to properties of the kernel of the operator $A$. In a more general setting, the kernel size allows for assessing accuracy bounds for solving ill-posed inverse problems in terms of $\mathcal{M}_1$ and the measurement model $F$. The main theorem contributes to assessing ill-posed inverse problems, by using the kernel size to provide upper and lower bounds on the optimality constant. The optimality constant includes the best worst-case noise, the average and the statistical reconstruction error for the reconstruction of (4.2). These bounds hold for any data distribution or measure $\mu_1$ on $\mathcal{M}_1$ and general noise models $(\mathcal{E}, \nu)$ which allow for a disintegration of measure. Moreover, these bounds are theoretical accuracy limits of a decoder's accuracy in all above mentioned settings. Moreover, explicit optimization problems yielding the decoders that attain the

optimality constant are derived. Furthermore, under specific conditions, despite the optimal map being a set-valued map, it is possible to show that it can be approximated by a neural network. In this ill-posed setting, we identify sufficient and necessary conditions on $(A, \mathcal{M}_1)$ such that the optimal map with worst-case noise can be approximated by a neural network. This makes the use of deep learning techniques for inverse problems plausible, at least in those scenarios where prior knowledge on $(A, \mathcal{M}_1)$ is available. Yet, we also remark that learning the optimal map can be challenging: obstacles remain when training a neural network that aims to minimise the best worst-case reconstruction. The proposed framework can be seen as a generalisation of previous concepts. In fact, the framework of the optimal map can be related to approximation theory. Moreover, the optimality constant generalises well-known quantities in approximation theory, the so-called widths, to assess the worst-case reconstruction error for ill-posed inverse problems. Furthermore, the optimal map with worst-case noise is related to the concept of fundamental decoders [21].

# Chapter 5

# Summary and Conclusion

This thesis explores the use of DL in inverse problems and aims at providing a theoretical basis for assessing the stability and accuracy of such methods. Particularly, fundamental trade-offs of decoders for inverse problems are investigated. The underlying aim is to provide a mathematical understanding for the use of deep neural networks relying on training data and training procedures for solving inverse problems. The approach of this thesis is twofold, firstly, providing theoretical results highlighting and possibly explaining DL in inverse problems and, secondly, illustrating these findings by numerical results. Moreover, these new findings are compared theoretically and numerically with the characteristics of reconstruction methods that do not involve training.

The first chapter is aimed at providing a brief overview of existing methods and theoretical frameworks for solving ill-posed inverse problems. It highlights the increasing use of data-driven methods such as deep learning for ill-posed inverse problems and provides an insight into some existing theoretical results in this area. Moreover, it motivates further by providing examples of applications where inverse problems arise that could be solved using learning based methods.

The main research work is presented in Chapters 2,3 and 4 of the thesis. Each chapter is based on a research paper that at the time of writing is in the process of review or being prepared for submission.

In the second chapter, a fully learned neural network approach for image reconstruction, which was introduced in a recent paper by Rosen et. al. and coined 'automated transform by manifold approximation' (AUTOMAP), is examined. Its potential benefits with respect to accuracy and disadvantages with respect to stability and robustness compared to standard methods for image reconstruction are investigated. We show that without further conditions on the sampling operator, such fully learned approaches to solving inverse problems become unstable. This result is somewhat interesting as the authors' claim that their work provides theoretical stability guarantees for AUTOMAP. With Theorem 2.2.1 and numerous experiments in Chapter 2, however, we can demonstrate that the guarantees provided in [192] only

hold in a limited context for a limited range of inverse problems and our results provide an explanation why these do not hold for general inverse problems. The conditions of Theorem 2.2.1 are shown to be satisfied in practical settings and the consequences of the theorem are demonstrated in various experiments. Thus, this chapter provides sufficient conditions for instability that also complement classical approaches providing conditions for a decoder to be stable. Moreover, as a result of this we establish a fundamental accuracy-stability tradeoff which, in the case of AUTOMAP, has a readily available practical application.

In the third chapter, we present a comprehensive mathematical analysis explaining different causes of AI generated hallucinations, commonly regarded as realistic-looking artefacts in the reconstructed image, and the links to instabilities. The relevance of AI generated hallucinations is illustrated in numerical experiments and imaging examples obtained from other publications. Our results establish four crucial issues for AI methods in finite dimensional linear inverse problems. Firstly, overly accurate AI methods will wrongly transfer details from one image to another reconstructed image creating a hallucination. Secondly, that there is an accuracy-hallucination trade-off. These two findings are theoretically illustrated by Theorems 3.4.2 and 3.4.4. Theorem 3.4.4 shows that hallucinations can occur with any probability distribution used on the set. Thirdly, there is an accuracy-stability trade-off. This relates to the findings presented in the second chapter, which are presented in numerous experiments and theoretically explained by Theorem 2.2.1. In the third chapter we extend these findings by establishing that optimising these trade-offs is hard through standard training processes. This difficulty is illustrated in Theorems 3.4.12 and 3.4.15. However, it should be noted that these results merely illustrate that there exist difficulties in obtaining an optimal map through training and do not provide an explicit solution as in how training procedures could be designed in order to obtain an optimal map. More insight into preventing undesirable effects, such as instabilities and hallucinations, can be gained from the conditions of Theorems 3.4.2, 3.4.4 and 3.4.7. These results show that one needs to use knowledge of the sampling model in order to avoid the conditions resulting in hallucinations and instabilities. Designing training procedures and choosing training sets accordingly could thus be a possibility for mitigating such effects.

