Average Kernel Sizes

—

Computable Sharp Accuracy Bounds for Inverse Problems*

Nina M. Gottschling † David Iagaru †‡ Jakob Gawlikowski †
Ioannis Sgouralis §

October 14, 2025

Abstract

The reconstruction of an unknown quantity from noisy measurements is a mathematical problem relevant in most applied sciences, for example, in medical imaging, radar inverse scattering, or astronomy. This underlying mathematical problem is often an ill-posed (non-linear) reconstruction problem, referred to as an ill-posed inverse problem. To tackle such problems, there exist a myriad of methods to design approximate inverse maps, ranging from optimization-based approaches, such as compressed sensing, over Bayesian approaches, to data-driven techniques such as deep learning, and variants in between the two techniques. For all stable approximate inverse maps, there are accuracy limits that are strictly larger than zero for ill-posed inverse problems, due to the accuracy-stability tradeoff [Gottschling et al., *SIAM Review*, 67.1 (2025), pp. 73-104] and [Colbrook et al., *Proceedings of the National Academy of Sciences*, 119.12 (2022), pp. e2107151119]. The variety of methods that aim to solve such problems with optimal accuracy begs for a unifying approach to help scientists choose the approximate inverse map that obtains this theoretical optimum. However, up to now there do not exist computable accuracy bounds to this optimum that are applicable to all inverse problems. In this work, we provide computable sharp accuracy bounds to the reconstruction error of solution methods to inverse problems. The accuracy bounds are method-independent and purely depend on the dataset of signals, the forward model of the inverse problem, and the noise model. The simplicity and mathematical rigor of our approach enable an efficient computation of accuracy bounds before designing and optimizing inverse maps. In order to facilitate the use of these bounds in scientific applications, we provide an algorithmic framework and an accompanying software library to compute these accuracy bounds. We demonstrate the applicability and validity of the algorithms on two challenging and relevant inverse problems from different domains: fluorescence localization microscopy and super-resolution of multi-spectral satellite data. Computing the accuracy bounds for a given problem before solving it enables a fundamental shift towards optimizing datasets and forward models. The accuracy bounds facilitate a theoretical and empirical understanding of the accuracy limits of different methods.

Keywords Inverse problems, optimization, accuracy bounds, uncertainty, imaging, deep learning

1 Introduction

In recent years, the intersection of artificial intelligence and inverse problems has revolutionized various scientific fields. Inverse problems, at their core, involve reconstructing an unknown signal, such as an

*NMG, DI and JG acknowledge the funding provided by the German Aerospace Center from the department EO Data Science at the Earth Observation Center. This work was supported by the Helmholtz Association's Initiative and Networking Fund on the HAICORE (at) FZJ partition.

†German Aerospace Center (DLR), Remote Sensing Technology Institute, Wessling, Germany.
‡Department of Applied Mathematics, École Polytechnique, Paris, France.
§Department of Mathematics, University of Tennessee, Knoxville, Tennessee.

image or features of an image, from indirect or incomplete measurements. This task is often ill-posed and computationally challenging [16].
Ill-posed finite dimensional inverse problems are ubiquitous in the computational sciences as they appear in multiple applications. An incomplete list of examples includes most types of computational imaging [3, 11, 22], matrix completion [25], parametric partial differential equations and system identification [1], phase retrieval [26, 38, 68], and quantized sampling [21]. Examples of applications are radar inverse scattering [20] or astronomy [34].
Traditional reconstruction methods, such as regularization inverse techniques [71] or compressed sensing [39], have achieved significant success. However, emerging AI and deep learning workflows have ushered in a new era of methods. AI-based methods and combinations of optimization-based iterative and AI methods, such as [32], have enabled remarkable improvements in reconstruction quality and speed in various inverse problems. These include, but are not limited to, the inverse problems of Magnetic Resonance Imaging (MRI) [48, 78], microscopy [44], and computer tomography (CT) [73]. For example, deep neural networks, trained on large datasets, can learn complex data-driven reconstruction methods that outperform classical methods in MRI [43].

1.1 What are accuracy limits in benchmark challenges?

After the successful rise of AI-based methods for solving inverse problems, a striking phenomenon has emerged in benchmark dataset challenges: despite a plethora of universal approximation theorems for neural networks [10, 12, 45, 54, 61, 74] and diverse architectures, training strategies, and theoretical approaches, the top-performing participants in benchmark challenges based on data sets report similar reconstruction errors. This accuracy plateau of different state-of-the-aft methods is apparent in different applied inverse problems: in the fast MRI challenge [48, 57, 76], in [33] on three benchmark problems - designed from image inpainting, x-ray tomography, and phase retrieval - and in the recently published benchmark dataset collection [77] covering five different inverse problems, linear inverse scattering, MRI, black hole imaging, full waveform inversion, and the discretized Navier Stokes equation. This begs the question: what dataset and inverse problem-dependent accuracy limits on benchmark dataset challenges are? In theory, deep neural networks are universal approximators that obtain vanishing errors. However, the authors of [2] point out that there is a gap between theoretical guaranties and the practical training and design of such networks. This observed gap between theory and practice even further motivates the question of what inverse problem- and dataset-dependent accuracy limits are. In fact, previous results even prove that there are accuracy bounds to these limits for any method that calculates solutions to inverse problems on measurable metric spaces [42, 63]. An *open question* remains whether there are computable accuracy bounds for methods computing solutions to inverse problems on finitely many samples and general datasets [24]. Computing method-independent accuracy bounds from datasets and corresponding inverse problems could enable dataset and inverse problem optimization by minimizing these accuracy bounds. This would be a crucial step to break the accuracy plateaus that are obtained in benchmark dataset challenges [33, 48, 57, 76, 77] and bridge the gap between theory and practice by providing limits to the accuracy that can be obtained on a given dataset.

2 Overview of the paper

In this section, we present a brief overview of the paper. We first formalize the main problem studied and then introduce three key challenges and discuss the related literature. Finally, we give a summary of our main result and discuss its scientific impact.

2.1 Problem outline

Following [42], we express an inverse problem using the following general framework:

$$\text{recover } x \in \mathcal{M}_1 \subset \mathbb{C}^{d_1} \text{ given a point } y = F(x, e) \in \mathbb{C}^{d_2} \text{ of } x \text{ and } e \in \mathcal{E} \subset \mathbb{C}^{d_3}. \tag{2.1}$$

Here, x represents the signal of interest, while e represents the noise. The sets $\mathcal{M}_1$ and $\mathcal{E}$ describe the signals of interest and noise, respectively. The function $F\colon \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2 \subset \mathbb{C}^{d_2}$ which describes the

forward model maps onto its image $\mathcal{M}_2 = F(\mathcal{M}_1 \times \mathcal{E})$ and is deliberately kept general to encompass many known models.

In applications, the inverse problem in (2.1) often does not have a unique solution and/or the solutions cannot be computed. However, applying the forward model to verify that a solution maps to a noisy measurement is often computable.

The goal of solving the inverse problem is to produce an approximation or set of approximations $\hat{x}$ to the true solution x. Any algorithm or method that yields an approximate inverse map from noisy measurements in the set $\mathcal{M}_2$ to an approximation to the true solution x, is denoted by $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ with $\phi(y) = \hat{x}$. Some examples are approximate inverse maps obtained from optimization-based inverse methods, such as compressed sensing [39] or total variation minimization [46, 66], combinations of optimization and data-driven inverse methods [6, 19], and (un-)supervised Deep Learning methods [13, 70, 78]. A valuable and comprehensive mathematical review of data-driven methods for solving inverse problems is provided in [6]. The reconstruction error of an approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ for approximately solving (2.1) is often quantified by the empirical expectation of the residual of the reconstruction and solutions of the $M \in \mathbb{N}$ samples $\mathcal{D} = \{(x_m, y_m)\}_{m=1}^M = \{(x_m, F(x_m, e_m))\}_{m=1}^M \subset \mathcal{M}_1 \times \mathcal{M}_2$, and can be denoted by

$$\mathcal{L}(\mathcal{D}, \phi, p) = \left(\frac{1}{M} \sum_{m=1}^{M} \|x_m - \phi(y_m)\|^p \right)^{1/p}, \tag{2.2}$$

for some $p \in (0, \infty)$ and some (pseudo) norm $\|\cdot\|$ on $\mathcal{M}_1$. For $p = 1$ this is the mean absolute error, and for $p = 2$ this is the mean squared error and is also known as empirical risk or mean squared loss, which are commonly used to evaluate the accuracy of models and also approximate inverse maps, see, e.g., [[17], Chapter 1] and [[58], Chapter 1].

A common objective in inverse problems for applications in various fields, is to find an approximate inverse map that minimizes the loss (2.2) in a given class of functions, such as finding the optimal architecture and parameter values for deep neural networks [6, 40].

The *accuracy limit* for any approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ is simply obtained by taking the infimum over all possible maps of the reconstruction error,

$$\inf_{\theta} \mathcal{L}(\mathcal{D}, \theta, p) \leqslant \mathcal{L}(\mathcal{D}, \phi, p), \tag{2.3}$$

where we take the infimum over all functions $\theta : \mathcal{M}_2 \to \mathbb{C}^{d_1}$. Such functions are kept as general as possible and do not necessarily need to be continuous. The accuracy limit in (2.3) in theory can be obtained by devising and computing infinitely many approximate inverse maps to solve (2.1). As this is impossible in practice, *the aim in this work is to find a data-driven bound for* (2.3) *that is independent of any special function class.* However, computing solutions to inverse problems can face challenges, some of which are described below.

2.2 Challenges for computing solutions to inverse problems

(i) No unique solution: In real-world scenarios, the measurements $y \in \mathcal{M}_2$ in (2.1) may only be partially available and corrupted by errors due to equipment constraints [24]. As a result, the same measurement can result from more than one unique signal $x \in \mathcal{M}_1$ and an instance of noise $e \in \mathcal{E}$. In fact, in most instances of inverse problems of interest given a measurement, there is no unique solution. For example, this occurs if the problem of recovering x given y in (2.1) is ill-posed or undersampled, unless further assumptions are made, such as the robust null space property or the restricted isometry property [30, 39]. If the forward model F is poorly conditioned or dimensionality reducing, then, given a measurement, there may not exist a unique solution. In particular, this means that for some $y \in \mathcal{M}_2$ the first component of the pre-image of the measurement y under F contains more than one element,

$$\left|\pi_1(F^{-1}(y))\right| = |\{x \in \mathcal{M}_1 : \exists e \in \mathcal{E} \text{ s.t } F(x, e) = y\}| > 1, \tag{2.4}$$

where $\pi_1 : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_1$ is the projection onto the first component $\pi_1(x, e) = x$ for $(x, e) \in \mathcal{M}_1 \times \mathcal{E}$. Note that (2.4) can be true if $\mathcal{M}_1 \subset \mathbb{C}^{d_1}$ is countable, countably finite, infinite, or uncountable.

(ii) Existence of accuracy bounds: For inverse problems where there are multiple solutions for one measurement, as in (2.4), we can only obtain approximate inverse maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$. This is equivalent to only being able to obtain a non-zero reconstruction error. In [41] we establish an accuracy-stability trade-off for approximate inverse maps that solve ill-posed inverse problems. In particular, in [41] we show that approximate inverse maps with smaller error and, thus, higher accuracy are prone to instabilities and incorrect transfer of details. This accuracy-stability trade-off that is also proved in [32] and implies that there exists a constant $K = K(F, \mathcal{M}_1, \mathcal{E}, p) > 0$ depending on F, $\mathcal{M}_1$, $\mathcal{E}$ and p that is independent of any approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ such that

$$K \leqslant \mathcal{L}(\mathcal{D}, \phi, p), \tag{2.5}$$

for all stable, or also Lipschitz continuous, $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$. Note that Lipschitz continuity is well-defined, as any normed space induces a metric space. For optimal inverse maps, previous works have provided fundamental upper and lower bounds to the average and worst-case reconstruction error [42]. By providing sharp and quantifiable lower and upper bounds to the reconstruction error of optimal inverse maps, this answers the question posed by M. Burger and T. Roith in *Learning in Image Reconstruction: A Cautionary Tale* [24],

> *"Can we ever quantify the reconstruction error [of ill-posed inverse problems]?"*

However, the provided average bounds in [42] are not computable and this remains a key challenge for many applications.

(iii) Universal approximators do not evade accuracy bounds: In theory, neural networks are universal approximators and have shown great success in obtaining a vanishing approximation error [10, 12, 45, 54, 61, 74]. As the authors of [2] point out, such approximation results typically do not provide theoretical guaranties on trainability. In fact, in [77] in a benchmark test in different realistic inverse problems, an accuracy-efficiency trade-off for diffusion models as approximate inverse maps is established. In particular, accuracy bounds to the best worst-case reconstruction error are established by some existing frameworks. For example, for noiseless linear inverse problems, there are the Gelfand widths [62] and the best k-term approximation [30]. Worst-case accuracy bounds for set-valued decoders on Banach spaces have been derived in [5] and then extended to metric spaces in [53]. For inverse problems with noise, optimal learning [18] can provide such worst-case accuracy bounds. For linear inverse problems, the notion of generalized instance optimality [23] provides point-wise upper bounds on the reconstruction error. Worst- and average- case accuracy bounds on measurable metric spaces are established in [42, 63].