Finally, in the fourth chapter, we investigate how DL based methods for solving inverse problems can perform better than standard methods. Thus, we establish fundamental and universal accuracy bounds for solving, possibly non-linear, ill-posed inverse problems with different noise models. To measure the accuracy of the best decoder we establish a universal optimality constant. The universal optimality constant can be chosen to be the best worst-case noise, the smallest average and the smallest statistical reconstruction error that a decoder for an ill-posed inverse problem can obtain. A key point in order for this optimality framework to be applicable to deep learning used for solving inverse problems, is that we have a fixed sampling operator $F$ and set $\mathcal{M}_1$ which we want to reconstruct, where no condition implying that $F$ is invertible on $\mathcal{M}_1$ is imposed, such as $\mathcal{N}(F) \cap (\mathcal{M}_1 - \mathcal{M}_1) = \{0\}$. For example, these could be a fixed image acquisition device on a specific set of data. The

main result is presented in Theorem 4.2.9, where the optimality constant is bounded from above and below by the kernel size, which only depends on the sampling operator $F$ and set $\mathcal{M}_1$ of the ill-posed inverse problem considered. Given this, the kernel size and optimality constant can provide a mean to assess and compare the accuracy of a decoder obtained by DL methods to standard methods by providing a lower bound on the accuracy. Moreover, we provide theoretical conditions for neural networks to approximate the optimal decoder for a given an ill-posed inverse problem and assess how training could yield an optimal map. In the last part of the fourth chapter, we provide theoretical motivation and foundations for the proposed framework. We establish lower bounds of optimality constant by well-known quantities from approximation theory, such as $n$-widths. We also relate the optimality constant and optimal map to optimal decoders, also referred to as instance optimal decoders. Moreover, there exist necessary and sufficient conditions for instance optimal decoders, such as the rNSP and NSP, and we relate these to the kernel size by showing that having zero kernel size is a necessary condition for the rNSP, Remark 4.4.9. Thus, we can show that the kernel size is in some sense more general than the rNSP. The results of this chapter provide many opportunities for further research. Especially a numerically feasible approximation of the kernel size would be useful for practical applications and, thus, this is discussed in the following.

## 5.1 Outlook

There are multiple interesting directions based on this thesis for future research. Possible directions for future work include the use the upper and lower bound on reconstruction accuracy in Theorem 4.2.9 to derive optimal sampling patterns and measures on $\mathcal{M}_1$ that minimize the kernel size and, thus, the reconstruction accuracy.

A further interesting perspective would be to establish a theorem relating the MMSE estimator and the optimal map provided by Theorem 4.2.9 with the relevant metrics induced by norms and $p = 2$. Related to this, a more practically relevant extension would be to provide means and methods to use the lower bounds for the optimality constant in learning problems. Here $\mathcal{M}_1$ can be replaced with the first component of the training set $\mathcal{T}_1$ in order to provide a lower bound for the training error. Moreover, for this appropriate algorithms and approximations for the disintegration of measure need to be established in order to compute the kernel size. Possible methods would be the interpretation of the disintegration of measure as a conditional probability and approximating the resulting posterior by methods currently used in Baysian DL, such as Monte Carlo or stochastic regularization techniques. A nice overview of this topic is given in the thesis [76] and a survey is given in [78].

Another interesting direction for future research is based on the current open problem to establish a framework for uncertainty quantification, [118]. This would entail extending the optimal map framework and kernel size, such that they can be applied for estimating errors

in linear and non-linear regression problems, [17, 149, 169]. If this can be achieved, based on the kernel size and Theorem 4.2.9, an optimal mean prediction accuracy, given by an extension of the optimality constant, and upper and lower bounds for this could be derived. Hence, this could provide a possibility for obtaining a baseline of uncertainty quantification for neural networks used for inverse problems and regression problems. An interesting application of this, is to provide uncertainty estimates for decoders of ill-posed inverse problems on large scales, such as SAR, [193] and other methods in earth observation.

Another interesting pssibility for further research is extending the optimality constant framework to encompass relative error rather than absolute error. Given the definition of the optimality constant, Definition 4.2.6, this would require further assumptions in order to maintain a well defined quantity. Possible conditions that could be investigated could entail that the decoder only outputs values close to zero on a set of measure zero or not at all.

# Bibliography


[1] D. ACHLIOPTAS, *Database-friendly random projections*, in Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2001, pp. 274–281.

[2] B. ADCOCK, A. C. HANSEN, AND B. ROMAN, *A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements*, arXiv:1403.6541, (2014).

[3] ——, *The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing*, in Compressed Sensing and Its Applications, Birkhäuser, 2015.

[4] J. ADLER AND O. ÖKTEM, *Solving ill-posed inverse problems using iterative deep neural networks*, Inverse Problems, 33 (2017), p. 124007.

[5] B. ADOCK AND A. C. HANSEN, *Compressive Imaging: Structure, Sampling, Learning*, CAMBRIDGE University Press, 2021.

[6] 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., (2020).

[7] L. ARDIZZONE, J. KRUSE, S. WIRKERT, D. RAHNER, E. W. PELLEGRINI, R. S. KLESSEN, L. MAIER-HEIN, C. ROTHER, AND U. KÖTHE, *Analyzing inverse problems with invertible neural networks*, arXiv preprint arXiv:1808.04730, (2018).

[8] S. ARRIDGE, P. MAASS, O. ÖKTEM, AND C.-B. SCHÖNLIEB, *Solving inverse problems using data-driven models*, Acta Numer., 28 (2019), pp. 1–174.

[9] A. ASPRI, Y. KOROLEV, AND O. SCHERZER, *Data driven regularization by projection*, Inverse Problems, 36 (2020), p. 125009.

[10] G. AUBERT AND J.-F. AUJOL, *A variational approach to removing multiplicative noise*, SIAM journal on applied mathematics, 68 (2008), pp. 925–946.