2.3 Contributions

In this work, we address the challenges that were previously mentioned for computing solutions to inverse problems. Challenge (i) - *that of inverse problems* (2.1) *that have no unique solution* - is a necessary assumption for our main result. This is encapsulated in our central assumption, (2.4), that is satisfied if the forward model of (2.1) is not injective on the product set of signals and noise $\mathcal{M}_1 \times \mathcal{E}$. Due to the non-injectivity of the forward model, there exist measurements $y \in \mathcal{M}_2$ such that the inverse problem (2.1) does not have a unique solution. From this the challenge (ii) - *that there exist accuracy bounds* - follows. Unfortunately, this leads to the challenge (iii) - *that universal approximators cannot evade the accuracy bounds*. In order to address these challenges, we develop a framework to compute accuracy bounds for the reconstruction error of inverse problems. In this work, we provide computable lower and upper accuracy bounds to the empirical reconstruction error in (2.2) of any inverse map solving an inverse problem (2.1).

We now present a nontechnical summary of our main result and refer to Section 3 for the formal statement. The accuracy bounds for any approximate inverse map in (2.5) can be computed with the average kernel size. Informally, the average kernel size is the average distance between samples of the x-components of $(x, e), (x', e') \in \mathcal{M}_1 \times \mathcal{E}$ that the forward model maps to the same measurement $y = F(x, e) = F(x', e')$ over all the measurement samples. If for each measurement sample, the average kernel size is computed with an equal number of x-components we refer to this as *computed with an equal number of samples*.

Main result 2.1 (Accuracy Bounds for Computing Solutions to Inverse Problems - Theorem 3.3)**.** *Let $p \in (0, \infty)$, then for an inverse problem as in* (2.1)*, we have that*

(i) half the average kernel size computed with an equal number of samples is a sharp lower bound to the reconstruction error of any approximate inverse map. Additionally, the average kernel size is a sharp upper bound to the reconstruction error of approximate inverse maps that attain the accuracy limit in (2.3)*,*

(ii) half the average kernel size is a sharp lower bound to the reconstruction error of any measurable approximate inverse map.

The average kernel size in Theorem 3.3 provides sharp accuracy bounds that depend only on samples from the sets $\mathcal{M}_1$ and $\mathcal{E}$, the forward model F and the evaluation (pseudo) norm for the reconstruction error in (2.2). For linear F the bounds depend on its kernel, and the kernel size generalizes the concept of kernel to non-linear forward models. We also provide a software library that enables users to apply the proposed approach with minimal implementation overhead on custom data sets and problems. In Section 4 in order to demonstrate the applicability of our results and accompanying software library, we apply the framework to two characteristic examples of inverse problems with non-injective forward models. As a characteristic example for an inverse problem with a non-injective and highly non-linear forward model, we use fluorescence localization microscopy. For a large-scale inverse problem with a linear and non-injective forward model, we take super-resolution of multi-spectral satellite data, which is a common large scale problem in geoscience [64]. Our library enables computing the accuracy bounds on super-resolution of multi-spectral satellite data and other inverse problems requiring the use of GPUs based on their size.

2.4 Outlook and impact

The lower bound to the reconstruction error can be computed before designing and engineering approximate inverse maps, such as training deep neural networks or optimization methods. This is because the bounds only depend on the forward model F, and the sets $\mathcal{M}_1$ and $\mathcal{E}$. Thus, the provided accuracy bounds can give researchers and practitioners insight into model development and understanding and significantly increase computational efficiency by elucidating possible accuracy measures. Such bounds can be used as early stopping thresholds during DNN training procedures. As the bounds can be computed for a given dataset, the accuracy bounds can be used to assess the quality of datasets to train approximate inverse maps. The bounds help to satisfy the principles of numerical analysis when solving inverse problems [2] - in particular, accuracy, computational efficiency, and sample complexity.

2.5 Outline

The remainder of this paper is as follows. In Section 3 we introduce the computational framework for inverse problems and corresponding forward problems. In Section 3.2 we introduce the algorithms for computing the accuracy bounds in Section 3.3. In sections 4.1 and 4.2 we describe the forward problems for localization microscopy and super-resolution of multi-spectral satellite data, as well as present numerical results validating our theoretical results. Additional information for this paper is provided in the supplementary material. The library accompanying the computational framework is available at the following Github repository.

3 Main results

In this section, we define a framework for inverse problems and present the corresponding main theoretical results. The section is divided into three subsections. Firstly, Section 3.1 formulates inverse problems, secondly, Section 3.2 introduces an algorithm for computing the lower bound and thirdly, Section 3.3 establishes inequalities using the lower bound provided by the introduced algorithm. The proofs of the main results are referred to Section 5.

3.1 Problem formulation

We define an inverse problem as a forward problem $(F, \mathcal{M}_1, \mathcal{E})$ equipped with the task from (2.1). The dimensions in (2.1) are $d_1, d_2, d_3 \in \mathbb{N}$ and $\mathbb{C}^{d_1}$ is equipped with a pseudo-norm $\|\cdot\|$ and $\mathbb{C}^{d_2}$ is equipped with a norm $|||\cdot|||$. We equip $[0, \infty)$ with the standard Borel sigma algebra. We equip $\mathbb{C}^{d_1}$, $\mathcal{M}_2$ and $\mathcal{E}$ with sigma algebras such that $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2$ is measurable. Throughout the text, we denote by $\pi_1 \colon \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_1$ the projection $(x, e) \mapsto x$. Finally, an approximate inverse map that, from noisy measurements in the set $\mathcal{M}_2$ computes an approximation to the true solution x, is denoted by

$$\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}, \quad y \mapsto \phi(y) = \hat{x}. \tag{3.1}$$

The reconstruction error of an approximate inverse map (3.1) for approximately solving (2.1) is often quantified by the empirical expectation of the residual of the reconstruction and a solution. For $M \in \mathbb{N}$ samples

$$\{(x_m, y_m)\}_{m=1}^M = \{(x_m, F(x_m, e_m))\}_{m=1}^M \subset \mathcal{M}_1 \times \mathcal{M}_2, \tag{3.2}$$

and for $p \in (0, \infty)$, the reconstruction error is defined by (2.2). Note that in (2.2) explicitly only the samples $\{(x_m, y_m)\}_{m=1}^M \subset \mathcal{M}_1 \times \mathcal{M}_2$ are needed and not the noise samples $\{e_m\}_{m=1}^M \in \mathcal{E}$.

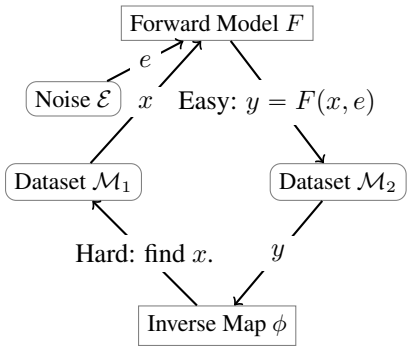


Figure 1: Inverse problem framework.

The forward model $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2$ in (2.1) is a measurable map onto its image, $\mathcal{M}_2 = F(\mathcal{M}_1 \times \mathcal{E})$. Examples of the forward model include, but are not limited to,

$$\begin{aligned} F(x, e) &= G(x) + e, \\ F(x, e) &= G(x) \odot e, \\ F(x, e) &= G(x) \odot e_1 + e_2, \end{aligned} \tag{3.3}$$

where $G \colon \mathbb{C}^{d_1} \to \mathbb{C}^{d_2}$ is a linear or non-linear forward model, and the dimension d_3 of the set $\mathcal{E}$ depends on the considered model in such a way that the point-wise multiplication $\odot$ is defined.

3.2 Algorithms for computing the average kernel sizes

In the following, we present two algorithms for computing the lower accuracy bound - the average kernel size - to the reconstruction error (2.2) and the optimal accuracy established in Theorem 3.3. The lower bound in Theorem 3.3 may be non-zero because of our central assumption that for some measurement there does not exist a unique solution. In particular, we assume that given a noisy measurement $y \in \mathcal{M}_2$, there can exist multiple feasible solutions $x \in \mathcal{M}_1$ that the forward model maps to the measurement with some realization of noise. Following [42] we denote the set of possible solutions $x \in \mathcal{M}_1$ corresponding to a given $y \in \mathcal{M}_2$ as the *feasible set* with

$$F_y := \pi_1(F^{-1}(y)) = \{x \in \mathcal{M}_1 : \exists e \in \mathcal{E} \text{ s.t. } F(x, e) = y\}. \tag{3.4}$$

This is the projection onto the first component of the pre-image of y. For many inverse problems (2.1) the feasible sets are non-singletons. This means that there are often practical cases where at least two signals $x, x' \in \mathcal{M}_1$ and noise realizations $e, e' \in \mathcal{E}$ map to the same noisy measurement $y = F(x, e) = F(x', e')$. Methods for solving ill-posed inverse problems (2.1) rely on exact or approximate inverse maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ that take noisy measurements $y \in \mathcal{M}_2$ as inputs and provide an approximation $\hat{x}$ to a solution $x \in F_y$. The core issue is that no exact or approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ can distinguish between elements in the feasible set F_y of a given measurement $y \in \mathcal{M}_2$. The basic observation that the feasible set (3.4) contains more than one element motivates the inequality in Theorem 3.3. Algorithm 1 is the first algorithm required to compute the lower bound. Algorithm 1 approximates the feasible sets (3.4) for a set of measurements, which is depicted in Figure 2.

Algorithm 1 requires a number $K \in \mathbb{N}$ of measurement samples in $\mathcal{M}_2$, a description of the dataset $\mathcal{M}_1 \times \mathcal{E}$ and a measurable forward model F. This is applicable to all the different types of forward

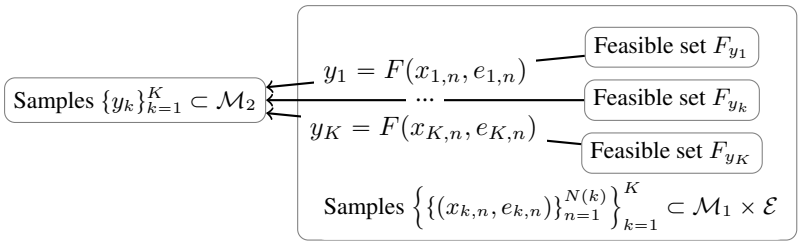


Figure 2: Allocation of x-components of samples into feasible sets with Algorithm 1. Note that the noise components can enlarge the feasible sets. For example, this occurs for additive and multiplicative noise models.

models in (3.3). For example, Section 4.1 and Section 4.2 demonstrate that Algorithm 2 is applicable to complicated forward models, such as those in localization microscopy, and for very large data dimensions $d_1, d_2 \in \mathbb{N}$, such as in multi-spectral satellite data. Given samples of measurements from an inverse problem (2.1) depicted in Figure 2, the feasible sets for each measurement sample are approximated from below in Algorithm 1. This step can be implemented in various ways, yet the samples of x-components all have to satisfy $F(x, e) = y$. This requires finding possible solutions and instances of noise that the forward model maps to the same measurement. The different possible samples of noise are a crucial part of any measurement and these are relevant in Algorithm 1, as they increase the possible samples of x-components that can map to this measurement.

Algorithm 1 Algorithm for approximating the feasible sets

Require: $K, N(K)_{\max} \in \mathbb{N}$, $\mathcal{M}_2$, $\mathcal{M}_1 \times \mathcal{E}$, F
  $\mathcal{D} = \varnothing$
  for $k \in \{1, \dots, K\}$ **do**
    $y_k \in \mathcal{M}_2$, ▷ The samples y_k can be inputs or sampled during the algorithm.
    $F_{y_k}^{N(k)} = \varnothing$
    $N(k) = 0$
    while $(x_{k,n}, e_{k,n}) \in \mathcal{M}_1 \times \mathcal{E}$ **do** ▷ Depending on problem acquire samples.
      if $F(x_{k,n}, e_{k,n}) = y_k$ **then** ▷ Implementation is forward model dependent.
        $F_{y_k}^{N(k)} \leftarrow F_{y_k}^{N(k)} \bigcup \{x_{k,n}\}$
        $\mathcal{D} \leftarrow \mathcal{D} \bigcup \{(x_{k,n}, y_k)\}$ ▷ This relates $(x_{k,n}, y_k)$ to (x_m, y_m) in (2.2).
      else if $\left|F_{y_k}^{N(k)}\right| \geqslant N(K)_{\max}$ **then** ▷ Threshold for $\mathcal{M}_1 \times \mathcal{E}$ infinite.
        break
      end if
    end while
    $N(k) = \left|F_{y_k}^{N(k)}\right|$
    $\{N(k)\}_{k=1}^K \leftarrow \bigcup_{k=1}^K N(k)$
    $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^K \leftarrow \bigcup_{k=1}^K F_{y_k}^{N(k)}$
    $M = \sum_{k=1}^K N(k)$
  end for
    return $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^K$, $\{N(k)\}_{k=1}^K$, $\mathcal{D} = \{x_m, y_m\}_{m=1}^M$

Algorithm 1 returns a list of approximations to feasible sets $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^K$ and their sizes $\{N(k)\}_{k=1}^K$ after sampling measurements and, then, elements in the feasible sets of these measurements. In order to compute the average kernel size, there are no restrictions on how the samples are acquired in Algorithm 1, such as restricting to Monte Carlo, Markov chain Monte Carlo, [27, 52] or to brute force root-finding.