[11] N. BAKER, F. ALEXANDER, T. BREMER, A. HAGBERG, Y. Y. KEVREKIDIS, H. NAJM, M. PARASHAR, A. PATRA, J. SETHIAN, S. WILD, AND K. WILLCOX, *Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence*, U.S. Department of Energy Advanced Scientific Computing Research, (2019).

[12] R. G. BARANIUK AND M. B. WAKIN, *Random projections of smooth manifolds*, Foundations of computational mathematics, 9 (2009), pp. 51–77.

[13] A. BASTOUNIS AND A. C. HANSEN, *On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels*, SIAM J. Imaging Sci., 10 (2017), pp. 335–371.

[14] C. BELTHANGADY AND L. A. ROYER, *Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction*, Nature methods, 16 (2019), pp. 1215–1225.

[15] M. BERTERO, P. BOCCACCI, AND C. DE MOL, *Introduction to inverse problems in imaging*, CRC press, 2021.

[16] S. BHADRA, V. A. KELKAR, F. J. BROOKS, AND M. A. ANASTASIO, *On hallucinations in tomographic image reconstruction*, IEEE transactions on medical imaging, 40 (2021), pp. 3249–3260.

[17] P. BINEV, A. COHEN, W. DAHMEN, R. DEVORE, V. TEMLYAKOV, AND P. BARTLETT, *Universal algorithms for learning theory part i: Piecewise constant functions.*, Journal of Machine Learning Research, 6 (2005).

[18] L. BLUM, F. CUCKER, M. SHUB, AND S. SMALE, *Complexity and real computation. springer-verlag, new york 1998.*

[19] L. BLUM, M. SHUB, AND S. SMALE, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, in The Collected Papers of Stephen Smale: Volume 3, World Scientific, 2000, pp. 1293–1338.

[20] T. BLUMENSATH AND M. E. DAVIES, *Sampling theorems for signals from the union of finite-dimensional linear subspaces*, IEEE Transactions on Information Theory, 55 (2009), pp. 1872–1882.

[21] A. BOURRIER, M. E. DAVIES, T. PELEG, P. PÉREZ, AND R. GRIBONVAL, *Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems*, IEEE Transactions on Information Theory, 60 (2014), pp. 7928–7946.

[22] S. BOYD, S. P. BOYD, AND L. VANDENBERGHE, *Convex optimization*, Cambridge university press, 2004.

[23] H. BREZIS AND H. BRÉZIS, *Functional analysis, Sobolev spaces and partial differential equations*, vol. 2, Springer, 2011.

[24] T. A. BUBBA, G. KUTYNIOK, M. LASSAS, M. MÄRZ, W. SAMEK, S. SILTANEN, AND V. SRINIVASAN, *Learning the invisibile: A hybrid deep learning-shearlet framework for limited angle computed tomography*, arXiv:1811.04602, (2018).

[25] C. L. BYRNE, *Iterative optimization in inverse problems*, CRC Press, 2014.

[26] A. CALDERÓN AND A. ZYGMUND, *On the existence of certain singular integrals*, (1952).

[27] A. P. CALDERÓN AND A. ZYGMUND, *On singular integrals*, American Journal of Mathematics, 78 (1956), pp. 289–309.

[28] D. CALVETTI, B. LEWIS, AND L. REICHEL, *On the regularizing properties of the GMRES method*, Numerische Mathematik, 91 (2002), pp. 605–625.

[29] E. J. CANDES, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on information theory, 52 (2006), pp. 489–509.

[30] E. J. CANDES, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 59 (2006), pp. 1207–1223.

[31] E. J. CANDÈS AND T. TAO, *Near optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[32] N. CARLINI, P. MISHRA, T. VAIDYA, Y. ZHANG, M. SHERR, C. SHIELDS, D. WAGNER, AND W. ZHOU, *Hidden voice commands*, in 25th USENIX Security Symp., 2016, pp. 513–530.

[33] N. CARLINI AND D. WAGNER, *Audio adversarial examples: Targeted attacks on speech-to-text*, in 2018 IEEE Security and Privacy Worksh., 2018, pp. 1–7.

[34] M. E. CELEBI AND K. AYDIN, *Unsupervised learning algorithms*, Springer, 2016.

[35] A. CHAKRABORTY, M. ALAM, V. DEY, A. CHATTOPADHYAY, AND D. MUKHOPADHYAY, *Adversarial Attacks and Defences: A Survey*, arXiv:1810.00069, (2018).

[36] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[37] ——, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Math. Program., 159 (2016), pp. 253–287.

[38] V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, AND A. S. WILLSKY, *The convex geometry of linear inverse problems*, Foundations of Computational mathematics, 12 (2012), pp. 805–849.

[39] J. T. CHANG AND D. POLLARD, *Conditioning as disintegration*, Statistica Neerlandica, 51 (1997), pp. 287–317.

[40] D. CHARALAMBOS AND B. ALIPRANTIS, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer-Verlag Berlin and Heidelberg GmbH & Company KG, 2013.

[41] A. S. CHAUDHARI, C. M. SANDINO, E. K. COLE, D. B. LARSON, G. E. GOLD, S. S. VASANAWALA, M. P. LUNGREN, B. A. HARGREAVES, AND C. P. LANGLOTZ, *Prospective deployment of deep learning in MRI: A framework for important considerations, challenges, and recommendations for best practices*, Journal of Magnetic Resonance Imaging, (2020).

[42] H. CHEN, Y. ZHANG, M. K. KALRA, F. LIN, Y. CHEN, P. LIAO, J. ZHOU, AND G. WANG, *Low-dose CT with a residual encoder-decoder convolutional neural network*, IEEE transactions on medical imaging, 36 (2017), pp. 2524–2535.