Remark 3.1 (Does the Distribution of Noise Matter?). In the while loop of Algorithm 1, samples $(x_{k,n}, e_{k,n}) \in \mathcal{M}_1 \times \mathcal{E}$ are obtained. As mentioned, there is no restriction on how these samples of x- and noise components are acquired. For example, consider the set of bounded noise in a closed ball around zero and

note that this set can be equipped with different distributions to model the noise. However, the closed ball around zero contains exactly the same instantiations of noise vectors, albeit with different frequencies. As Theorem 3.3 is an analytical result, independent of any probabilistic setting, it is valid in both cases.

The output of Algorithm 1 are the input of Algorithm 2 to compute the accuracy bounds: (half) the average kernel size.

Definition 3.2 (Average Kernel Size)**.** Let $p \in (0, \infty)$. Let $(F, \mathcal{M}_1, \mathcal{E})$ be a forward problem. Fix $K, N(K)_{\max} \in \mathbb{N}$ and apply Algorithm 1 to the forward problem $(F, \mathcal{M}_1, \mathcal{E})$ to obtain $\left\{\{x_{k,n}\}_{n=1}^{N(k)}\right\}_{k=1}^{K} = \left\{F_{y_k}^{N(k)}\right\}_{k=1}^{K}$, $\{N(k)\}_{k=1}^{K}$, $\mathcal{D}$.The average kernel size is defined as

$$\operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{N(k)^2} \sum_{n'=1}^{N(k)} \sum_{n=1}^{N(k)} \|x_{k,n} - x_{k,n'}\|^p \right)^{1/p}. \tag{3.5}$$

The average kernel size is the average distance between the samples of x-components of $(x, e), (x', e') \in \mathcal{M}_1 \times \mathcal{E}$ that the forward model maps to the same measurement $F(x, e) = F(x', e') = y_k$ on all the measurement samples $\{y_k\}_{k=1}^{K}$. For linear inverse problems, this is the average size of the kernel or null space of the forward model intersected with the dataset $\mathcal{M}_1 - \mathcal{M}_1 = \{x - x' : x, x' \in \mathcal{M}_1\}$. This is the motivation to refer to the lower bound as the average kernel size for general non-linear forward models. Half of the average kernel size, the output of Algorithm 2, provides a lower bound to the reconstruction error for any approximate inverse map computed with the same data. This is because, in order to prove the inequality in Theorem 3.3, the triangle inequality and the allocation of data points that the forward model maps to the same measurements are used. Thus, the lower bound in Theorem 3.3 computed with Algorithm 2 is independent of the applied approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$. In total, Algorithm 2 only requires an order of the pseudo-norm $p \in (0, \infty)$, a list of approximations to feasible sets $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^{K} = \left\{\{x_{k,n}\}_{n=1}^{N(k)}\right\}_{k=1}^{K}$ and their sizes $\{N(k)\}_{k=1}^{K}$. Finally, Algorithm 2 computes the average kernel size (3.5).

Algorithm 2 Algorithm for computing the average kernel size with feasible sets

Require: $p \in (0, \infty)$, $K \in \mathbb{N}$, $\{F_{y_k}^{N(k)}\}_{k=1}^{K}$, $\{N(k)\}_{k=1}^{K}$
  $\operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K = 0$
  for $k \in \{1, \dots, K\}$ **do**
    $v_k = 0$
    if $N(k) \neq 0$ **then**
      while $x_{k,n}, x_{k,n'} \in F_{y_k}$ **do**
        $v_k \leftarrow v_k + \|x_{k,n} - x_{k,n'}\|^p$
      end while
      $v_k = \frac{1}{N(k)^2} v_k$
    else if $N(k) = 0$ **then**
      $v_k = 0$
    end if
    $\operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K \leftarrow \operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K + v_k$
  end for
  $\operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K = \left(\frac{1}{K} \operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K\right)^{1/p}$
  return $\operatorname{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K$

Note that the average kernel size depends on the parameter $p \in (0, \infty)$ and this parameter is fixed by the relevant loss function. Moreover, the average kernel size depends on the number of measurement samples and samples in the feasible sets. This means that the values of the loss and average kernel size may vary, but optimizing these is out of the scope of this work and highly dependent on the considered inverse problem. However, even in this general setting it is possible to prove sharp accuracy bounds for computing solutions to inverse problems, as our main result, Theorem 3.3, shows.

3.3 Sharp accuracy bounds for computing solutions to inverse problems

Our main theorem provides a lower bound to the reconstruction error (2.2) that any measurable approximate inverse map for any inverse problem of the form (2.1) can attain. In particular, the lower bound is independent of the approximate inverse map applied - or method to obtain an approximate solution - and depends only on the forward model F, the data set $\mathcal{M}_1$ and the set of noise $\mathcal{E}$, as well as evaluation tools, such as the pseudo-norm $\|\cdot\|$ on $\mathbb{C}^{d_1}$ and norm $|||\cdot|||$ on $\mathbb{C}^{d_2}$ and sample number $K \in \mathbb{N}$.

Theorem 3.3 (Accuracy Bounds for Computing Solutions to Inverse Problems)**.** *Let $p \in (0, \infty)$. Let $(F, \mathcal{M}_1, \mathcal{E})$ be a forward problem and $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ be a measurable approximate inverse map. Fix $K, N(K)_{\max} \in \mathbb{N}$ and apply Algorithm 1 to the forward problem $(F, \mathcal{M}_1, \mathcal{E})$ to obtain $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^{K}$, $\{N(k)\}_{k=1}^{K}$, $\mathcal{D}$. Input p, K, $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^{K}$ and $\{N(k)\}_{k=1}^{K}$ into Algorithm 2 to obtain the average kernel size. Then*

*(i) if $N(k) = N(K)$ for all $k \in \{1, \ldots, K\}$ we have that **half the average kernel size is a lower bound** to the reconstruction error of any approximate inverse map ϕ on the dataset $\mathcal{D} = \{(x_m, y_m)\}_{m=1}^{M}$,*

$$\frac{1}{2}\,\mathrm{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K \leqslant \mathcal{L}(\mathcal{D}, \phi, p),$$

where $M = KN(K)$. Additionally, we have that

$$\frac{1}{2}\,\mathrm{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K \leqslant \inf_{\phi} \mathcal{L}(\mathcal{D}, \phi, p) \leqslant \mathrm{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K,$$

where the infimum is taken over all approximate inverse maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$. The approximate inverse map

$$\theta(y_k) := \operatorname*{argmin}_{z \in \mathbb{C}^{d_1}} \frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - z\|^p,$$

is one map that attains the infimum.

*(ii) **half the average kernel size is a lower bound** to the reconstruction error of any measurable approximate inverse map ϕ on the dataset $\mathcal{D} = \{(x_m, y_m)\}_{m=1}^{M}$,*

$$\frac{1}{2}\,\mathrm{Kersize}(F, \mathcal{M}_1, \mathcal{E}, p)_K \leqslant \mathcal{L}(\mathcal{D}, \phi, p),$$

where $M = \sum_{k=1}^{K} N(k)$.

Part (i) of Theorem 3.3 in combination with Algorithm 2 and Algorithm 1 provides a practical recipe for quantifying the lower accuracy bound to the reconstruction error of any approximate inverse map. This is useful in various settings. For example, for deep learning engineers that aim to fit an optimal model or network architecture as an approximate inverse map for a given inverse problem on a given dataset with supervised training. Part (i) of Theorem 3.3 provides an upper bound, in terms of the average kernel size, on the accuracy of the reconstruction for the optimal approximate inverse maps that attain the infimum in the accuracy limit (2.3). Moreover, in Section **SM** 1 we show that the lower bound is sharp - in the sense that for a specific forward model, dataset, and set of noise, as well as approximate inverse map, the lower is attained. Additionally, if $N(k) \neq N(k')$ for some $k, k' \in \{1, \ldots, K\}$ with $k \neq k'$, then part (ii) of Theorem 3.3 holds for all approximate inverse maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ that are measurable.

Remark 3.4 (Application to Finite Datasets $\mathcal{M}_1$). Consider $M' \in \mathbb{N}$ and $\mathcal{D}' = \{(x_m, y_m)\}_{m=1}^{M'} \subset \mathcal{M}_1 \times \mathcal{M}_2$ with a given forward model F and noise $\mathcal{E}$. Then, Theorem 3.3 in combination with Algorithm 2 and Algorithm 1 now enable to compute accuracy bounds before even fitting a model to a given dataset. The relation between the dataset $\mathcal{D}' = \{(x_m, y_m)\}_{m=1}^{M'}$ of size $M' \in \mathbb{N}$ and the allocation to feasible sets $\{F_{y_k}^{N(k)}\}_{k=1}^{K}$ and dataset $\mathcal{D}$ is provided by 1. In particular, by setting $I = |\pi_2(\mathcal{D}')|$, $\mathcal{M}_2 = \pi_2(\mathcal{D}')$ and $\mathcal{M}_1 = \pi_1(\mathcal{D}')$ and letting $N(K)_{\max} = |\mathcal{M}_1| \in \mathbb{N}$ the full dataset can be allocated into feasible sets and a

larger dataset of size $M \geqslant M'$. This means that the output dataset $\mathcal{D}$ of Algorithm 1 is a super-set of the input data, $\mathcal{D}' \subset \mathcal{D}$. However, note that depending on the set of noise and forward model, the size of the feasible sets may be one. In this case, the average kernel size is zero and, in order to find x-components that are mapped onto a given measurement by the forward model, dataset enlargement techniques may be needed.

3.4 Fast lower bound

For forward models $F : \mathcal{M}_1 \times \mathcal{E} \to \mathcal{M}_2$ of the form $F(x, e) = Ax + e$, with $A \in \mathbb{C}^{d_2 \times d_1}$ and $(x, e) \in \mathcal{M}_1 \times \mathcal{E}$, Theorem 3.3 in combination with Algorithm 2 for a dataset of size M can be optimized with an additional symmetry assumption from the computational complexity of $\mathcal{O}(M^2)$ down to $\mathcal{O}(M)$. The symmetry assumption in (3.6) is equivalent to assuming that the reflection of a vector around the orthogonal component to the kernel of the forward model is a realistic data point. If the forward model F is a linear map on $\mathcal{M}_1 \times \mathcal{E}$, we can compute the projection onto the kernel of F with its Moore-Penrose inverse, [14, 50]. In the following $P_{\mathcal{N}(F)} : \mathcal{M}_1 \times \mathcal{E} \to \mathbb{C}^{d_1}$ denotes the projection onto the kernel of the forward model F and is computed by $P_{\mathcal{N}(F)} = \mathbb{1} - F^\dagger F$, where $F^\dagger$ is the Moore-Penrose inverse of F.

Theorem 3.5 (Fast Lower Bound for Functions on the Measurement Data)**.** *Let $p \in (0, \infty)$. Let $(F, \mathcal{M}_1, \mathcal{E})$ be a forward problem with $F(x, e) = Ax + e$ with $A \in \mathbb{C}^{d_2 \times d_1}$ for $(x, e) \in \mathcal{M}_1 \times \mathcal{E}$. Let $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ be an approximate inverse map. Let $\mathcal{D}' = \{(x_m, y_m)\}_{m=1}^{M'} \subset \mathcal{M}_1 \times \mathcal{M}_2$. Extend $\mathcal{D}'$ to $\mathcal{D}$ by adding*

$$(x', y_m) = \left(\pi_1\left((x_m, e_m) - 2P_{\mathcal{N}(F)}(x_m, e_m)\right), y_m\right) \text{ to } \mathcal{D}, \tag{3.6}$$

for all $x_m \in \pi_1(\mathcal{D}')$ with $e_m \in \mathcal{E}$ such that $F(x_m, e_m) = y_m$. Then,

*(i) we have that the **average symmetric kernel size** is a lower bound to the reconstruction error of any approximate inverse map ϕ on the dataset $\mathcal{D} = \{(x_m, y_m)\}_{m=1}^{M}$, with $M = 2M'$*

$$\mathrm{SKersize}(F, \mathcal{M}_1, \mathcal{E}, p)_M := \left(\frac{1}{M'} \sum_{m=1}^{M'} \|v_m\|^p\right)^{1/p} \leqslant \mathcal{L}(\mathcal{D}, \phi, p),$$

with $\|v_m\| = \|\pi_1(P_{\mathcal{N}(F)}(x_m, e_m))\|$ for $m \in \{1, \ldots, M'\}$.

(ii) additionally, we have that

$$\mathrm{SKersize}(F, \mathcal{M}_1, \mathcal{E}, p)_M \leqslant \inf_{\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}} \mathcal{L}(\mathcal{D}, \phi, p) \leqslant 2\,\mathrm{SKersize}(F, \mathcal{M}_1, \mathcal{E}, p)_M,$$

where the infimum is taken over all approximate inverse maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$.

4 Computing accuracy bounds in applied inverse problems

In the following, we apply the framework from Section 3.1 to two common inverse problems that arise in localization microscopy and super-resolution of multi-spectral satellite data. First, we formulate the inverse problems in the proposed framework. Then, we empirically verify Theorem 3.3 and Theorem 3.5 by computing the lower and upper accuracy bounds based on the average (symmetric) kernel size and, respectively, the loss.

4.1 Accuracy bounds in microscopy

Localization microscopy aims to localize point-like light-sources at the highest possible resolution that attains the system's inherent accuracy limits [51, 67]. With the average kernel size, the upper and lower bounds of the optimum localization error in microscopy can be quantified. In previous work in localization microscopy, the inverse of the Fischer information matrix is computed as a lower bound to the variance of an estimator for the localization [59]. The work of [59] uses the classical Cramer-Rao bound [35, 47, 65, 75], which bounds the variance of an unbiased estimator by the inverse Fischer information matrix. For the derivation of this lower bound in [59] the following assumptions are made. Firstly,

the estimator is unbiased (i.e., an estimation procedure whose mean produces the correct result), second, the spatial coordinates of the detected photons are independent and identically distributed random variables, and thirdly, the counting process and the random variables that describe the spatial coordinates are mutually independent. Unfortunately, these assumptions are not satisfied in many applications of interest to practitioners [51]. For example, as there may exist multiple possible locations corresponding to a measurement data sample, obtaining an unbiased estimator may be impossible, even with a very large sample size. As Theorem 3.3 holds under more general assumptions, part (i) of Theorem 3.3 provides a lower bound to the localization accuracy of any estimator in a broader context. Additionally, the best estimators for a given dataset can be determined using part (ii) of Theorem 3.3.

4.1.1 Localization microscopy as an inverse problem

In the following, we give a brief conceptual introduction to single-molecule localization microscopy as an inverse problem. For further details, see the reviews on [51, 72]. In localization microscopy, one objective is to recover the position of a particle in a volume $\mathcal{V} \subset \mathbb{R}^3$ and the background photon flux $C \in [0, C_{max}]$ and the particle photon emission rate $h \in [0, h_{max}]$ given a detection discretized into $d_2^x \times d_2^y$ pixels for $d_2^x, d_2^y \in \mathbb{N}$. In summary, one tries to reconstruct $\Theta = (x, y, z, C, h)$ from noisy measurements of pixel-wise intensities at pixels $(p_x, p_y) \in \mathcal{I}_{d_2} = \{1, \dots, d_2^x\} \times \{1, \dots, d_2^y\}$. Hence, the set of signals is given by $\mathcal{M}_1 = \mathcal{V} \times [0, C_{max}] \times [0, h_{max}] \subset \mathbb{R}^5$, the set of noise is given by $\mathcal{E} = \mathcal{B}(0, \mathbb{1}\epsilon_1) \times \mathcal{B}(0, \mathbb{1}\epsilon_2) \subset \mathbb{R}^{2(d_2^x \times d_2^y)}$ and the forward model is given by

$$\begin{aligned} F : \mathcal{M}_1 \times \mathcal{E} &\to [0, I_{max}]^{d_2^x \times d_2^y} \\ (\Theta, e) &\mapsto F(\Theta, e) = w(\Theta, e) := \big(w_{(p_x,p_y)}(\Theta, e)\big)_{(p_x,p_y)\in\mathcal{I}_{d_2}} . \end{aligned} \tag{4.1}$$

The measured intensity $w_{(p_x,p_y)}(\Theta, e)$ for $(p_x, p_y) \in \mathcal{I}_{d_2}$ results from a mixture of additive and multiplicative noise and a non-linear forward model, for details, see Section **SM** 2.1. In this setting, the inverse problem in the framework of (2.1) is given by

$$\text{recover } \Theta \text{ given data } w(\Theta, e) = F(\Theta, e) \text{ of } \Theta \in \mathcal{M}_1 \subset \mathbb{R}^5 \text{ and } e \in \mathcal{E} \subset \mathbb{R}^2. \tag{4.2}$$

The key point here is that the inverse problem of localization microscopy in (4.2) satisfies our central assumption (2.4). In particular, two different $\Theta, \Theta' \in \mathcal{M}_1$ with $\Theta \neq \Theta'$ and noise samples $e, e' \in \mathcal{E}$ can be mapped to the same noisy measurement

$$F(\Theta, e) = F(\Theta', e'). \tag{4.3}$$

Now (4.3) states that two correct results can lead to exactly the same noisy measurement. As there is more than one possible correct result, and hence, the average kernel size is non-zero, and, hence, there is no estimation procedure that can yield a mean which is the unique correct result. As (2.4) is satisfied, this motivates computing the accuracy bounds, the average kernel sizes, and validating the lower bound in Theorem 3.3.

4.1.2 Numerical experiments for computing the average kernel size

For the inverse problem of fluorescence microscopy (4.2) any method providing an estimator of $\hat{\Theta}$ is an approximate inverse map $\hat{\Theta} : \mathcal{M}_2 \to \mathbb{R}^5$ taking noisy measurements $w \in \mathcal{M}_2$ as input and producing approximations to the location and background photon flux and particle photon emission rate Θ, $\phi(w) = \hat{\Theta}$. We adapt the algorithm for allocating samples to the feasible sets Algorithm 1 for the particular forward model of fluorescence microscopy (4.1). In particular, the condition of assigning samples to feasible sets $F(\Theta, e) = w(\Theta, e)$, is met by obtaining MCMC samples where the noise samples are within the desired noise model. Details on the MCMC sampling are given in Section **SM** 2.3. We use samples from four imaging set-ups, where the different set-ups $\{A_1, A_2, A_3, A_4\}$ correspond to different background photon fluxes and particle photon emission rates. The best set-up is A_1, which has a low background photon flux and a high particle photon emission rate, and the worst is A_4. Using these samples, we apply Algorithm 1 and Algorithm 2 to compute the average kernel size and the root mean squared error (RMSE) loss for different estimators from these simulated localization microscopy datasets.

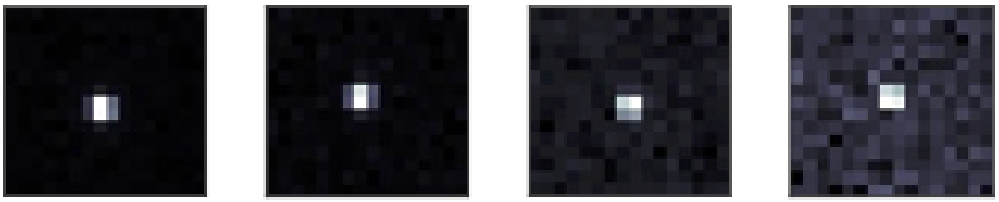
Estimators	A1		A2		A3		A4	
Dataset Size	25	25×1501	25	25×1501	25	25×1501	25	25×1501
Mean	2.917	2.634	3.905	3.965	6.559	6.296	12.195	11.635
Median	2.634	2.634	3.965	3.965	6.294	6.294	11.632	11.632
MAP	3.611	3.611	5.089	5.089	9.318	9.318	16.923	16.923
ML	4.248	3.719	5.338	5.812	9.453	8.153	17.009	16.675
$\frac{1}{2}$ Kersize		1.863		2.804		4.451		8.225

Table 1: RMSE in $[nm]$ with for different estimators on the dataset consisting of the ground truth and on the dataset output of Algorithm 1 for microscope set-ups $\{A_1, A_2, A_3, A_4\}$. The last row is the corresponding lower accuracy bound, half the average kernel size $\frac{1}{2}\,\mathrm{Kersize}(F, \mathcal{M}_1, \mathcal{E}, 2)_K$, for microscope set-ups $\{A_1, A_2, A_3, A_4\}$ with $K = 25$ and $N(K) = 1501$.

For each set-up $\{A_1, A_2, A_3, A_4\}$ we consider $K = 25$ different noisy measurements $\{w_k\}_{k=1}^{25} \subset \mathcal{M}_2 = F(\mathcal{M}_1 \times \mathcal{E})$. For each example measurement, we approximate the feasible sets with $N(25) = 1501$ MCMC samples. We use a pseudo-norm, which is the ℓ_2 norm composed with the projection onto the first two spatial components $\pi_{1,2}(\Theta) = \Theta_{x,y} = (x, y)$, to compute the average kernel size as in (3.5). We compute the RMSE loss $\mathcal{L}(\mathcal{D}, \phi, 2)$ from (2.2) for the mean, median, maximum a posteriori, and maximum likelihood estimators to validate the lower accuracy bound in part (i) of Theorem 3.3. The scatter plots of the RMSE against half the average kernel size in Figure 3 validate Theorem 3.3 for each measurement and microscope set-up. Moreover, the mean and median estimators are consistently optimal and within the accuracy bounds of Theorem 3.3, (ii). Additionally, we compute the RMSE on the input datasets and the output datasets of Algorithm 1 in Table 1. Table 1 shows that the loss remains consistently in the same range for different set-ups and estimators across the input datasets and the output datasets of Algorithm 1. A convergence study of the number of samples in the feasible set is provided in Figure **SM** 2.

4.2 Accuracy bounds for super-resolution of multi-spectral satellite data

Single image super-resolution converts a low-resolution image with coarse details to a corresponding high-resolution image with higher visual quality and refined details. Even if prior information is utilized, there can exist multiple high-resolution solutions for the same low-resolution image. Thus, the inverse problem of super-resolution satisfies our central assumption (2.4). Additionally, assessing the quality of the output is not straightforward, there are many different quantitative metrics - such as the peak signal to noise ratio (PSNR) or the structural similarity index (SSIM) - and these metrics only loosely correlate to human perception [4]. Another key challenge is to determine the limit to the super-resolution factor. This is particularly difficult in super-resolution of multi-spectral satellite data due to dataset sizes and the resulting need for quantitative methods. A resulting challenge is to choose the optimal fraction between high and low resolution for different senors and selections of spectral bands. This question is unanswered. However, the question that is tackled in [8], is what the best spatial resolution for obtaining the optimum accuracy on a down-stream task, such as land cover classification. The authors [8] claim that the spatial resolution with the maximal local variance is optimal. The underlying assumption is that scenes are assumed to be composed of objects with different reflectance values on a continuous background and that if the spatial resolution is that of the objects the local variance will be maximal [36]. Similarly, in [55], as early as 1911, and [69] in 1938, the authors empirically investigate the relation between local variance and spatial resolution. Part (i) of Theorem 3.5 provides a lower bound depending on the super-resolution factor to the super-resolution error of any method on a given dataset. Additionally, the best super-resolution methods on a given dataset can be determined with the upper bound, in part (ii) of Theorem 3.5. In the following, we only compute the RMSE to validate the accuracy bounds provided by Theorem 3.5. However, as mentioned, determining the best super-resolution method on a given dataset is dependent on the norm and evaluation score and beyond the scope of this work.



(a) Different noisy measurements per microscope set-up $\{A_1, A_2, A_3, A_4\}$ with increasing background photon flux and decreasing particle photon emission rate from left to right.

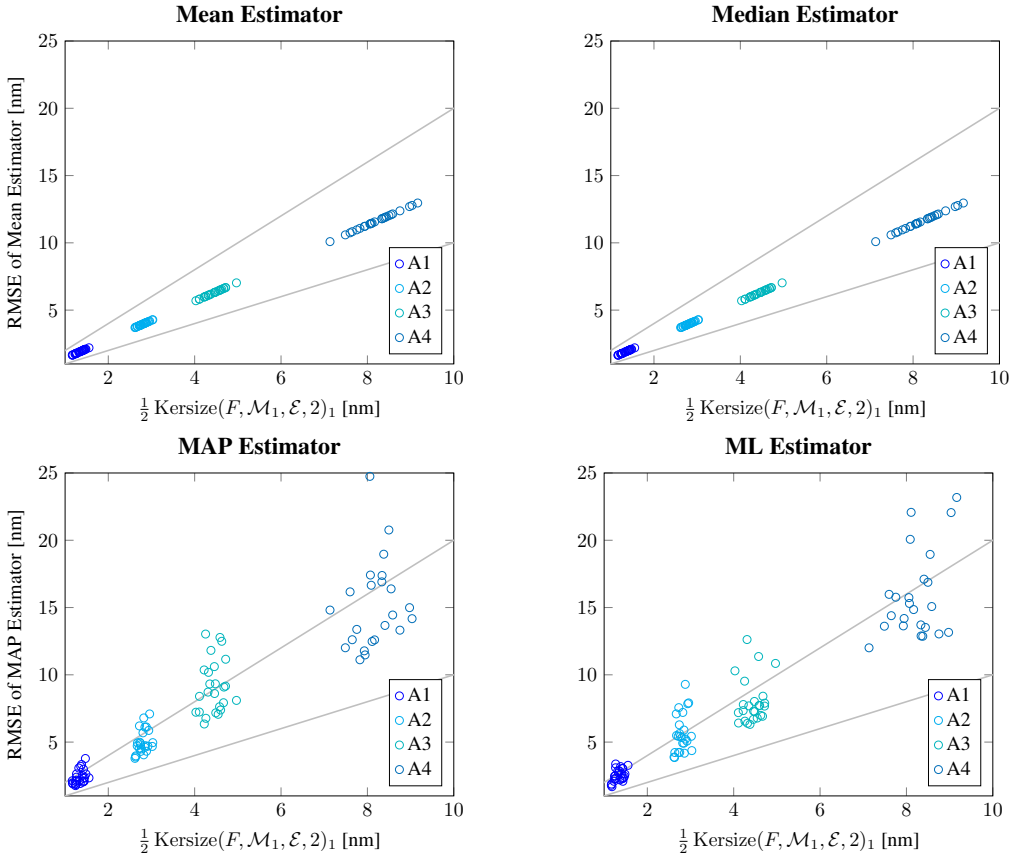
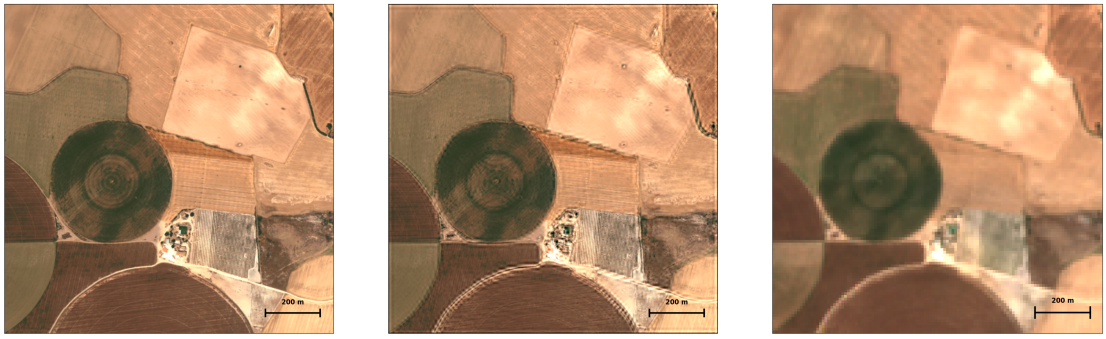