[43] C. Q. CHOI, *7 revealing ways AIs fail: Neural networks can be disastrously brittle, forgetful, and surprisingly bad at math*, IEEE Spectrum, 58 (2021), pp. 42–47.

[44] A. COHEN, W. DAHMEN, AND R. DEVORE, *Compressed sensing and best k-term approximation*, Journal of the American mathematical society, 22 (2009), pp. 211–231.

[45] J. P. COHEN, M. LUCK, AND S. HONARI, *Distribution matching losses can hallucinate features in medical image translation*, in International conference on medical image computing and computer-assisted intervention, Springer, 2018, pp. 529–536.

[46] 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*, Proceedings of the National Academy of Sciences, 119 (2022), p. e2107151119.

[47] A. CONMY, S. MUKHERJEE, AND C.-B. SCHÖNLIEB, *StyleGAN-induced data-driven regularization for inverse problems*, arXiv preprint arXiv:2110.03814, (2021).

[48] R. CONT AND P. TANKOV, *Retrieving lévy processes from option prices: Regularization of an ill-posed inverse problem*, SIAM Journal on Control and Optimization, 45 (2006), pp. 1–25.

[49] Z. DAI, Z. YANG, F. YANG, W. W. COHEN, AND R. R. SALAKHUTDINOV, *Good semi-supervised learning that requires a bad GAN*, Advances in neural information processing systems, 30 (2017).

[50] M. DASHTI AND A. M. STUART, *The Bayesian approach to inverse problems*, in Handbook of Uncertainty Quantification, Springer, 2017.

[51] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 57 (2004), pp. 1413–1457.

[52] K. DE HAAN, Y. RIVENSON, Y. WU, AND A. OZCAN, *Deep-learning-based image reconstruction and enhancement in optical microscopy*, Proceedings of the IEEE, 108 (2019), pp. 30–50.

[53] A. DESHMANE, V. GULANI, M. A. GRISWOLD, AND N. SEIBERLICH, *Parallel MR imaging*, Journal of Magnetic Resonance Imaging, 36 (2012), pp. 55–72.

[54] R. DEVORE, B. HANIN, AND G. PETROVA, *Neural network approximation*, Acta Numerica, 30 (2021), pp. 327–444.

[55] R. DEVORE, G. KERKYACHARIAN, D. PICARD, AND V. TEMLYAKOV, *Approximation methods for supervised learning*, Foundations of Computational Mathematics, 6 (2006), pp. 3–58.

[56] R. A. DEVORE, R. HOWARD, AND C. MICCHELLI, *Optimal nonlinear approximation*, Manuscripta Mathematica, 63 (1989), pp. 469–478.

[57] D. L. DONOHO, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[58] ——, *For most large underdetermined systems of linear equations the minimal l1-solution is also the sparsest solution*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 59 (2006), pp. 797–829.

[59] J. DUGUNDJI, *An extension of tietze's theorem.*, Pacific Journal of Mathematics, 1 (1951), pp. 353–367.

[60] W. E, J. HAN, AND A. JENTZEN, *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, Commun. Math. Stat., 5 (2017), pp. 349–380.

[61] A. EFTEKHARI AND M. B. WAKIN, *New analysis of manifold embeddings and signal recovery from compressive measurements*, Applied and Computational Harmonic Analysis, 39 (2015), pp. 67–109.

[62] H. W. ENGL AND C. W. GROETSCH, *Inverse and ill-posed problems*, vol. 4, Elsevier, 2014.

[63] H. W. ENGL AND P. KÜGLER, *Nonlinear inverse problems: theoretical aspects and some industrial applications*, in Multidisciplinary methods for analysis optimization and control of complex systems, Springer, 2005, pp. 3–47.

[64] H. W. ENGL, K. KUNISCH, AND A. NEUBAUER, *Convergence rates for Tikhonov regularisation of non-linear ill-posed problems*, Inverse problems, 5 (1989), p. 523.

[65] C. L. EPSTEIN, *Introduction to the mathematics of medical imaging*, SIAM, 2007.

[66] EUROPEAN COMMISSION, *Europe fit for the digital age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence*. URL: `https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682`, April 2021.

[67] K. EYKHOLT, I. EVTIMOV, E. FERNANDES, B. LI, A. RAHMATI, C. XIAO, A. PRAKASH, T. KOHNO, AND D. SONG, *Robust physical-world attacks on deep*

*learning visual classification*, in IEEE Conf. on Computer Vision and Pattern Recognition, 2018, pp. 1625–1634.

[68] E. A. F. KNOLL, *Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge.*, Magnetic resonance in medicine, (2020), pp. 3054–3070.

[69] Q. FAN, T. WITZEL, A. NUMMENMAA, K. R. VAN DIJK, J. D. VAN HORN, M. K. DREWS, L. H. SOMERVILLE, M. A. SHERIDAN, R. M. SANTILLANA, J. SNYDER, ET AL., *MGH–USC human connectome project datasets with ultra-high b-value diffusion MRI*, Neuroimage, 124 (2016), pp. 1108–1114.

[70] FDA, *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*, FDA News Release, April 11 2018, (2018).

[71] J. A. FESSLER, *Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms*, IEEE signal processing magazine, 37 (2020), pp. 33–40.

[72] M. A. FIGUEIREDO AND R. D. NOWAK, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.

[73] S. G. FINLAYSON, J. D. BOWERS, J. ITO, J. L. ZITTRAIN, A. L. BEAM, AND I. S. KOHANE, *Adversarial attacks on medical machine learning*, Science, 363 (2019), pp. 1287–1289.

[74] S. FOUCART AND H. RAUHUT, *A mathematical introduction to compressive sensing*, Birkhauser, 2013.

[75] A. FROMMER AND P. MAASS, *Fast CG-based methods for Tikhonov–Phillips regularization*, SIAM Journal on Scientific Computing, 20 (1999), pp. 1831–1850.

[76] Y. GAL ET AL., *Uncertainty in deep learning*, (2016).

[77] R. GARCIA AND F. CRESPON, *Radio tomography of the ionosphere: Analysis of an underdetermined, ill-posed inverse problem, and regional application*, Radio Science, 43 (2008), pp. 1–13.

[78] J. GAWLIKOWSKI, C. R. N. TASSI, M. ALI, J. LEE, M. HUMT, J. FENG, A. KRUSPE, R. TRIEBEL, P. JUNG, R. ROSCHER, ET AL., *A survey of uncertainty in deep neural networks*, arXiv preprint arXiv:2107.03342, (2021).

[79] L. GEERTS-OSSEVOORT, E. DE WEERDT, A. DUIJNDAM, G. VAN IJPEREN, H. PEETERS, M. DONEVA, M. NIJENHUIS, AND A. HUANG, *Compressed SENSE: Speed done right. Every time*, Philips FieldStrength Magazine, 2018 (2018), pp. 1–16.

[80] M. GENZEL, J. MACDONALD, AND M. MARZ, *Solving inverse problems with deep neural networks-robustness included*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2022).

[81] R. GIRYES, Y. C. ELDAR, A. M. BRONSTEIN, AND G. SAPIRO, *Tradeoffs between convergence speed and reconstruction accuracy in inverse problems*, IEEE Transactions on Signal Processing, 66 (2018), pp. 1676–1690.

[82] I. J. GOODFELLOW, J. SHLENS, AND C. SZEGEDY, *Explaining and harnessing adversarial examples*, in Proceedings of the Int. Conf. on Learning Representations, 2015.

[83] N. M. GOTTSCHLING, V. ANTUN, B. ADCOCK, AND A. C. HANSEN, *The troublesome kernel: why deep learning for inverse problems is typically unstable*, arXiv preprint arXiv:2001.01258, (2020).

[84] P. GRANGEAT, *Tomography*, John Wiley & Sons, 2013.

[85] R. GRIBONVAL AND M. NIELSEN, *Sparse representations in unions of bases*, IEEE transactions on Information theory, 49 (2003), pp. 3320–3325.

[86] R. GRIBONVAL AND M. NIKOLOVA, *On bayesian estimation and proximity operators*, Applied and Computational Harmonic Analysis, 50 (2021), pp. 49–72.

[87] M. GUERQUIN-KERN, L. LEJEUNE, K. P. PRUESSMANN, AND M. UNSER, *Realistic analytical phantoms for parallel magnetic resonance imaging*, IEEE Trans. Med. Imaging, 31 (2011), pp. 626–636.

[88] S. F. GULL AND G. J. DANIELL, *Image reconstruction from incomplete and noisy data*, Nature, 272 (1978), pp. 686–690.

[89] H. GUPTA, K. H. JIN, H. Q. NGUYEN, M. T. MCCANN, AND M. UNSER, *CNN-based projected gradient descent for consistent CT image reconstruction*, IEEE transactions on medical imaging, 37 (2018), pp. 1440–1453.

[90] J. HADAMARD, *Sur les problèmes aux dérivées partielles et leur signification physique*, Princeton university bulletin, (1902), pp. 49–52.

[91] ——, *Lectures on Cauchy's problem in linear partial differential equations*, Courier Corporation, 2003.

[92] 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 (2018), pp. 3055–3071.

[93] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surv. Math. Ind, 3 (1993), pp. 253–315.

[94] M. HANKE-BOURGEOIS, *Conjugate gradient type methods for ill-posed problems*, Longman Scientific & Techn., 1995.

[95] P. C. HANSEN AND D. P. O'LEARY, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM journal on scientific computing, 14 (1993), pp. 1487–1503.

[96] B. HE AND X. YUAN, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149.

[97] D. HEAVEN ET AL., *Why deep-learning AIs are so easy to fool*, Nature, 574 (2019), pp. 163–166.

[98] N. J. HIGHAM, *Accuracy and stability of numerical algorithms*, SIAM, 2002.

[99] D. P. HOFFMAN, I. SLAVITT, AND C. A. FITZPATRICK, *The promise and peril of deep learning in microscopy*, Nature Methods, 18 (2021), pp. 131–132.

[100] A. HOFINGER AND H. K. PIKKARAINEN, *Convergence rates for linear inverse problems in the presence of an additive normal noise*, Stochastic Analysis and Applications, 27 (2009), pp. 240–257.

[101] Y. HUANG, A. PREUHS, G. LAURITSCH, M. MANHART, X. HUANG, AND A. MAIER, *Data consistent artifact reduction for limited angle tomography with deep learning prior*, in Int. Works. on Machine Learning for Medical Image Reconstruction, Springer, 2019, pp. 101–112.

[102] 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 Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 145–153.

[103] Y.-M. HUANG, M. K. NG, AND Y.-W. WEN, *A new total variation method for multiplicative noise removal*, SIAM Journal on imaging sciences, 2 (2009), pp. 20–40.

[104] D. HYDE, E. MILLER, D. BROOKS, AND V. NTZIACHRISTOS, *New techniques for data fusion in multimodal FMT-CT imaging*, in 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2008, pp. 1597–1600.

[105] S. JÉGOU, M. DROZDZAL, D. VAZQUEZ, A. ROMERO, AND Y. BENGIO, *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation*, in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 11–19.

[106] 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 (2017), pp. 4509–4522.

[107] F. JOHN, *A note on "improper" problems in partial differential equations*, Communications on Pure and Applied Mathematics, 8 (1955), pp. 591–594.