Figure 3: RMSE of the $(x, y)-$location from different estimators against half of the average kernel size computed with feasible sets of size $N(1) = 15001$ with $K = 1$ for 25 different noisy measurements per microscope set-up $\{A_1, A_2, A_3, A_4\}$. The lower and upper gray lines ($y = x$ and $y = 2x$) visualize the lower and upper accuracy bounds. All RMSE values are above the computed lower accuracy bounds and validate Theorem 3.3, (i). The mean and median estimator's RMSE are below the upper accuracy bound. The different microscope set-ups $\{A_1, A_2, A_3, A_4\}$ highlight the increase of the kernel size with worsening imaging settings - a higher background photon flux and decreasing particle photon emission rate.

4.2.1 Spatial super-resolution of multi-spectral satellite data as an inverse problem

In super-resolution of multi-spectral satellite data the aim is to enhance spatial resolution of multi-dimensional arrays encoding the reflectance values of different spectral bands at possibly different spatial resolutions. For fourfold super-resolution, the task is to obtain a higher resolution satellite image $x \in [0, r_{max}]^{d^s \times 4d_1^x \times 4d_1^y}$ where $d^s \in \mathbb{N}$ is the number of spectral bands, $d_1^x, d_1^y \in \mathbb{N}$ are the number of pixels, and $r_{max} > 0$ is the maximum reflectance value, from a lower resolution image $y \in [0, r_{max}]^{d \times d_1^x \times d_1^y}$. We consider the simplest possible forward model that band-wise maps the higher resolution to lower resolution data with bilinear interpolation and additive noise $y_s = DS(x_s) + e_s$, for $s \in \{1, \dots, d^s\}$ where $e_s \in \mathcal{B}(0, \mathbb{1}\epsilon) \subset \mathbb{R}^{d_1^x \times d_1^y}$, $x \in [0, r_{max}]^{4d_1^x \times 4d_1^y}$ and $y \in [0, r_{max} + \epsilon]^{d_1^x \times 4d_1^y}$. For this down-sampling, each pixel in the output only depends on its neighboring pixels within a distance of



(a) High Resolution (b) Reflected Version (c) Low Resolution

Figure 4: RGB example visualizations of different versions of a multi-spectral satellite Croplands image from the NAIP data set. The images correspond to (a) the original high-resolution image, (b) a high-resolution version reflected around the orthogonal component of the downsampling maps kernel (according to 3.6), and (c) the low-resolution image computed with the down sampling operator.

two pixels. Hence, the set $\mathcal{M}_1 = [0, r_{max}]^{d^s \times 4d_1^x \times 4d_1^y}$ and the forward model are given by

$$\begin{aligned} F : \mathcal{M}_1 \times \mathcal{B}(0, \mathbb{1}\epsilon) &\to [0, r_{max} + \epsilon]^{d^s \times d_1^x \times d_1^y} \\ (x, e) &\mapsto F(x, e) = DS(x) + e. \end{aligned} \tag{4.4}$$

As the inverse problem of super-resolution satisfies our central assumption (2.4), we validate the lower bound Theorem 3.5 by computing the average symmetric kernel size and the loss in the symmetrized dataset.

4.2.2 Numerical experiments

We compute the average symmetric kernel size from Theorem 3.5. Here we use the Moore-Penrose inverse to compute the projection onto the kernel of the down-sampling forward model, see Section **SM** 4 for further details. In the numerical experiments, in addition to our library, we use the computational framework to evaluate methods for super-resolution of multi-spectral satellite data from [9], see Section **SM** 3.3 for further details. We use four different multi-spectral satellite image datasets: one from the National Agriculture Imagery Program (NAIP), one from SPOT imagery that was obtained from the worldstat dataset and the SPAIN- URBAN and CROPS datasets. We use a dataset consisting of four different datasets from the opensr-test benchmark dataset collection. The datasets consist of high-resolution aerial imagery at 2.5 m ground sampling distance captured in the visible and near-infrared spectrum (RGBNIR) and all Sentinel-2 L1C and L2A spectral bands. The satellite images cover diverse land cover classes ranging from urban areas over crop fields to forests. For fluorescence microscopy we compute the RMSE loss $\mathcal{L}(\mathcal{D}, \phi, 2)$ from (2.2) for different super-resolutions maps to validate the lower accuracy bound in part (i) of Theorem 3.5. We use the approximate inverse map introduced by [37], referred to as Diffusion (Open SR) map in the following and bilinear upscaling and bicubic interpolations from the pytorch library [60]. The scatter plots of the RMSE against half the average symmetric kernel size in Figure 6 validate part (i) of Theorem 3.5 for each super-resolution map, each low-resolution measurement and land cover class. The assumption that the satellite image reflected around the orthogonal component of the kernel of the forward model is a realistic image in Theorem 3.5 is visually validated in Figure 4. Moreover, all considered super-resolution maps are consistently optimal and within the accuracy bounds of Theorem 3.5, (ii). Additionally, as a loss we compute the RMSE on the whole input datasets and the whole output datasets of Algorithm 1 in Table 2. Table 2 shows that the RMSE remains consistently in the same range for different land cover classes and super-resolution methods across the input datasets and the output datasets of Algorithm 1. In Section **SM** 3.4 we show that the RMSE on the dataset of the selected methods is in the range of the RMSE obtained in previous work. By computing the average symmetric kernel size, we can provide a theoretically sound and empirically validated bound to modeling accuracy. These accuracy bounds, computed before modeling, can help determine the optimal spatial and temporal resolution of earth observation images, which are fundamental bottlenecks in downstream applications [15, 29, 49, 64].

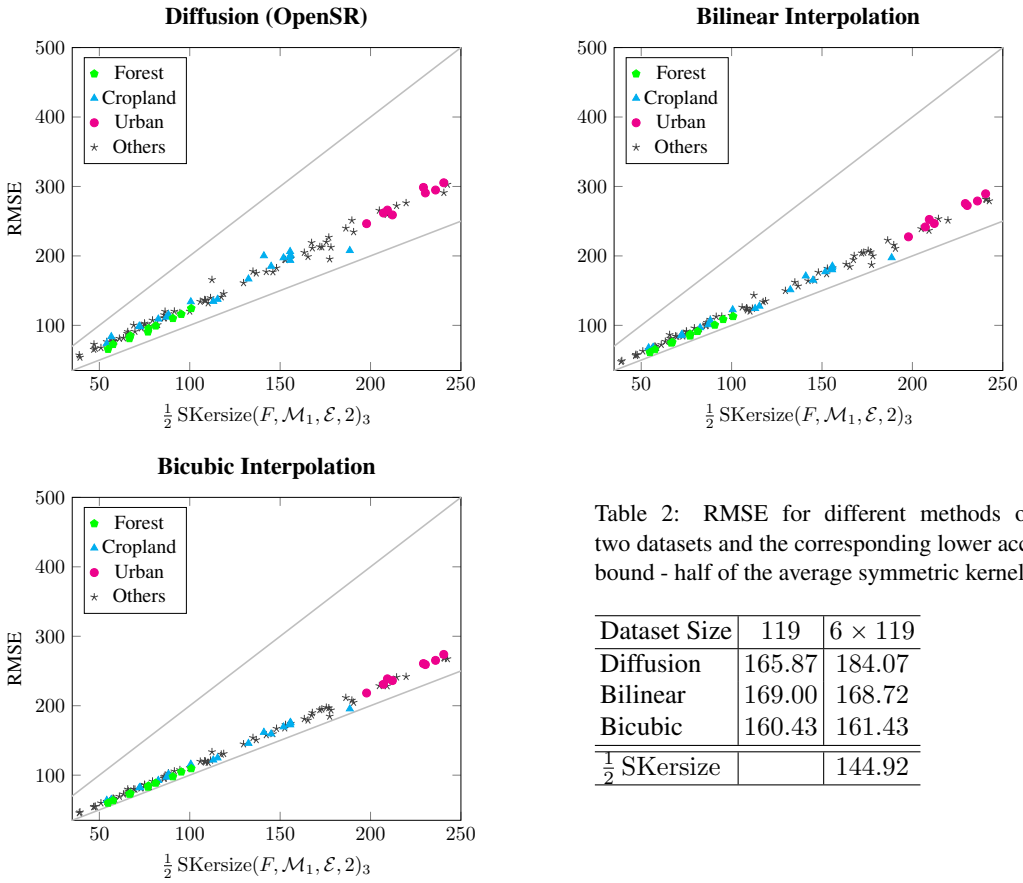


Table 2: RMSE for different methods on the two datasets and the corresponding lower accuracy bound - half of the average symmetric kernel size.

Dataset Size	119	6×119
Diffusion	165.87	184.07
Bilinear	169.00	168.72
Bicubic	160.43	161.43
$\frac{1}{2}$ SKersize		144.92

Figure 6: RMSE of different super-resolution methods against the average symmetric kernel size for triplets, $M = 3$, of 119 of images from the NAIP, SPAIN Crops and Urban and SPOT datasets. The lower and upper gray lines ($y = x$ and $y = 2x$) visualize the lower and upper accuracy bounds. All RMSE values are within the computed accuracy bounds. The separation into urban, cropland and forest images highlight the increase of the kernel size with increasing detail content of the images. The columns in Table 2 show the RMSE for different super-resolution methods on the dataset consisting of the ground truth and low resolution input with $M = 119$ and an enlarged dataset with $2M = 6 \times 119$. The output of Algorithm 1, with feasible sets of size $N(K) = 3$, is used in addition to Theorem 3.5 to obtain an enlarged dataset of size $2M = 6 \times 119$.

5 Proofs of the main results

We now prove the main results.

5.1 Proof of Theorem 3.3

Proof of Theorem 3.3. We start with the proof of part (i). Let $\{y_k\}_{k=1}^K \subset \mathcal{M}_2 = F(\mathcal{M}_1 \times \mathcal{E})$. Now obtain $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^K$ from Algorithm 1. For each distinct $y_k \in \mathcal{M}_2$ we obtain a lower approximation to the feasible set $\pi_1(F^{-1}(y_k))$ defined in (3.4) with $F_{y_k}^{N(k)} = \{x_{k,n}\}_{n=1}^{N(k)}$. In the following, we assume without loss of generality that the size of each feasible set satisfies $N(k) > 0$ for all $k \in \{1, \dots, K\}$. Let $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ be an approximate inverse map. Now using the inequality $\|x + x'\|^p \leqslant 2^{p-1}(\|x\|^p + \|x'\|^p)$, we obtain for fixed $k \in \{1, \dots, K\}$ with $y_k \in \mathcal{M}_2$ and fixed $n, n' \in \{1, \dots, N(k)\}$ with $x_{k,n}, x_{k,n'} \in F_{y_k}^{N(k)}$ that

$$\begin{aligned} \|x_{k,n} - x_{k,n'}\|^p &= \|x_{k,n} - \phi(y_k) + \phi(y_k) - x_{k,n'}\|^p \\ &\leqslant 2^{p-1} \left(\|x_{k,n} - \phi(y_k)\|^p + \|\phi(y_k) - x_{k,n'}\|^p\right). \end{aligned} \tag{5.1}$$

Now, averaging over all $n, n' \in \{1, \dots, N(k)\}$, and thus all $x_{k,n}, x_{k,n'} \in F_{y_k}^{N(k)}$, yields

$$\begin{aligned} &\frac{1}{N(k)^2} \sum_{n=1}^{N(k)} \sum_{n'=1}^{N(k)} \|x_{k,n} - x_{k,n'}\|^p \\ &\qquad \leqslant \frac{2^{p-1}}{N(k)^2} \sum_{n=1}^{N(k)} \sum_{n'=1}^{N(k)} \left(\|x_{k,n} - \phi(y_k)\|^p + \|\phi(y_k) - x_{k,n'}\|^p\right). \end{aligned} \tag{5.2}$$