[108] ——, *Numerical solution of problems which are not well posed in the sense of Hadamardd*, in Fritz John, Springer, 1985, pp. 411–424.

[109] W. B. JOHNSON AND J. LINDENSTRAUSS, *Extensions of lipschitz mappings into a Hilbert space 26*, Contemporary mathematics, 26 (1984).

[110] J. P. KAIPIO, V. KOLEHMAINEN, M. VAUHKONEN, AND E. SOMERSALO, *Inverse problems with structural prior information*, Inverse problems, 15 (1999), p. 713.

[111] E. KANG, J. MIN, AND J. C. YE, *A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction*, Medical physics, 44 (2017), pp. e360–e375.

[112] N. KERIVEN AND R. GRIBONVAL, *Instance optimal decoding and the restricted isometry property*, in Journal of Physics: Conference Series, vol. 1131, IOP Publishing, 2018, p. 012002.

[113] A. KIRSCH ET AL., *An introduction to the mathematical theory of inverse problems*, vol. 120, Springer, 2011.

[114] F. KNOLL, T. MURRELL, A. SRIRAM, N. YAKUBOVA, J. ZBONTAR, M. RABBAT, A. DEFAZIO, M. J. MUCKLEY, D. K. SODICKSON, C. L. ZITNICK, ET AL., *Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge*, Magnetic resonance in medicine, 84 (2020), pp. 3054–3070.

[115] K. KUNISCH AND T. POCK, *A bilevel optimization approach for parameter learning in variational models*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 938–983.

[116] A. KURAKIN, I. GOODFELLOW, AND S. BENGIO, *Adversarial examples in the physical world*, in Int. Conf. on Learning Representations, 2017.

[117] P. P. LAISSUE, R. A. ALGHAMDI, P. TOMANCAK, E. G. REYNAUD, AND H. SHROFF, *Assessing phototoxicity in live fluorescence imaging*, Nature methods, 14 (2017), pp. 657–661.

[118] B. LAKSHMINARAYANAN, A. PRITZEL, AND C. BLUNDELL, *Simple and scalable predictive uncertainty estimation using deep ensembles*, Advances in neural information processing systems, 30 (2017).

[119] A. S. LEONOV, *A posteriori accuracy estimations of solutions to ill-posed inverse problems and extra-optimal regularizing algorithms for their solution*, Numerical Analysis and Applications, 5 (2012), pp. 68–83.

[120] ——, *Locally extra-optimal regularizing algorithms and a posteriori estimates of the accuracy for ill-posed problems with discontinuous solutions*, Computational Mathematics and Mathematical Physics, 56 (2016), pp. 1–13.

[121] R. M. LEWITT, *Reconstruction algorithms: transform methods*, Proceedings of the IEEE, 71 (1983), pp. 390–408.

[122] C. LI AND B. ADCOCK, *Compressed sensing with local structure: uniform recovery guarantees for the sparsity in levels class*, Appl. Comput. Harmon. Anal., 46 (2019), pp. 453–477.

[123] B. LIANG, H. LI, M. SU, P. BIAN, X. LI, AND W. SHI, *Deep text classification can be fooled*, in The 27th Int. Joint Conf. on Artificial Intelligence, 2017.

[124] E. H. LIEB AND M. LOSS, *Analysis*, vol. 14, American Mathematical Soc., 2001.

[125] K. LØNNING, P. PUTZKY, J.-J. SONKE, L. RENEMAN, M. W. CAAN, AND M. WELLING, *Recurrent inference machines for reconstructing heterogeneous MRI data*, Med. Image Anal., 53 (2019), pp. 64–78.

[126] A. LUCAS, M. ILIADIS, R. MOLINA, AND A. K. KATSAGGELOS, *Using deep neural networks for inverse problems in imaging: beyond analytical methods*, IEEE Signal Processing Magazine, 35 (2018), pp. 20–36.

[127] S. LUNZ, O. ÖKTEM, AND C.-B. SCHÖNLIEB, *Adversarial regularizers in inverse problems*, Advances in neural information processing systems, 31 (2018).

[128] J. MA AND M. MÄRZ, *A multilevel based reweighting algorithm with joint regularizers for sparse recovery*, arXiv preprint arXiv:1604.06941, (2016).

[129] A. MALLIK, *Statistical rethinking: A Bayesian course with examples in R and Stan*, Technometrics, 63 (2021), pp. 440–441.

[130] M. MARDANI, Q. SUN, D. DONOHO, V. PAPYAN, H. MONAJEMI, S. VASANAWALA, AND J. PAULY, *Neural proximal gradient descent for compressive imaging*, Advances in Neural Information Processing Systems, 31 (2018), pp. 9573–9583.

[131] M. T. MCCANN, K. H. JIN, AND M. UNSER, *Convolutional neural networks for inverse problems in imaging: A review*, IEEE Signal Process. Mag., 34 (2017), pp. 85–95.

[132] M. T. MCCANN AND M. UNSER, *Biomedical image reconstruction: From the foundations to deep neural networks*, arXiv preprint arXiv:1901.03565, (2019).

[133] M. MOHRI, A. ROSTAMIZADEH, AND A. TALWALKAR, *Foundations of machine learning*, MIT press, 2018.

[134] S. MOOSAVI-DEZFOOLI, A. FAWZI, O. FAWZI, AND P. FROSSARD, *Universal adversarial perturbations*, in IEEE Conf. on computer vision and pattern recognition, July 2017, pp. 86–94.

[135] S. MOOSAVI-DEZFOOLI, A. FAWZI, AND P. FROSSARD, *DeepFool: A simple and accurate method to fool deep neural networks*, in IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 2574–2582.

[136] U. MOSCO, *Convergence of convex sets and of solutions of variational inequalities*, Advances in Mathematics, 3 (1969), pp. 510–585.