Now we use the fact that for a constant $c > 0$, we have that $\frac{1}{N(k)} \sum_{n=1}^{N(k)} c = c$ and that $\phi(y_k)$ is independent of $n \in \{1, \dots, N(k)\}$, to obtain

$$\begin{aligned} &\frac{1}{N(k)^2} \sum_{n=1}^{N(k)} \sum_{n'=1}^{N(k)} \|x_{k,n} - x_{k,n'}\|^p \\ &\qquad \leqslant \frac{2^{p-1}}{N(k)^2} \sum_{n=1}^{N(k)} \sum_{n'=1}^{N(k)} \left(\|x_{k,n} - \phi(y_k)\|^p + \|\phi(y_k) - x_{k,n'}\|^p\right) \\ &\qquad = \frac{2^p}{N(k)} \sum_{n=1}^{N(k)} \|x_{k,n} - \phi(y_k)\|^p. \end{aligned} \tag{5.3}$$

Now we sum over all $k \in \{1, \dots, K\}$ and divide by $2^p K$, to obtain

$$\frac{1}{2^p K} \sum_{k=1}^{K} \frac{1}{N(k)^2} \sum_{n=1}^{N(k)} \sum_{n'=1}^{N(k)} \|x_{k,n} - x_{k,n'}\|^p \leqslant \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N(k)} \sum_{n=1}^{N(k)} \|x_{k,n} - \phi(y_k)\|^p. \tag{5.4}$$

In the following we prove for $M = \sum_{k=1}^K N(k)$, $\left\{F_{y_k}^{N(k)}\right\}_{k=1}^K$ and $\mathcal{D} = \{(x_m, y_m)\}_{m=1}^M$ from Algorithm 1 and for any $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$, that the right hand side of (5.4) satisfies

$$\frac{1}{K} \sum_{k=1}^{K} \frac{1}{N(k)} \sum_{n=1}^{N(k)} \|x_{k,n} - \phi(y_k)\|^p \leqslant \mathcal{L}(\mathcal{D}, \phi, p)^p = \frac{1}{M} \sum_{m=1}^{M} \|x_m - \phi(y_m)\|^p. \tag{5.5}$$

We now prove that if $N(k) = N(K)$ for all $k \in \{1, \dots, K\}$, we have equality in (5.5). If $N(k) = N(K)$ for all $k \in \{1, \dots, K\}$, then $M = \sum_{k=1}^K N(k) = KN(K)$. Multiplying the reconstruction error over the dataset $\mathcal{D}$ from Algorithm 1 by M yields,

$$M\mathcal{L}(\mathcal{D}, \phi, p)^p = \sum_{m=1}^{M} \|x_m - \phi(y_m)\|^p. \tag{5.6}$$

Now we introduce two bijective maps to reassign the summation indices. Later we will use these introduced maps on the indices to prove part (ii), where $N(k) \neq N(k')$ for some $k, k' \in \{1, \dots, K\}$ with $k \neq k'$. We define a surjective function that maps the indices in $\{1, \dots, M\}$ to the index of the corresponding feasible set by

$$\begin{aligned} g : \{1, \dots, M\} &\to \{1, \dots, K\} \\ m \mapsto g(m) &= \operatorname*{argmin}_{k \in \{1, \dots, K\}} \left(2 - \mathbb{1}_{0 < m - \sum_{j=0}^{k-1} N(j)}(k) - \mathbb{1}_{0 \leqslant \sum_{j=1}^{k} N(j) - m}(k)\right). \end{aligned} \tag{5.7}$$

For any $m \in \{1, \dots, M\}$ we have that $g(m) = k \in \{1, \dots, K\}$. Moreover, as by definition $M = \sum_{k=1}^{K} N(k)$, we have that m and $k = g(m)$ are such that

$$0 \leqslant \sum_{j=0}^{k-1} N(j) < m \leqslant \sum_{j=1}^{k} N(j) \leqslant M. \tag{5.8}$$

From (5.8) for any $k \in \{1, \dots, K\}$ we have that

$$\varnothing \neq g^{-1}(k) = \left(\sum_{j=0}^{k-1} N(j), \sum_{j=1}^{k} N(j)\right] \cap \mathbb{N} \subset \{1, \dots, M\}. \tag{5.9}$$

Hence, g is surjective. Now we show that for $k' \neq k$, $g^{-1}(k') \cap g^{-1}(k) = \varnothing$. Let $k' \neq k$, without loss of generality we have that $k < k'$ and $k' = k + 1$. Then, we have that $(\sum_{j=0}^{k-1} N(j), \sum_{j=1}^{k} N(j)] \cap (\sum_{j=0}^{k} N(j), \sum_{j=1}^{k+1} N(j)] = \varnothing$ and, thus $g^{-1}(k') \cap g^{-1}(k) = \varnothing$. Now we define a bijective function depending on the function g to reassign the indices of samples to samples allocated to the feasible sets by

$$\begin{aligned} u : \{1, \dots, M\} &\to \bigcup_{k=1}^{K} \{k\} \times \{1, \dots, N(k)\} \\ m \mapsto u(m) &= \left(g(m), m - \sum_{k=1}^{g(m)-1} N(k)\right). \end{aligned} \tag{5.10}$$

Note that $u : \{1, \dots, M\} \to \bigcup_{k=1}^{K} \{k\} \times \{1, \dots, N(k)\}$ is surjective, as the function g defined in (5.7) is surjective and as, by definition, we have that

$$\begin{aligned} &\{1, \dots, N(g(m))\} \\ &\quad = \left\{ n \in \mathbb{N} : \quad n = m' - \sum_{k=1}^{g(m)-1} N(k) \text{ for } m' \in \left(\sum_{k=0}^{g(m)-1} N(k), \sum_{k=1}^{g(m)} N(k)\right] \cap \mathbb{N} \right\}. \end{aligned}$$

Moreover, the function u is injective, as for $m \neq m' \in \{1, \dots, M\}$, $u(m) \neq u(m')$. Now as $u : \{1, \dots, M\} \to \bigcup_{k=1}^{K} \{k\} \times \{1, \dots, N(k)\}$ is bijective, we can reassign the indices in the summation over $m \in \{1, \dots, M\}$ with u. Then, we obtain

$$\begin{aligned} M\mathcal{L}(\mathcal{D}, \phi, p)^p &= \sum_{m=1}^{M} \|x_m - \phi(y_m)\|^p = \sum_{m=1}^{M} \|x_{u(m)} - \phi(y_{u(m)})\|^p \\ &= \sum_{k \in \{1, \dots, K\}} \sum_{n=1}^{N(k)} \|x_{k,n} - \phi(F(x_{k,n}, e_{k,n}))\|^p. \end{aligned} \tag{5.11}$$

Dividing by $M = KN(K)$ yields (5.5). Combining (5.4) and (5.5) and taking the p-th root yields the desired inequality.

In order to prove the second part of (i) we define the map

$$\theta : \{y_k\}_{k=1}^K \to \mathbb{C}^{d_1}$$
$$y_k \mapsto \theta(y_k) := \operatorname*{argmin}_{z\in\mathbb{C}^{d_1}} \frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - z\|^p.$$

For fixed $k \in \{1, \dots, K\}$ now define

$$f_k : \mathbb{C}^{d_1} \to [0, \infty)$$
$$z \mapsto f_k(z) := \frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - z\|^p.$$

We define

$$r_k = \sup_{n,n'\in\{1,\dots,N(K)\}} \|x_{k,n} - x_{k,n'}\|^p.$$

In the following we prove that

(I) f_k is continuous,

(II) $\operatorname{argmin}_{z\in\mathbb{C}^{d_1}} f_k(z) = \operatorname{argmin}_{z\in\mathcal{B}(x_{k,n},2r_k)} f_k(z)$ for all $n \in \{1, \dots, N(K)\}$.

In order to prove (I) - that f_k is continuous - we first prove that $f_k^{1/p}$ is continuous. By Minkowski's inequality for the counting measure and dividing by $N(K)^{1/p}$, we obtain

$$\begin{aligned} f_k(z)^{1/p} &= \left(\frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - z\|^p \right)^{1/p} \\ &\leqslant \left(\frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - z'\|^p \right)^{1/p} + \left(\frac{1}{N(K)} \sum_{n=1}^{N(K)} \|z' - z\|^p \right)^{1/p} \\ &\leqslant f_k(z')^{1/p} + \|z' - z\|. \end{aligned} \tag{5.12}$$

Reversing the role of $z, z' \in \mathbb{C}^{d_1}$ and subtracting the resulting inequalities leads to

$$|f_k(z)^{1/p} - f_k(z')^{1/p}| \leqslant \|z' - z\|. \tag{5.13}$$

Thus, $f_k^{1/p}$ is continuous and, hence, $f_k = (f_k^{1/p})^p$ is continuous. Now we proceed to prove (II). Let $n \in \{1, \dots, N(K)\}$. Fix $z \in \mathbb{C}^{d_1} \backslash \mathcal{B}(x_{k,n}, 2r_k)$. Then, for $n' \in \{1, \dots, N(K)\}$ we have that

$$\|x_{k,n'} - z\| \geqslant \|z - x_{k,n}\| - \|x_{k,n} - x_{k,n'}\| > 2r_k - r_k = r_k. \tag{5.14}$$

Summing over $n' \in \{1, \dots, N(K)\}$ and dividing by $N(K)$, proves that $f_k(z) \geqslant r_k^p$ for $z \in \mathbb{C}^{d_1} \backslash \mathcal{B}(x_{k,n}, 2r_k)$. On the other hand, $f_k(x_{k,n}) \leqslant r_k^p$, which proves that f_k cannot have minimizers outside the ball $\mathcal{B}(x_{k,n}, 2r_k)$ for any $n \in \{1, \dots, N(K)\}$. This proves (II). Now recall the that $\theta(y_k) = \operatorname{argmin}_{z\in\mathcal{B}(x_{k,n},2r_k)} f_k(z)$ for any $n \in \{1, \dots, N(K)\}$. As $\mathcal{B}(x_{k,n}, 2r_k)$ is closed and bounded, it is compact by Heine-Borel. Thus, as f_k is continuous by the Extreme Value Theorem there exists a minimum that is attained on $\mathcal{B}(x_{k,n}, 2r_k)$. Hence, the map θ defined in (5.12) has non-empty values. We now prove that θ is an optimal map. Let $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ be a map, then by the minimizing definition of θ, we have that

$$\frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - \theta(F(x_{k,n}, e_{k,n}))\|^p \leqslant \frac{1}{N(K)} \sum_{n=1}^{N(K)} \|x_{k,n} - \phi(F(x_{k,n}, e_{k,n}))\|^p.$$

Summing over $k \in \{1, \dots, K\}$ and dividing by K yields,

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N(K)}\sum_{n=1}^{N(K)}\|x_{k,n}-\theta(F(x_{k,n},e_{k,n}))\|^p \\ \leqslant \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N(K)}\sum_{n=1}^{N(K)}\|x_{k,n}-\phi(F(x_{k,n},e_{k,n}))\|^p.$$

Now using the bijective index map $u : \{1, \dots, M\} \to \bigcup_{k=1}^{K}\{k\} \times \{1, \dots, N(k)\}$, that $M = KN(K)$, taking the p-th root and, then, taking the infimum over maps $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ yields

$$\left(\frac{1}{M}\sum_{m=1}^{M}\|x_m-\theta(F(x_m,e_m))\|^p\right)^{1/p} \\ \leqslant \inf_{\phi:\mathcal{M}_2\to\mathbb{C}^{d_1}}\left(\frac{1}{M}\sum_{m=1}^{M}\|x_m-\phi(F(x_m,e_m))\|^p\right)^{1/p} = \inf_{\phi:\mathcal{M}_2\to\mathbb{C}^{d_1}}\mathcal{L}(\mathcal{D},\phi,p).$$

The reverse inequality trivially holds and thus

$$\left(\frac{1}{M}\sum_{m=1}^{M}\|x_m-\theta(F(x_m,e_m))\|^p\right)^{1/p} = \inf_{\phi:\mathcal{M}_2\to\mathbb{C}^{d_1}}\mathcal{L}(\mathcal{D},\phi,p).$$

By the minimization property of θ we have that for any $n' \in \{1, \dots, N(K)\}$ that

$$\frac{1}{N(K)}\sum_{n=1}^{N(K)}\|x_{k,n}-\theta(F(x_{k,n},e_{k,n}))\|^p \leqslant \frac{1}{N(K)}\sum_{n=1}^{N(K)}\|x_{k,n}-x_{k,n'}\|^p.$$

Summing over $n' \in \{1, \dots, N(K)\}$ and dividing by $N(K)$ yields that

$$\frac{1}{N(K)}\sum_{n=1}^{N(K)}\|x_{k,n}-\theta(F(x_{k,n},e_{k,n}))\|^p \leqslant \frac{1}{N(K)^2}\sum_{n=1}^{N(K)}\sum_{n'=1}^{N(K)}\|x_{k,n}-x_{k,n'}\|^p.$$

Summing over $k \in \{1, \dots, K\}$, dividing by K and using the bijective index map u and that $M = KN(K)$ on the left hand side of the inequality yields that

$$\frac{1}{M}\sum_{m=1}^{M}\|x_m-\theta(F(x_m,e_m))\|^p \leqslant \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N(K)^2}\sum_{n=1}^{N(K)}\sum_{n'=1}^{N(K)}\|x_{k,n}-x_{k,n'}\|^p.$$

Thus, taking the p-th root and using the definition of the average kernel size and the above, we obtain

$$\frac{1}{2}\operatorname{Kersize}(F,\mathcal{M}_1,\mathcal{E},p)_K \leqslant \inf_{\phi:\mathcal{M}_2\to\mathbb{C}^{d_1}}\mathcal{L}(\mathcal{D},\phi,p) \leqslant \frac{1}{2}\operatorname{Kersize}(F,\mathcal{M}_1,\mathcal{E},p)_K.$$

In Section **SM** 1, we show that the lower accuracy bound is sharp. In particular, we show with a linear algebra example that for a specific forward model, dataset and set of noise as well as approximate inverse map the lower accuracy bound is attained.

We now prove the inequality (5.5) for part (ii) by using conditional probability densities and utilize these as disintegrations of a measure with respect to a measurable map on the indices as in [28]. This is the only part of the proof where we use that the forward model and approximate inverse map is measurable.

We equip $\{1, \dots, M\}$ with the sigma algebra of the power set $\mathcal{G} = 2^{\{1,\dots,M\}}$. Equip the set $\{1, \dots, K\}$ with the sigma algebra of the power set $2^{\{1,\dots,K\}}$ and note that by definition of $\mathcal{G}$ the function $g : \{1, \dots, M\} \to \{1, \dots, K\}$ is measurable. We define a measure on $\mathcal{G}$ by

$$\mathbb{P}(A) = \frac{|A|}{M}, \tag{5.15}$$

for $A \in \mathcal{G}$ with $A \neq \varnothing$ and $\mu(A) = 0$ otherwise. The elementary definition of disintegration and conditioning applied to the measure on $\mathcal{G}$, as in [28], is

$$\mathbb{P}_k(A) = \mathbb{P}(A|g^{-1}(k)) = \frac{\mathbb{P}(A \cap g^{-1}(k))}{\mathbb{P}(g^{-1}(k))} = \frac{|A \cap g^{-1}(k)|}{|g^{-1}(k)|}. \tag{5.16}$$

Following [28] we can check the following

(i) $\mathbb{P}_k(\cdot)$ is a probability measure on $\{1, \dots, M\}$ for all $k \in \{1, \dots, K\}$.

(ii) The measure $\mathbb{P}_i(\cdot)$ concentrates on the set $g^{-1}(k)$,

$$\mathbb{P}_k\left(\{\{1, \dots, M\}\} \backslash g^{-1}(k)\right) = 0. \tag{5.17}$$

(iii) For $A \subset \{1, \dots, M\}$

$$\mathbb{P}(A) = \sum_{k=1}^{K} \mathbb{P}(g^{-1}(k))\mathbb{P}_k(A). \tag{5.18}$$

The key point in proving (iii) is that g is surjective and that for $k \neq j$, $g^{-1}(k) \cap g^{-1}(j) = \varnothing$. For (i) let $A = \{1, \dots, M\} \in \mathcal{G}$ and $k \in \{1, \dots, K\}$, then as $A \cap g^{-1}(k) = g^{-1}(k)$

$$\mathbb{P}_k(A) = \frac{|A \cap g^{-1}(k)|}{|g^{-1}(k)|} = 1. \tag{5.19}$$

For (ii), as $\left(\{1, \dots, M\} \backslash g^{-1}(k)\right) \cap g^{-1}(k) = \varnothing$, we have that

$$\mathbb{P}_k\left(\{1, \dots, M\} \backslash g^{-1}(k)\right) = \frac{\left|\left(\{1, \dots, M\} \backslash g^{-1}(k)\right) \cap g^{-1}(k)\right|}{|g^{-1}(k)|} = 0. \tag{5.20}$$

Then, for (iii), recall that for $k \neq j$, $g^{-1}(k) \cap g^{-1}(j) = \varnothing$. Thus, for $A \subset \{1, \dots, M\}$ we have that $|(A \cap g^{-1}(k)) \cap (A \cap g^{-1}(j))| = 0$. Hence, we obtain

$$\sum_{k=1}^{K} \mathbb{P}(g^{-1}(k))\mathbb{P}_k(A) = \frac{1}{M} \sum_{k=1}^{K} |A \cap g^{-1}(k)| = \frac{|A|}{M} = \mathbb{P}(A). \tag{5.21}$$

Now let $f : \{1, \dots, M\} \to [0, \infty)$ be measurable and note that g is measurable, as $\{1, \dots, M\}$ is equipped with the sigma algebra of the power set. Then, by Theorem 1.6.12 [7], using the push-forward of $\mathbb{P}$ with respect to g yields

$$\mathbb{E}(f) = \sum_{m \in \{1, \dots, M\}} f(m)\mathbb{P}(m) = \sum_{k \in g(\{1, \dots, M\})} f(g^{-1}(k))(g_* \mathbb{P})(k). \tag{5.22}$$

Note that $f(g^{-1}(k)) = \mathbb{1}_{g^{-1}(k)}(m) f(m)$ and that by (i) and (ii) we have

$$\sum_{m \in g^{-1}(j)} \mathbb{P}(\{m\}|g^{-1}(j)) = 1. \tag{5.23}$$

Now we insert $1 = \sum_{m \in g^{-1}(k)} \mathbb{P}(\{m\}|g^{-1}(k))$ into (5.22) to obtain,

$$\begin{aligned} \mathbb{E}(f) &= \sum_{k \in g(\{1, \dots, M\})} f(g^{-1}(k))(g_* \mathbb{P})(k) \\ &= \sum_{k \in g(\{1, \dots, M\})} \sum_{m \in g^{-1}(k)} \mathbb{P}(\{m\}|g^{-1}(k)) \mathbb{1}_{g^{-1}(k)}(m) f(m)(g_* \mathbb{P})(k) \\ &= \sum_{k \in \{1, \dots, K\}} \frac{1}{N(k)} \sum_{m \in g^{-1}(k)} f(m)(g_* \mathbb{P})(k), \end{aligned} \tag{5.24}$$

where in the last equality we used that $f \geqslant 0$ and $\mathbb{P}(\{m\}|g^{-1}(k)) = \frac{1}{N(k)}$ for $m \in g^{-1}(k)$. Now we have that

$$\begin{aligned}(g_*\mathbb{P})(k) = \mathbb{P}(g^{-1}(k)) = \frac{|g^{-1}(k)|}{M} &= \frac{N(k)}{\sum_{j=1}^{K} N(j)} \\ &\geqslant \frac{N(k)}{K \min_{j\in\{1,\dots,K\}} N(j)} \geqslant \frac{1}{K},\end{aligned} \tag{5.25}$$

with equality if $N(k) = N(K)$ for all $k \in \{1, \dots, K\}$ and note that by assumption $N(k) > 0$. Thus, inserting (5.25) in (5.24) and noting that $\mathbb{P}(m) = \frac{1}{M}$ in (5.22) we obtain

$$\mathbb{E}(f) = \frac{1}{M} \sum_{m\in\{1,\dots,M\}} f(m) \geqslant \frac{1}{K} \sum_{k\in\{1,\dots,K\}} \frac{1}{N(k)} \sum_{m\in g^{-1}(k)} f(m), \tag{5.26}$$

with equality if $N(k) = N(K)$ for all $k \in \{1, \dots, K\}$. Now we define a special $f : \{1, \dots, M\} \to [0, \infty)$ with $f := \|x - \phi(F(x, e))\|^p$ in order to obtain our desired result,

$$\begin{aligned}f : \{1, \dots, M\} &\to [0, \infty) \\ m \mapsto f(m) = \|x - \phi(F(x, e))\|^p(m) &:= \|x_m - \phi(F(x_m, e_m))\|^p.\end{aligned} \tag{5.27}$$

Then, f is measurable as $\{1, \dots, M\}$ is equipped with the sigma algebra of the power set $2^{\{1,\dots,M\}}$ and as a composition of measurable functions, the norm, the approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ and the forward model F. Here we need that the approximate inverse map $\phi : \mathcal{M}_2 \to \mathbb{C}^{d_1}$ and the forward model F are measurable. Hence, by inserting this choice of f into (5.26) we obtain

$$\begin{aligned}\mathbb{E}(f) &= \frac{1}{M} \sum_{m\in\{1,\dots,M\}} \|x_m - \phi(F(x_m, e_m))\|^p \\ &\geqslant \frac{1}{K} \sum_{k\in\{1,\dots,K\}} \frac{1}{N(k)} \sum_{m\in g^{-1}(k)} \|x_m - \phi(F(x_m, e_m))\|^p.\end{aligned} \tag{5.28}$$

Now we can reassign the indices in the inner summation over $m \in g^{-1}(k)$ on the left hand side of (5.28) by the function $u : \{1, \dots, M\} \to \bigcup_{k=1}^{K} \{k\} \times \{1, \dots, N(k)\}$ defined in (5.10). Then, reassigning the indices of the samples from Algorithm 1 in the inner summation over $m \in g^{-1}(k)$ on the left hand side of (5.24) with u yields

$$\begin{aligned}\mathbb{E}(f) &= \frac{1}{M} \sum_{m\in\{1,\dots,M\}} \|x_m - \phi(F(x_m, e_m))\|^p \\ &\geqslant \frac{1}{K} \sum_{k\in\{1,\dots,K\}} \frac{1}{N(k)} \sum_{m\in g^{-1}(k)} \|x_{\pi_1(u(m))} - \phi(F(x_{\pi_1(u(m))}, e_{\pi_1(u(m))}))\|^p \\ &= \frac{1}{K} \sum_{k\in\{1,\dots,K\}} \frac{1}{N(k)} \sum_{n=1}^{N(k)} \|x_{k,n} - \phi(F(x_{k,n}, e_{k,n}))\|^p.\end{aligned} \tag{5.29}$$

□

5.2 Proof of Theorem 3.5

Proof. The proof of Theorem 3.5 closely follows the proof of Theorem 3.3. We start with proving part (i). We merely change the application of the triangle inequality in (5.1) to obtain the desired result. In particular, using the inequality $\|x+x'\|^p \leqslant 2^{p-1}(\|x\|^p+\|x'\|^p)$, we obtain for fixed $m \in \{1, \dots, M'\}$ with $y_m \in \mathcal{M}_2$. Let $e_m \in \mathcal{E}$ be given by $e_m = y_m - F(x_m, 0)$ and note that $y_m = F(x_m, e_m)$. Now define $v_m = \pi_1(P_{\mathcal{N}(F)}(x_m, e_m))$ and $v_m^{\perp} = \pi_1(P_{\mathcal{N}^{\perp}(F)}(x_m, e_m))$ and by the properties of the Moore-Penrose

inverse, see Section **SM** 4, we have that $x_{m'} = v_m^{\perp} - v_m \in F^{-1}(y_m)$ with $m' = m'(m) = M' + m$. Then, we have that

$$\begin{aligned} \|2v_m\|^p &= \|x_m - \phi(y_m) + \phi(y_m) - (v_m^{\perp} - v_m)\|^p \\ &\leqslant 2^{p-1} \left(\|x_m - \phi(y_m)\|^p + \|\phi(y_m) - x_{m'}\|^p \right). \end{aligned} \tag{5.30}$$

Averaging over $m \in \{1, \dots, M'\}$ yields

$$\frac{2^p}{M'} \sum_{m=1}^{M'} \|v_m\|^p \leqslant \frac{2^{p-1}}{M'} \sum_{m=1}^{M'} \left(\|x_m - \phi(y_m)\|^p + \|\phi(y_m) - x_{m'(m)}\|^p \right). \tag{5.31}$$

Reassigning the indices with $m' : \{1, \dots, M'\} \to \{1, \dots, M\}$ where $m'(m) = M' + m$ yields

$$\frac{2^p}{M'} \sum_{m=1}^{M'} \|v_m\|^p \leqslant \frac{2^p}{M} \sum_{m=1}^{M} \|x_m - \phi(y_m)\|^p. \tag{5.32}$$

To prove part (ii) the analogous steps as in the proof of part (i) of Theorem 3.3 can be used. Here the approximate feasible sets need to be replaced by $\{x_m, x_m - 2P_{\mathcal{N}(F)}x_m\}$ for each $m \in \{1, \dots, M'\}$. □

6 Discussion

Inverse problems with non-injective forward models occur often in the sciences and engineering yielding reconstruction problems with no unique solution. The myriad of techniques to design approximate inverse maps that can tackle such problems beg for a unifying framework of evaluation to enable a theoretical and empirical understanding of the method-independent accuracy bounds of methods for designing approximate inverse maps. In this work we have provided sharp accuracy bounds - the average (symmetric) kernel size - to the reconstruction error of any approximate inverse map and algorithms to compute these bounds. Furthermore, we have provided a framework that enables scientists to apply the algorithms to given datasets with low computational and implementation overhead. The lower accuracy bound in part (i) of Theorem 3.3 holds for any possible approximate inverse map, such as deep neural networks and stochastic estimators, and only depend on the forward model F, the dataset of signals $\mathcal{M}_1$ and the set of noise $\mathcal{E}$. For linear forward models with additive noise, the bounds, (half) the average kernel size, depend on the kernel. The average kernel size generalizes the notion of the kernel to non-linear forward models by the notion of non-injectivity. In particular, we prove part (i) of Theorem 3.3 for measurable forward models, but the sigma algebras on the domain and image of the forward model can be chosen such that it is measurable - rendering the inequalities valid under assumptions that are satisfied in a large number of applications.