[137] T. MOTZKIN, E. G. STRAUS, AND F. VALENTINE, *The number of farthest points.*, Pacific journal of Mathematics, 3 (1953), pp. 221–232.

[138] M. J. MUCKLEY, B. RIEMENSCHNEIDER, A. RADMANESH, S. KIM, G. JEONG, J. KO, Y. JUN, H. SHIN, D. HWANG, M. MOSTAPHA, ET AL., *State-of-the-art Machine Learning MRI Reconstruction in 2020: Results of the Second fastMRI Challenge*, arXiv preprint arXiv:2012.06318, (2020).

[139] S. MUKHERJEE, S. DITTMER, Z. SHUMAYLOV, S. LUNZ, O. ÖKTEM, AND C.-B. SCHÖNLIEB, *Learned convex regularizers for inverse problems*, arXiv preprint arXiv:2008.02839, (2020).

[140] S. MUKHERJEE, P. NIYOGI, T. POGGIO, AND R. RIFKIN, *Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization*, Advances in Computational Mathematics, 25 (2006), pp. 161–193.

[141] F. NATTERER AND F. WÜBBELING, *Mathematical methods in image reconstruction*, SIAM, 2001.

[142] A. NGUYEN, J. YOSINSKI, AND J. CLUNE, *Deep neural networks are easily fooled: High confidence predictions for unrecognizable images*, in IEEE Conf. on Computer Vision and Pattern Recognition, 2015, pp. 427–436.

[143] M. NIKOLOVA, *Model distortions in Bayesian MAP reconstruction*, Inverse Problems & Imaging, 1 (2007), p. 399.

[144] A. ODENA, *Semi-supervised learning with generative adversarial networks*, arXiv preprint arXiv:1606.01583, (2016).

[145] F. O'SULLIVAN, *A statistical perspective on ill-posed inverse problems*, Statistical science, (1986), pp. 502–518.

[146] T. PELEG, R. GRIBONVAL, AND M. E. DAVIES, *Compressed sensing and best approximation from unions of subspaces: Beyond dictionaries*, in 21st European Signal Processing Conference (EUSIPCO 2013), IEEE, 2013, pp. 1–5.

[147] D. L. PHILLIPS, *A technique for the numerical solution of certain integral equations of the first kind*, Journal of the ACM (JACM), 9 (1962), pp. 84–97.

[148] A. PINKUS, *Approximation theory of the MLP model in neural networks*, Acta numerica, 8 (1999), pp. 143–195.

[149] M. A. POOLE AND P. N. O'FARRELL, *The assumptions of the linear regression model*, Transactions of the Institute of British Geographers, (1971), pp. 145–158.

[150] C. POON, *On the role of total variation in compressed sensing*, SIAM J. Imaging Sci., 8 (2015), pp. 682–720.

[151] K. P. PRUESSMANN, M. WEIGER, M. B. SCHEIDEGGER, AND P. BOESIGER, *SENSE: sensitivity encoding for fast MRI*, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 42 (1999), pp. 952–962.

[152] C. QIAO, D. LI, Y. GUO, C. LIU, T. JIANG, Q. DAI, AND D. LI, *Evaluation and development of deep neural networks for image super-resolution in optical microscopy*, Nature Methods, 18 (2021), pp. 194–202.

[153] A. RAJ, Y. BRESLER, AND B. LI, *Improving robustness of deep-learning-based image reconstruction*, in International Conference on Machine Learning, PMLR, 2020, pp. 7932–7942.

[154] G. RAMACHANDRAN AND A. LAKSHMINARAYANAN, *Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms*, Proceedings of the National Academy of Sciences, 68 (1971), pp. 2236–2240.

[155] Z. RAMZI, J. STARCK, AND P. CIUCIU, *XPDNet for MRI reconstruction: An application to the 2020 fastMRI challenge*, in 2021 ISMRM annual meeting, no. Abstract, vol. 275, 2021.

[156] S. RAVISHANKAR, J. C. YE, AND J. A. FESSLER, *Image reconstruction: From sparsity to data-adaptive methods and machine learning*, Proceedings of the IEEE, 108 (2019), pp. 86–109.

[157] C. REBUFFEL, M. ROBERTI, L. SOULIER, G. SCOUTHEETEN, R. CANCELLIERE, AND P. GALLINARI, *Controlling hallucinations at word level in data-to-text generation*, Data Mining and Knowledge Discovery, 36 (2022), pp. 318–354.

[158] S. H. RUDY, S. L. BRUNTON, J. L. PROCTOR, AND J. N. KUTZ, *Data-driven discovery of partial differential equations*, Science Advances, 3 (2017).

[159] A. SANAAT, I. SHIRI, S. FERDOWSI, H. ARABI, AND H. ZAIDI, *Robust-Deep: A method for increasing brain imaging datasets to improve deep learning models' performance and robustness*, Journal of digital imaging, (2022), pp. 1–13.

[160] A. SAWATZKY, C. BRUNE, F. WUBBELING, T. KOSTERS, K. SCHAFERS, AND M. BURGER, *Accurate EM-TV algorithm in PET with low SNR*, in 2008 IEEE nuclear science symposium conference record, IEEE, 2008, pp. 5133–5137.

[161] J. SCHLEMPER, J. CABALLERO, J. V. HAJNAL, A. PRICE, AND D. RUECKERT, *A deep cascade of convolutional neural networks for MR image reconstruction*, in Int. Conf. on Information Processing in Medical Imaging, Springer, 2017, pp. 647–658.

[162] C. SCHWAB AND J. ZECH, *Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ*, Analysis and Applications, 17 (2019), pp. 19–55.

[163] Z. SHEN, *Deep network approximation characterized by number of neurons*, Communications in Computational Physics, 28 (2020).

[164] J. R. SHEWCHUK ET AL., *An introduction to the conjugate gradient method without the agonizing pain*, 1994.

[165] J. SHI AND S. OSHER, *A nonlinear inverse scale space method for a convex multiplicative noise model*, SIAM Journal on imaging sciences, 1 (2008), pp. 294–321.

[166] E. Y. SIDKY, I. LORENTE, J. G. BRANKOV, AND X. PAN, *Do CNNs solve the CT inverse problem?*, IEEE Transactions on Biomedical Engineering, 68 (2020), pp. 1799–1810.

[167] J. SIETSMA AND R. J. DOW, *Creating artificial neural networks that generalize*, Neural networks, 4 (1991), pp. 67–79.

[168] S. SMALE, *Mathematical problems for the next century*, The mathematical intelligencer, 20 (1998), pp. 7–15.

[169] G. K. SMYTH, *Nonlinear regression*, Encyclopedia of environmetrics, 3 (2002), pp. 1405–1411.

[170] A. SRIRAM, J. ZBONTAR, T. MURRELL, C. L. ZITNICK, A. DEFAZIO, AND D. K. SODICKSON, *GrappaNet: Combining Parallel Imaging With Deep Learning for Multi-Coil MRI Reconstruction*, in Proceedings of the IEEE/CVF Conf. on Comp. Visi. and Patt. Recogn., 06 2020.

[171] R. STRACK, *Imaging: AI transforms image reconstruction*, Nature Methods, 15 (2018), p. 309.

[172] A. M. STUART, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[173] J. SUN, H. LI, Z. XU, ET AL., *Deep ADMM-Net for compressive sensing MRI*, Advances in neural information processing systems, 29 (2016).

[174] C. SZEGEDY, W. ZAREMBA, I. SUTSKEVER, J. BRUNA, D. ERHAN, I. J. GOODFELLOW, AND R. FERGUS, *Intriguing properties of neural networks*, in Int. Conf. on Learning Representations, 2014.

[175] A. N. TIKHONOV, *On the stability of inverse problems*, in Dokl. Akad. Nauk SSSR, vol. 39, 1943, pp. 195–198.

[176] A. N. TIKHONOV, *On the solution of ill-posed problems and the method of regularization*, in Doklady Akademii Nauk, vol. 151, Russian Academy of Sciences, 1963, pp. 501–504.

[177] A. N. TIKHONOV, V. I. ARSENIN, V. ARSENIN, ET AL., *Solutions of ill-posed problems*, Vh Winston, 1977.

[178] Y. TRAONMILIN AND R. GRIBONVAL, *Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all*, Applied and Computational Harmonic Analysis, 45 (2018), pp. 170–205.

[179] M. UNSER, *A unifying representer theorem for inverse problems and machine learning*, Foundations of Computational Mathematics, (2020), pp. 1–20.

[180] M. UNSER, J. FAGEOT, AND H. GUPTA, *Representer theorems for sparsity-promoting l2 regularization*, IEEE Transactions on Information Theory, 62 (2016), pp. 5167–5180.

[181] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solutions*, SIAM Journal on Scientific Computing, 31 (2008), pp. 890–912.

[182] P. VINCENT, H. LAROCHELLE, Y. BENGIO, AND P.-A. MANZAGOL, *Extracting and composing robust features with denoising autoencoders*, in Proceedings of the 25th international conference on Machine learning, 2008, pp. 1096–1103.

[183] G. WANG, J. C. YE, AND B. DE MAN, *Deep learning for tomographic image reconstruction*, Nature Machine Intelligence, 2 (2020), pp. 737–748.

[184] Y. WANG, A. S. LEONOV, D. V. LUKYANENKO, AND A. G. YAGOLA, *General Tikhonov regularization with applications in geoscience*, CSIAM Trans. Appl. Math, 1 (2020), pp. 53–85.

[185] E. WEINAN AND B. YU, *The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems*, Commun. Math. Stat., 6 (2018), pp. 1–14.

[186] D. YAROTSKY, *Optimal approximation of continuous functions by very deep ReLU networks*, arXiv:1802.03620, (2018).

[187] X. YUAN AND S. PANG, *Structured illumination temporal compressive microscopy*, Biomedical Optics Express, 7 (2016), pp. 746–758.

[188] J. ZBONTAR, F. KNOLL, A. SRIRAM, T. MURRELL, Z. HUANG, M. J. MUCKLEY, A. DEFAZIO, R. STERN, P. JOHNSON, M. BRUNO, ET AL., *fastMRI: An open dataset and benchmarks for accelerated MRI*, arXiv preprint arXiv:1811.08839, (2018).

[189] G. L. ZENG, *Medical image reconstruction: a conceptual tutorial.*, Springer, 2010.

[190] G. ZHANG, C. YAN, X. JI, T. ZHANG, T. ZHANG, AND W. XU, *Dolphinattack: Inaudible voice commands*, in Proc. of the 2017 ACM SIGSAC Conf. on Computer and Commun. Security, 2017, pp. 103–117.

[191] X.-L. ZHAO, F. WANG, AND M. K. NG, *A new convex optimization model for multiplicative noise and blur removal*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 456–475.

[192] B. ZHU, J. Z. LIU, S. F. CAULEY, B. R. ROSEN, AND M. S. ROSEN, *Image reconstruction by domain-transform manifold learning*, Nature, 555 (2018), p. 487.

[193] X. X. ZHU AND R. BAMLER, *Tomographic sar inversion by l_{1}-norm regularization—the compressive sensing approach*, IEEE transactions on Geoscience and Remote Sensing, 48 (2010), pp. 3839–3846.