In Section 4, we demonstrate the applicability of the algorithms and validate the lower accuracy bounds for two different inverse problems. In *localization microscopy,* our lower accuracy bounds from Theorem 3.3 improve upon the well-known Cramer-Rao lower bounds [59] to the variance of estimators, by providing bounds to the localization accuracy under experimentally verified assumptions. Together with the provided implementations, this provides a general framework applicable to localization microscopy for the optimization and design of experiments and estimators - which are approximate inverse maps. For *super-resolution of multi-spectral satellite data,* Theorem 3.5 provides theoretically funded and computationally feasible quantification of accuracy bounds to the empirical error obtained by any super-resolution map. We empirically validated the lower accuracy bound, the average symmetric kernel size, from Theorem 3.5 on four different satellite imaging modalities and for three different approximate inverse maps. Additionally, with respect to the RMSE, the three methods empirically are all within the upper and lower accuracy bounds and, hence, optimal. While we compute and validate the accuracy bounds for the most common loss, the RMSE, the provided bounds and algorithms are applicable for any other evaluation (pseudo) norm of interest to applicants. In fact, our main results, part (i) of Theorem 3.3 and Theorem 3.5, can be generalized to all evaluation functions and scores that satisfy the triangle inequality.

6.1 Relation to previous work

Many frameworks for the classification of the lowest achievable worst-case reconstruction error for problems, similar to the inverse problem in (2.1), exist. For example, Gelfand widths for noiseless linear forward models [62], best k-term approximation [30], and worst-case bounds for set-valued approximate inverse maps on Banach spaces have been established in [5] and, then, extended to metric spaces in [53]. For a forward models with noise, [18] have established the notion of optimal learning and upper bounds for the worst-case accuracy based on the Chebycheff radius. Generalized instance optimality of approximate inverse maps and the corresponding upper bounds on the reconstruction error have been presented in [23]. In [31] the authors propose a framework to measure the optimal accuracy for non-linear methods of approximation. See also [56] for a survey of early work on optimal recovery. For normed spaces, worst- and average- case accuracy bounds are established in [63]. In many of these works, the underlying spaces are infinite or finite dimensional Banach spaces, and the reconstruction error is measured by a given norm and (upper) bounds on the worst-case error are provided. However, these results have not been sufficient for evaluating approximate inverse maps, such as stochastic estimators or deep neural networks, in order to be applicable in computationally challenging settings - with large datasets and non-linear forward models with a mixture of additive and multiplicative noise. For example, in [2] B. Adcock and N. Dexter state that despite the ability of deep neural networks to approximate functions relevant to scientific computing, a large gap between expressivity and practical performance remains. Previous work has not provided a feasible approach and a corresponding software library for computing accuracy bounds on a general class of inverse problems and large datasets requiring the use of GPUs. In this work, by providing computable and quantifiable lower and upper accuracy bounds to the average reconstruction error of any approximate inverse maps, we provide a computable answer to the question posed by M. Burger and T. Roith in [24].

6.2 Conclusion

A fundamental limitation of this work is that the accuracy bounds in Theorem 3.3 only hold on a dataset processed by Algorithm 1. In order to evaluate approximate inverse maps on an extension of this dataset requires convergence estimates of the average kernel size as the number of samples increases. The average kernel size could be shown to converge to a factor times the average kernel size on measurable metric spaces defined in [42]. Ideally, this result is proved for exchangeable samples, as in many applications there is no guarantee that the samples are independently distributed. A further limitation of the proposed framework is that Algorithm 1 requires possible solutions of a given inverse problem and in some applications it may not be possible to acquire these. However, a benefit of this work is that neither the distribution nor the number of possible solutions influence the validity of the accuracy bounds in Theorem 3.3 and Theorem 3.5. The accuracy bounds in Theorem 3.5 only require one possible solution for linear forward models with additive noise, thereby enabling a computationally efficient estimation of accuracy bounds. In the realm of deep learning, yet another limitation of this work is, that Algorithm 1 requires the existence of training data with ground truth labels and is not immediately applicable to a self-supervised setting. Establishing a generalization of the accuracy bounds, the average (symmetric) kernel size, is an intriguing question for future work and would enable guarantees for generalization of approximate inverse maps obtained by optimization on a finite dataset to an uncountable set.

References

[1] B. Adcock, S. Brugiapaglia, and C. G. Webster. *Sparse polynomial approximation of high-dimensional functions*, volume 25. SIAM, 2022.

[2] B. Adcock and N. Dexter. The gap between theory and practice in function approximation with deep neural networks. *SIAM Journal on Mathematics of Data Science*, 3(2):624–655, 2021.

[3] B. Adcock and A. C. Hansen. *Compressive Imaging: Structure, Sampling, Learning*. Cambridge University Press, Cambridge, UK, 2021.

[4] S. Anwar, S. Khan, and N. Barnes. A deep journey into super-resolution: A survey. *ACM computing surveys (CSUR)*, 53(3):1–34, 2020.

[5] V. V. Arestov. Optimal recovery of operators and related problems. *Collection of Papers from the All-Union School on the Theory of Functions*, pages 3–20, 1986.

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

[7] R. B. Ash. *Real analysis and probability: probability and mathematical statistics: a series of monographs and textbooks*. Academic press, 2014.

[8] P. M. Atkinson and P. J. Curran. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogrammetric Engineering & Remote Sensing*, 63(12), 1997.

[9] C. Aybar, D. Montero, S. Donike, F. Kalaitzis, and L. Gómez-Chova. A comprehensive benchmark for optical remote sensing image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[10] F. Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017.

[11] H. H. Barrett and K. J. Myers. *Foundations of image science*. John Wiley & Sons, 2013.

[12] C. Beck, A. Jentzen, and B. Kuckuck. Full error analysis for the training of deep neural networks. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 25(02):2150020, 2022.

[13] C. Belthangady and L. A. Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, 16(12):1215–1225, 2019.

[14] A. Ben-Israel and T. N. Greville. *Generalized inverses: theory and applications*. Springer Science & Business Media, 2006.

[15] J. A. Benediktsson, J. Chanussot, and W. M. Moon. Very high-resolution remote sensing: Challenges and opportunities [point of view]. *Proceedings of the IEEE*, 100(6):1907–1910, 2012.

[16] M. Bertero, P. Boccacci, and C. De Mol. *Introduction to inverse problems in imaging*. CRC press, 2021.

[17] P. J. Bickel and K. A. Doksum. *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman and Hall/CRC, 2015.

[18] P. Binev, A. Bonito, R. DeVore, and G. Petrova. Optimal learning. *Calcolo*, 61(1):15, 2024.

[19] L. Borcea, J. Garnier, A. V. Mamonov, and J. Zimmerling. Waveform inversion with a data driven estimate of the internal wave. *SIAM Journal on Imaging Sciences*, 16(1):280–312, 2023.

[20] B. Borden. Mathematical problems in radar inverse scattering. *Inverse Problems*, 18(1):R1, 2001.

[21] P. T. Boufounos and R. G. Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pages 16–21, 2008.

[22] C. A. Bouman. *Foundations of Computational Imaging: A Model-Based Approach*. SIAM, Philadelphia, PA, 2022.

[23] 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(12):7928–7946, 2014.

[24] M. Burger and T. Roith. Learning in image reconstruction: A cautionary tale. *SIAM News, Collections*, 57(08), 2024.

[25] E. J. Candes and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[26] E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[27] G. Casella and C. Robert. Monte carlo statistical methods. *University of Florida*, 2008.

[28] J. T. Chang and D. Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

[29] S. Chen, Y. Ogawa, C. Zhao, and Y. Sekimoto. Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195:129–152, 2023.

[30] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[31] A. Cohen, R. Devore, G. Petrova, and P. Wojtaszczyk. Optimal stable nonlinear approximation. *Foundations of Computational Mathematics*, 22(3):607–648, 2022.

[32] 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(12):e2107151119, 2022.

[33] E. S. Crafts and U. Villa. Benchmarking diffusion annealing-based bayesian inverse problem solvers. *IEEE Open Journal of Signal Processing*, 2025.

[34] I. Craig and J. Brown. *Inverse problems in astronomy*. Adam Hilger Ltd., Accord, MA, 1986.

[35] H. Cramér. *Mathematical methods of statistics*, volume 9. Princeton university press, 1999.

[36] P. Curran and H. Williamson. Airborne mss data to estimate glai. *International Journal of Remote Sensing*, 8(1):57–74, 1987.

[37] S. Donike, C. Aybar, L. Gómez-Chova, and F. Kalaitzis. Trustworthy super-resolution of multispectral sentinel-2 imagery with latent diffusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:6940–6952, 2025.

[38] A. Fannjiang and T. Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29:125–228, 2020.

[39] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Birkhauser, 2013.

[40] 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.

[41] N. M. Gottschling, V. Antun, A. C. Hansen, and B. Adcock. The troublesome kernel: On hallucinations, no free lunches, and the accuracy-stability tradeoff in inverse problems. *SIAM Review*, 67(1):73–104, 2025.

[42] N. M. Gottschling, P. Campodonico, V. Antun, and A. C. Hansen. On the existence of optimal multi-valued decoders and their accuracy bounds for undersampled inverse problems. *arXiv preprint arXiv:2311.16898*, 2023.

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

[44] D. P. Hoffman, I. Slavitt, and C. A. Fitzpatrick. The promise and peril of deep learning in microscopy. *Nature Methods*, 18(2):131–132, 2021.

[45] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[46] 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(1):20–40, 2009.

[47] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[48] 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(6):3054–3070, 2020.

[49] J. Kong, Y. Ryu, S. Jeong, Z. Zhong, W. Choi, J. Kim, K. Lee, J. Lim, K. Jang, J. Chun, et al. Super resolution of historic landsat imagery using a dual generative adversarial network (gan) model with cubesat constellation imagery for spatially enhanced long-term vegetation monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 200:1–23, 2023.

[50] Y. Korolev and J. Latz. Inverse problems, 2020. Lecture notes, Michaelmas Term 2020, University of Cambridge.

[51] A. Lee, K. Tsekouras, C. Calderon, C. Bustamante, and S. Pressé. Unraveling the thousand word picture: an introduction to super-resolution data analysis. *Chemical reviews*, 117(11):7276–7330, 2017.

[52] J. S. Liu and J. S. Liu. *Monte Carlo strategies in scientific computing*, volume 10. Springer, 2001.

[53] G. G. Magaril-Il'yaev and K. Y. Osipenko. Optimal recovery of functionals based on inaccurate data. *Mat. Zametki*, 50(6):85–93, 1991.

[54] V. Maiorov and A. Pinkus. Lower bounds for approximation by mlp neural networks. *Neurocomputing*, 25(1-3):81–91, 1999.

[55] W. Mercer and A. Hall. The experimental error of field trials. *The Journal of Agricultural Science*, 4(2):107–132, 1911.

[56] C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. *Optimal estimation in approximation theory*, pages 1–54, 1977.

[57] 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.

[58] K. P. Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.

[59] R. J. Ober, S. Ram, and E. S. Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, 2004.

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[61] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep relu neural networks. *Neural Networks*, 108:296–330, 2018.

[62] A. Pinkus. *N-widths in Approximation Theory*, volume 7. Springer Science & Business Media, 2012.

[63] L. Plaskota. *Noisy information and computational complexity*, volume 95. Cambridge University Press, 1996.

[64] D. A. Quattrochi and M. F. Goodchild. *Scale in remote sensing and GIS*. CRC press, 1997.

[65] C. R. Rao, C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.

[66] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[67] I. Sgouralis, L. W. Xu, A. P. Jalihal, Z. Kilic, N. G. Walter, and S. Pressé. Bnp-track: a framework for superresolved tracking. *Nature Methods*, 21(9):1716–1724, 2024.

[68] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[69] H. F. Smith. An empirical law describing heterogeneity in the yields of agricultural crops. *The Journal of Agricultural Science*, 28(1):1–23, 1938.

[70] R. Strack. Imaging: AI transforms image reconstruction. *Nature Methods*, 15(5):309, 2018.

[71] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Sov Dok*, 4:1035–1038, 1963.

[72] L. von Diezmann, Y. Shechtman, and W. Moerner. Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking. *Chemical reviews*, 117(11):7244–7275, 2017.

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

[74] D. Yarotsky. Error bounds for approximations with deep relu networks. *Neural networks*, 94:103–114, 2017.

[75] S. Zacks. The theory of statistical inference. *(No Title)*, 1971.

[76] 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.

[77] H. Zheng, W. Chu, B. Zhang, Z. Wu, A. Wang, B. T. Feng, C. Zou, Y. Sun, N. Kovachki, Z. E. Ross, et al. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. *arXiv preprint arXiv:2503.11043*, 2025.

[78] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487, 03 2018.